

TÜRKİYE CUHURİYETİ
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLERİ ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI

ELEKTRONİK BELGE YÖNETİM SİSTEMLERİ (EBYS) ÇERÇEVESİNDE
METİN MADENCİLİĞİ

Yüksek Lisans Tezi

Deniz Begüm KÜÇÜK

Ankara, 2025

**TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLERİ ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI**

**ELEKTRONİK BELGE YÖNETİM SİSTEMLERİ (EBYS)
ÇERÇEVESİNDE METİN MADENCİLİĞİ**

Yüksek Lisans Tezi

Deniz Begüm KÜÇÜK

**Tez Danışmanı
Pof. Dr. Tülay OĞUZ**

Ankara, 2025

**TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI**

**ELEKTRONİK BELGE YÖNETİM SİSTEMLERİ (EBYS)
ÇERÇEVESİNDE METİN MADENCİLİĞİ**

Yüksek Lisans Tezi

Tez Danışmanı

Prof. Dr. Tülay OĞUZ

TEZ JÜRİSİ ÜYELERİ

Adı ve Soyadı

1-Prof. Dr. Tülay OĞUZ

2-Prof. Dr. Sacit ARSLANTEKİN

3-Doç. Dr. Mustafa BAYTER

Tez Savunma Tarihi

4.07.2025

**TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE,**

Prof. Dr. Tülay OĞUZ danışmanlığında hazırladığım “**Elektronik Belge Yönetim Sistemleri (EBYS) Çerçevesinde Metin Madenciliği**” adlı yüksek lisans tezimdeki bütün bilgilerin akademik kurallara ve etik davranış ilkelerine uygun olarak toplanıp sunulduğunu, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallarına uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul edeceğimi beyan ederim.

4.07.2025

Deniz Begüm KÜÇÜK

Bitanecik anneme ve canım babama...



“per aspera ad astra.”

"zorluklardan yıldızlara"

İÇİNDEKİLER

İTHAF.....	i
İÇİNDEKİLER.....	ii
KISALTMALAR.....	vii
ŞEKİLLER.....	vii
TABLolar.....	viii
I. BÖLÜM.....	1
GİRİŞ.....	1
I. 1. Önem	3
I. 2. Amaç	10
I. 3. Problem ve Hipotez	3
I. 4. Sınırlılıklar	14
I. 5. Yöntem	15
I. 5. 1. Araştırmanın Modeli.....	16
I. 5. 2. Verilerin Toplanması	17
I. 5. 3. Verilerin Çözümlemesi	21
II. BÖLÜM.....	22
ELEKTRONİK BELGE YÖNETİM SİSTEMLERİNDEN DİJİTAL DÖNÜŞÜME.....	22
II. 1. Elektronik Belge Yönetim Sistemleri: Tanım ve İşlevler	22
II. 1. 1. Türkiye’de EBYS: Yasal Düzenlemeler ve Uygulama Standartları.....	32
II. 1. 2. EBYS Bileşenleri.....	39
II. 2. Dijital Dönüşüm.....	54
II. 2. 1. Literatür Değerlendirmesi.....	59
II. 2. 2. Türkiye’de Dijital Dönüşüm.....	66
II. 2. 3. Dijital Dönüşüm Kullanım Alanları ve Etkileri.....	73
III. BÖLÜM	
METİN MADENCİLİĞİNİN TEMELLERİ: KÖKENİ, YÖNTEMLERİ VE UYGULAMA ALANLARI.....	83
III. 1. Metin Madenciliğine Genel Bakış.....	83
III. 2. Veri Madenciliği: Tanım, Amaç ve Veri Ön İşleme Aşamaları.....	89

III. 2. 1. Veri Ön İşleme Süreci.....	93
III. 2. 1. 1. Veri Temizleme (Data Cleaning): III. 2. 1. 2. Veri Dönüştürme (Data Transformation).....	94
III. 2. 1. 3. Veri İndirgeme (Data Reduction).....	94
III. 2. 2. Veri Madenciliği Teknikleri.....	94
III. 2. 2. 1. Sınıflandırma (Classification).....	94
III. 2. 2. 2. Regresyon(Regression).....	97
III. 2. 2. 3. Kümeleme (Clustering).....	97
III. 2. 2. 4. Bağımlılık Modelleme/Birliktelik Kuralı Madenciliği (Dependency Modelling /Association Rule Mining).....	98
III. 2. 2. 5. Anormallik Tespiti (Anomaly Detection).....	99
III. 2. 2. 6. Özetleme (Summarization).....	99
III. 2. 3. Veri Madenciliği Algoritmaları.....	100
III. 2. 3. 1. Algoritma seçimini etkileyen temel faktörler.....	102
III. 2. 3. 1. 1. Etiketli/Etiketsiz Veri Durumu.....	102
III. 2. 3. 1. 2. Veri Boyutu ve Karmaşıklığı.....	102
III. 2. 3. 1. 3. Doğrusal/Doğrusal Olmayan İlişkiler.....	102
III. 2. 3. 1. 4. Veri Dağılımı ve Şekli.....	103
III. 2. 3. 1. 5. Gürültülü Veri ve Aykırı Değerler.....	103
III. 2. 3. 1. 6. Dengesiz Veri Setleri.....	103
III. 2. 4. Veri Madenciliği Kullanım Alanları.....	103
III. 2. 4. 1. Pazarlama.....	104
III. 2. 4. 2. Sağlık Hizmetleri.....	106
III. 2. 4. 3. Eğitim ve Araştırma.....	107
III. 2. 4. 4. Finansal Hizmetler.....	107
III. 2. 4. 5. Güvenlik.....	107
III. 3. Metin Yoğunluğunda Anlam Arayışı: Metin Madenciliği İhtiyacı.....	108
III. 3. 1. Metin Madenciliği.....	111
III. 3. 1. 1. Metin Madenciliği Süreci: Ön İşleme ve Vektörleştirme.....	114
III. 3. 1. 1. 1. Metin Temizleme.....	115

III. 3. 1. 1. 1. 1. Noktalama İşaretlerini ve Rakamları Filtreleme.....	115
III. 3. 1. 1. 1. 2. Normalizasyon.....	115
III. 3. 1. 1. 2. Tokenizasyon.....	116
III. 3. 1. 1. 2. 1. Kelime Tokenizasyonu.....	116
III. 3. 1. 1. 2. 2. Karakter Tokenizasyonu.....	116
III. 3. 1. 1. 2. 3. Cümle Tokenizasyonu.....	116
III. 3. 1. 1. 2. 4. N-gram Tokenizasyonu.....	116
III. 3. 1. 1. 3. Durdurma Kelimeleri Filtreleme.....	117
III. 3. 1. 1. 4. Lemmatizasyon.....	117
III. 3. 1. 1. 4. 1. Kelime Türü (Part-of-Speech (POS)).....	117
III. 3. 1. 1. 4. 2. Morfolojik Analiz (Morphological Analysis).....	118
III. 3. 1. 1. 4. 3. Sözlük Arama (Dictionary Lookup).....	118
III. 3. 1. 1. 5. Kök Çıkarma.....	118
III. 3. 1. 2. Metin Madenciliği Teknikleri/Yöntemleri.....	124
III. 3. 1. 2. 1. Bilgi Çıkarma.....	125
III. 3. 1. 2. 2. Bilgi Erişim.....	126
III. 3. 1. 2. 3. Özetleme.....	128
III. 3. 1. 2. 4. Kümeleme.....	130
III. 3. 1. 2. 5. Kategorizasyon.....	132
III. 3. 1. 2. 6. Konu Takibi (Topic Tracking).....	134
III. 3. 1. 2. 7. Doğal Dil İşleme.....	135
III. 3. 1. 2. 8. Soru Yanıtlama.....	136
III. 3. 1. 2. 9. Trend Analizi.....	137
III. 3. 1. 2. 10. Kavram Bağlantısı (Concept Linkage).....	138
III. 3. 1. 2. 11. Birliktelik Kuralı Madenciliği.....	138
III. 3. 1. 2. 12. Bilgi Görselleştirme.....	139
III. 3. 1. 3. Metin Madenciliği Kullanım Alanları.....	140
III. 3. 1. 3. 1. Pazar Analizi.....	141

III. 3. 1. 3. 2. Eğitim ve Araştırma.....	142
III. 3. 1. 3. 3. Müşteri İlişkileri Yönetimi.....	144
III. 3. 1. 3. 4. İnsan Kaynakları Yönetimi.....	145
III. 3. 1. 3. 5. Yaşam Bilimleri.....	146
III. 3. 1. 3. 6. Sosyal Medya.....	146
III. 3. 1. 3. 7. Güvenlik.....	147
IV. BÖLÜM	
EBYS'LERDE METİN MADENCİLİĞİ.....	148
IV. 1. EBYS'lerde Metin Madenciliği Kullanım Alanları: Sorunlar ve Çözümler.....	157
IV. 1. 1. Bilgi Tanımlama: Metin Madenciliği ve Ontoloji Tabanlı Yaklaşımlar.....	158
IV. 1. 1. 1. Sorun.....	158
IV. 1. 1. 2. Çözüm.....	168
IV. 1. 2. İş Zekâsı: Metin Madenciliği ile Bilgi Keşfi ve Öngörü Sağlama.....	187
IV. 1. 2. 1. Sorun.....	187
IV. 1. 2. 2. Çözüm.....	189
IV. 1. 3. Güvenlik: Metin Madenciliği ile Siber Güvenlik ve Bilgi Güvenliği Uygulamaları.....	202
IV. 1. 3. 1. Sorun.....	202
IV. 1. 3. 2. Çözüm.....	204
V. BÖLÜM	
SONUÇ VE ÖNERİLER.....	212
V. 1. Gelecek Çalışmalar İçin Öneriler.....	225
KAYNAKÇA.....	227
ÖZET.....	250
ABSTRACT.....	251

KISALTMALAR

BBY: Bilgi ve Belge Yönetimi

BoW: Bag of Words

DDO: Dijital Dönüşüm Ofisi

EBYS: Elektronik Belge Yönetim Sistemi

IDF: Inverse Document Frequency

KDD: Knowledge Discovery in Databases

KEP: Kayıtlı Elektronik Posta

KWIC: Key Word in Context

RSS: Really Simple Syndication / Rich Site Summary

SDP: Standart Dosya Planı

TDF: Ters Doküman Frekansı

TF: Terim Frekansı

TF-TDF: Terim Frekansı-Ters Doküman Frekansı

ŞEKİLLER

Şekil 1.	33
Şekil 2.....	34
Şekil 3.....	34
Şekil 4.....	37
Şekil 5.....	39
Şekil 6.....	52
Şekil 7.....	72
Şekil 8.....	87
Şekil 9.....	93
Şekil 10.....	122
Şekil 11.....	153
Şekil 12.....	162
Şekil 13.....	163
Şekil 14.....	167
Şekil 15.....	174
Şekil 16.....	176
Şekil 17.....	177
Şekil 18.....	178
Şekil 19.....	179
Şekil 20.....	183
Şekil 21.....	194
Şekil 22.....	196
Şekil 23.....	199
Şekil 24.....	200
Şekil 25.....	201
Şekil 26.....	207
Şekil 27.....	211

TABLÖLÄR

Tablo 1..... 64
Tablo 2..... 88
Tablo 3..... 155
Tablo 4..... 173



I. BÖLÜM

GİRİŞ

Metin madenciliği en yalın ifadesiyle yapılandırılmamış verilerden bilgi çıkarma (Information Extraction) ve bilgi keşfi (Knowledge Discovery) sürecini sağlayan bir tekniktir. Bilgi çıkarma kısaca “varlıkların, ilişkilerin ve en zorlayıcı olan olayların (kimin kime ne yaptığı) çıkarılması da dahil olmak üzere, bazı ilgi alanlarıyla ilgili bilgiler için metnin taranması işlemidir. Anahtar kelime aramalarından daha derin bir analiz gerektirir” (Hobbs, 2002, s. 260) bilgi keşfi ise, “büyük hacimli yapılandırılmış veya yapılandırılmamış verilerden bilgi oluşturulmasıdır” (Zhu ve diğerleri, 2013, s. 203) Yapılandırılmamış veri, insanlar tarafından kolaylıkla anlaşılıp yorumlanabilen ancak bilgisayar tarafından anlaşılması ve anlamlandırılması zor olan verilerdir. Metin madenciliği yöntemi ile yapılandırılmamış veriler kolay bir şekilde anlamlandırılarak bilgi çıkarma, bilgi keşfi süreçlerinin otomatikleştirilmesi sağlanmış ve bu sayede iş yükü azaltılarak emekten, paradan ve zamandan tasarruf etmek mümkün hale gelmiştir.

Organizasyonlar günümüz teknolojisinde iş ve işlem süreçlerini ağı dayalı elektronik sistemlerde yürütebilmekte ve yönetebilmektedir. Kuruluşların, faaliyetleri sonucu oluşan belgeler ya da üretilen dokümanlar dijital ortama taşınabilmekte ya da doğrudan dijital ortamda oluşturulabilmektedir. Bu noktada ise Elektronik Belge Yönetim Sistemlerinin (EBYS) varlığı ortaya çıkmaktadır. EBYS, “idarelerin faaliyetlerini yerine getirirken hazırladıkları ve faaliyetlerinin delili olabilecek eYazışma Teknik Rehberi’ne uygun belgelerin içerik, metadata, format ve ilişkisel özelliklerini koruyan, belgelerin ait olduğu fonksiyon veya işlem için delil teşkil eden ve aidiyet zinciri içerisindeki yönetimini elektronik ortamda sağlayan sistem”dir (Resmî Yazışmalarda Uygulanacak

Usul ve Esaslar Hakkında Yönetmelik, 2020, s. 1). Çalışmamızda ise EBYS terimi; tüm organizasyonların faaliyetlerine yönelik olarak kişiselleştirilebilen, süreçlerin otomatikleştirildiği, dahili ve harici verilerin kullanılarak organizasyonda kullanılabilecek her açıdan (bilgi erişim, bilgi çıkarımı, bilgi keşfi, bilgi güvenliği, siber güvenlik vb.) değer yaratan belge tabanlı bir yönetim sistemi olarak tasarlanmıştır. Bu noktada çalışma, tek bir EBYS uygulamasını ele almamakta olup, bütünsel ve organizasyon gereklerine yönelik olarak kişiselleştirilebilir bir çerçeve sunmaktadır.

EBYS'ler belgeye dayalı sistemler olduğundan çok fazla yapılandırılmamış veriyi bünyesinde taşımaktadır. Yapılandırılmamış verilerin işlenmesini işletmeler açısından değerlendiren Chen (2001, s. 2), "İnternet, web patlaması ve birçok arama motorunun başarısının ardından işletmeler için çarpıcı bir öge haline gelmiştir; bilgi olgusaldır, ancak yapılandırılmamıştır ve çoğu durumda metinseldir" saptamasında bulunmuştur. Kurum ve kuruluşlarda büyük miktarda varlık gösteren yapılandırılmamış metinsel veriler, bilgi yönetimi bağlamında kurum ve kuruluşun geleceğine yön vermede, fırsat, tehdit ve risklerin belirlenmesinde önemli bir yere sahiptir.

Yapılandırılmamış ya da metinsel verilerin bilgi yönetimi açısından sağladığı yararları Chen şöyle özetlemektedir: Bilgi çıkarımsaldır, soyuttur ve iş kararlarını desteklemek için gereklidir (Chen, 2001, s. 2). Bilgiyi anlamlandırmak insanlar tarafından kolaylıkla gerçekleştirilebilen bir eylemken bu eylemi, bilgisayarın gerçekleştirmesini sağlamak kolay değildir. Bilginin bilgisayar tarafından anlamlandırılabilmesi için birkaç adımlı önışlem sürecinin gerçekleşmesi gerekmektedir. "Bilgi kavramı birçok disiplinde ve iş uygulamasında yaygın hale gelmiştir. Örneğin, bilgi bilimciler; taksonomileri, konu başlıklarını ve sınıflandırma şemalarını bilginin temsili olarak kabul eder" (Chen, 2001, s. 2).

Bu çalışmada da Bilgi ve Belge Yönetimi (BBY) bağlamında ve EBYS çerçevesinde; Metin Madenciliği tekniğinin kullanım biçimleri belirlenmeye çalışılacaktır. Çalışmamız metin madenciliğini EBYS süreçlerine entegre ederek; bilgi erişim, bilgi çıkarma, bilgi keşfi, ilişki ve örüntülerin tespiti gibi konularda verimi üst noktaya ulaştırmayı; EBYS uygulamalarındaki sorunları aşmayı böylelikle etkin bir bilgi yönetim süreci gerçekleştirmeyi amaçlayan bir mimari ortaya koymaktadır. Örneğin EBYS’lerde; Bilgi varlıklarının otomatik bir şekilde tanımlanması, ontoloji tabanlı erişim sisteminin geliştirilmesi, meta verilerin otomatik çıkarılması, Standart Dosya Planına (SDP) uygun şekilde konu başlıklarının otomatik bir şekilde atanması, hassas verilerin tespiti gibi noktalarda metin madenciliği tekniğinin kullanılabilirliği kavramsal olarak irdelenecektir. Bu bağlamda bilgi haritalarında metin madenciliğini kullanarak örtük bilginin anlamlandırılması, iş zekasını (business intelligence) arttırarak fırsat, tehdit ve risklerin ortaya çıkarılması, siber güvenlikte güvenlik açıklıklarının belirlenmesi, güvenlik açığı veri tabanlarından bilgi çıkarılması, elektronik belgelere dayalı erişim kontrol politikalarının oluşturulması, belgelerin güvenlik düzeyinin belirlenmesi, yetkili erişimin sağlanması ve benzeri yararların mevzuatla da ilişkilendirilerek ortaya konulması amaçlanmaktadır.

I. 1. Önem

Kurum ve kuruluşlar iş ve işlemlerini gerçekleştirirken fonksiyonlarına göre çeşitli konu ve biçimlerde belgeler/dokümanlar üretmektedir ve ürettikleri belgelerin/dokümanların büyük çoğunluğu yapılandırılmamış verilerden oluşmaktadır. Günümüz dünyasında verilerin %80’inden fazlası yapılandırılmamış haldedir (Harbert, 2021). Yapılandırılmamış veriler, içerisinde keşfedilmeyi bekleyen ilişki, örüntü ve

bilgiler ile doludur. Organizasyonlarda belge/doküman yönetimi için son derece önem taşıyan, yapılandırılmamış metinsel veriler, metin madenciliği ile işlenerek, organizasyonlara değer katabilir. Ancak yapılan literatür taramaları ortaya koymaktadır ki, öngörülerin oluşmasında önem arz eden metin madenciliğinin uygulanması ve faydaları tam olarak araştırılmamıştır (Khan ve Vorley, 2017, s. 4).

Bazı bilim insanları bu konular üzerine araştırmalar yapmış olsa da bilgi yönetimi ve metin madenciliği yönlerini bütünleştiren kapsayıcı bir çalışma bulunmamaktadır (Schieber ve Kruses, 2014, s.7). Bu noktada yapılan çalışma, bir belge yönetimi sürecinin tamamını göz önüne alarak, bütünsel bir yaklaşım ile; hangi aşamalarda, hangi amaçlarla metin madenciliğinin uygulanabileceğini; metin madenciliği ile mevcut ve potansiyel sorunların üstesinden ne şekilde gelinebileceğini EBYS'ler üzerinden açıklamaktadır. Gerçekleştirilen literatür incelemesi sonucunda, “EBYS’lerde Metin Madenciliğini” bütünsel şekilde ele alan bu çalışmanın, alanda benzeri görülmemiş bir yaklaşım sunduğu düşünülmektedir; ancak benzer çalışmaların varlığı ihtimali tamamen dışlanmamaktadır. Bu noktada çalışma, bu konudaki literatür boşluğunu doldurarak, literatüre özgün bir katkı sağlamaktadır.

Yapılacak çalışma ile kuruluşlarda organizasyon yapısına, kurumun iş ve fonksiyonlarına bağlı olarak gerekli bilgi çıkarımı buna bağlı örtük bilgilerin anlamlandırılması akabinde bu iş ve fonksiyonlara bağlı fırsat, tehdit ve riskler göz önünde bulundurularak kurumların gerekli önlemleri alabilmesine, iş zekâsını arttırmasına, ürün ve hizmetlerini geliştirmesine, kurumsal politikalar oluşturmaya özetle kurumun varlığına, geleceğine ve ülke içi katkılarına ışık tutmaktadır.

Kurum ve kuruluşlar iş ve işlemlerini gerçekleştirirken kurumun fonksiyonuna göre çeşitli konu ve biçimlerde belgeler üretmekte ve üretilen bu belgelerin büyük çoğunluğu yapılandırılmamış verilerden oluşmaktadır. “İnternetteki belgeler (örneğin,

haber makaleleri, başyazılar, araştırma yazıları, raporlar ve resmî belgeler), projeleri etkili bir şekilde yönetmek ve kontrol etmek için yararlı kaynaklar olarak kullanılabilir” (Pathirage ve diğerleri, 2007 aktaran Moon ve diğerleri, 2018, s. 4792). “Bu durumda, çeşitli metin verilerinden çıkarılan bilgiler, karar vermede ve iş stratejisi geliştirmede önemli bir rol oynayabilir” (Qady ve Kandil, 2010 aktaran Moon ve diğerleri, 2018, s. 4792)

Veri madenciliği ve metin madenciliği, karar vericilerin iş zekâsı elde etmesine ve bilinçli kararlar vermesine yardımcı olabilir, ancak bunlardan birini kullanmak bize yalnızca kısmi bir resim verir. Veri madenciliği uygulaması, yalnızca sayılarla cevaplanamayan sorulara yol açabilir. Bu nedenle karar vericiler, sayıların açıklamalarını bulmak için metinsel verileri incelemek üzere metin madenciliğine ihtiyaç duyacaktır. (Li ve Brook, 2006, s. 1416).

“Yapılandırılmamış metni (veya metin madenciliği teknolojisini) iş zekasına entegre etmeye yönelik artan ilgi, bir dizi faktörden kaynaklanmaktadır” (Li ve Brook, 2006, s. 1410). Dolayısı ile kurumlarda çoğunlukla bulunan bu yapılandırılmamış metinsel veriler metin madenciliği yöntemi kullanılarak kurumlara bilgi yönetimi açısından birçok fayda sağlayacaktır.

Metin Madenciliğinin sağladığı bilgi çıkarma ve bilgi keşfi sayesinde kurumların oluşturacağı bilgi haritalama (information mapping) yolu ile iç ve dış kaynaklarının analizini genişleterek örtük bilgilere erişilmesi, organizasyonun iş zekasını arttırması beklenmektedir. Kurumsal stratejiler ve pazarlama bağlamında değerlendirildiğinde kuruma ivme kazandıracak bu yöntem ve tekniklerin önemi ön plana çıkmaktadır. EBYS'lere entegre şekilde çalışacak, belgeleri ve belgelerdeki bilgi parçalarını çekerek

anlamlandıran, kuruma özgü bilgi haritası; karar almalarda kurumun ve EBYS'lerin vazgeçilmez bir parçası olacaktır.

EBYS'lerde organizasyonlar iş ve işlemleri süresince çok sayıda belge üretmektedirler. Elektronik ortamda oluşturulan bu belgeleri temsil etmek amacıyla belgelerin özelliklerine ait metadatalar oluşturulmaktadır. Metadatalar, elektronik belgelerin özelliklerini tanımlayan, içerik, bağlam ve yapıları hakkında bilgi sağlayan verilerdir. Belgenin oluşturulduğu zaman, yaratıcı gibi özellikler otomatik olarak belirlenmesine karşın konu metadataları belgenin tümünü okuyup belgeyi en iyi tanımlayan şekilde ifade edilmesi gerekmektedir. Bu durumda kurumsal düzeyde yapılacak analizler için konu üst verilerinin standart bir yaklaşımla oluşturulması önem kazanmaktadır.

Kurumsallık sürekliliği gerektirir ve sürekliliği sağlamak için iş ve işlemlerin standartlaştırılması gerekmektedir. Kurumsallaşmada önemli sorunlardan biri, çalışanların zamanla değişebilmesidir; dolayısı ile 5-20 yıllık yapılacak bir analiz için oluşturulacak konu üst verilerinin de bir standarda bağlanması gerekmektedir. Konu üst verilerinin metin madenciliği ile otomatik oluşturulması, sadece standartlaşma sağlamayacak aynı zamanda çalışanları da iş yükünden kurtarmış olacaktır.

TS 13298/2015 standardında EBYS'lerde metadatalar ile ilgili önem şu şekilde ortaya konmuştur:

- EBYS sistemi, kurumlararası belge paylaşımını sağlamak amacıyla sistemdeki herhangi bir belgeyi ilgili metadata setiyle birlikte Elektronik Yazışma Paketi (EYP) kurallarına uygun olarak gönderebilmeli ve alabilmelidir.
- EBYS, bünyesindeki elemanlara ait metadata bilgileri üzerinden arama yapabilmelidir. Arama kapsamına elektronik belge yöneticisinin tanımladığı metadata

elemanları yanı sıra kullanıcı tarafından tanımlanan metadata bilgileri de dahil edilmelidir.

- Elektronik belgelerin korunmasına yönelik metadatalar EBYS içerisinde tanımlanabilir olmalıdır.
- EBYS, sistem içerisinde tanımlanacak metadata elemanları için herhangi bir kısıtlama getirmemelidir.
- Metadata elemanlarının içerdiği bilgiler, EBYS'nin fonksiyonlarını şekillendirmelidir. (örnek: yetkili erişim için)
- EBYS içerisinde farklı belge türleri için farklı metadata elemanları tanımlanabilmelidir.
- Metadata elemanlarına ait bilgiler mümkün olduğu ölçüde sistem tarafından kontrol edilmelidir. Yanlış veri girişleri engellenmeli ve kullanıcılar uyarılmalıdır.

Yukarıda sunulan bilgiler ışığında metadata oluşturulmasının EBYS'lerde belge paylaşımı, belgeye erişim, standartlaşma gibi açılardan ne kadar önemli olduğu açıkça görülebilmektedir. Özellikle son maddede vurgulanan, “üst veri elemanlarına ait bilgilerin mümkün olduğu ölçüde doğru girilmesi” söylemi bizi üst verilerin standartlaştırılarak otomatik hale getirilmesine yönlendirmiştir. Bu noktada motivasyonumuz ise metin madenciliğinin uygulanabilirliği olmuştur.

“Artık günümüzde, kara, deniz, hava ve uzayın yanı sıra ‘siber ortam’ da yeni bir mücadele alanı olarak ortaya çıktı” (Çifci, 2013, s. 2 aktaran Özdemirci ve Torunlar, 2018, s.79) Dolayısı ile kurumun teknik önemler içerisinde ele alması gereken siber güvenlik kavramı hem yazışmaların hem depolamanın dijitalleştiği günümüzde üzerinde önemle durulması gereken ve en üst düzeyde tedbirler alınması gerektiren bir olgu olarak karşımıza çıkmaktadır.

Metin madenciliği ile güvenliğı saęlama konusu ile ilgili olarak Ignaczak ve arkadaşları (2021) çalışmalarında siber güvenlik ve bilgi güvenliğı kullanım alanlarından řu řekilde bahsetmiştir:

- Metin madenciliğı, güvenlik açıklarını belirlemek için kaynak kodlarını analiz edebilir veya varlıklardaki yeni zayıflıkları keşfetmek için güvenlik açığı veritabanlarından bilgi çıkarabilir.
- Metin madenciliğı, bir önem puanı hesaplayarak veya yeni güvenlik açıklarını özelliklerine göre kategorize ederek güvenlik açığı keşfini destekler.
- Tehditler hakkında uyarılar vermek için metin madenciliğı kullanılır. Önlemenin aksine tespit, bir kuruluřa kendisini gelecekte etkileyebilecek kötü niyetli davranışlar hakkında bilgi saęlar.
- Eriřim kontrolü, doęal dilde yazılmış politikalara dayalı bir mekanizmada erişim kontrol kurallarının oluşturulması veya elektronik belgelere dayalı erişim kontrol politikalarının oluşturulması gibi faaliyetlerin otomasyonunu saęlamak için metin madenciliğini kullanır.
- Metin madenciliğı, bir belgenin içeriğini analiz edebilir ve otomatik olarak güvenlik düzeyi atayabilir ayrıca e-posta hesaplarını izleyebilir ve veri sızıntısını tahmin etmek için kullanıcı davranışındaki deęişiklikleri belirleyebilir.
- Güvenlik kontrolleri, içeriğı korumak ve oluşuma yetkisiz erişimi önlemek için metin madenciliğı uygulayabilir.
- Bir olayı tespit etmek kuruluşlar için çok zor olabilir, ancak metin madenciliğı farklı genel veri kaynaklarını izleyebilir ve bir kuruluşla ilgili olaylar hakkında uyarılar verebilir. Bir olay tespitinden sonra, bir kuruluşun olayın ayrıntılarını analiz etmesi ve belgelemesi gerekir ve metin madenciliğı uygulaması öncelik sırasını kolaylaştırabilir.

- Metin madenciliğine dayalı sistemler, bir kuruluşa gönderilen kimlik avını engelleyebilir. Veri sızıntısı sistemlerinde, metin madenciliği, sınıflandırılmış bilgileri harici ağlara iletme girişimini algılayabilir ve bir veri ihlali önleyebilir.

“Belgenin imzacısı olan yetkili makam, belgeye ait gizlilik derecesinin belirlenmesinden sorumludur” (İletişim Başkanlığı Elektronik Belge Yönetim Sistemi Uygulama Yönergesi). EBYS’ler üzerinde metin madenciliği ile güvenlik düzeyi atarken Gizlilik Dereceli Belgelerde Uygulanacak Usul ve Esaslar Hakkında Yönetmelik esas alınarak bir uygulama gerçekleştirilebilir. Bunun yanında süreç otomatikleşirken, veri işleyenlerin yapacağı hatalar en aza indirgenmiş olacaktır. Unutulmamalıdır ki kurumlarda en zayıf halkayı insanlar oluşturmaktadır. Bu bağlamda metin madenciliği ile yapılacak otomatikleştirme işlemi kurumun güvenlik tedbirlerini de arttıracaktır.

Tüm bunlar göz önüne alındığında yapılacak tez çalışması; kurum ve kuruluşların EBYS’lerinde metin madenciliği kullanarak bilgi güvenliğinin oldukça önemli bileşenlerinden birini oluşturan yetkili erişimin sağlanması, bilgi haritaları oluşturularak kuruma dair fırsat, tehdit, risk ve politikaların belirlenmesi; gizlilik derecelerinin, konu numaralarının, konu metadatalarının belirlenirken kişiler tarafından yapılan hataların en az seviyeye indirilerek otomatikleştirilmesi, bilgi güvenliği ve siber güvenliğin sağlanması konularında yararları ortaya koyarak alanımızda ve belge yönetim sistemlerinde metin madenciliğinin benimsenmesini ve EBYS’lerin gelişimi adına yasal düzenlemelerin gelişimine dikkat çekmek hedeflenmektedir.

I. 2. Amaç

EBYS'nin öncelikli işlevi, kurumların yasal yükümlülüklerini, üstlendiği işlevleri yerine getirirken gerekli olan belgelerin e-ortamda üretim sürecini gerçekleştirmektir. Yani belge öncelikle yasal ve idari amaçlar için üretilerek yönetim eyleminin gerçekleştirilmesi sağlanmaktadır (Özdemirci, Savaş, Akdoğan, 2017, s. 37). EBYS'lerdeki bu yönetim eyleminin en verimli şekilde gerçekleşebilmesi için ise bilgi teknolojilerinin benimsenerek etkin bir belge yönetim süreci gerçekleştirmek gerekmektedir. Böylece kurum ve kuruluşların karar verme süreçlerinde doğru adımlarla ilerleyerek kurumsal sürekliliği sağlanacaktır. Bu da sadece kuruma değil aynı zamanda ülkenin gelişimine de katkı sağlayacak önemli bir noktadır.

Özetlemek gerekirse;

Belge Yönetimi, organizasyonların sürekliliğini sağlayan ve geleceğini oluşturan önemli bir süreçtir. İçerdiği bileşenler dolayısı ile birçok alanla iç içe olmakta ve dokunmakta, yeni teknolojilerle kendisini geliştirmesi beklenmektedir. Belgesini yönetemeyen bir kurumun sürekliliğini sağlaması, kurumuyla ilgili kararlar alabilmesi, iş ve işlemleri yönetebilmesi dolayısı ile ayakta kalması olası gözükmemektedir.

Türkiye'de 5070 sayılı Elektronik İmza Kanunu'nda yapılan düzenlemeler ile EBYS'lerin kullanımı zorunlu hale gelmiştir. Kurumlar iş ve işlemlerini sürdürürken oluşturduğu belgeleri bu kanunlarla elektronik ortama taşımış ve iş süreçlerini EBYS'ler üzerinden yürütmektedirler. Organizasyonların oluşturdukları bu belgelerin büyük çoğunluğu yapılandırılmamış verilerin oluşturduğu göz önüne alındığında EBYS'lerde metin madenciliği kullanımı kaçınılmaz hale gelmektedir.

Bu çalışmada metin madenciliğinin EBYS'lerin hangi alanlarında kullanılabileceğini ve ne gibi yararlar sağlayabileceğini belirlemek amaçlanmaktadır.

Aynı zamanda uygulanabilirlik, yasal düzenlemelerle olan ilişkisi kurularak açıklanmaya çalışılacaktır. Çalışmanın başlangıcında gözlemlerimize dayalı olarak EBYS’lerde metin madenciliğinin uygulanabileceği varsayılan konular şu şekilde belirlenmiştir: Konu Metadataları Çıkarma, Bilgi Haritaları, Bilgi Güvenliği, Siber Güvenlik ve Konu Başlığı.

Çalışmanın sonunda ise bu konular, kullanım alanları göz önüne alınarak üç ana başlık altında detaylandırılmıştır. Bunlar;

- Belgelerin tanımlanmasını, konu metadatalarının oluşturulmasını, sınıflandırılmasını, kategorize edilmesini ve ontoloji tabanlı arama motorlarının geliştirilmesini içeren “**Bilgi Tanımlama**”,
- Kurumlarda etkili karar alınmasını sağlayacak olan, bilgi çıkarımı, bilgi keşfi, örtük bilgilerin ve gizli ilişki ve desenlerin ortaya çıkarılmasını sağlayan bilgi haritalarını içeren “**İş Zekâsı**”,
- Güvenlik açıkları, güvenlik kontrolleri ve veri ihlallerinin tespit edilmesini içeren “Siber Güvenlik” ile hassas bilgilerin tespit edilmesi ile yetkili erişimi sağlayan “Bilgi Güvenliği” konularını içeren “**Güvenlik**” başlıklarıdır.

Çalışma EBYS’ler üzerinden “bilgi tanımlama”, “iş zekası” ve “güvenlik” konularındaki mevcut ve potansiyel sorunları ele alarak aşağıdaki sorulara çözüm üretmeyi amaçlamaktadır:

- Metin madenciliği teknikleri EBYS’lerde hangi aşamalarda kullanılmalı,
- Ne amaç ile kullanılmalı ve
- Bunların kuruma ve personele sağlayacağı katkı/katkılar nedir?

Çalışma ortaya koymaktadır ki, metin madenciliği EBYS’ler için kullanım alanı oldukça geniş olan belge yönetiminin önemli konularından yetkili erişim, gizlilik derecelendirme, fırsat-tehdit-risklerin belirlenmesi, veri ihlalinin engellenmesi, olay

tespiti gibi konularda bir otomatikleşme ve standart sağlayarak kurumun sürekliliğini sağlayacak ve kişiler tarafından yapılan yanlışların önüne geçerek emek, zaman ve maliyetten kazanacaktır.

Bu bağlamda düşünüldüğünde; organizasyonlar belge yönetimi kapsamında yeni teknoloji ve gelişmelerle sistemlerini kurmaları ve bunun yönetiminde yer alacak bilgi uzmanlarının bu yeni teknolojilere hâkim olması gerekmektedir. Yeni teknolojiler ile BBY’de uzman olan kişiler bilgilerini diğer disiplinler ile harmanlayarak yazılımcılar ve veri analistleri ile aynı dili konuşacak seviyeye gelecek ve belge yönetimini ileri düzeye taşıyacaklardır. Tüm bunların bir standarda oturması için ise EBYS’ler ile ilgili daha detaylı yasal düzenlemeler yapılarak alınacak kararlarda disiplinlerarası bir zihin ve alanında uzman kişiler tarafından düzenlenmesi yerinde bir karar olacaktır.

Çalışmanın temel amacı; *EBYS’lerde metin madenciliği kullanım alanlarını belirleyerek sistemlerin iyileştirilmesini, kurumlarda belge yönetiminin niteliğinin artırılması ve EBYS’ler hakkında yapılmış ve yapılacak olan mevzuatın geliştirilmesine katkı sağlamaktır.*

Metin madenciliği tekniği BBY alanı için uygulama alanı oldukça geniş, ilişkili ve kendi yapısı içerisinde popüler ve gelişmiş bir yöntem olmasına rağmen bilgi uzmanları tarafından günümüzde tam olarak benimsenmemiştir (Lamba & Madhusudhan, 2022, s. 5)

Yapılacak çalışmada, belge yönetimi bağlamında EBYS’lerde metin madenciliği uygulamasının; kurum ve kuruluşlara sağlayacağı katma değerleri ortaya koyarak, kurum ve kuruluşların ilerlemesini ve bilgi uzmanlarının bilinçlendirilip metin madenciliği tekniğini benimseyerek bu gelişmelerde aktif rol oynamasının sağlanması; bu gelişmeler nihayetinde ülkemiz içerisinde EBYS’lerin kullanımını geliştirecek ve kurumlara

dolayısı ile de ÷lkemize katma değęer saęlamasına olanak tanıyacak yasal dñzenlemeler ve kanunların oluřmasına/geliřmesine de ön ayak olması beklenmektedir.

I. 3. Problem ve Hipotez

Yapılan çalıřmaya dair problemler genelden özele řu řekilde sıralanabilir:

Kurumlarda belge yñnetiminde kullanılan sistemlerde yeni teknolojilerin (yapay zekâ, doęal dil iřleme, metin madencilięi vb.) kullanılmamasının sistem üzerinde oluřturacaęı temel sorunundan hareketle;

Kullanıcı taraflı hataların bulunması,

Üst verilerin oluřturulmasında niteliksiz uygulamaların gerçekteřtirilmesi,

Belgelere eriřmede uzun vakitler harcanması,

Fırsat, risk ve tehditlerin belirlenmesi ve sistem içi gñvenlięin saęlanması hususlarında bahsedilen teknolojilerin kullanılmaması,

EBYS'lerde veri iřleyenlerin iř ve iřlemlerini yñrñtñrken konu bařlıklarını ve konu metadatalarını belirlemede yanlışlıklar yaparak kurum içi bir standart saęlayamaması ve geriye dñnñk bir çalıřmada, belgeye eriřimde sorunlar çıkararak kuruma; emek, zaman ve maliyet aęısından zarara uęratması,

Konu bařlıkları, konu metadataları, yetkili eriřim, belgelerin gizlilik derecelerinin belirlenmesi noktalarında kiřilerin inisiyatifine bırakılarak bir standart saęlanamaması ve bu durumun kurumların sñreklilięini tehdit etmesi,

Sistemlere ait tarama motorlarının ontoloji tabanlı yerine sñzdizimsel olması,

Ülkemizde EBYS'lerin zorunlu tutulmasına karşın EBYS'leri konu edinen mevzuat incelendiğinde içerik açısından yetersiz kalması ve bunun sonucunda kurumların; belge yönetimi için kullanılacak teknikler ve bunların sağlayabileceği yararlar hakkında habersiz olmaları.

Bu doğrultuda problemimiz;

“Kurumlarda belge yönetiminde kullanılan sistemlerde kişiler tarafından gerçekleştirilen iş ve işlem süreçlerinin hatalara açık olması ve bu açıkların giderilmesinde metin madenciliği tekniklerinin kullanılmamasıdır”; şeklinde belirtilebilir.

Bu probleme çözüm bulmak amacıyla temel hipotezimiz ve alt hipotezimiz ise şu şekilde verilebilir:

Temel hipotezimiz şöyledir:

H0: EBYS'lerde insan kaynaklı hataların giderilmesinde metin madenciliği tekniği kullanılabilir

Bu temel hipoteze bağlı olarak alt hipotezimiz de,

H1: EBYS'lerde metin madenciliği tekniği kullanılarak, sistemin gelişimi bilgi güvenliği, siber güvenlik ve kurumsal politikaların oluşturulmasında yarar sağlanabilir.

olarak oluşturulmuştur.

I. 4. Sınırlılıklar

Gerçekleştirilen tez çalışması Belge Yönetimi bağlamında EBYS'lerde Metin Madenciliği tekniğinin kullanım alanlarını ve faydalarını kapsamaktadır.

Çalışma; Ontoloji Tabanlı Arama Motoru Geliştirme- Konu Metadatalarını Belirleme- Konu Numarası Verme- Bilgi Haritaları ile Fırsat Risk ve Tehditleri

Belirleme- Metinsel Verilerden Bilgi Çıkarma ve Keşfi- Hassas Verilerin Tespiti ile Bilgi Güvenliğini Sağlama ve Siber Güvenlik konuları bağlamında inceleme yapmakta olup, bu konuları üç ana başlıkta incelemektedir. Bunlar; Bilgi Tanımlama, İş Zekâsı ve Güvenliktir. Çalışmada arşiv yönetim süreçleri, dolayısı ile de ayıklama ve imha süreçleri kapsam dışıdır. Bunun yanı sıra çalışma bir uygulama içermemektedir. Bu nedenle çalışmada EBYS’lerde metin madenciliği, algoritmalar üzerinden değil teknikler üzerinden işlenmektedir. Tezde yer alan Python kodları, bu tekniklerin EBYS’ler üzerinde doğrudan çalışan kapsamlı bir uygulama niteliğinde olmayıp, ele alınan **tekniklerin işleyişini ve somutlaştırılmasını desteklemek** amacıyla sunulmuştur.

I. 5. Yöntem

“Yöntem, “belli bir problemi çözmek için izlenen düzenli yol” şeklinde tanımlanabilir. Araştırma yöntemi, problemin çözümü için gerekli verilerin toplanması, analizi ve yorumlanmasını içerir. Araştırma yöntemi bir başka araştırmacıların aynı çalışmayı gerektiğinde aynen uygulayabilmesine olanak sağlayacak kapsam ve ayrıntıda verilmelidir. Yöntem bölümünde “araştırmanın modeli”, “veriler ve toplanması” ile “verilerin çözümü ve yorumlanması” açıklanır. Araştırmanın uygulamalı olduğu durumlarda yöntem bölümüne “evren ve örneklem” başlığı eklenir.” (Eymen, 2007, s. 21)

Çalışma, temel olarak; belgenin yaşam döngüsü göz önüne alınarak hangi aşamalarda hangi metin madenciliği tekniklerinin kullanılabileceğini, EBYS’lerdeki mevcut ve potansiyel sorunlar üzerinden çözmeye çalışmaktadır. Bu noktada çalışmada

kullanılan yöntem, betimsel araştırma yöntemi olup, bu kapsamda özellikle; doküman analizi, literatür taraması ve sistematik literatür taraması kullanılmaktadır. Araştırmanın modeli tekil araştırma modeli olup, araştırma da bir uygulama gerçekleşmeyeceği için “evren ve örneklem” yer almamaktadır. Araştırma yönteminin detaylarına aşağıdaki başlıklarda yer verilmektedir.

I. 5. 1. Araştırmanın Modeli

“Araştırma modeli araştırma amacına uygun olarak verilerin toplanması ve çözümlenebilmesi için gerekli koşulların tanımlanmasıdır. Bu koşullar, “tekil” ve “ilişkisel” olmak üzere iki farklı araştırma modeli kullanılarak açıklanabilir.

Tekil araştırma modelinde ilgilenilen olay ve konulara ilişkin değişkenler ayrı ayrı tanımlanır. Bu tanımlama geçmiş ya da şimdiki zamanla sınırlı olabileceği gibi, zamanın bir fonksiyonu olarak gelişimsel de olabilir.” (Eymen, 2007, s. 22)

İlişkisel araştırma yöntemi ise iki veya daha fazla değişken arasındaki ilişkileri ortaya çıkarmak için gerçekleştirilen bir araştırma modelidir (Şen, 2015).

Çalışma, “EBYS’ler çerçevesinde Metin Madenciliği”ni incelemekte olup, iki veya daha fazla değişken arasındaki ilişkinin ortaya konması söz konusu değildir. Dolayısıyla belirlenen araştırma modeli, tekil araştırma modelidir.

I. 5. 2. Verilerin Toplanması

Araştırmanın kuramsal kısmında yer alan kavramsal içerikler, mevzuat, kılavuz, standart örneklerini tespit etmek üzere literatür taraması yapılmış ve konuya ilişkin daha önce yazılmış olan literatüre dayalı betimsel araştırma yöntemi kullanılmıştır.

“Betimsel araştırma yöntemi, herhangi bir durum, olay ve problemi etraflica tanımlamak, yorumlamak ve irdellemek için kullanılır ve ölçütler belirleyerek incelenen olaylar ve değişkenler arasında ilişkinin varlığı ve derecesi sorgulanır.” (Aydoğdu ve diğeri, 2017, s.558)

“Betimleme ya da tarama araştırmaları olayların, objelerin, varlıkların, kurumların, grupların ve çeşitli alanların ‘ne’ olduğunu betimleyen incelemelerdir. Bu tip incelemeler, mevcut durumları, şartları ve özellikleri aynen ortaya koymaya çalışır. Bu araştırmalarda ilişkiler; inanışlar, görüşler, davranışlar; uygulanmaktaki süreçler, etkiler ve gelişmekte olan yön ve eğilimler üzerinde durulur. Betimleme araştırmaları mevcut olayların daha önceki olay ve şartlarla ilişkilerini de dikkate alarak, durumlar arasındaki etkileşimi açıklamaya çalışır.” (Kaptan, 2000, s. 175 aktaran Özkol, 2023, s. 10).

Temelinde betimsel araştırma yöntemi, çalışılan konunun niteliklerini betimlemeyi amaçlamaktadır (QuestionPro, 9.06.2025). Bu çalışmada, metin madenciliğinin farklı alanlardaki çözüm ve uygulamaları, belge yönetimi bağlamında EBYS'lere uyarlanarak, metin madenciliğinin EBYS'ler üzerinde “uygulanabilirliğini göstermek” ve EBYS'lerdeki sorunla çözüm getirebilmek amaçlanmaktadır. Dolayısı ile çalışmamızda, ele alınan konu manipüle edilmeden, literatürdeki mevcut kaynaklar

kullanılarak, tanımlanmış ve tasvir edilmiştir (Salomão, 2023; Bingöl Üniversitesi, [Tarih Yok], s. 3).

Çalışmada kullanılan ve betimleyici araştırmanın bir türü olarak karşımıza çıkan bir araştırma yöntemi de karşılaştırmalı araştırmadır. Çalışma, farklı alanlarda uygulanan metin madenciliği çalışmalarını bilgi/belge yönetimine uyarlanabilirliğini incelemesi açısından, bir nevi farklı alanlardaki (bağlamlardaki) konuları karşılaştırma amacı taşımaktadır. (Bingöl Üniversitesi, [Tarih Yok], s. 4).

Bu doğrultuda çalışmada verileri toplamak için, araştırma konusuna ve alt konulara yönelik bilgilenme amacıyla bu konularla ilgili çalışmalar incelenmiş, literatür taraması yapılmıştır. Literatür taraması araştırma problemi ile ilgili bilginin literatürün özeti, sentezi ve incelenmesidir (Balcı, 2016, s. 75.) Bu sayede benzer problemi araştıran çalışmalar görülmekte, ilgili konu anlaşılmakta ve son gelişmeler takip edilmektedir.

Literatür taraması sayesinde, bir konudaki bilgi kaynakları incelenerek, mevcut bilgilere genel bir bakış sağlanmakta ve araştırılan konuya dair teori, yöntem ve boşluklar tartışılmaktadır. Bu süreçte, analiz, sentez ve eleştirel düşünme önem arz etmektedir (Best Edit & Proof, 2021). Çalışma, literatürde yer alan metin madenciliği uygulamalarını eleştirel biçimde inceleyerek ve sentezleyerek belge yönetimine yeni bir bakış açısı kazandırmaktadır.

Çalışma literatür taraması sırasında araştırmanın kavramsal çerçevesini oluşturmak için metin analitiği, metin analizi, metin madenciliği, EBYS, metin madenciliği ve bilgi yönetimi, metin analitiği ve bilgi yönetimi, metin analizi ve bilgi yönetimi, ontoloji tabanlı erişim sistemleri, metin madenciliği ve belge yönetimi, metin madenciliği ve siber güvenlik, metin madenciliği ve metadata, metin madenciliği ve bilgi güvenliği terimleri İngilizceleriyle birlikte arama motoru, (Google), Ulusal Tez Merkezi, akademik araştırma motoru (Google akademik) ve Ankara Üniversitesi Kütüphane ve

Dokümantasyon Daire Başkanlığının sunduğu veri tabanlarından araştırılmıştır. Ortaya çıkan yayın sayısı oldukça fazladır. Bunlar arasından doğrudan yapılacak çalışma için anlatımımızı ilgilendirecek, olgusal veriler seçilerek kullanılmıştır.

Çalışmada kullanılan kaynaklar aşağıdaki gibidir:

- Bilgi Dünyası Dergisi
- Bilgi Yönetimi Dergisi
- Google
- Google Akademik
- IEEE
- Information Sciences
- ScienceDirect
- Scopus
- Springer Link
- T.C. Dijital Dönüşüm Ofisi
- T.C. Resmî Gazete
- TSE
- UMI ProQuest Digital Dissertations
- Web of Science
- YÖK Ulusal Tez Kataloğu

Araştırmada kullanılan yöntem, nitel veri toplama tekniklerinden biri olan belge (doküman) analizi/incelemesidir. Nitel araştırmalarda belge incelemesinin belgeler, arşiv kayıtları ve çeşitli materyalin araştırma konusu veriyi toplama ve çözümle işlemi olduğunu söyleyen Gürbüz ve Şahin (2017), belge incelenmesinde önemli noktalardan birisinin araştırma konusuyla ilgili veriyi elde edebilecek tarzda tarama işlemi yapmak ve veri kaydı için bir sistem geliştirmek olduğunu ifade etmektedirler.

“Doküman analizi, yazılı belgelerin içeriğini titizlikle ve sistematik olarak analiz etmek için kullanılan bir nitel araştırma yöntemidir (Wach, 2013 aktaran Kiral, 2020, s.173). Doküman analizi, basılı ve elektronik materyaller olmak üzere tüm belgeleri incelemek ve değerlendirmek için kullanılan sistemli bir yöntemdir. Nitel araştırmada kullanılan diğer yöntemler gibi doküman analizi de anlam çıkarmak, ilgili konu hakkında bir anlayış oluşturmak, ampirik bilgi geliştirmek için verilerin incelenmesini ve yorumlanmasını gerektirmektedir” (Corbin & Strauss, 2008 aktaran Kiral, 2020, s.173).

Doküman analizi, başlı başına bir araştırma yöntemi olup, araştırılan konuya dair literatür taramasını da içermektedir. Bu yönü ile literatür taraması, doküman analizi sürecindeki önemli bir veri toplama tekniği olarak yerini almaktadır (Sak ve diğerleri, 2021, s. 244; Özkan, 2023, s. 3)

Doküman analizi kısaca, çeşitli verilerin; toplanması, gözden geçirilmesi, sorgulanması ve analizini gibi süreçleri içeren bilimsel bir araştırma yöntemidir (Özkan, 2023, s. 2). Doküman analizini gerçekleştirirken dikkat edilmesi gereken diğer aşamalar ise; “Doküman analizi, uygun dokümanları bulma, bu dokümanların orijinalliğini kontrol etme, kodlama ve kataloglama konusunda bir sistematik oluşturma ile veri analizi yapma aşamalarından oluşmaktadır” (Merriam, 2009 Aktaran Sak ve diğerleri, 2021, s. 242). Analizin bu süreçlere uyarak gerçekleştirilmesi çalışmanın sistematiki açısından önemlidir.

Çalışmada sistematikliği bir adım ileriye taşıyan diğer yöntem Sistematik Literatür Taraması (SLT)dir. “SLT, bir araştırma sorusu veya ilgili konu hakkında birden fazla araştırma çalışmasından elde edilen bulguların toplamının, eleştirel olarak değerlendirmenin, birleştirmenin ve sunmanın sistematik bir yoludur” (Pati ve Lorusso, 2017, s. 1 aktaran Yıldız, 2022, s. 372).

Çalışmanın temel yaklaşımı, EBYS’lerde metin madenciliğinin çeşitli kullanım alanlarının saptanması, kurum ve kuruluşlara bu teknolojilerin kullanılmamasının yarattığı eksiklikleri öngörerek öneriler sunmaktır. Çalışmamız, mevcut literatürü pasif bir şekilde özetlemenin çok ötesinde olup, metin madenciliği teknolojisini temel alarak, EBYS’lerdeki mevcut ve potansiyel sorunlara yönelik çözümler üretmeyi ve farklı alanlardaki metin madenciliği uygulamalarını belge yönetimine uyarlayarak EBYS’leri en iyi versiyonuna getirmeyi amaçlamaktadır. Bu süreçte çalışmada; ilişki kurma, analiz ve sentez ile betimsel araştırma kapsamında, yoğun bir şekilde doküman analizi ve sistematik literatür taraması yöntemleri kullanılmaktadır.

Araştırmanın biçimsel düzenlemeleri, Ankara Üniversitesi Sosyal Bilimler Enstitüsü’nün 2025 tarihli “Tez Yazım Yönergesi” esas alınarak yapılmıştır. Çalışmanın atıfları ve kaynakçası ise American Psychological Association (Amerikan Psikoloji Derneği) (APA) sistemi doğrultusunda hazırlanmıştır.

I. 5. 3. Verilerin Çözümlemesi

“Toplanan ham veriler, çoğu kez, toplandıkları şekliyle fazla bir şey anlatmazlar. Bunların nitel ve nicel analiz teknikleri kullanılarak çözümlenmesi ve yorumlanması gerekir. Bu nedenle ham verilerin nasıl toplanacağı kadar nasıl çözümleneceği ve yorumlanacağının da önceden belirlenmesi önemlidir.” (Eymen, 2007, s. 25)

Bu bağlamda çalışma bilgi tanımlama, iş zekası ve güvenlik başlıkları altında kategorize edilmiş olup, metin madenciliğinde farklı alanlara uygulanan çalışmalar bu başlıklar doğrultusunda sınıflandırılmış ve sistematik şekilde; ilişkiler kurulmuş, karşılaştırmalar yapılmış ve eleştirel bakış açısı ile ele alınmıştır.

Sınıflandırılan kaynaklar ve EBYS'ler arasında kurulan ilişkiler sayesinde; sorunlar tespit edilmiş, uyarlamalar gerçekleştirilmiş ve çözümler üretilmiştir. Bunlar dışında bilgileri anlamlı hale getirebilmek amacıyla oluşturulan tablo ve şemalarla anlatım güçlendirilmiş ve ilişkiler ortaya konularak verilerin çözümlemesi gerçekleştirilmiştir.



II. BÖLÜM

ELEKTRONİK BELGE YÖNETİM SİSTEMLERİNDEN DİJİTAL DÖNÜŞÜME

Bu bölümde, teknolojinin gelişimiyle birlikte kurumların faaliyetlerini fiziksel ortamdan elektronik ortama taşımasını sağlayan Elektronik Belge Yönetim Sistemleri (EBYS) detaylı olarak ele alınacaktır. İlk olarak, EBYS'nin tanımı, temel bileşenleri ve Türkiye'deki yasal düzenlemeler ile uygulama standartları açıklanacaktır. Ardından, EBYS'lerin dijital dönüşüm sürecindeki kritik rolü vurgulanarak, bu dönüşümün ne olduğu, dijitalleşme ve dijitalizasyon terimleriyle arasındaki farklar ortaya konulacaktır. Bölümün ilerleyen kısımlarında, dijital dönüşümün dünyada ve Türkiye'deki gelişimi, başlıca kullanım alanları ve kurumlar üzerindeki hem olumlu hem de potansiyel olumsuz etkileri değerlendirilecektir. Bu kapsamlı inceleme, ileriki bölümlerde EBYS'ler aracılığıyla metin madenciliği tekniklerinin nasıl başarılabileceği, hangi noktalarda geliştirilebileceği ve ne gibi amaçlara hizmet edebileceği konularına zemin hazırlayacaktır.

II. 1. Elektronik Belge Yönetim Sistemleri: Tanım ve İşlevler

Teknolojinin gelişmesi ile kurumların fiziksel ortamda gerçekleştirdiği iş ve işlemler elektronik ortama taşınmıştır. Elektronik işlemlerin yaygınlaşması ise kuruma/personele/kullanıcıya; hızlı, verimli, entegre ve daha az maliyetli belge yönetim süreçlerini sağlayacak bilgi yönetim sistemlerini kazandırmıştır. Bu gelişmeler doğrultusunda ortaya çıkan Elektronik Belge Yönetim Sistemleri (EBYS) e-devlet yapısını oluşturan önemli bir adım olup, sistemlerin bütünsel çalışmasını gerektiren

uluslararası/ulusal/kurumsal prosedürler, politikalar, mevzuat ve standartları gerektirmektedir.

En genel ifadesiyle bir EBYS çeşitli iş ve işlem süreçleri sonucunda oluşturulan ve kurumlararası bir iletişim aracı olarak da görülebilecek ‘belgelerin’ yönetilmesini sağlayan sistemlerdir. Bir EBYS için belgenin ne anlama geldiği, ‘doküman’ ve ‘kayıt’ farkını da ortaya koyarak aşağıdaki gibi açıklanmaktadır:

Belge: Kurum, kuruluş veya birey tarafından resmi sorumlulukların yerine getirilmesi için alınmış, ya da üretilmiş, içerik, ilişki ve formatı ile ait olduğu fonksiyon için delil teşkil eden kayıtlı bilgidir. Belgenin resmi bir niteliğe sahip olması onu kalıcı ve değiştirilemez kılmaktadır. Bu ayrım ise belgeyi dokümandan ayıran en büyük özelliktir. Her doküman bir belge değildir ama her belge bir doküman görevi görebilmektedir (Kandur, 2006, s. 9; Aydın ve Özdemirci, 2011, s. 107; Okuonghae ve Bakare-Fatungases, 2023, s. 38).

Doküman: Doküman çok daha geniş anlamlı bir kavramdır. Doküman, resmî belge niteliği taşımamakta olup bu özelliği ile kanıtlayıcı bir niteliği de bulunmamaktadır. Ancak dokümanlar, kurumsal aktivitelerin gerçekleştirilmesi amacıyla gelecekte hazırlanacak resmi yazılarda kullanıcılara rehberlik ederek, kullanıcıların bilgi amaçlı olarak kullanabilecekleri kaynakları oluşturmaktadır (Kandur, 2006, s. 9; Aydın ve Özdemirci, 2011, s. 107).

Kayıt: Kayıt kavramına gelince, doküman kavramının aksine, son derece dar kapsamlı bir kavramdır. Kayıt, herhangi bir konudaki bir bilginin varlığına veya yokluğuna işaret etmektedir. Belge veya dokümanların üretim aşamasında kullanılan bir girdidir. Belge veya doküman ile ifade edilen konulara ait daha önceden üretilmiş herhangi

bir belge ya da dokümanın olup olmadığına işaret eder. Bir başka deyişle “kayıt” kavramı, üretilecek olan belge veya dokümanlara değil geçmişte üretilmiş olan belge ya da dokümanlara ait bir kavramdır (Kandur, 2006, s. 9).

EBYS, organizasyonların iş ve işlemleri sonucunda ve organizasyon yapısı doğrultusunda belgenin yaşam döngüsüne dair (ürettikleri/aldıkları belgelerin tanımlanması, düzenlenmesi, dağıtılması, kullanılması ve saklanması gibi) eylemlerin gerçekleştirildiği bir sistemdir. Bu eylemler gerçekleştirilirken ise dikkat edilmesi gereken bazı zorunluluklar bulunmaktadır. Bu zorunluluklar şu şekilde sıralanmaktadır;

- Belgelerin, etkili bir şekilde oluşturulmasını, erişilmesini ve kullanılmasını sağlayarak, alanında uzman kişiler tarafından tanımlanması ve tasnif edilmesi,
- Yetkili erişimin ve gizlilik derecelerinin amacına ve mevzuata uygun yürütülmesi,
- Bilgi güvenliği ve siber güvenlik noktalarında sistemselsel ve insan kaynaklı tedbirlerin sağlanması.

EBYS’lerin temelini oluşturan bilgi yönetimi kavramı, organizasyonlardaki kolektif bilginin (paylaşılan, yeniden kullanılan, sürekli geliştirilen ve bireysel olmayan bilgi) belirlenmesini ve kullanılmasını sağlayarak, organizasyonların karar almasını ve rekabet avantajı elde etmesini sağlamaktadır (Svård, 2017, s. 30). Bu noktada EBYS’leri de içerisine alan, bilgi odaklı yönetim sistemleri, iş süreçlerinde verimliliği artırma ve rekabet avantajı sağlama konularında belirleyici araçlardır (Khan ve Vorley, 2017, s.3).

Sprague çalışmasında (1995), (Parrish and Courtney, 2007, s. 224) Elektronik Belge Yönetiminin her terimini ayrı bir bileşen kabul ederek aşağıdaki gibi tanımlamaktadır:

Elektronik: *Bilişim teknolojilerinin kullanılmasıdır.*

Belge: *İnsanın anlayabileceği şekilde yapılandırılmış, bir konu ile ilgili bilginin bir bütün olarak saklanması ve işlenmesidir.*

Yönetim: *örgütsel bir amacı ilerletmek amacıyla belgelerle yapılabilecek tüm faaliyetleri ifade etmektedir.*

Elektronik belge, fiziksel ortamda oluşturulmuş bir belgenin dijitalleştirilmesini değil, bilgisayar ve yazılım yardımı ile elektronik ortamda oluşturulan ve saklanan veri/bilgi/belgeleri ifade etmektedir (Eroğlu ve Külcü, 2014, s. 331). Elektronik belgeler, resmi veya ticari iş ve işlemlerin kanıtını oluşturmaktadır. Bu nedenle bu belgelerin, orijinalliğinin, doğruluğunun ve erişilebilirliğinin sağlanması ve doğru bir şekilde yönetilmesi gerekmektedir (Borglund ve diğerleri, 2009b, s. 51). Bu gerekliliklerin yerine getirilmesi ise EBYS'ler ile sağlanmaktadır. Yazıcı ve Özdemirci (2019, s. 85) EBYS'yi, organizasyonel faaliyetlerin, en az maliyetle ve en verimli şekilde sürdürülmesine olanak sağlayan temel e-devlet hizmeti olarak tanımlamaktadır. EBYS'lerin bünyesinde barındırdığı iç ve dış entegrasyonlar sayesinde, iş ve işlemler tek bir noktadan yürütülebilmektedir. Bu sayede kurumlarda bürokratik işlemlerin hızlanması, süreçlerin etkin yönetilmesi, veri tekrarının önlenmesi gibi faydalar sağlanarak; kurumlararası standartlaşmaya olanak veren etkin, etkili, verimli ve kullanışlı bir e-devlet yapısı ortaya çıkmaktadır.

The National Archives and Records Administration (NARA) EBYS'yi; elektronik, mikroform, kâğıt gibi çeşitli formatlarda oluşturulmuş belgelerin otomatik bir şekilde yönetilmesi olarak tanımlamaktadır (NARA, 11.12.2023). Benzer şekilde Glossary of Terms de EBYS'yi; bilgisayar ve uygulama yazılımları ile format ayırt etmeksizin farklı belge türlerinin, elektronik bir şekilde yönetildiği bir sistem veya süreç olarak tanımlamaktadır (Griffin ve diğerleri, 2009, s. 16). EBYS ile ilgili yapılan çalışmalar, elektronik ortamda belgelerin üretilmesi ve yönetimini yaygınlaştırmayı amaçlasa da; ülkemizde ilgili mevzuat incelendiğinde hizmete özel gizlilik derecesi dışındaki gizlilik dereceli belgelerin fiziksel ortamda üretilmesi gibi istisnalar da bulunmaktadır. Bu nedenle EBYS'ler elektronik ortamda oluşturulmuş belgelerin yönetiminin yanı sıra dijitalleştirilmiş belgelerin yönetimini de kapsayabilmektedir.

02.02.2015 tarihli ve 29255 sayılı Resmî Gazete'de yayımlanan Resmi Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelikte EBYS kısaca, idare faaliyetlerinin delili olabilecek resmi belgelerin; içerik, metadata, format ve ilişkisel özelliklerinin, elektronik ortamda korunduğu ve yönetildiği bir sistem olarak tanımlanmaktadır. Tanım özellikle, EBYSlerin hesap verilebilir yapısını ön plana çıkararak, belgelerin iş ve işlemler sürecinde her türlü eylemin ispatı olduğunu, dolayısı ile de hukuki açıdan delil oluşturmasını vurgulamaktadır. Eylemlerin kanıtlanması; kurumun, çalışanların ve verisi işlenen ilgili kişinin haklarını koruma açısından önem taşımakta olup, EBYS'lerin önemli bir özelliğini oluşturmaktadır.

TS 13298 numaralı Elektronik Belge Yönetimi Standardında da (2009) benzer şekilde EBYS; belgelerin yaşam döngüsü boyunca, üretiminden tasfiyesine kadar tüm süreçlerde; içerik, format ve ilişkisel özelliklerinin korunarak yönetilmesi olarak tanımlanmaktadır. İletişim Başkanlığı Elektronik Belge Yönetim Sistemi Uygulama Yönergesinde de bu tanıma bağlı kalınarak süreçlerin e-yazışma teknik rehberine uygun

gerçekleştirilmesi gerektiği belirtilmektedir (T.C. Cumhurbaşkanlığı İletişim Başkanlığı...).

Avustralya Ulusal Arşivi de (2005) tanımlamasında aynı unsurları kullanmaktadır, bu tanıma göre EBYS'ler; iş süreçlerinin kanıtlanmasını sağlayarak, belgenin yaşam döngüsünü otomatik bir şekilde yöneten ilişkisel bir sistem olarak açıklamaktadır (Avustralya Ulusal Arşivi, 2005 aktaran Özdemirci ve diğerleri, 2013, s. 14). EBYS'lerde belgenin yaşam döngüsü gerçekleşirken kurumların, veri işleyenlerin ve ilgili kişilerin hakları korunmaktadır. EBYS'ler bunu; iş etkinliklerini kanıtlamak amacıyla kullanılan elektronik imza, elektronik mühür, zaman damgası gibi entegrasyonlar ile sağlamaktadır. Bu entegrasyonlar belgelerin resmi bir statü kazanması için önem taşımaktadır. Özellikle dijitalleştirilmiş belgelerin, doğruluğu ve bütünlüğünün korunduğu ve bir değişime maruz kalmadığını ortaya koymak için, belgeye zaman damgasının işlenmesi gerekmektedir (Aydın, 2010, s. 44).

Ülkemizde yayımlanan birçok mevzuatta, EBYS'lerin iş süreçlerinde, eylemleri ispatlama amacı taşıdığı ortaya çıkmaktadır. Bu mevzuat göz önüne alındığında; (5070 sayılı Elektronik İmza Kanunu, Kamu Kurum ve Kuruluşları Arasında Elektronik Ortamdaki Belge Paylaşımında Kullanılan Kurumsal Şifreleme ve Elektronik Mühür Sertifikalarına İlişkin Usul ve Esaslar, Resmî Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik)

- bilgi ve belge paylaşımlarında elektronik imza, elektronik mühür ve zaman damgası kullanımlarının güvenli ve etkin bir belge yönetim süreci oluşturduğu,
- belgenin bütünlüğünü, inkâr edilemezliğini ve kaynağını belirlemede önemli rol oynadığı

- bilgi güvenliği, veri güvenliği noktalarında temel bir rol üstlendiği belirtilmektedir.

Arıcı ve Kandur (2016, s. 76) çalışmalarında EBYS'lerin karar alma süreçlerindeki işlevselliğini ortaya çıkararak; EBYS'leri, iç ve dış bilgi kaynaklarının depolandığı bir veri havuzu olarak tanımlamaktadırlar. İçerisinde bu kadar çok ve çeşitli verilerin bulunduğu bir sistemin ise, doğru bilgi yönetim stratejileri ile yönetilerek, organizasyonlar için değerli bilgiler elde edilebileceğini belirtmektedirler. EBYS'ler ile kurumsal zekanın, iş zekasının geliştirilmesi; kurumlar içinde fırsat, tehdit ve risklerin ortaya konmasında, bunlara yönelik önlemler alınmasında ya da ürün/hizmet geliştirilmesinde son derece önem arz etmektedir. Belgelerin etkin yönetimi kurumlara nerede olduğunu, hedefleri doğrultusunda nereye gideceğini/gitmesi gerektiğini, gideceği yolda neler yapması gerektiğini gösteren; kurumsal politikalar oluşturulmasını sağlayan, bilgi boşluklarını tamamlayan ve kurumlara karar vermede yol gösteren önemli bir araçtır. EBYS'ler ise bu süreci ilk olarak kurum bazında sonrasında ise ülkeler için en üst performans düzeyine çıkarmayı hedefleyen sistemlerdir.

EBYS'ler; teknolojik, organizasyonel ve piyasa değişikliklerine yanıt olarak kurumsal bilgi ve altyapı yönetiminin yanı sıra; yönetim uygulamalarını da kapsayan dinamik ve sürekli bir yaşam döngüsüdür (Paivarinta, Munkvold, 2005 aktaran Arıcı ve Kandur, 2016, s. 69). Sağlıklı bir kurumsal yönetim için sağlıklı bir belge yönetimi gerekmektedir. Kurumun sürekliliği ve gelişimi buna bağlıdır. EBYS'ler için önem arz eden; yetkilendirme, metadata, tasnif ve gizlilik gibi işlemlerinin belirli standartlar ve kurallar aracılığıyla oluşturulması; bilgi güvenliği yönetim sisteminin kurulması ve bu bağlamda yapılacak tüm işlemler, EBYS kurulmadan önce oluşturulan politikalar doğrultusunda gerçekleştirilmelidir. Böylelikle EBYS'lerin işleyişi kişilere bağlı olarak değil kurumsal bir anlayışla süreklilik kazanır.

Public Record Office tarafından yayımlanan Functional Requirements for Electronic Records Management Systems (İşlevsel Gereksinimler için Elektronik Belge Yönetim Sistemleri) adlı çalışmada, EBYS'lerin bilgiyi etkili bir şekilde yöneterek, organizasyonlara bilgi üstünlüğü sağlama ve operasyonel gereksinimleri destekleme özelliklerine vurgu yapılmaktadır. Bu özellikler, organizasyonlarda kurumsal hafıza oluşturma ve süreklilik için bilgi gereksinimlerini desteklemektedir (Public Record Office, 1999, s. 5). EBYS'lerin yönetilmesinde ulusal/uluslararası/kurumsal politikalar son derece önemlidir. Belgelerin etkin yönetimi için sistem ve iş süreçlerinin her aşaması sistem kurulmadan önce belirlenmelidir. Belirlenen bu politikalar belge yönetiminin ve dolayısı ile kurumun sürekliliğini sağlar.

Belgelere etkin erişim; iyi planlanmış ve tasarlanmış sistem ve süreçler ile gerçekleşmektedir. Devlet Belge Ofisi Manitoba Arşivleri (Government Records Office Archives of Manitoba) tarafından yayımlanan Belge ve Bilgi Yönetimi Terimleri Sözlüğünde (Glossary of Records and Information Management Terms) “Elektronik Doküman ve Belge Yönetim Sistemi (Electronic Document and Records Management System)- dijital belgeleri, belgelerin gerekli yapısını, içeriğini ve bağlamını sağlayan bir sistemde yakalayan ve düzenleyen; belgelerin değiştirilemeyecek şekilde korunmasını (özgünlük) sağlayan; ilgili belgeleri birbirine bağlayan; erişimi mümkün kılan; belgeleri saklayan; belgelere erişimi kontrol eden ve iş ve belge yönetimi kurallarına göre belgelerin imha edilmesine izin veren bir belge tutma sistemi (recordkeeping system)” olarak oldukça detaylı bir şekilde tanımlamaktadır (Government Records Office Archives of Manitoba, s. 7, 2023). Bu tanımda diğer incelenen tanımlardan farklı olarak Bilgi ve Belge Yöneticilerinin uzmanlık alanına giren ve EBYS'lerin temel fonksiyonlarından birini oluşturan “bilgi erişim”den bahsedilmektedir. Özellikle tanımda geçen “...belgelerin gerekli yapısını, içeriğini ve bağlamını sağlayan bir sistemde yakalayan ve

düzenleyen...” ve “...ilgili belgeleri birbirine bağlayan...” ifadeleri, bilgi/belge/veri erişimi için önemlidir. Günümüzde EBYS’ler incelendiğinde bu sistemlerin bilgileri ne derece etkin yakaladığı, düzenlediği ve birbirleriyle ilişkilendirdiği tartışılır bir konudur. Bu sistemlerin bahsedilen noktalarda niteliğinin artırılmasını sağlayacak olan bilgi ve belge yöneticileridir.

Belgelerin/bilgilerin/verilerin sistemde yakalanması/düzenlenmesi/ilişkilendirilmesi için;

- bu belgelerin ontoloji tabanlı konu metadatalarının oluşturulması,
- kurumun fonksiyonlarına uygun dosya planının oluşturulması,
- bu plan uygulanırken sınıflama (classification) yönteminin yanında, birbirine benzer bilgi/belge/verilerin kaçırılmamasını sağlayacak kümeleme (clustering) yönetiminin de beraber kullanılması,
- sistem içerisinde yapılacak taramaların sözdizimsel tarama yerine ontoloji tabanlı gerçekleştirilmesini sağlamak bazı yararlı uygulamalardır.

EBYS uygulamaları, organizasyonlardaki değişimle ilgili olup, uzun bir süreci oluşturmaktadır (Borglund ve diğerleri, 2009a, s. 60). Günümüzde değişim kaçınılmazdır, bu değişimi başarılı bir şekilde yönetebilmek ise, organizasyonların örgütsel değişim yeteneklerine bağlıdır. Bu noktada üst yönetim desteği ve personelin farkındalığı EBYS’lerin başarısında önemli bir role sahiptir (Svård, 2017, s. 30).

Tüm bu bilgiler ışığında EBYS’lerin özelliklerini şu şekilde maddelemek mümkündür:

- EBYS’ler organizasyonların iş ve işlemleri sırasında ürettikleri/aldıkları belgelerin yönetilmesi sürecini içermektedir.

- EBYS’ler belgenin yaşam döngüsünü içeren tanımlama, düzenlenme, dağıtım, kullanım ve saklama eylemlerini gerçekleştirmektedir.
- EBYS’ler alanında uzman kişilerce yönetilmesi gereken sistemlerdir.
- Güvenlik noktasında önem arz eden sistem odaklı (siber güvenlik) ve insan odaklı (bilgi güvenliği) gereksinimler, EBYS bileşenlerini içermektedir.
- EBYS’ler kurumun fonksiyonları ve organizasyon yapısı hakkında bilgi vermektedir.
- EBYS’ler; kurumların, veri işleyenlerin, veri sorumlusunun eylemlerinin değiştirilemez şekilde bütünlük ve güvenliğini sağlayarak kayıt altına almakta ve iş etkinliklerini kanıtlayarak hem kurumun hem kurum çalışanlarının hem de verisi işlenen ilgili kişinin haklarını korumaktadır.
- EBYS’ler iç ve dış entegrasyonlara sahiptir. Bu entegrasyonlar sayesinde diğer kurumlarla birlikte çalışabilirlik sağlanarak e-devlet yapısına geçiş desteklenmektedir.
- EBYS’ler kurumlarda bir standart oluşturulmasını sağlayarak kurumda sürekliliği desteklemektedir.
- EBYS’ler verimli, kullanışlı ve daha az maliyete sahip belge yönetim sürecini sağlamaktadır.
- EBYS’ler ile kurumsal zekâ ve iş zekası desteklenerek fırsat, risk ve tehditler belirlenmektedir. Böylece kurumun karar verme mekanizmasını geliştirmek ve uygun kurumsal politikaların oluşturulması sağlanmaktadır.
- EBYS’ler işlemleri otomatikleştirilerek zamandan, emekten ve maliyetten tasarruf edilmesini sağlamaktadır.

- EBYS’ler kurumsal hafızayı oluşturarak yeni ürün/hizmet ve politikaların geliştirilmesine yardımcı olmaktadır. Bu sayede hem kuruma hem de ülkeye katma değer sağlamaktadır.

II. 1. 1. Türkiye’de EBYS: Yasal Düzenlemeler ve Uygulama Standartları

Bu bölümde, EBYS’ler ile ilgili başlıca yasal düzenlemeler kronolojik olarak ele alınmış ve literatürden saptanan uygulamadaki aksaklıklarla ilişkilendirilerek sorunlara dikkat çekilmeye çalışılmıştır.

Ülkemizde elektronik ortamda belge yönetim için 2004 yılında 5070 sayılı Elektronik İmza Kanunu’nun yürürlüğe girmesiyle birlikte ilk adım atılmış, kuruluşlarda EBYS’lerin kurulması ile ilgili çalışmalar başlatılmıştır (Menteş ve Solhan, 2019, s. 13). Elektronik İmza Kanunu’nda hukukî olarak güvenli elektronik imzanın, ıslak imza ile eşdeğer olduğu açık bir şekilde belirtilmektedir. Bunun yanı sıra, EBYS’ler ile entegre çalışan e-imzaların; belgelerin ve verilerin bütünlüğünün korunmasında, imzayı atan kişinin eyleminde ispat teşkil etmesinde ve imzayı atan kişinin kimliğinin belirlenebilmesinde önemli bir işlevi olduğu belirtilmektedir. Bu yönü ile e-imza organizasyonlarda, sistemlerin ve eylemlerin hızlı ve güvenilir bir işlenmesini sağlamaktadır. Bu EBYS’ler için önemli bir husustur. Çünkü hızlılık ve güvenlik EBYS’lerin önemli hedeflerini oluşturmaktadır. e-İmza ile birlikte sorumluluk üstlenme, inkar edilemezlik, işlemlerin hızlanması gibi noktalarda verimlilik arttırılmış bulunmaktadır. Solhan (2016, s. 55) çalışmasında e-imzalar ile ilgili olarak, kamu kurumlarında sıklıkla mevzuat dışı uygulamalara rastlanabildiğini belirtmektedir. Bu uygulamalara örnek olarak; kurumlarda elektronik ortamda üretilen belgelerin basılı kopyalarının alınarak fiziksel ortamda saklanması verilmektedir. Bu uygulama da

görülmektedir ki 2004 yılında yayımlanan Elektronik İmza Kanunu, uzun yıllar anlaşılamamış ya da benimsenememiştir. Kurumlarda yapılan bu yanlış uygulamalar, kurumları sadece zaman, depolama ve kâğıt israfı konularında zarara uğratmaktadır. Elektronik İmza Kanunu EBYS'ler açısından son derece öneme sahip olup, kurumlara elektronik imzanın hukukî ve teknik yönleri ile ilgili yol gösterici şekilde dikkate alınması ve uyulması gereken bir kaynaktır.

Şekil 1.

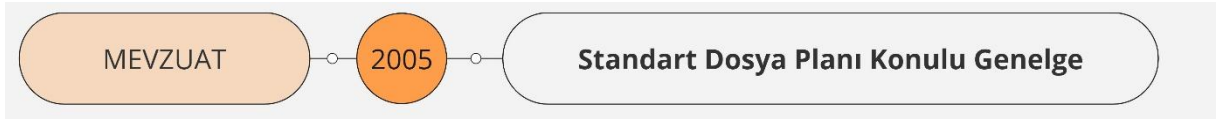
2004 yılına ait mevzuat



2005 yılında ise Standart Dosya Planı Konulu 2005/7 Genelgesi ile hem elektronik hem de fiziksel ortamda kurum içi ve kurumlararası bir standart sağlamak amacıyla kurumlara dosyalama sistemi sunulmuştur (Standart Dosya Planı..., 2005). Bu düşünceyle kurumların hem kendi içlerinde ürettiği belgeler hem de diğer kurumlar tarafından edinilen belgeler belli bir standarda dayanmakta; yapılacak işlemler hızlı, etkin ve etkili bir şekilde yürütülmekte; belgeye, bilgiye, veriye erişimde yüksek verimlilik sağlanmaktadır. Belgelerin düzenlenmesinde kişisel değil kurumsal teknik ve yöntemler kullanılmaktadır. Böylece çalışanın inisiyatifine bağlı kalmaksızın belli bir standart ve düzen çerçevesinde belgelerin düzenlenmesi ve erişimi sağlanmaktadır. Bu, çalışanın kurumdan ayrılması durumunda, iş süreçlerinin aksamadan ilerlemesini, dolayısı ile de sürekliliği sağlamaktadır. Fakat süreklilik ve standart sağlama açısından önem arz eden Standart Dosya Planının; EBYS'ler için uygunluğu ve kurumların bu planı EBYS'lerde ne kadar etkin ve verimli kullanabildiği tartışmalı bir konu olup, konu ile ilgili sorun ve çözüm önerileri 4. Bölümde ele alınmaktadır.

Şekil 2.

2005 yılına ait mevzuat



2011 yılında 28036 sayılı Kayıtlı Elektronik Posta Sistemine İlişkin Usul ve Esaslar Hakkında Yönetmelik yayımlanmıştır. Yönetmeliğin amacı; “Kayıtlı Elektronik Posta (KEP) sisteminin hukukî ve teknik yönleri ile işleyişine ilişkin usul ve esasları düzenlemek”tir (KEP, 2011). Yönetmelik KEP ile ilgili; hesap başvurusu, hesabın kullanıma açılması ve kapatılması sistemin kullanımı gibi sistem içerisinde yapılabilecek işlemler hakkında bilgiler vermektedir. KEP ile elektronik iletilerin gönderimi ve teslimatı sırasında yapılan işlemler, hukuki delil sağlamakta ve KEP kullanıcılarının eylemleri ispat edilerek olası durumlarda kişilerin haklarının korunması sağlanmaktadır. Böylelikle gönderilen bir iletinin gönderildiği, alındığı, içeriğinin değişmediği ve belirli bir tarihte iletiliği ispatlanabilir hale gelmiştir. Bu anlamda KEP, EBYS’lerin güvenlik boyutunda önemli yer tutmaktadır. Bunun yanı sıra kurum içi ve kurumlararası belge, bilgi, veri paylaşımında birlikte çalışılabilirlik sağlanarak EBYS’ler içinde hızlı, etkin, etkili, güvenilir ve verimli bir bilgi paylaşım süreci yaşanmasına hizmet etmektedir.

Şekil 3.

2011 yılına ait mevzuat



e- Yazışma Projesi konulu 2017/21 sayılı genelge yine EBYS’ler için öneme sahip bir kaynaktır. Bu genelgedeki en dikkat çeken nokta

“... kamu kurum ve kuruluşları arasındaki resmi yazışmaların elektronik ortamda yürütülmesini amaçlayan e- Yazışma Projesi kapsamında, kamu kurum ve kuruluşlarının 31/7/2017 tarihine kadar Elektronik Belge Yönetim Sistemine (EBYS) geçmesi talimatlandırılmış...”

ibaresidir (e- Yazışma Projesi..., 2017) Bu adımla birlikte artık EBYS’lerin yaygınlaştırılması ve geliştirilmesi hedeflenerek, iyi yapılandırılmış bir e-devlet sistemine geçişin hedeflendiği anlaşılmaktadır. Genelgede 2/2/2015 tarihli ve 29255 sayılı Resmi Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmeliğe ve e- Yazışma Teknik Rehberine atıfta bulunmaktadır. 2015 yılında yayımlanan 29255 sayılı yönetmelikte bilgi, belge ve doküman alışverişini hızlı ve güvenli yürütebilmek amacıyla fiziksel ve elektronik ortamda imzalanarak yapılan resmi yazışmalara ilişkin kurallar sunulmaktadır (Resmi Yazışmalarda Uygulanacak..., 2015). Genelgede ise “idareler arasında resmi yazışma kapsamında iletilecek e-belgelerin oluşturulmasında e-Yazışma Teknik Rehberinde tanımlanan kurallara uyulmasının zorunlu” olduğu belirtilmektedir. Buradan da Resmi Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik ile e- Yazışma Teknik Rehberi’nin birbirini tamamlayan yasal enstrümanlar olduğu ve söz konusu genelgenin EBYS’lerin yaygınlaşmasında, kullanımında ve ulusal bir standardının oluşturulmasında itici bir rol oynadığı söylenebilir. Yönetmeliğin güncel versiyonu 10/6/2020 tarihli ve 31151 sayılı Resmî Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik’tir. Yeni yönetmelikte ilk göze çarpan, resmi yazışmaların elektronik ortamda gerçekleşeceği ve sadece zorunlu/olağanüstü durumlarda el yazısı ile imzalanan belgeler ile gerçekleştirileceği bilgisidir (Resmi Yazışmalarda Uygulanacak..., 2020). 2015 yılındaki yönetmelik ile karşılaştırıldığında artık elektronik ortamın istisnai durumlar hariç tüm resmi yazışmalar için tek ortam olarak kabul edildiği görülmektedir. 2015’te yayımlanan yönetmelikte “fiziksel ve elektronik ortamda

imzalanarak yapılan” resmi yazışmalardan bahsedilirken, 2020 yılında daha net bir ifadeyle resmi yazışma ortamının elektronik ortam olduğu ortaya konmaktadır. Bu yönetmelik, resmi yazışmalara ilişkin kurallar belirlemek ile kalmayıp, güvenli ve verimli bir yazışma süreci ortaya koymak amacıyla yapılan yenilik ve düzenlemeleri de açıklamaktadır. Buna ek olarak yine söz konusu yönetmelikte de e- Yazışma Rehberinin üzerinde önemle durulmakta ve resmi yazışmaların e- Yazışma Rehberine uygun şekilde hazırlanması gerektiğinden bahsedilmektedir. e- Yazışma Rehberinin günümüzdeki son sürümü 2.0’dır. Söz konusu rehber yeni ihtiyaçlara yanıt vermek ve gelişmeleri yansıtmak amacıyla Dijital Dönüşüm Ofisi tarafından güncellenmektedir. “Bu rehber kamu kurum ve kuruluşları arasında elektronik ortamda yapılacak resmi yazışmalarda kullanılacak e- Yazışma Paketinin mantıksal yapısına ve teknik mimarisine ilişkin bilgiler içermektedir” (e-Yazışma Teknik Rehberi 2.0, 2020). Rehberde gerçekleştirilmesi gereken entegrasyonlardan (Devlet Teşkilatı Merkezi Kayıt Sistemi (DETSİS), Hizmet Envanteri Yönetim Sistemi, KEP, e- Yazışma Uygulama Arayüzü (API)), e-Yazışma Paketi ve bileşenlerinden (Belgeye İlişkin Bileşenler ve Paket Yapısına İlişkin Bileşenler) ve son olarak Diyagram Açıklamalarından detaylı bir şekilde bahsedilmektedir. Bu rehber EBYS’ler bağlamında, kurumlararası entegrasyonu dolayısı ile birlikte çalışabilirliği sağlayarak; e-devlet düzeyinde hızlı, verimli, güvenli, etkin bir belge yönetimi sürecini yaşatmayı hedeflemektedir. E-Devlete geçişte önemli yer tutan rehberin hedef kitlesi, tüm kamu kurum ve kuruluşları ile kamuya EBYS çözümü sunan özel sektör firmalarıdır. Bu nedenle söz konusu rehber sadece bilgi verici değil aynı zamanda yazılım geliştirme noktasında da verdiği detaylı teknik bilgi ile sistemin geliştirilmesini desteklemektedir.

Şekil 4.

2015-2020 Yıllarına Dair Mevzuat ve Rehber



25/4/2022 tarihli ve 5529 sayılı Gizlilik Dereceli Belgelerde Uygulanacak Usul ve Esaslar Hakkında Yönetmelik EBYS’lerin kullanımının düzenlenmesine yönelik esasları belirten bir yönetmeliktir.

Yönetmeliğin amacı, “muhteviyatına veya güncelliğine göre gizlilik dereceli belgeler için uygulanacak usul ve esasları belirlemek ve gizlilik dereceli belgelerde uygulama birliğini sağlamaktır”. Yönetmeliğin 3d) maddesi bilmesi gereken prensibi, erişim yetkisinin belirlenmesi bakımından önemlidir. Buna göre ilke; “bir belgeyi sadece görev ve sorumlulukları gereği öğrenmekle, incelemekle, gereğini yerine getirmekle ve korumakla sorumlu bulunanların yetkileri düzeyinde bilgi sahibi olması ve nüfuz etmesi” olarak tanımlanmıştır. Gizlilik esasları ise 4.madde ile düzenlenmiştir. Buna göre;

Çok Gizli, Gizli ve Hizmete Özel olmak üzere üç adet millî gizlilik derecesi bulunur:

a) Çok Gizli: Açıklanması veya yetkisiz kişilerce öğrenilmesi hâlinde Devletin dış ilişkilerine, millî savunmasına, millî güvenliğine ve müttefiklerle olan faaliyetlerine önemli derecede zarar verebilecek belgeler için kullanılır.

b) Gizli: İzinsiz açıklanması veya yetkisiz kişilerce öğrenilmesi hâlinde Devletin menfaatlerine, güvenlik, istihbarat ve teknoloji faaliyetlerine zarar verebilecek belgeler için kullanılır.

c) Hizmete Özel: İzinsiz açıklanması veya yetkisiz kişilerce öğrenilmesi hâlinde herhangi bir idari faaliyete, gerçek veya tüzel kişiye, idari soruşturmaya, adli soruşturmaya ve kovuşturmaya zarar verebilecek belgeler için kullanılır.

Yönetmelikte belgelerin gizlilik derecelerinin; hangi içeriğe uygun olarak verildiği açıklanmakta ve yanı sıra gizlilik derecelerinin nasıl kullanılacağı da belirtilmektedir. Yönetmeliğin 6.maddesi gizlilik derecelerinin süresini düzenlemektedir. Örneğin. *Çok Gizli belgeye gizlilik derecesinin verilmesi sırasında gizlilik derecesinin düşürüleceği, kaldırılacağı veya belgenin imha edileceği tarih ya da olay belge üzerinde belirtilebilir.*

Gizlilik derecesinin düşürülmesi, devam ettirilmesi, kaldırılması ve belgenin imhası kararlarında neyin esas alınacağı bir değerlendirme komisyonu tarafından gerçekleştirilmektedir. Yönetmeliğin 7. maddesi bu konuyu açıklamaktadır. Örneğin: *Çok Gizli belgenin gizlilik derecesinin aynı şekilde devam etmesi, düşürülmesi, kaldırılması ile Çok Gizli belgenin imha edilmesi kararı “Çok Gizli Gizlilik Dereceli Belgeleri Değerlendirme Komisyonu” tarafından verilir.*

Bunlar dışında söz konusu yönetmelikte gizlilik dereceli belgelere ilişkin fiziksel ve teknik güvenlik tedbirleri ile bu belgelere ilişkin işlemler açıklanmaktadır (Gizlilik Dereceli..., 2022). Burada EBYS’ler ile ilgili en önemli nokta hizmete özel gizlilik dereceli belgelerdir. Çünkü gizlilik derecelerinden sadece hizmete özel gizlilik dereceli belgeler EBYS’lerde işlem görmektedir. Ayrıca yönetmeliği desteklemek amacıyla Gizlilik Dereceli Belgelerde Uygulanacak Usul ve Esaslar Kılavuzu başlıklı bir kaynak

da hazırlanmıştır. Kılavuzda gizlilik dereceleri ile ilgili ülkemizdeki mevzuat çalışmalarına kısaca değinilmiş neden yeni bir mevzuata ihtiyaç duyulduğu açıklanmış ve yönetmelikte yer alan hükümler oldukça detaylı biçimde ve görsel öğelerle desteklenerek, ilgili birim ve yöneticilere yardımcı olmak amaçlanmıştır. Kılavuz; gizlilik derecelerinin önemini, süreçlere katkılarını, yanlış kullanımına ilişkin örnekleri ve bunun sonucunda oluşabilecek durumları açıklamaktadır. Gizlilik derecelerinin uygulanması sürecinde, EBYS’lerin nasıl yapılandırılması gerektiği, nelerin göz önüne alınarak geliştirilmesi gerektiği detaylı belirtilmektedir (Gizlilik Dereceli Belgelerde..., 2022) Bu bağlamda gizlilik dereceleri ile ilgili hazırlanmış yönetmelik ve kılavuz; EBYS’ler için son derece önemli olan yetkili erişimin sağlanması ve belgelerin gizlilik ve bütünlüğünün korunması bakımından önemli enstrümanlardır.

Şekil 5.

2022 Yılına Dair Mevzuat ve Rehber



II. 1. 2. EBYS Bileşenleri

Günümüzde bilgi; kişiler, organizasyonlar ve devletler için bir güç haline gelmekte ve bilgi toplumunun ham maddesi olarak tüm işler bilgi üzerinden yürütülmektedir. Bilginin sürekli artışı ve yaratmış olduğu değer, ulusal ve uluslararası düzeyde rekabet seviyesini arttırmaktadır. Bilginin verimli kullanımı; kişilere, organizasyonlara, devletlere ekonomik bir değer yaratmaktadır. “Ancak verilerin değerli

olması için kaydedilmesi gerekmektedir” (Mukred ve diğerleri, 2022). Bu değeri ve yaratılan rekabet ortamında bilgi üstünlüğünü sağlamak için ise etkin bir bilgi yönetimi gerekmektedir (Önaçan ve Özkanlı, 2012, s. 2).

Bilgi teknolojilerinin sürekli olarak gelişimi ile veri/bilgi/belge üretimi kontrol edilemez şekilde artmakta ve fiziksel ortamda veri/bilgi/belge yönetimini olanaksız kılmaktadır. Kuruluşlar için doğru ve zamanında bilgi, güç haline gelmektedir (Mukred ve diğerleri, 2022). Bilginin güç olduğu ve bu güce sahip olma noktasında bilgi yönetiminin hayati öneme sahip olduğu günümüzde; iş akışı yönetimi, raporlama, entegrasyon, güvenlik, yazılım, sürdürülebilirlik, bütünlük, kullanılabilirlik bileşenlerini içeren iyi organize edilmiş ve tasarlanmış bir EBYS kullanımı zorunlu hale gelmektedir. Bu bölümde EBYS tanımının yanı sıra EBYS işlevleri ve gerekliliklerini daha iyi ortaya koyabilmek adına EBYS bileşenleri işlenmekte olup, bileşenlerin temel olarak neden EBYS için neden önemli olduğu açıklanmaktadır. EBYS’lerin nasıl geliştirilmesi gerektiği, eksiklikleri ve çözümü 4. Bölüm olan EBYS’lerde Metin Madenciliği başlığında açıklanmaktadır.

Bir kuruluştaki bilgi ve belge yönetimi işlevi, o kurumdaki personelden, hizmeti alan kişiye kadar büyük bir çoğunluğu kapsamaktadır. Kuruluşun iş süreçleri ve faaliyetleri sonucunda, birçok personel veri işleyen ve belge yaratıcısı olarak karşımıza çıkmaktadır. Bu noktada veri/bilgi/belgelerin gerçek, güvenilir ve eksiksiz bir şekilde, en üst kalitede elde edilmesi kuruluşun alacağı kararlar ve sunacağı hizmetlerin kalitesini de doğru orantılı bir şekilde oluşturacak/etkileyecektir (Svård, 2017, s. 15). Bu nedenle kuruluştaki veri işleyen konumundaki personellere büyük bir görev düşmektedir. Bulunulan rekabet ortamında bilginin önemini; erişimi, bütünlüğü ve doğruluğu koruyarak güvenliğini sağlamak oluşturmaktadır. Veri/bilgi/belgelerin eksiksiz, kanıtlanabilir,

güvenilir, değiştirilemez olabilmesi ve gerçekliğini koruyabilmesi için organizasyonel (idari) ve sistemsel (teknik) önlemler alınması gerekmektedir. Bu noktada da karşımıza önemli EBYS bileşenlerinden **bilgi güvenliği** ve **siber güvenlik** kavramları çıkmaktadır. Güvenlik noktasında bir EBYS yazılımı kurulması veri/bilgi/belgelerin güvenliği için yeterli bir önlemleri oluşturmamaktadır. Nitekim bilgi güvenliğinde en zayıf halkayı insan oluşturmakta olup açıkların büyük kısmı insan hatalarından ortaya çıkmaktadır (Önaçan, Medeni, Özkanlı, 2012, s. 16). Organizasyonel bağlamda insan faktörü zayıf belge yönetimi davranışına neden olmaktadır (Borglund ve diğerleri, 2009a s. 60). Bu nedenle veri gizliliği ve güvenliğinin sağlanabilmesi için; belge yönetimi dolayısı ile de EBYS konularında personelin eğitim sürecinden geçmesi önemlidir. Bu farkındalığı sağlayabilmek, kuruluşlarda bilgi güvenliği noktasında risklerin kabul edilebilir seviyeye çekilmesine yardımcı olmaktadır (Kantürk, Şahinaslan ve Borandağ, s. 598 aktaran Aydın, Saydam, 2019, s. 5). Bu süreçlerin gerçekleşebilmesi ve EBYS'nin verimli kullanımı için ise üst yönetimin süreci desteklemesi gerekmektedir (Borglund ve diğerleri, 2009a, s. 60).

Bilgi güvenliği konusuna değinirken ISO 27001 Bilgi Güvenliği standardından bahsetmek konunun önemini ortaya koymakta önem taşımaktadır. Bu standart EBYS özelinde hazırlanmamış olmasına rağmen EBYS için önem arz etmektedir. Bilgi Güvenliği Yönetim Sisteminin (BGYS) temel amacını, bilginin; gizlilik, bütünlük ve erişilebilirliğinin sağlanması oluşturmaktadır. Bu amaç, EBYS amaçları ile örtüşmekte olup, EBYS'lerde bilgi güvenliğinin sağlanması EBYS'nin ve kurumun sürekliliği için önem arz etmektedir. Bu nedenle EBYS'nin anlamlı olabilmesi en başta bilgi güvenliğinden geçmektedir. Bilgi güvenliği sağlandığı takdirde belge yönetiminde yapılan faaliyetler anlam kazanmakta ve doğru, güvenilir bilgiye erişim sağlanabilmektedir. Tabii bilgiye erişim noktasında da dikkat edilmesi gereken önemli

ayrıntılar gizlidir. Bilgiye erişimin sağlanması önemlidir fakat herkes bu bilgiye erişmemelidir. Kurumsal mahremiyet, kişisel veriler ve gizlilik noktalarında EBYS üzerinde yetkili erişim sağlanması gerekmektedir. 27001 standardında erişim kontrolü, kullanıcı erişim yönetimi, sistem ve uygulama erişim kontrolü başlıklarında detaylı şekilde yetkili erişimin üzerinde durulmakta olup diğer başlıklarda da konuyla ilişki kurulmaktadır. 27001 bilgi güvenliği standardında aşağıda belirtilen hususları kapsayacak şekilde oluşturulacak kurumsal politikalar, bilgi güvenliği sistemi oluşturma süreçlerinde (kurulma, uygulama, güncelliğini ve kontrollerini sağlama) kurumsal sürekliliği sağlayacak bir mimari ortaya koymaktadır. Bir bilgi yönetim sisteminde genel hatlarıyla aşağıdaki akışlar yer almaktadır: (ISO 27001 Bilgi Güvenliği Standardı)

- kurumun sahip olduğu bilgi varlıklarını belirlemek,
- bir varlık envanteri hazırlamak,
- kurumun sahip olduğu bilginin önem derecesini belirleyerek sınıflandırma ve etiketleme yapmak,
- ortamda depolanan bilginin yetkisiz erişimlerden engelleyerek bütünlüğünü korumak,
- kriptografik kontroller ile bilginin gizliliği ve bütünlüğü için önemler almak,
- fiziksel olarak yetkisiz erişimi engelleyecek güvenli alanlar yaratmak,
- varlıklara zarar gelmesi durumunda (kaybolması, hasar görmesi, çalınması vb.) kurumun iş ve işlemlerini kesintisiz yürütebilmek,
- gelebilecek zararlara karşı işletim, yazılım, ağ kontrollerini sağlamak,
- veri kaybına karşı yedekleme, hesap verilebilirlik noktasında ise olayları kaydederek kanıt üretmek,
- sistemi sürekli olarak güncel tutmak, düzenli olarak bakımını sağlamak ve iyileştirmek.

Gör÷lmektedir ki bgys ile EBYS arasında önemli bir ilişki bulunmaktadır. Bu ilişki bilgi güvenliğinin gizlilik, bütünlük, doğruluk, kanıtlanabilirlik, erişilebilirlik amaçları ile ortaya çıkmaktadır. Nitekim etkin bir bilgi yönetimi için en başta EBYS'nin bu özelliklere sahip olması gerekmektedir. Bunun sebebi ise; ancak veriler doğru ise kurumlar doğru kararlar alabilmekte ve yetkili erişim sağlandıysa bütünlük korunabilmektedir. Bu da aynı şekilde hem doğru kararlar alabilmeyi hem de kurumun alacağı stratejilerde kurumsal mahremiyeti sağlayabilmektedir. Tüm bunların sonucunda ise (güvenlik, doğruluk, bütünlük sağlandığı noktada) kişi ve kurumlar, faaliyetleri sonucu eylemlerini kanıtlayarak, hesap verebilirlik özelliğini sağlayabilmektedir.

Luciana Duranti bir EBYS'de güvenilirlik ve özgünlüğün temel ilke olduğunu belirterek, güvenliğin EBYS için olmazsa olmaz bir bileşen olduğunu desteklemektedir (Meijer, 2001 citing Duranti, 1998 aktaran Parrish and Courtney, 2007, s. 225). Çalışmasında Yalçinkaya (2016, s. 83-84), bilgi güvenliğinin EBYS için önemini ele alırken; EBYS'lerin bilgi güvenliği ve gizlilik noktalarında kritik bir konumda olduğunu, hukuki mevzuatla uyumlu bir şekilde oluşturulan EBYS'nin ise güvenilirliği artırarak sahtecilik riskinin azalacağını, bunun da tüm kademelerde demokrasi, doğruluk ve netlik noktalarında hissedileceği belirtmektedir. EBYS'lerin güvenlik kapsamında, sahip olduğu süreçlerin sadece yetki verilen kullanıcılar tarafından yönetilmesi ve sistem erişiminin, elektronik sertifikalar (e-imza gibi) kullanılarak gerçekleştirilebilmesi erişilebilirlik, gizlilik ve doğrulukta önemli tedbirler yaratmaktadır (Eroğlu, Külcü, s. 332, 2014). Umut ve Külcü (2014) ise yine çalışmalarında belgelerin bütünlüğünün korunması ile ilgili yapılan düzenleme, aktarma, depolama, ayıklama ve imha işlemlerinin eby kapsamındaki başlıca önemli işlemler olduğunu belirtmektedir (Odabaş, 2008, s. 130; Odabaş, 2009, s.415; aktaran Umut, Külcü, 2014, s.104). Belge

yönetiminin sağladığı önemli faydalardan bütünlüğün korunması özelliği, veri kaybı ve tutarsızlıkların önüne geçerek kurumlar için, iş süreç ve faaliyetlerinde birer zorunluluk haline gelmektedir (Hase ve Galt, 2011 aktaran Demirtel, Gökkurt Bayram, 2014, s. 93). Bilgi yönetimi ile ilgili güvenlik önemleri alınmadan yapılandırılacak arşivleme ve güvenli erişime ilişkin teknolojik altyapının, EBYS'ler için sağlıklı bir süreç yaratacağı düşünülmemektedir (Aydın, Özdemirci, 2011, s. 106). Güvenlik tedbirleri alınmadan tasarlanmış bir politika ve sistem; belge yönetimini ve kurumun sürekliliğini sekteye uğratmaktadır. Belgelerin doğruluğu sağlanamaz ve erişim yetkileri ortaya konamazsa, kurum doğru kararlar alamayacak ve bilgilerin yetkisiz kişilerin eline geçme riski artacaktır. Bilgilerin yetkisiz erişime maruz kalması ise, organizasyonların, çekişmeli rekabet ortamında varlığını kaybetmesine neden olmaktadır.

Güvenliğin sağlanmasında bir diğer önemli husus ise, kurumun ve/veya kişilerin eylemlerini ortaya koyarak faaliyetlerin kanıtlanması ve hesap verilebilirliğin sağlanmasıdır. Belgeler kurumsal faaliyet ve işlemlerin kanıtını oluşturmakta olup belgelerin orijinalliğini, doğruluğunu, bütünlüğünü ve güvenilirliğini korumak için doğru şekilde yönetilmesi gerekmektedir (Uluslararası Standartlar Organizasyonu, 2001a; Reed, 2005 aktaran Borglund ve diğerleri, 2009a, s. 51). 2007 yılında yapılan bir çalışmada (Volonino, 2003 aktaran Parrish and Courtney, s. 225 ,2007) elektronik belgelerin kanıt özelliğine dikkat çekerek, bilgi ve iletişim sistemlerinin artık denetimlerde, soruşturmalarda veya davalarda elektronik kanıt için üreme alanı olduğu belirtilmektedir.

Bilgi güvenliği, güvenlik hususunda daha çok insan kaynaklı oluşabilecek risklerin (kişisel veri, hassas veri, kurumsal bilgiler vb.) önlenmesi ile ilgiliyken, siber güvenlik ise donanımsal, sistemsel riskleri önlemeye yönelik tedbirleri kapsamaktadır.

Kurumun sađlıklı veri/bilgi/belge sürecini yönetmesi için her iki güvenlik aşamalarının eksiksiz yerine getirmesi önemlidir. Günümüzde savaşlar artık (en azından bilgi toplumu seviyesinde olanlar için) veri/bilgi/belgeleri ele geçirerek gerçekleşmektedir. Bir devlete, kuruma, kişiye zarar vermek için o hedefin vurulması gerekmemektedir (Tamtürk, s.862, 2017). Bu nedenle riskleri tespit etmek ve kabul edilebilir seviyeye indirmek için alınması gereken tedbirler önemlidir. Organizasyonel bağlamda alınabilecek tedbir ve bunların önemi bilgi güvenliği bileşeninde ele alınmıştı. Sistemsel bağlamda alınabilecek önlemler ise EBYS'lerin bir diğer bileşeni olan siber güvenlik konusu çerçevesinde açıklanmaktadır. Daha önce de belirtildiği gibi bilgi güçtür ve artık ülkeler, organizasyonlar gücünü ortaya koyabilmek için sınırları değil bilgiyi tutan kaynağı ele geçirmeye çalışmaktadır. Bu nedenledir ki veri/bilgi/belgelerin üretildiği ve depolandığı yazılım ve donanımın korunması çok önemli olup, gelişen bilgi teknolojilerine de uyum sağlayacak şekilde güncel kalması gerekmektedir (Aydın, 2005, s.94). Aydın (2005) e-belgelerin depolanmasında birinci ve en önemli hususun donanım olduğunu aynı zamanda yazılımın yaşam döngüsünün çoğunlukla iki yıldan az olduğunu belirtmektedir. Kuruluşun hangi yazılımı seçeceği önemli bir aşamadır. Her kuruluşun farklı iş süreçleri, faaliyetleri ve beklentileri bulunmaktadır. Bu nedenle, bu özelliklerin ortaya konarak kuruluşun yapısına uygun bir yazılım seçmesi önemlidir. Nitekim yazılım seçme noktasında doğabilecek riskler nedeniyle, içlerinde Türkiye'nin de yer aldığı bazı ülkeler, bu süreçte EBYS sertifikalandırmaları gerekmektedir (Katuu, 2012, s. 460-464; aktaran Yalçınkaya, 2016, s. 77). Kullanım aşamasında ise bu sistemlerin bakımı ve güncellemesi yapılmadığı takdirde, süreçler işlenemez hale gelerek, belgelere erişim sağlanamamakta hatta belgeler yetkisiz kişilerin eline geçebilmektedir. Yalçınkaya (2016, s. 73) altyapının EBYS için çok önemli olduğu ve organizasyonların EBYS ile ilgili planlama yaparken teknolojik altyapıyı iyi bir şekilde değerlendirmeleri gerektiğini belirtmektedir. Bu

noktada ise; sunucu, ağ yapısı, donanım vb.'lerin analiz edilmesinin EBYS başarısını etkilemede temel kaynak olduğunu eklemektedir. Bu gelişen bilgi teknolojileri karşısında bilgi güvenliğini en verimli şekilde tutabilmek için; yazılım-donanım güncellemeleri, depolamadaki kısıtlamalar ve bilginin potansiyel değerini içeren koruma planına sahip olması önemlidir (Minnesota Historical Society, 2004, s. 2; aktaran Aydın, Özdemirci, 2011, s. 109).

Siber güvenlikle ilgili ülkemizde ulusal çapta alınacak tedbirleri içeren 2019/12 Sayılı Cumhurbaşkanlığı Bilgi ve İletişim Güvenliği Tedbirleri Genelgesi ve bu genelgeyi esas alarak oluşturulan Dijital Dönüşüm Ofisi'nin 2020 yılında yayımlamış olduğu Bilgi ve İletişim Güvenliği Rehberi bulunmaktadır. Söz konusu genelgede güvenlik açığı bulunması durumunda, yasal sorumlulukları doğurabilecek, yetkisiz şekilde erişilen içeriğin, organizasyonlara maddi ve/veya manevi çok büyük zarar vereceği veri/bilgi, şeklinde tanımlanan kritik verinin, milli yazılım ve donanımlar ile korunarak bu verilere gizlilik, bütünlük ve erişilebilirliğin en iyi seviyede sağlanması amaçlanmaktadır. Bu da güvenli bir ağda, erişimin kontrollü sağlanması ve log kayıtlarının değiştirilemez yapıda kullanılmasını gerektirmektedir. Genelgede bu konuları sağlayacak tedbirler ele alınmaktadır. Bu tedbirler genel olarak aşağıdaki gibi sıralanmaktadır;

- gizlilik dereceli bilgi ve belgelerin iletiminde yerli ve milli sistemlerin kullanılması,
- bu haberleşmelerin kişisel veya yetkilendirilmemiş hiçbir cihaz ve ortamda gerçekleştirilmemesi,
- güvenli yazılımların geliştirilmesi ve üretici veya tedarikçiden; kamu kurum ve kuruluşlarında kullanılacak yazılım ve donanımın kullanım amacına uymayan hiçbir güvenlik açığının olmadığına dair taahhütname alınması,

- kuruluşların aldıkları siber bildirim sonucu önlemler alması,
- kritik verilerin iletiminde radyolink haberleşmesi kullanılmaması ancak mecburi durum söz konusu olduğunda milli kripto sistemlerinin kullanılmasıdır.

Bilgi ve İletişim Güvenliği Rehberi ise genelgeyi temel alan fakat genelgenin oldukça detaylandırılmış tedbir ve tedbirleri uygulamaya yönelik uygulamalarını içeren bir çalışmadır. Rehber bilgi güvenliği ile ilgili ulusal ve uluslararası standart ve mevzuatı da esas alarak kamu; kurum ve kuruluşları ile kritik altyapı hizmeti veren işletmelere yönelik kritik bilgi/verilerin gizlilik, bütünlük ve erişilebilirliğini koruyacak tedbirler ile ilgili yapılması gereken faaliyetleri ortaya koymaktadır.

Bilgi güvenliğinde sıkça geçirdiğimiz orijinallik, güvenilirlik, bütünlük ve bunlara ek olarak kullanılabilirlik, bir belgenin belge niteliği kazanabilmesi için gerekli özellikleri oluşturmaktadır. Belgenin sahip olması gereken bu özellikleri International Council on Archives'den aktaran Svärd (2017, s. 16) aşağıdaki gibi tanımlamıştır.

- **Orijinallik:** Belgenin iddia edildiği gibi olduğu, onu yaratan veya gönderen kişi tarafından yaratıldığı veya gönderildiği ve gerçekleştiği iddia edilen zamanda yaratıldığı veya gönderildiğini ifade etmektedir.
- **Güvenilirlik:** Belgelerin, tasdik ettikleri işlem(ler)in tam ve doğru bir temsili olduğunu ifade etmektedir.
- **Bütünlük:** Belgenin eksiksiz, değiştirilmemiş ve yetkisiz değişikliklere karşı korunmuş olduğunu ifade etmektedir. Bu özellik “dokunulmazlık” olarak da adlandırılmaktadır.
- **Kullanılabilirlik:** Belgenin; bulunabilir, erişilebilir, korunabilir ve yeniden kullanılabilir olduğunu ifade etmektedir.

Aslında söylenebilir ki, EBYS bileşenleri olarak ortaya konan özellikler en başta bir belgenin belge olabilmesi için sahip olması gereken niteliklerin korunmasını sağlamaktadır. Tüm EBYS bileşenleri bir bütünsel bir yaklaşımla ortaya konulduğunda belgenin orijinalliği, güvenilirliği ve bütünlüğü sağlanmış olacaktır.

Bu üç özelliğin yanında sistemin kurulması aşamasında önemli bir bileşen de kullanılabilirliktir. Sistemde her önlem harika bir şekilde alınmış ve planlanmış olsa bile o sistemin verimini ve kurumsal işleyişi etkileyen önemli bir nokta kullanılabilir bir sistem oluşturulmasıdır. “Kullanılabilir bir sistem; en az çaba, girişim ve zaman harcanarak kullanıcıyı amacına ulaştıran verimli, etkili/yararlı, kolay öğrenilen, kullanımı kolay hatırlanan, hataları önleyen, hata toleransı yüksek ve kullanımı kolay sistemlerdir” (Alfred, Brusaw ve Oliu, 2000; Bağış, 2002; Battleson, Booth ve Weintrop, 2001; Benyon, Turner ve Turner, 2006; Bevan, 1995; Booth, 1989; Campbell ve Aucoin, 2003; Dee ve Allen, 2006; Doll ve Torkzadeh, 1991; Head, 1999; Hix ve Hartson, 1993; IEEE Std.610.12, 1990; ISO 9241-11, 2018; Kanis, 1998; Lecerof ve Paterno, 1998; Logan, 1994; Nielsen, 1993; Norman, 2013; Preece, 1993; Quesenbery, 2001; Shackel ve Richardson, 1991; Squires ve Preece, 1999; Wickens vd.,2003 aktaran Koyuncu Tunç, Külcü, 2020, s. 276).

Bu noktada kullanılabilirliğin verimini arttıran hususlar aşağıdaki gibidir:

- Sistem arayüzünün, kullanıcıların davranışları izlenerek; kullanıcı eğilimine göre tasarlanması,
- Çeşitli değerlendirme yöntemleri ile sistemin sürekli olarak geliştirilmesi,
- Engelli bireylerde göz önüne alınarak (sesli komut, punto büyüklüğü, renk seçimleri vb.) düzenlenmesi,

- Kurumsal iş ve süreçlerin analiz edilerek işlemleri en kolay, hızlı ve kullanıcının işini kolaylaştıran şekilde ayarlanması kullanılabilirliği sağlayarak iş süreçlerinde verimi arttırmaktadır.

Kamu kurum ve kuruluşlarında kullanımı zorunlu olan EBYS yazılımının kullanılabilirliğinin incelenmesi, ölçütlerin belirlenmesi ve standartlaştırılarak farkındalık yaratılması sistemin ve iş süreçlerinin akışı konusunda oldukça önemlidir (Koyuncu Tunç, Külcü, 2020, s. 272). Bu nedenle aslında bu süreçlerin satın alma öncesinde gerçekleştirilerek, bir değerlendirmeye varılması önemlidir. Bu yolla, sistemin kurumsal kültüre uygun olup olmadığı belirlenerek, satın alma sonrasında çıkabilecek ek maliyet konusunda önlem alınabilmektedir (Koyuncu Tunç, Külcü, 2020, s. 294). Aksi taktirde kurum personelleri sistemde süreçleri gerçekleştirme konusunda sorunlar yaşayacak, bu da kurumu zaman, para ve emek noktasında verimsizliğe düşürecektir. Sistemin kullanılabilirlik derecesi kullanıcıların sistemi benimsemesiyle doğru orantılıdır (Kaya ve Eren, 2019, s. 88). Nitekim Svärd (2017, 72) bir belediye ile yaptığı röportaja yer verdiği çalışmada, belediyede kullanılan sistemin kullanıcı dostu olmaması nedeniyle; sistemin personel tarafından kullanılmamaya başladığını tespit etmektedir.

Günümüzde sürekli artan verinin/bilginin elektronik ortamda düzenlenmemesi; işlerin yavaşlamasına, bilgilerin kaybolmasına ve/veya erişilmemesine ve kurumsal kararlar alınamamasına yol açarak kuruma ciddi zarar vermesi söz konusudur. Bir elektronik belgenin anlamı ve faydası onun kullanılabilirliği ile ilgilidir (A National Electronic Commerce Coordinating Council E-Sign Policy Workgroup [NECC E-Sign Policy Workgroup] 2001, s. 9; aktaran Aydın, Özdemirci, 2011, s. 109). Bilgi güçtür fakat erişemediğiniz bilgi güç değildir. Bu durum tam tersi kurumun varlığını tehdit

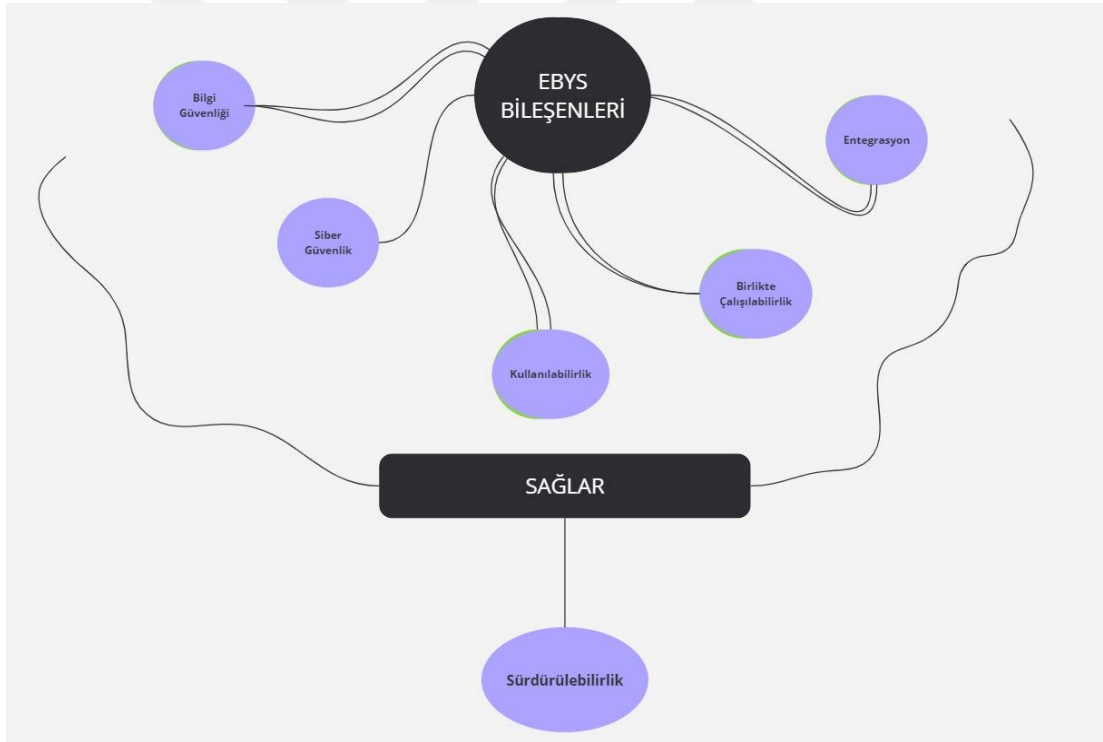
etmektedir. Bilgi ancak zamanında erişilebilir ve kullanılabilir haldeyse bir kaynak olarak kabul edilebilir bunu da bilgileri güvenilir şekilde yöneten kullanılabilir sistemler ve bilgi yönetim stratejileri sağlayabilmektedir (Sv rd, 2017). e-belgelerin kullanımının saęlanması, erişim unsurlarının EBYS ile iyi entegrasyonunu gerektirmektedir. Entegrasyon ve birlikte çalışılabilirlik iş süreçlerinin daha hızlı yürüt lmesine yardımcı olan, kullanılabilirlik için de önemli yeri olan iki EBYS bileşenini oluşturmaktadır. Bunlar sayesinde sistemler etkileşimli bir şekilde çalışarak, iş yük  azaltmaktadır. Aynı işlemin birden çok sistem üzerinden gerçekleştirilmesi, veri tekrarına neden olmaktadır. Entegre ve birlikte çalışabilir sistemlerse bu sorunların ön ne geçilmesinde önem arz etmektedir. Entegrasyon ve birlikte çalışılabilirlik tam olarak aynı anlama gelmese de birbirini önemli derecede etkileyen ve birbirinden ayrılmaz iki kavramı oluşturmaktadır. Bu iki kavramın EBYS’lerdeki önemine geçmeden önce kavramsal olarak bir açıklama getirmek konu için daha anlaşılır olacaktır. T.C. Kalkınma Bakanlığı e- Dönüş m T rkiye kapsamında hazırlamış olduęu çalışmasında (2012) birlikte çalışabilirlięi řu şekilde tanımlamaktadır: “bir sistemin ya da sürecin, ortak standartlar çerçevesinde bir dięer sistemin ya da sürecin bilgisini ve/veya işlevlerini kullanabilme yeteneęi” (Birlikte Çalışılabilirlik Esasları Rehberi S r m 2.1., 2012, s.4). Yine aynı rehberde entegrasyon terimi “kamu hizmetlerinin elektronik ortamda birlikte çalışacak ve ortak bir çözüm oluşturunca şekilde sunulması” olarak tanımlanmış olup, kavramın temel olarak devletin hizmet sunumunda hizmetler arası etkin veri paylaşımını kapsadığını belirtmektedir (Birlikte Çalışılabilirlik Esasları Rehberi S r m 2.1., 2012, s. 18). Entegrasyon EBYS için verimlilięi arttıran ve etkin bir bilgi yönetim süreci yür t lmesini saęlayan bir fakt r olarak karřımıza çıkmaktadır. “...EBYS bir organizasyonun t m sistemleri ile entegre olabilen ve iş yapma şeklini deęiřtiren bir sistemdir” (Yalçınkaya, s. 68, 2016).

Kuruluşların bilgiyi doğru yönetmek ve yönlendirmek için kurumsal iş süreci analizlerine yönelik farklı bilgi sistemlerinin entegrasyonuna karar vermesi gerekmektedir. Geleneksel ve birbirleri ile uyumsuz veri sistem modelleri, kurumların uzaklaşması gereken, verimsiz bilgi yönetim süreci yaşatmaktadır (Svård, 2017, s. 5). Sistem entegrasyonu sistemlerin birbirleri ile iletişim kurmasını sağlamaktadır. Böylece verimsiz bilgi yönetim süreci yaşatan uyumsuz veri sistemi modellerinin ortadan kalkması sağlanmaktadır (Rockley ve diğerleri, 2003; aktaran Svård, 2017, s. 31). EBYS'nin kurumun iç ve dış bilgi sistemleri ile entegrasyonu sistemin işlerliği noktasında önem kazanmaktadır. Kurumun sahip olduğu bilgi varlıkları ve faaliyetleri doğrultusunda oluşturulmuş öğrenci bilgi sistemi, maaş bilgi sistemi, personel bilgi sistemi gibi iç entegrasyonların yanında; e-imza, elektronik mühür, zaman damgası, kayıtlı elektronik posta gibi dış entegrasyonlar da EBYS için önem arz etmektedir (Özdemirci, 2019, s. 7). Doğru zamanda doğru bilgiye ulaşma kurumların gelişimi ve varlığını devam ettirebilmesi bakımından oldukça önemlidir. Bu nedenle iyi tasarlanmış bir bilgi yönetim sistemi, kurumlar için hayatidir. Kurumun bilgi yönetim sisteminin farklı sistemlerle entegre çalışması; farklı uygulamaların bütün değer zincirindeki süreçleri bütünleştirmesi ve doğru iş süreçlerinin yaşanmasını sağlamaktadır (Themistocleous, Irani, Kuljis, and Love, 2004; aktaran Svård, 2017, s. 31). Günümüzde kurumların bir iş için birden çok kez veri girişi sağlaması, entegre sistemlerin olmaması nedeniyledir. Bu hem iş süreçlerinde karmaşıklıklar yaratmakta hem de süreçleri uzatmaktadır. Süreçler kurumlararası veya departmanlarda sınırlarının ötesine uzandığından, bilgi sistemleri entegrasyon ile bilgi paylaşımını en verimli hale ulaştıracak şekilde uygulanmalıdır. Bu sayede kurumlar içinde, kurumlar arasında veya vatandaşlar ile ilişkilerde potansiyel işlevsellik yakalanmakta ve birlikte çalışılabilir sistemler ortaya konmaktadır (Svård, 2017, s. 6). Bahsedildiği gibi

sistemlerin birbiri ile konuşabilmesi ve bilgi alışverişinin sorunsuz gerçekleşmesi birlikte çalışan sistemler ile gerçekleşmektedir (Eroğlu ve Külcü, 2014, s. 330) Yalçinkaya (2016, s. 70) sistemler ve platformların birlikte çalışabilirlik ilkelerine uyumlu olması gerektiğini belirtmektedir. “...Devlet bilgi kaynaklarının birleştirilmesinin ve özerk bilgi sistemlerinin birlikte çalışmasının e-Devlet kalkınma hedeflerine ulaşılması için çok önemlidir...” (Headayetullah & Pradhan, 2010; aktaran Svärd, 2017, s. 4). Bu ortaya koymaktadır ki, aslında kurumların, entegre ve birlikte çalışabilir tasarladıkları EBYS’ler etkin ve verimli bir e-devlet yönetimini ortaya koymaktadır.

Şekil 6.

EBYS Bileşenleri



Bu bölümde son olarak bahsedilecek bileşen sürdürülebilirliktir. Sürdürülebilirlik EBYS’de; güvenliğin, kullanılabilirliğin, entegrasyonun ve birlikte çalışırılığın bir bütün olarak sağladığı bir bileşen olarak ortaya çıkmaktadır.

Sürdürülebilirlik bir kurum için en önemli hedeflerin başında gelmektedir. EBYS’de şu ana kadar yapılması gerektiği bahsedilen tüm faaliyetler sistemin ve bunun sonucunda ise kurumun sürekliliğini sağlamaktadır. Süreklilik o sistemi ve kurumu ayakta tutmakta ve varlığının devamlılığını sağlamaktadır.

Ulusal ve uluslararası mevzuat ve standartlar göz önüne alınarak oluşturulmuş bir kurumsal politika ile tüm aşamalar (güvenlik, kullanılabilirlik, entegrasyonlar, birlikte çalışır sistem, süreklilik vb.) detaylı planlanmalı ve ardından sistem tasarlanmaya başlanmalıdır. Tüm bileşenler kendi içerisinde bir yarar sağlasa da bileşenler bütünsel açıdan incelendiğinde sürdürülebilirliği sağlamaktadır. Bileşenlerden herhangi birinin aksaması ise sürekliliğe bir tehdit oluşturmaktadır. Güvenlik noktasında veri/bilgi/belgelerin bütünlüğünün, erişilebilirliğinin ve doğruluğunun korunması o kurumun iş akışlarında ihtiyaç duyduğu belgeye ihtiyaç duyduğu an sorunsuz erişebilmesini sağlamaktadır. Bu sayede kurum iş sürekliliğinin yanında kurumsal karar alma noktasında verimli bir süreç geçirecek ve kurumsal sürekliliğini de sağlamış olacaktır. Yine bilgi güvenliği noktasında yazılım ve donanımın sürekli olarak kontrol edilmesi ve teknoloji akışına göre güncel kalabilmesi sistemin sürekliliğini sağlamakta bu da etkileşimli bir şekilde önce bilgilerin ardından ise kurumun sürekliliğini sağlamaktadır. Sistemin kullanılabilir olması, entegrasyonlarının sağlanması ve birlikte çalışır sistemlerin kurulması iş süreçlerinin daha hızlı ve etkin gerçekleşmesine, daha hızlı karar alınmasına, iş yükünün azalmasına ve veri tekrarının önüne geçilmesini sağlayarak yine sürekliliğe fayda sağlamaktadır.

II. 2. Dijital Dönüşüm

EBYS'ler e-devlete giden yolda önemli bir adımı oluşturup, dijital dönüşümün sağlanmasında önemli bir aktördür. Bu başlıkta incelenecek dijital dönüşüm konusu; EBYS'ler ile neler başarılabilir, hangi noktalarda geliştirilebilir, amaçları nelerdir sorularının cevapları ile ilgili, ileriki bölümlerde yol gösterecek olup, “dijital dönüşüm nedir, dünyada ve Türkiye’de gelişimi nasıl olmuştur, olumlu ve olumsuz etkileri nelerdir, organizasyonlardaki işlevleri nedir” açıklanmaktadır.

Dijital dönüşüm; en temel anlamı ile yapay zekâ, makine öğrenimi, metin madenciliği, nesnelerin interneti ve doğal dil işleme gibi yeni dijital teknolojiler kullanılarak bir kuruluşun operasyonlarının, süreçlerinin ve hizmetlerinin geliştirilmesi, sunulması ve bundan katma değer elde edilmesidir. Çalışmanın ilerleyen kısımlarında dijital dönüşüm kavramı detaylı olarak açıklanacaktır; ancak öncesinde dijital dönüşüm ile sıklıkla karıştırılan: Dijitalleşme (digitization) ve dijitalizasyon (digitalization) terimlerine değinmek faydalı olacaktır:

Dijitalleşme (digitization); analog bilgileri, bilgisayarların depolayabileceği, işleyebileceği ve iletebileceği bir dijital formata dönüştürme sürecidir. (Verhoef ve ark., 2021; aktaran Hsiao, 2023). Örneğin bilgi merkezlerinde fiziksel olarak bulunan bilgi kaynaklarının bilgisayarca okunabilir şekilde aktarılmasıdır.

Dijitalizasyon (digitalization), mevcut iş süreçlerini değiştirmek için dijital teknoloji ve yeteneklerini uygulama sürecidir (Hsiao, 2023). Örneğin bir kuruluştaki yeni dijital teknolojileri kullanarak fırsat, tehdit ve risklerin belirlenmesi ve bu bilgilerin karar alma sürecinde uygulanmasıdır.

Dijital dönüşüm (digital transformation) ise dijital teknolojilerin yalnızca bir kuruluşun işine değil, aynı zamanda bir bütün olarak ekonomiye ve topluma getirdiği köklü bir değişimdir (Kim, Choi ve Lew, 2021; aktaran Hsiao, 2023). Vial'ın (2019) kavramsal tanımına göre dijital dönüşüm, "bilgi, bilgi işlem, iletişim ve bağlantı teknolojilerinin kombinasyonları yoluyla bir varlığın özelliklerinde önemli değişiklikleri tetikleyerek bu varlığı iyileştirmeyi amaçlayan bir süreçtir" (Vial, 2019; aktaran Hsiao, 2023). İş süreçlerinde, operasyonel rutinlerde, kurumsal yeteneklerde ve hatta iş piyasalarında büyük değişikliklere yol açmaktadır (Singh, Sharma ve Dhir, 2021; aktaran Hsiao, 2023). Dijital teknolojilerin fikir yaratmadan ürün/hizmet üretimine kadar her aşamada kullanılarak kurum geleneğini, kurumlararası rekabeti, bireylerin yetenek ve davranışlarını dönüştürmesi ve ekonomik, toplumsal vb. alanlarda büyük değişiklikler yaratılması bir dijital dönüşüm örneğidir.

Yapılan tanımlar ortaya koymaktadır ki; çoğu kez birbiri yerine kullanılan dijitalleşme, dijitalizasyon ve dijital dönüşüm terimleri aslında birbirini takip eden ve birbirini etkileyen aşamaları ifade etmektedir. Verhoef ve diğerleri de (2021) çalışmalarında bu görüşü destekleyerek, dijital dönüşümü dönemsel olarak değerlendirmekte ve bilişimleştirme (informatization), dijitalizasyon ve dijital dönüşüm olmak üzere üç aşamaya bölmektedir (Verhoef et al.; aktaran Ding ve diğerleri, 2024, s.3).

Türkçe literatürde dijitalleştirme ve sayısallaştırma terimleri aynı şekilde tanımlanmaktadır. Biz ise çalışmanın bundan sonraki kısmında kavramsal tutarlılığı sağlamak adına dijitalleştirme terimini tercih etmekteyiz.

Dijitalleştirme işlemi fiziksel olarak elde bulunan bilgi kaynaklarının elektronik ortamda 0'lı ve 1'li sayılar yolu ile ifade edilmesidir. “Dijitalleştirme kavramı ilk olarak 1703 yılında Wilhelm'in İkili Aritmetiğin açıklamasında ele alınmış ve kavramsallaştırılmıştır” (Ding ve diğerleri, 2024, s.3). Bu noktada Wilhelm'in İkili Aritmetiğinde bahsedilen ikili sayılar, sayıların 2 tabanında yazılması ile oluşturulmakta ve böylelikle tüm sayılar 0 ve 1 rakamları kullanılarak ifade edilmektedirler. (Vikipedi, İkili Sayı Sistemi) Dijitalleşme ile sayısal ortamda oluşturulmamış analog olarak var olan veri/bilgi/belgelerin bilgisayarca anlaşılabilir formatlara dönüştürülmesi söz konusudur. Bu işlem temel olarak, iş süreçlerinin yeni dijital teknolojiler ile uygulanmasını sağlayan dijitalizasyon sürecine hazırlık olarak ele alınabilmektedir. Fakat burada önemli olan nokta dijitalizasyon ile dijital dönüşümün farkını ortaya koyabilmektir. Dijital dönüşümde de dijitalizasyonda olduğu gibi ürün ve hizmetlere yeni dijital teknolojilerin uygulanması söz konusudur. Burada iki kavramın ayrımını; dijital dönüşümün bu teknolojileri ürün ve hizmetlerine uygulayarak; toplumsal, ekonomik vb. açılardan da büyük dönüşümler yaratması oluşturmaktadır. Dolayısıyla dijitalleşme ve dijitalizasyon süreçlerinde; kuruluşlar veri/bilgi/belgelerini dijitalleştirerek, iş süreçlerinin tamamını elektronik ortamda yeni dijital teknolojileri kullanarak gerçekleştirmektedir. Bu sayede kurumsal gelenek form değiştirerek, ulusal düzeyde bir dönüşüm ve entegrasyon sağlanmaktadır. Organizasyonel, toplumsal ve ekonomik boyutları büyük oranda etkileyen bu dönüşüme uyum sağlanması ve beceriler kazanılması ile de dijital dönüşüm gerçekleştirilmiş olmaktadır.

Dijital dönüşümün toplumsal, ekonomik, kurumsal açılardan yaratmış olduğu köklü değişim, Dördüncü Sanayi Devrimi (Fourth Industrial Revolution) veya endüstri 4.0 (I4.0) olarak adlandırılan yeni toplumsal evrede üretim ilişkilerinin yaratacağı köklü değişime de altyapı hazırlamaktadır. Hatta bu iki olgunun birbirini besleyen yapıda

olduğunu söylemek de mümkündür. Nitekim endüstri 4.0 kapsamında, kuruluşlar yaygın biçimde dijitalleşmeye ve endüstriyel yetenekleri geliştirmek için yapay zeka, arttırılmış gerçeklik, büyük veri, nesnelerin interneti, bulut bilişim, veri yönetimi gibi dijital teknolojilerin kullanımına yönelmiş (Luthra ve diğerleri; aktaran aktaran Nadeem ve diğerleri, 2024, s.2) ve bu teknolojileri üretime uygulayarak değer yaratılması sağlanmıştır. Endüstri 4.0, dijital cihazlar, operasyonlar, makineler ve alıcıların yanı sıra müşteriler ve tedarikçiler de dahil olmak üzere birden fazla paydaş arasında gerçek zamanlı veri paylaşımını mümkün kılan, birbirine bağlı dijital ve sanal teknolojileri kapsamaktadır (de Sousa Jabbour ve diğerleri; Li, Hou ve Wu aktaran; Nadeem ve diğerleri, 2024, s.2). Bu bağlamda “kapsamlı dijitalleştirme, otomasyon, kişiselleştirme, optimizasyon, adaptasyon, insan-makine etkileşimi, otomatik iletişim, veri paylaşımı, katma değerli işler ve hizmetler Endüstri 4.0'ın baskın özellikleri” (Lu; aktaran Nadeem ve diğerleri, 2024, s.2) ni oluşturmaktadır. Tüm bunlar göz önüne alındığında endüstri 4.0 ile dijital dönüşümün birbirinden ayrı düşünülmesi söz konusu olamamakta, ikisi de birbirinin gelişimini sağlayan ve olumlu anlamda büyük değişimleri oluşturarak etkin, etkili ve verimli süreç ve hizmetleri yaratmaktadır.

Ding ve arkadaşları (2024, s. 3) Çin’de dijital dönüşüm ve inovasyon performansı arasındaki ilişkiyi inceledikleri ve dijital dönüşümün inovasyon performansını iyileştirip iyileştirmediğini ortaya koydukları çalışmalarında, dijital dönüşümün; dijital çağın ekonomik gelişimine özgü bir olgu olduğunu ve ekonomik paydaşların geleneksel ekonomik biçimlerinin dijitalleşme yoluyla yenilenmesi ve daha gelişmiş bir ekonomik forma kavuşmasını sağladığını belirtmektedir. Dijital çağda bireylerin, organizasyonların ve ülkelerin elindeki en değerli güç veridir. Verinin en yüksek potansiyeli, ürün ve hizmet üretme ve ekonomik değer yaratmada yatmaktadır. Günümüzde makine öğrenmesi, veri madenciliği, yapay zekâ vb. yeni dijital teknolojileri kullanarak veriler

anlamlandırılabilen ve analiz edilebilmektedir. Bu teknolojiler kullanıcı/müşteri bilgileri, tedarik zinciri ve üretim hatları gibi çok farklı kaynaklardan veri akışı sağlamak ve bizi büyük veriye götürmektedir. Bu varlık değeri kullanılarak; daha inovatif performanslar ortaya konulmaktadır. Bu yolla, kurumsal iş süreçlerinde, geliştirilecek hizmetlerde/ürünlerde ve kurumsal politikalarda dijital teknolojiler uygulanarak, katma değer yaratılabilmektedir. Böylece verinin bilgiye; bilginin ise hizmete/ürüne dönüşümü sağlanmaktadır. Yeni dijital teknolojilerin yarattığı sonuçlar açısından, sorunlarla baş edebilmek ve kazanımları ürün ve hizmetlerde kullanmak dijital dönüşümün gerçekleşmesi için önem arz etmektedir. “Sosyal medya, mobil uygulamalar, bulut bilişim ve nesnelerin interneti gibi dijital teknolojiler, birçok sektörde köklü değişimlere yol açmaktadır. Bu durum, kuruluşları hayatta kalabilmek veya pazardaki hâkimiyetlerini korumak için dijital dönüşüme uyum sağlamaya mecbur bırakmaktadır” (Margiono, 2021, s. 315). Bunun en yakın ve belki de en etkili örneği COVID-19 pandemi döneminde görülmektedir. Her türden organizasyon, varlıklarını sürdürebilmeleri için dijital teknolojileri ürün ve hizmetlerine uyarlayarak veya var olan dijital hizmetlerini geliştirerek ürünlerini/hizmetlerini müşterilerine/kullancılara ulaştırmıştır. Pandemi sürecinde dijital dönüşümün olumlu gelişimi, dijitalleşmeye adaptasyon ve ürün ve hizmetler için dijital teknolojilerin sürekli geliştirilmesi ile görülebilmektedir.

“Dijital dönüşüm hem toplumsal bir olgu hem de iş geliştirme için bir yöntem olarak düşünülebilmektedir” (Magnusson, Elliot and Hagberg, 2022, s. 316). Dijital dönüşümde organizasyonların, ürün ve hizmetlerini yeni teknolojiler ile sunması iş geliştirme boyutunu oluştururken; bu hizmetleri sunan personel ve hizmetten/üründen faydalanan müşterinin/kullanıcının, bu teknolojilere uyum sağlayabilme ve kullanabilme becerileri ise toplumsal boyutu oluşturmaktadır. Bu noktada ise teknoloji okuryazarlığı ve dijital vatandaşlık gibi kavramlar ortaya çıkmaktadır.

Türkiye'de dijital dönüşümün aşamalarını ve mevcut durumunu değerlendirmeden önce, literatürde tanımlandığı şekliyle dijital dönüşümü tanımlamak ve etkilerini tartışmak gerekmektedir. Bu, Türkiye'deki dijital dönüşüm sürecini tartışmak için bir temel sağlayacaktır.

II. 2. 1. Literatür Değerlendirmesi

Yabancı literatür incelendiğinde, dijital dönüşümün kuruluşlarda yenilikçi fikirlerin değerlendirilmesinden iş süreçlerine, ürün/hizmet geliştirmeden rekabeti arttırmaya kadar tüm alanlarda uygulandığı; kamu kurumlarında ise üretilen hizmetlerin vatandaş ile buluşmasında, politikalar oluşturulmasında ve bütüncül, birlikte çalışabilir sistemlerin inşasında önemli bir potansiyele sahip olduğu yönünde tartışmaların yer aldığı görülmektedir. Bunun yanı sıra, bu dönüşümün kuruluşların ötesine geçtiği ve meydana getirdikleri sosyal ve ekonomik boyutların üzerine de önemle durulduğu gözlemlenmektedir.

Hsiao (2023) çalışmasında dijital dönüşümle ilgili olarak literatürde henüz net bir tanımın oluşturulmadığını fakat yapılan çalışmalarda dijital teknolojiler ile iş modeli arasındaki bağlantıya odaklanıldığını belirtmektedir. Bu noktada iş modeli için dinamik yeteneklerin geliştirilmesinde kullanılan; "dijital algılama" (digital sensing), "dijital yakalama" (digital seizing) ve "dijital dönüştürme" (digital transforming) (Warner & Wäger, 2019; Kraus et al., 2021, Imran et al., 2021; aktaran Hsiao, 2023) terimlerinin önem taşıdığından söz etmektedir.

Dijital dönüşüm ile tedarikçiler ve müşterilerden elde edilen kaynakları entegre ederek iş süreçleri dijitalleştirilmekte (Hsiao, 2023), yine aynı kaynaklar kullanılarak elde edilen verilerin analizi ile değer yaratılmakta ve yakalanmaktadır. Yaratılan bu değer,

yine yeni teknolojiler ile inovatif ürün ve hizmetlere dönüştürülmekte; organizasyonlarda, ülkelerde katma değer sağlamaktadır. Bu bağlamda birçok araştırmacının da önemini vurguladığı iş süreçlerinin dijitalleşme yoluyla yeniden yapılandırılması (Hsiao, 2023), süreçlerde değer yaratarak ürün/hizmetlerin veriminde artış sağlamaktadır. Ortaya iyi bir ürün ve/veya hizmet konulabilmesi iyi bir iş modeli oluşturulmasını gerektirmektedir. Dijital dönüşüm, teknolojilerin ürün/hizmetlere uygulanmasından önce iş modellerinde uygulanarak süreçlerde yenilik ve değer yaratmaktadır.

Dijital dönüşümün iş süreçlerinde, karar almada ve ürün/hizmetlerde yarattığı bu etki ekonomik ve rekabet gücü yaratma açılarından da dikkat çekmektedir. Dijital dönüşümde yeni teknolojilerin kullanılması ile üretilen hizmet/ürünün verimi ve kalitesi arttırılmakta, yenilikçi ürün/hizmetler ortaya konulmakta ve yeni yeteneklerin ortaya çıkardığı yeni istihdam alanları yaratılabilmektedir. Bu teknolojilerin etkin kullanılmasının, ürün/hizmetlerin verimliliği üzerinde önemli etkisi bulunmaktadır (Ding ve diğerleri, 2024, s. 18). Bu bağlamda dijital dönüşümün rekabet gücü yaratması önemli bir olgudur. Aşağıda dijital dönüşümün rekabet gücünü vurgulayan bazı tanımlar yer almaktadır.

- Nedeem ve arkadaşları (2024) çalışmasında, dijital dönüşümün iş modelini tamamen yeniden şekillendirme potansiyeline sahip olduğunu, yeni çalışma, etkileşim ve katılım yolları sunmanın yanı sıra paydaş değerini, **kurumsal rekabet gücünü ve genel iş performansını** artırdığını savunmaktadır. (Depaoli, Za ve Scornavacca, Grant ve diğerleri, Yadav ve Pavlou; aktaran Nadeem ve diğerleri, 2024, s.2).
- Temelde dijital dönüşüm, veri yönetim sürecini güçlendirebilecek, kaynak dağılımının verimliliğini artırabilecek ve yeni dijital teknolojilerin yardımıyla bir

organizasyonun **temel rekabet gücünü artırabilecek** bir süreçtir (Wang ve diğerleri, 2023, s. 1).

- “Dijital dönüşüm, organizasyonun faaliyetlerini daha verimli ve etkili bir şekilde tasarlamak, daha iyi performans ve **rekabet avantajı elde etmek** amacıyla dijital teknolojiler tarafından süreçlerin, ürünlerin, iş modellerinin ve insan davranışlarının modifiye edilmesini ön plana çıkarmaktadır” (Singh et al., 2021; Wenger et al., 2021; Li, 2022; aktaran Hsiao, 2023). Başka bir deyişle, dijital dönüşüm; ekonomiler, toplumlar, kurumlar ve bireyler için sonuçları derinden şekillendirerek; dijital teknolojilerin ötesine geçmektedir (Singh, Klarner, & Hess, 2020; Konopik et al., 2022; aktaran Hsiao, 2023).

Yine literatür, araştırmacıların dijital dönüşümü tanımlarken genellikle yeni teknolojilere öncelik verdiğini ortaya koymaktadır. Hai, Van ve Tuyet (2024) de çalışmalarında, birçok çalışmada ele alınan dijital dönüşüm unsurlarının dört faktörü içerdiğini belirtmiştir. Bunlar: Nesnelerin İnterneti, büyük veri, sibernetik ve birlikte çalışabilirliktir (Hai, Van, Tuyet, 2021, s. 24). Bu durum, dijital dönüşüm literatüründe yeni teknolojilere yapılan vurguya işaret etmektedir.

Nadeem ve arkadaşları (2024, s. 2) çalışmalarında, dijital teknolojilerin literatürde öne çıkan rolüne ilişkin kapsamlı bir genel bakış sunmaktadır:

Dijital dönüşüm kavramı, siber ve fiziksel ortamları birbirine bağlayarak yeniden canlandırmaktadır. Dijital Dönüşüm, bilgi, bilgisayar, iletişim ve ağ teknolojilerinin bir araya getirilmesiyle, bir varlığın özelliklerinde önemli değişiklikler tetikleyerek onu iyileştirmeyi amaçlayan bir süreç olarak tanımlanmaktadır. Dijital dönüşüm, Nesnelerin İnterneti (IoT), yapay zekâ, büyük veri analitiği,

bulut bilişim, blockchain, platformlar ve siber-fiziksel sistemler (CPS) gibi dijital teknolojileri içermektedir. Son teknolojiler olan CPS ve IoT, ürün ve hizmetleri, pazarları, tedarik zincirlerini, iş modellerini, insan becerilerini, çalışma ortamlarını, pazarları ve tüm ekonomiyi değiştirmektedir.

Yeni dijital teknolojilerin, gelenek açısından yarattığı dönüşüm, yine dijital dönüşüm literatüründe sıklıkla dile getirilen bir hususu ortaya koymaktadır. Yapay zeka, makine öğrenmesi, veri madenciliği gibi dijital teknolojiler, sosyal yaşamda köklü bir değişim yaratma potansiyeline sahiptir (Hai, Van, Tuyet, 2021, .s 24). Çünkü dijital dönüşüm, organizasyonların iş süreçleri ve hizmetleri/ürünleri ile ilgili olduğu kadar kullanıcılar/müşteriler de ilgilidir. Her ürün ve hizmetin bir kullanıcı kitlesi bulunmaktadır. Dolayısı ile bu hizmet ve ürünlerin topluma ulaştırılması önemlidir. Toplumun dönüşen ürün ve hizmetlere erişebilme becerileri yok ise sadece organizasyonel bir dijital dönüşümün faydasından bahsedilemez. Dijital dönüşüm sürecine paralel olarak toplumun da dijital teknoloji becerileri edinmesi, yapılan dijital dönüşüm çalışmalarını anlamlı kılacaktır. Süreçler bu şekilde yürütüldüğü takdirde, bu teknolojilerin sosyal yaşamda bir değişim yaratacağı söylenebilmektedir. Bu düşüncüyü destekleyen Hsiao, (2023) çalışmasında dijital teknolojilerin toplumu ve işletmeleri dönüştürerek hem iç hem de dış olarak yeni zorluklarla karşı karşıya kaldığını belirtmektedir (Magistretti et al., 2021; Van Veldhoven & Vanthienen, 2022 aktaran; Hsiao, 2023). Başka bir ifadeyle dijital dönüşüm, günlük yaşamın her alanını etkileyen değişimler olarak anlaşılabilmektedir (Stolterman, Erik, and Fors, 2004; aktaran Hai, Van, Tuyet, 2021, .s 24). Nitekim Margiono (2021, s. 315), dijital teknolojilerin yarattığı dönüşümü, çalışmasında şu şekilde ele almaktadır; “dijital dönüşüm, "bilgi, bilgisayar, iletişim ve ağ teknolojileri birleşimleriyle bir varlığın özelliklerinde önemli değişiklikler

tetikleyerek onu iyileştirmeyi amaçlayan bir süreçtir". Benzer şekilde Liu ve arkadaşları (2023) çalışmalarında, dijital dönüşümün, yeni nesil dijital teknolojilerin, geleneksel çok yönlü kurumsal unsurlarla derinlemesine entegrasyonu olduğunu ve kurumsal performansı artırmak için yeni bir yol sunduğunu ileri sürmektedir. Sonuç olarak dijital dönüşümün dijital teknolojiler ile ilişkisi şu şekilde ele alınabilir: "Dijital dönüşüm, değişen liderlik bağlamlarına yanıt vermek için firmalar, ekosistemler, endüstriler veya alanlar içerisinde dijital teknolojileri kullanarak mevcut iş süreçlerini, kültürleri, yetenekleri ve deneyimleri oluşturma veya uyarlama sürecidir. Dijital çağdaki bu değişim dijital geçiş olarak görülmektedir" (Luong, 2020; aktaran Hai, Van, Tuyet, 2021, .s 24; Warner & Wäger, 2019; Wenger et al., 2021; Kraus et al., 2022; Konopik et al., 2022; aktaran Hsiao, 2023).

Hai, Van ve Tuyet (2021) çalışmalarında dijital dönüşüme ilişkin literatür taramasından elde ettikleri tanımlar aşağıdaki gibidir.

Tablo 1.

Dijital dönüşüm terimine ilişkin tanımlar

DİJİTAL DÖNÜŞÜM TERİMİNE İLİŞKİN TANIMLAR	
Stolterman & Fors (2004)	Dijital dönüşüm, dijital teknolojinin insan hayatının her alanında neden olduğu veya etkilediği değişimlerdir.
Martin (2008)	Dijital dönüşüm genel olarak, bilgi ve iletişim teknolojisinin sadece önemsiz bir otomasyon gerçekleştirmek yerine, iş dünyasında, devlette ve toplumda temelde yeni yetenekler yaratmak için kullanılması olarak anlaşılmaktadır.
Westerman, et al.	Dijital dönüşüm (DT), işletmelerin performansını veya erişimini önemli ölçüde artırmak için teknolojinin kullanılması anlamına gelir. Bu konu dünya çapındaki şirketler arasında popülerlik kazanıyor. Çeşitli sektörlerden yöneticiler, analitik, mobilite, sosyal medya ve akıllı gömülü cihazlar gibi dijital gelişmeleri kullanmanın yanı sıra müşteri ilişkilerini, iç süreçleri ve değer önerilerini değiştirmek için ERP gibi geleneksel teknolojilerin kullanımını geliştiriyor.
McDonald & Rowsell- Jones (2012)	...dijital dönüşüm, kaynakların yalnızca dijitalleştirilmesinin ötesine geçerek dijital varlıklardan değer ve gelir yaratılmasına yol açmaktadır.
Fitzgerald, Kruschwitz, Bonnet & Welch (2013)	Sosyal medya, mobil cihazlar, analitik ve gömülü cihazlar gibi yeni dijital teknolojilerin kullanımı, müşteri deneyimini geliştirmek, operasyonları kolaylaştırmak veya yeni iş modelleri oluşturmak gibi önemli iş iyileştirmeleri sağlayabilir.
PwC (2013)	Dijital dönüşüm, bir bütün olarak toplum üzerinde önemli bir etkiye sahip olan yeni internet tabanlı teknolojilerin uygulanması yoluyla iş dünyasının kapsamlı bir şekilde elden geçirilmesi anlamına gelmektedir.
Mazzone (2014)	Dijital dönüşüm, bir şirketi, iş modelini, fikri, süreci veya metodolojiyi hem stratejik hem de taktiksel olarak dijital olarak geliştirmeye yönelik kasıtlı ve sürekli bir süreçtir.
Digital Economy (2015)	Dijitalleşme, ekonomi ve toplumun tüm sektörlerinin kapsamlı bir şekilde birbirine bağlanması anlamına gelir ve ilgili bilgilerin toplanmasını, analiz edilmesini ve eyleme dönüştürülebilir içgörülere dönüştürülmesini sağlar. Bu değişiklikler avantajlar ve fırsatlar sunarken, aynı zamanda yeni zorluklar da ortaya çıkarmaktadır.
Schweer & Sahl (2016)	Dijital dönüşüm, ekonominin tüm sektörlerinin tutarlı bir şekilde ağıyla bağlanması ve oyuncuların dijital ekonominin yeni gerçeklerine uyum sağlaması olarak tanımlanmaktadır. Ağıyla bağlı sistemlerdeki kararlar, veri alışverişi ve analizi, seçeneklerin hesaplanması ve değerlendirilmesinin yanı sıra eylemlerin başlatılması ve sonuçların ortaya konmasını içerir.
Kotarba (2018)	Dijital dönüşüm, teknolojik ilerleme ve inovasyonun hızlı temposu nedeniyle iş modellerinin uyarlanması veya değiştirilmesi anlamına gelir ve bu da tüketici ve sosyal davranışlarda değişikliklere yol açar.
OECD (2019)	Dijital dönüşüm, ekonomilerin ve toplumların dijitalleşmesinin ve sayısallaşmasının bir sonucudur.
Vietnamese Government (2019)	Dijital dönüşüm, sosyo-ekonomik hayatımızın tüm yönlerini bütüncül ve kapsamlı bir şekilde değiştirmek, yaşama, çalışma ve birlikte ilişki kurma biçimimizi yeniden şekillendirmek için veri ve dijital teknolojilerin kullanılmasıdır.
Kozarkiewicz (2020)	Dijital dönüşüm, dijital teknolojilerin endüstri ve toplumda yıkıcı değişikliklerin yaratılması ve güçlendirilmesinde merkezi bir rol oynadığı bir süreçtir. Dikkatli bir planlama ve yürütme gerektiren karmaşık ve devam eden bir süreçtir.

(Hai, Van, Tuyet, 2021, ss. 23-24)

Genel olarak dijital dönüşüm tanımlarından yola çıkarak bir değerlendirmeye varıldığında dijital dönüşüm; dijitalleşme ve dijitalizasyon süreçlerinin devamı

niteliğinde, dijitalizasyonun sonraki aşaması olarak değerlendirilebilmektedir. Dijitalleşme fiziksel ortamda yer alan kaynakların bilgisayarca okunabilir formata dönüştürülmesi işlemidir. Dijitalleştirilen veya dijital ortamda üretilmiş verilerin anlamlandırılmasından bilgi elde edilmektedir. Bu bilgilerin, hizmete/ürüne dökülmesine kadar tüm aşamaları kapsayan iş süreçlerinin; yapay zeka, makine öğrenmesi, doğal dil işleme, veri madenciliği gibi dijital teknolojiler ile geliştirilmesi dijitalizasyon sürecini ifade etmektedir. Bu sürecin yarattığı büyük dönüşümler ise dijital dönüşümü sağlamaktadır. Yaratılan dönüşümler ekonomik düzeyde verimliliğin artırılması, rekabet gücünde artış ve istihdam yaratma gibi faydaları ortaya koymaktadır.

Dijital çağ, bilginin büyük önem taşıyan bir güç olduğunun altını çizmektedir. Kuruluşlar; dahili ve harici kaynaklarından elde ettikleri verileri yakalayıp, dijital teknolojiler ile anlamlandırmakta, bunun sonucunda ise, kurumsal iş süreç ve modellerini geliştirmektedirler. Böylece yine dijital teknolojileri kullanarak, elde ettikleri bilgilerden inovatif bir değer yaratmaktadırlar. Bu bağlamda Hsiao'nun (2023) çalışmasında belirttiği gibi “dijital dönüşüm, yeni iş fırsatları yaratmak için dijital teknolojileri kullanarak kuruluşların faaliyet gösterme ve müşterilere değer sunma biçiminde köklü bir değişikliği” (Hsiao, 2023) ifade etmektedir.

Bu dönüşüm sadece ekonomik açıdan gerçekleşmemektedir. Margiono'ya (2021, s. 315) göre dijital dönüşüm, endüstriyel düzeyde yenilikçi organizasyonları diğerlerinden ayırmakta, sosyal düzeyde ise insanların etkileşim ve yaşam biçimlerini değiştirmektedir. Daha önce de ifade edildiği gibi organizasyonlar, hizmetlerini kullanıcılarla/müşterilerle buluşturmak için üretmekte ve geliştirmektedir. Dijital dönüşüm yeni yetenekler, deneyimler ve uyarlamalar gerektirmektedir. Bu ise yeni bir toplumsal algılayış ve düzenin sağlanması ile gerçekleşebilmektedir. Bu nedenle toplumun yeni becerilere açık bir şekilde; dijital okuryazarlık, medya okuryazarlığı, ağ

okuryazarlığı ve dijital vatandaş becerilerine sahip olması önem taşımaktadır. Ancak bu dönüşüm ile kuruluşların sunduğu ürünlerle/hizmetlerle etkileşim; anlamlı ve verimli hale gelebilmektedir.

II. 2. 2. Türkiye’de Dijital Dönüşüm

Dijital dönüşümün ekonomik, sosyal vb. alanlarda yarattığı büyük dönüşüm ülke içerisinde de entegre ve birlikte çalışır sistemler oluşmasını sağlayarak, e-devlet yapısını geliştirmekte ve açık, hesap verilebilir, etkileşimli bir yapı oluşturmaktadır. Türkiye’de dijital dönüşüme ilişkin literatür incelendiğinde, tarihsel gelişimin ve kamu sektöründeki faaliyetlerin en sık ele alınan konular olduğu görülmektedir. Bu nedenle Türkiye’de dijital dönüşüme dair literatür tarihsel gelişim şeklinde ele alınmaktadır.

'Dijital dönüşüm' terimi 1960'lardan beri kullanılmaktadır (Vial, 2019, 118; aktaran Binici, 202, 10). Uysal, Kurban ve Çıgman (2023, s. 213-214) çalışmalarında, dijital dönüşüm kavramının, kamu yönetiminde ancak, yakın zamanda uygulanabildiğini belirtmektedir. Çalışmalarında dijital dönüşüm kavramı; karar alma ve iş akışları da dâhil olmak üzere kurumsal süreçlerin dijitalleştirilmesi ve zaman ve mekân tasarrufu yoluyla hizmet sunumunun kolaylaştırılması olarak tanımlanmaktadır.

Türkiye’de, e-Devlet politikası 20-30 yıldır yürürlükte olup, ABD'deki başlangıcından bu yana küresel olarak benimsenmektedir. İlk uygulamaları vergi beyannamesi ile ilgili olan e-devlet uygulamaları, ABD’de 1985 yılında başlamıştır (Demirhan, 2011: 24-25; aktaran Karasoy ve Babaoğlu, 2020, s. 120). Bilgi teknolojilerinin gelişimi ve yayılmasıyla ise hizmetlerin kapsamı artmaya başlamıştır.

E-devlet uygulamaları, belge çevirmenleri ve basit bilgi sayaçları ile başlayarak kademeli olarak gelişmiş, otomasyon çalışmaları ise 1970'lerden sonra artmıştır. Süreçleri hızlandırmayı ve kolaylaştırmayı amaçlayan bu uygulamaların dönüşümü; teknolojinin sürekli bir şekilde gelişiminin sonucu olarak mümkün olmuştur (Çarıkçı, 2010; aktaran Karasoy ve Babaoğlu, 2020, s. 119).

Türkiye'deki süreç incelendiğinde, telekomünikasyon altyapısının güçlendirilmesine yönelik faaliyetlerin 1980'li yıllarda başladığı görülmektedir (Karasoy ve Babaoğlu, 2020, s. 116-117). “Bu dönemdeki ilk çalışma, 1983 yılında Bilim ve Teknoloji Yüksek Kurulu’nun kurulmasıdır” (Boyalı, 2023, s. 174). Ancak, e-Devlet'e geçişin kurumsallaşması, gerçekçi hedefleri olan programların uygulandığı 1999-2000 yıllarına kadar başlamamıştır. Bu dönemde Türkiye'nin Avrupa Birliği'ne aday ülke konumunu sürdürmesi en büyük itici güç olmuştur. İlk yıllarda çeşitli faaliyetler farklı kamu kurumları tarafından koordinasyonsuz bir şekilde yürütülmüştür. Ancak ilerleyen yıllarda bu kurumlar arasında, koordinasyonun geliştirilmesi için çaba gösterilmiştir. Özellikle kamu hizmetlerini tek bir platform altında birleştirmek için web tabanlı teknolojiler kullanılmıştır. Ancak mevcut e-Devlet süreci sadece devlet içinde bir dijitalleşme projesi değildir. Aynı zamanda hizmet sunumunun ötesine geçerek, ülke için katma değer üretecek bir süreçtir (Karasoy ve Babaoğlu, 2020, s. 116-117). Zaten dijital dönüşüm literatüründen de görüldüğü üzere, dijital dönüşüm bir dijitalleştirme çalışmasından öte; yeni dijital teknolojileri kullanarak, üretilecek hizmet ve ürünlerde katma değer sağlamayı içermektedir.

Türkiye ilk e-Devlet stratejisini 2002 yılında geliştirmiş olsa da daha önceki strateji belgelerinde de kamu sektöründe bilgi teknolojilerinin kullanımına yönelik hedefler, öncelikler ve araçlar yer almaktadır. Ancak strateji belgelerinin geneline dağılmış olan bu unsurların tutarlı bir şekilde entegre edildiğini söylemek her zaman

mümkün değildir. Uysal, Kurban ve Çığman (2023, s. 214) çalışmalarında bu analizin sübjektif değerlendirmeleri içermediğini belirtmektedir. Kamu sektöründe bilgi teknolojilerinin kullanımı ilk olarak Dördüncü Beş Yıllık Kalkınma Planında ele alınmış olup, bu gelişmeler ileride yayımlanan kalkınma planlarında da kendini göstermektedir. Bu kalkınma planlarından; Sekizinci Beş Yıllık Kalkınma Planı, 2023 Türkiye vizyonunu belirlemesi ile ön plana çıkmaktadır (Uysal, Kurban ve Çığman, 2023, s. 214-215).

1999'da Avrupa'da etkisini gösteren dijital dönüşüm, e- Avrupa Girişimi ile somutlaşarak (Karasoy ve Babaoğlu, 2020, s. 120), Türkiye'deki dijital dönüşüm politikası için rehber niteliği taşımaktadır. Nitekim e-Avrupa + Eylem Planı ve e-Türkiye Girişimi (2001), e-Avrupa Girişimindeki izleri taşıyan, dijital dönüşümdeki önemli adımları göstermektedir (Boyalı, 2023, s. 174). Akabinde Türkiye'de 2002 yılında, bu dönüşüme ait ilk bütüncül e-Devlet/e-dönüşüm stratejisi ortaya konulmuştur (Uysal, Kurban ve Çığman, 2023, s. 215). Önceki yıllarda bilgi toplumuna geçiş süreci için Vergi Dairesi Tam Otomasyon Projesi' (VEDOP), Merkezi Nüfus İdaresi Sistemi' (MERNIS) gibi dijital dönüşüme yönelik uygulamalar gerçekleştirilmiş olsa da Türkiye'de ilk politika 2002 yılında Avrupa birliğinin gerçekleştirmiş olduğu e-Avrupa+ Girişimine üyelik sonucunda hazırlanmıştır (Saylan, 2009 aktaran Karasoy ve Babaoğlu, 2020, s. 122). Daha sonra hızlıca dönüşümle ilgili aşağıda yer alan gelişmeler yaşanmıştır:

- 3 ocak 2003 58. Hükümet tarafından hazırlanan Acil eylem Planı'nda E-Dönüşüm Türkiye Projesine yer verilmesi,
- 2003/12 sayılı başbakanlık genelgesi ile e-Dönüşüm Türkiye Projesi'nin amaçları, kurumsal yapısı ve uygulama esasları belirlenmesi ve sürecin yönetimi için DPT

- Müsteşarlığının görevlendirilmesi, (Karasoy & Babaoğlu, 2020, s. 122)
- 2003 yılında DPT bünyesinde Bilgi Toplumu Dairesi (BTD) kurulması, (Karasoy & Babaoğlu, 2020, s. 122)
- 2003/48 sayılı Başbakanlık Genelgesi'yle e-Dönüşüm Türkiye Projesi Kısa Dönem
- Eylem Planı uygulamaya konması ve e-Dönüşüm Türkiye İcra Kurulu oluşturulması, (Karasoy & Babaoğlu, 2020, s. 123)
- 2003-2004 e-Dönüşüm Türkiye projesi
- 2005 e-Dönüşüm TR projesi
- 2005 eylem planı ve uygulamaları, (Karasoy & Babaoğlu, 2020, s. 122)
- 2005-2006 yıllarında Bilgi Toplumu Stratejisi ve Eylem Planı hazırlanarak 2006-2010 döneminde uygulanmak üzere 2006 yılı temmuz ayında 2006/38 sayılı Yüksek Planlama Kurulu kararı ile yürürlüğe konması, (Karasoy & Babaoğlu, 2020, s. 122)
- 2007/7 sayılı Başbakanlık Genelgesi ile Dönüşüm Liderleri Kurulu oluşturulmuştur. Kurul, 2007/7 sayılı Başbakanlık Genelgesi ile yeniden yapılandırılmış ve kamu kurumu niteliğindeki meslek kuruluşları ve sivil toplum kuruluşlarının temsilcileri de kurul bünyesine dahil edilmiştir (Karasoy & Babaoğlu, 2020, s. 122)

2014-2018 yıllarını kapsayan Onuncu Kalkınma Planında e-Devlet konusunun bütüncül bir yaklaşımla ele alınması ve bu konuda bir eksiklik hissedildiği dile getirilmiştir (Kalkınma Bakanlığı, 2013: 55 aktaran Uysal, Kurban ve Çığman, 2023, s.

216). Uysal Kurban ve ıĖman (2023, s. 216) alıřmalarında bu konuda atılan ilk adımın Bilgi Toplumu Stratejisi ve Eylem Planı olduėunu belirtmektedir. Bundan bir yıl sonra 2016-2019 Ulusal e-Devlet Stratejisi ve Eylem Planı ilk bütüncül e-Devlet stratejisi olarak yayımlanmıştır.

Ülkemizde bütünsel bir yaklaşımla sürdürülmeyen, farklı kurumlarda ayrı ayrı yönetilen, dijital dönüşüm süreci, 2018 yılında T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi'nin (DDO) kurulmasıyla son bulmaktadır. Yine aynı yıl e-Devlet Kapısı Projesi ile vatandaşlara; hızlı, entegre ve etkili bir hizmet sunmayı amaçlayan e-devlet yapısı faaliyete geçirilmiştir.

DDO günümüzde de (2025) alıřmalarına aktif bir şekilde devam ederek Türkiye'de önemli projeler yaratmaktadır. Karasoy ve Babaoėlu (2020, s. 126) alıřmalarında DDO'nin 11 tane projesi olduėunu belirtmektedir (Karasoy ve Babaoėlu, 2020, s. 126). řu anda DDO, 24 tane proje yürütmekte olup; bunlara ilaveten dört tane de tamamlanmış projeleri bulunmaktadır. Sağlık, dijital okuryazarlık, eğitim, yapay zekâ, siber güvenlik gibi birçok alana katkı sunan bu projeler aşağıda sıralanmaktadır (Dijital Dönüşüm Ofisi, 2025).

DDO Yürütölen Projeler:

1. AçıkVeri
2. Dijital Türkiye
3. e-Yazışma
4. KamuNet
5. KAYSİS
6. Devlet Teřkilatı Merkezi Kayıt Sistemi (DETSİS)
7. TekDurak

8. Türk Beyin Projesi (TBP)
9. Ulusal Veri Sözlüğü
10. Siber Kümelenme
11. Siber Zekâ Bilgi Yarışması
12. Kamu Veri Alanı
13. Türksat-5A'nın Yolculuğu Dijital Resim Sergisi
14. Türkiye TechnoHub Platformu
15. Dijital İnovasyon İş Birliği Platformu
16. Dijital Dönüşüm Yetenek Merkezi (DDYM)
17. Yerli Siber Güvenlik Ürünlerinin Yaygınlaştırılması Platformu
18. Siber Güvenlik Meslek Yüksekokulları
19. DIGIATHON
20. Dijital Genç Yapay Zekâ Ekosistemi
21. Sayısal Göz Projesi
22. Kamuda Açık Kaynak Kodlu Yazılım Kullanımı
23. DEP4ALL
24. Dijital Dönüşüm Liderleri Gelişim Programı

DDO Tamamlanmış Projeler:

1. Fikir Maratonu
2. HackZeugma
3. 81 ilde 81 Siber Kahraman
4. HackIstanbul

II. 2. 3. Dijital Dönüşüm Kullanım Alanları ve Etkileri

COVID-19 pandemi süreci dijital dönüşüm hareketine büyük bir hız kazandırmıştır. Bu süreçte, insanların büyük bir kısmı çevrimiçi hale getirmiş ve dijital teknolojilerin kullanımı; kamusal, özel ve sosyal hayatın sürekliliğini garanti altına almak için vazgeçilmez hale gelmiştir (The Digital Regulation Platform, 2023). Wang ve arkadaşları (2023) çalışmasında dijital dönüşümün belirsiz koşullar altında organizasyonlara daha iyi verimlilik sağladığını belirtmektedir (Wang ve diğerleri, 2023, s. 2). Pandemi sürecindeki gelişmeler ise bu görüşü doğrular niteliktedir. Nitekim salgın sırasında organizasyonlar varlıklarını sürdürebilmek ve hizmetlerini/ürünlerini kullanıcılara/müşterilere sunabilmek için süreçlerini dijitalleştirmiş ve/veya mevcut dijital süreçlerini geliştirmişlerdir. Bu süreç organizasyonların yanı sıra toplumun hizmetten/üründen yararlanma alışkanlıklarını da dönüştürmüştür. Yaşamın her alanında bir dönüşüm süreci yaşanmakta ve süreçler evrimleşmektedir, dolayısı ile bu süreçler insanları etkilemekte ve sadece belirli bir kuruluşla sınırlı kalmamaktadır (Iivari, Sharma and Ventä-Olkkonen, 2020; aktaran Hai, Van, Tuyet, 2021, .s 27). Bu bağlamda dijital dönüşüm, bireysel düzeyden (çevrimiçi öğrenme, çalışma ve alışveriş yapma) tüm uluslara (dijital ekonomilere, yönetişime, topluma doğru geçiş), organizasyonlardan (yeni iş modelleri, yeni hizmetler, yeni teslimat yolları arayışı) tüm endüstrilere (işlem otomasyonuna geçiş, yapay zekanın avantajlarını keşfetme), yerel otoritelerden ulusal hükümetlere (daha şeffaf ve verimli yönetişim, dijital kamu hizmetleri sunma) kadar önemli dönüşümler yaşatmaktadır (The Digital Regulation Platform, 2023). Wang ve arkadaşlarının dijital dönüşümün belirsiz koşullarda daha iyi verim sağladığını ifade etmektedir. Pandemi süreci bitmiş olmasına rağmen o dönemde alınan verimlilik ile günümüzde organizasyonların; iş süreçlerini, toplantılarını, eğitimlerini, ürünlerini/hizmetlerini yeni dijital teknolojiler ile geliştirmeye devam ettirmesi, dijital

dönüşümün belirsiz koşullardaki etkisini daha iyi ortaya koymaktadır. Toplumun ise bu hizmet ve ürünleri dijital ortam aracılığı ile elde etmesi/kullanması göz önüne alındığında pandemi döneminin dijital dönüşümü hızlandırdığı ve toplum açısından da organizasyonlar açısından da verimliliğini sürdürdüğü ortaya çıkmaktadır. Artık bir seçenek olmaktan çıkan dijital dönüşüm, tüm kamu ve özel kuruluşlar için vazgeçilmez bir strateji haline gelmekte ve organizasyonların dijital dönüşüme hazır olma durumunu aydınlatmaktadır (Hai, Van, Tuyet, 2021, .s 27).

“Organizasyonlar, yapay zekadan blok zincirine ve nesnelerin internetine (IoT) kadar dijital dönüşüme bağlı olarak yeni çalışma şekillerine adapte olmak zorundadır.” (Hai, Van, Tuyet, 2021, .s 21). Günümüz teknolojilerinin geldiği noktaya ve bu mevcut durumdaki sosyolojik yapı göz önüne alındığında geleneksel yöntemlerde ısrarcı olmak, ürün ve hizmetlerin etkin sunulamaması ve rakip organizasyonlara rekabette güçsüzleşmeye neden olmaktadır. Bu durumun ise kurum içi/şirket içi bir süreksizlik yaratması söz konusudur. Dijital dönüşüm, her büyüklükteki organizasyonlar için, özellikle de dijital dönüşüm olmadan silinme riskiyle karşı karşıya olan işletmeler için hayati önem taşımaktadır (Thomas, 2019; aktaran Hai, Van, Tuyet, 2021, .s 21). “Günümüzde dijital dönüşüm bir kuruluşu her yönüyle etkilemektedir. Bu nedenle, kuruluşlar gerekli değişimi sağlamak için sürekli olarak bunun farkında olmalıdır. Bu bir varış noktası değil, yeni teknolojilerin kurumun ve toplumun değerini artırmak için sürekli olarak gelişeceği bir yolculuktur” (Boulton; aktaran Hai, Van, Tuyet, 2021, .s 25). Organizasyonlar sürekliliğini koruyabilmek için bazı gereklilikler bulunmaktadır. Bunlar: Bilgilerin taze tutulması, bilginin etkin kullanılması, yeni teknolojiler ile ürünler/hizmetler geliştirilmesi, yönetici desteğinin alınması ve çalışanlara yönelik dijital teknolojiler ile ilgili farkındalık ve eğitimlerde bulunması ve iş süreçleri ile kurumsal politikaları kapsayan süreçlere yeni teknolojilerin dahil edilmesidir. Dünyadaki en önemli

1000 şirketin yarısından fazlası, dijital dönüşümü karar ve iş süreçlerinin temeli haline getirmektedir (Nambisan, Wright ve Feldman; aktaran Ding et al., 2024, s.2).

Bu bağlamda, dijitalin tüm alanlardaki güçlü etkisi ortadadır. Bu etkinin gelecekte de artması beklenmektedir. Bu nedenle gelişmekte olan ülkelerin bu etkiyi dikkate alması gerekmektedir (Lan, 2019; aktaran Hai, Van, Tuyet, 2021, .s 21). Ülkeler kendi dijital dönüşüm politikalarını özenle oluşturmalı ve hem organizasyonlara hem de topluma benimsetmek için düzenlemeleri etkili bir şekilde gerçekleştirmelidir. Dijital dönüşüm için benzersiz ve evrensel bir şekilde oluşturulmuş tek bir stratejik plan yoktur. Her ülkenin kendine özgü siyasi, sosyal ve ekonomik yapısı bulunmaktadır. Bu bağlamda; bir dijital dönüşüm politikası; ülkelerin liderliği, kültürü, ekosistemin genel karmaşıklığı ve diğer faktörler de dahil edilecek şekilde hazırlanmalıdır (The Digital Regulation Platform, 2023). “Dijital dönüşüm stratejisi; deneyimlerin, fikirlerin, araçların ve mevcut kaynakların benzersiz bir karışımını temsil etmektedir” (The Digital Regulation Platform, 2023). Ülkede oluşturulacak dijital dönüşüm politikasını, ülke içerisinde birçok etken oluşturduğu gibi; dijital dönüşümde ülke içerisindeki iş süreçlerinden politikalara, ekonomik yapıdan sosyal düzene kadar her alanı dönüştürmektedir. “...Bazı yazarlar net bir görüş belirtmese de süreçler göstermektedir ki; dijital dönüşüm, yenilikçilik ve organizasyonlar ile bireylerin kapasitesi ile ilişkili bir süreçtir (Berghaus & Back, 2016; Warner, Wäger. 2019; aktaran Hai, Van, Tuyet, 2021, .s 24). Dolayısı ile söz konusu gelişmelerin, ülkeden ülkeye farklılık göstermesi doğaldır. Bu neden ile kurulacak politikaların da bu kıstaslar göz önüne alınarak hazırlanması gerekmektedir. “Dijital teknolojilerin kullanımındaki son patlama, bu konuyu tartışmaların önüne çıkartmıştır. Dijital dönüşüm herkesi ve çoğu faaliyet alanını etkilemektedir” (Kozarkiewicz, 2020; aktaran Hai, Van, Tuyet, 2021, .s 21). “Dijital teknolojinin gücü bir organizasyonun her alanına uygulanabilmektedir” (Tanguy et al., 2017; aktaran Hai, Van, Tuyet, 2021, .s 21).

Bu uygulama alanlarında hazırlanan politikalarsa, tıpkı ülke politikalarında olduğu gibi; sektörden sektöre, organizasyonların büyüklüğüne, verilen hizmetlere yönelik birçok kriter göz önüne alınarak, kurumsal yapıyla bütünleşir şekilde hazırlanmalıdır.

“Dijital dönüşüme ilişkin ilk çalışmalar genellikle teknoloji perspektifinden başlamakta ve dijital teknolojinin; üretim, operasyon veya hizmetler gibi ticari operasyonlarda uygulanmasına odaklanmaktadır” (Lee vd., 2015; Karagiannaki vd., 2017; aktaran Liu et al., 2023). Yapılan son çalışmalarda, dijital dönüşümün hem dijital teknolojilerin uygulanmasını hem de organizasyonların yeni nesil dijital teknolojileri kullanarak; üretim, operasyon ve ürünlerini/hizmetlerini yeniden düzenleme ve iyileştirme sürecini içerdiğini belirtmektedir (Nambisan et al., 2017; Vial, 2019; Hilali et al., 2020; AlNuaimi et al., 2022; aktaran Liu et al., 2023). Burada ise dijital dönüşüm çalışmalarının ilk başlarda sonuç odaklı başlamasına rağmen bu algının değişerek günümüzde süreç odaklı ele alınması göze çarpmaktadır. Dolayısı ile günümüz çalışmalarında dijital dönüşüm; “iş süreçlerinde, operasyonel rutinlerde, kurumsal yeteneklerde ve hatta işletmenin pazarlarında büyük değişikliklere yol açmaktadır” (Singh, Sharma ve Dhir, 2021; aktaran Hsiao, 2023). Tüm bunlar ortaya koymaktadır ki; teorik olarak dijital dönüşüm; yeni dijital teknolojileri; üretim yönetimi, organizasyon operasyonu, araştırma ve yenilik gibi çeşitli iş yönleriyle birleştirebilmektedir (Wang ve diğerleri, 2023, s. 2) Bununla birlikte dijital dönüşümün, ticari faaliyetlerde kullanımından önce, bu faaliyetlerin geliştirildiği iş süreçlerinde, politikalarda, fikir üretmede kısaca kurumsal yönetimin her alanında yer alması gerektiği ön plana çıkmaktadır.

“Örneğin, "iş süreci yeniden mühendisliği" (business process re-engineering) veya "kurumsal kaynak planlaması" (enterprise resource planning) organizasyon içerisindeki iş akışlarının ve süreçlerinin

analiz ve tasarımı odaklanırken, "hizmet tasarımı" (service design) ise müşteriler için hizmet yeniliği yaratmak adına kaynakların entegre edildiği sürecin yeniden tasarlanması üzerinde durur. Bu dönüşümler, kaynaklar (işleten ve işletilen kaynaklar), aktörler (tedarikçiler ve müşteriler), kurumlar ve değer gibi unsurları içeren organizasyon içerisinde süreçleri teknoloji kullanarak yeniden yapılandırmayı amaçlar" (Hsiao, 2023).

Bununla birlikte, Liu ve arkadaşları (2023) çalışmalarında, organizasyonların iş süreçlerine, dijital dönüşümü daha derin bir şekilde entegre ederek, kurumsal performans üzerindeki etkinin artacağını belirtmektedir. Bu etkileri ise aşağıdaki gibi sıralamaktadır:

- Maliyetlerin artması,
- Verimliliğin artması,
- Yenilikçiliğin teşvik edilmesi,
- Gelişmiş kurumsal performansın temelini oluşturması

Yine aynı çalışmada, kurumsal performansta dijital dönüşümden kaynaklanan üç ana gelişme ise;

- İşletme maliyetlerinin azaltılması,
- İşletme verimliliğinin artırılması,
- İnovasyonun teşvik edilmesi şeklinde ifade edilmektedir.

Wang ve arkadaşları da (2023, s. 2) entegrasyon etkisi ile ilgili çalışmasında etkileri şu şekilde belirtmektedir;

- Paydaşlar arasındaki algıyı ve onlardan bilgi edinilmesini iyileştirebilmekte, böylece organizasyonun bilgi açıklama kalitesini artırabilmektedir.

- Paydaşlar organizasyondan yeterli bilgi alabilmektedir. Böylece bilgi asimetrisi azalmakta (daha doğru, şeffaf ve güncel bilgi artmakta) ve organizasyonların çevresel, sosyal ve kurumsal yönetim (environmental, social, and governance (ESG)) sorumlulukları daha etkin bir şekilde sağlanmaktadır.
- Son olarak, bir organizasyonun finansal performansını iyileştirebilmektedir.

Çalışmalar ortaya koymaktadır ki; dijital dönüşümün iş süreçlerinin her adımında (araştırmalarda, fikir üretmede, politikalarda, ürün geliştirmede) etkin yer alması, organizasyona ekonomik anlamda değer katmaktadır. Bu durum, paydaşlar arasında sağlıklı bir bilgi paylaşımıyla da bilgi asimetrisini azaltarak; iş süreçlerini kolaylaştırmayı, hızlı ve etkin iletişim sağlamayı, iş süreçlerinin ve onun doğurduğu hizmetlerin/ürünlerin kalitesini arttırmayı sağlamaktadır.

Dijital dönüşüm organizasyonlara ekonomik büyüme ve piyasa sonuçları üzerinden olumlu sonuçlar doğursa da (Economic Impact of Broadband, Digitization and ICT Regulation (itu.int); aktaran The Digital Regulation Platform, 2023) tek olumlu etkisi organizasyonlar üzerine olmamaktadır. Organizasyonların dijital dönüşüm hareketleri ülkenin dijital dönüşüm politikası ile şekillenerek bütüncül, birlikte çalışır bir yapı kurulabilirse, bu kamu kurum ve kuruluşlarının ve dolayısı ile ülkenin yönetim noktasında faydalarını ortaya çıkarmaktadır. The Digital Regulation Platform'da (2023) bu faydalar şu şekilde sıralanmaktadır;

- Şeffaflığı ve hesap verebilirliği artırma,
- Bürokrasiyi, yolsuzluğu, vergi kaçakçılığını azaltma,
- Vatandaşların hükümetlerle etkileşimini kolaylaştırma,

- Toplum için sađlık ve eđitim hizmetlerinin sunulmasını iyileřtirme,
- Sosyal katılımı ve iletiřimi kolaylařtırma
- Refahı artırma

Sađlanacak faydalar gz nne alındıđında daha aık, adil, retken bir yapı kurabilmek adına devletlerin dijital dnřm noktasında teřvik ve politikaları ciddi nem tařımaktadır.

Organizasyonlara ekonomik katma deđer sađlama ve devletin ynetim yapısına yaptıđı katkılar yanında dijital dnřm, yeni dijital teknolojiler ile sreleri entegre ederek veri/bilgi/hizmet/rn noktalarında da ekonomik ynn dıřında bir katma deđer sađlamaktadır. Bu teknolojileri iř srelerinde kullanarak; karar vermede, yeniliki rn ve hizmetler ortaya koymada dijital dnřm yeni bir soluk getirmektedir. Ding ve arkadaşları (2024, s. 2) alıřmasında dijital dnřmn deđer yaratmada farkını řu řekilde ortaya koymaktadır;

Dijital dnřm;

- Dijitalleřmeyi endstriyel kalkınmayla birleřtirmektedir,
- İnovasyon ekosisteminin yeniden inřasını teřvik etmektedir,
- İřletmelerin sindirme, zmseme ve yeniden icat etme kapasitesini arttırmaktadır,
- Orijinal retim yntemlerini dijital teknoloji ve akıllı retimle deđerıřtirmenin tesine geerek kurumsal performans ve rekabet avantajını arttırmaktadır.

“Arařtırmacılar, firmaların dijital dnřmnn makroekonomik dzeyde verimlilik ve performans zerindeki olumlu etkilerini dođrulamaktadır” (Shahbaz, et al.,

2019; aktaran Hai, Van, Tuyet, 2021, .s 27). Hai, Van ve Tuyet (2021, s. 27) çalışmasında bu olumlu etkilei şu şekilde sıralamaktadır;

- Geleneksel iş akışlarının dijital süreçlerle değiştirilmesi,
- Başarıya odaklanmak yerine yeni bir geliştirme stratejisi araştırmaya daha fazla zaman ayrılması,
- Modern ve profesyonel çalışma modelinin değiştirilmesi, çalışanların manevi yaşamının sürekli iyileştirilmesi,
- İş akışı verimliliğinin artırılması ve teknik hataların en aza indirilmesi,
- Yeni hizmet ve teknolojilerin hızlı ve esnek bir şekilde uygulanması,
- İş kalitesi ve performansının artırılması,
- İşgücü verimliliğinin artırılması ve ürünlerin kalitesinin ve görünümünün iyileştirilmesi,
- Faydalancı memnuniyetini sürekli artırılması,
- Yeniden yatırım yeteneğinin artırılması,

“...Ulusal dijital dönüşüm, sağlık, eğitim, ulaşım, enerji, çevre, yönetim vb. gibi çeşitli alanları kapsayan ve çok sayıda paydaşa ve bunların çıkarlarına ilişkin karmaşık bir süreçtir. Önemli yatırım gerektirir, karmaşık etik ve yasal sorunları ele almayı ve sürekli değişen dijital ortamda alakalı kalmak için sürekli uyum sağlamayı zorunlu kılar. Süreçte yol almak, uygun araçların kullanılmasını ve kapsamlı ve koordineli yaklaşımları gerektirir” (The Digital Regulation Platform, 2023).

Bu nedenle dijital dönüşümde uygun politikaların oluşturulmasının oldukça önem taşımaktadır. Dijital dönüşüm her zaman olumlu etkiler göstermeyebilmektedir. Yanlış

uygulanan dijital dönüşüm politikaları olumsuz sonuçlar doğurabilmektedir. Bu olumsuzluklar aşağıda sıralanmaktadır;

- İşgücü kesintileri, siber güvenlik ve veri gizliliği sorunları,
- Çevrimdışı veya dijital olarak okuma yazma bilmeyen nüfusları daha da marjinalize etme ve olumsuz çevresel etkilere katkıda bulunma (örneğin artan e-atık hacmi, enerji tüketimi, donanım ürünleri için gereken doğal kaynakların madenciliği ve çıkarılması) (The Digital Regulation Platform, 2023),
- Bir işletmenin organizasyon yapısı dijital dönüşümle uyumlu olmaması durumunda çok sayıda yönetim maliyeti ve öğrenme maliyeti ortaya çıkması ve kurumsal performansın iyileştirilmesine zarar vermesi (Ekata, 2012; Hajli ve diğerleri, 2015; Ertani ve diğerleri, 2021; aktaran Liu et al., 2023).

“Bunlara ek olarak, küçük ve orta ölçekli işletmeler dijital dönüşüm sürecinde işgücünde beceri eksikliği, başarılı bir dijital dönüşümü sağlayacak kadar güçlü bir bilgi teknolojisi platformu eksikliği, özellikle de yeterli dijital düşünceye sahip liderlik eksikliği ve kültür zorlukları, işletmelerde dijital teknoloji gibi engellerle karşı karşıyadır. Ancak, dijital dönüşüme uyum sağlamak için işletmeler dijital dönüşüme giderek daha fazla katılmakta, bulut bilişim teknolojisi, ağ güvenliği ve yapay zekaya yatırım yapmaktadır” (Anh, 2020; aktaran Hai, Van, Tuyet, 2021, .s 26).

İşletmeler dışında kamu sektöründe de dijital dönüşüm sorunları mevcuttur.

Hai, Van ve Tuyet (2021) çalışmasında bu konuyu şu şekilde dile getirmektedir:

“Kamu sektörü kurumlarının bilgi ve iletişim teknolojileri sektöründeki faaliyetleri, e-devlet geliřtirmelerine uygulanarak idari reformlara katkıda bulunmuřtur, ancak sonuçlar etkileyici deęildir, iřlenen başvuru sayısı dūřüktür. Çevrimiçi iřlem hacmi hala dūřüktür, aę üzerinden iřlem iřleme hala sınırlıdır; ulusal veri tabanları yavaşça devreye alınmıřtır. Ayrıca, devlet kurumları arasında veri baęlantısı ve paylařımı hala sınırlıdır ve alıřma modelini ve yöntemini deęiřtirmek için ileri dijital teknolojinin uygulanması fazla deęildir. Bilgi ve iletişim teknolojilerinin uygulama ve geliřtirmelerinin oęu gerek anlamda dijital dnūřüm deęildir, henüz veri ve dijital teknolojilere dayalı modeller, üretim süreçleri, ürünler açısından ıęır açan bir dnūřüm yaratmamıřtır (Anh, 2020; aktaran Hai,Van, Tuyet, 2021, .s 26). Genel olarak, kamu sektöründe nispeten olumlu taraflar da vardır, bilgi ve iletişim teknolojilerinin uygulanması ve geliřtirilmesine dikkat edilmiř ve dnūřtürölmesi gereken birok alana yayılmıřtır, halkın yařam kalitesinin yükseltilmesine, sosyal uçurumların azaltılmasına, özellikle eęitim ve saęlık alanlarında katkı saęlamıřtır (Anh, 2020; aktaran Hai,Van, Tuyet, 2021, .s 26). Onlarca yıllık gerilięin ardından, hükümetler dijital kimlik, nüfus hakları ve dijital yařam tarzları alanındaki geliřmeler sayesinde hızlı dijital dnūřümü teřvik etmek için yoğun bir řekilde alıřmaktadır ve dijital dnūřümü ekonomiyi canlandırmanın anahtarı olarak görmektedir” (Hieu,2020; aktaran Hai,Van, Tuyet, 2021, .s 26).

III. BÖLÜM

METİN MADENCİLİĞİNİN TEMELLERİ: KÖKENİ, YÖNTEMLERİ VE UYGULAMA ALANLARI

III. 1. Metin Madenciliğine Genel Bakış

Çalışmanın bu bölümünde metin madenciliğinin ne olduğu, kapsamı, teknik ve yöntemleri ile uygulama alanları, literatüre dayalı olarak verilmektedir. Bu bölümde sunulan kavramsal çerçevenin bir sonraki bölümde irdelenecek olan -metin madenciliğinin belge yönetim sistemlerinde uygulanabilirliği ve nasıl bir değer yaratabileceği- sorularını yanıtlamaya katkı sağlayacağı düşünülmektedir.

Günümüzde verinin değeri tartışılmaz bir noktadır. Fakat aslında veri tek başına değerli değildir. Veriyi değerli yapan, veri analitiğidir; bir başka deyişle veriyi anlamlı, yararlanılabilir bilgiye dönüştürebilmektir. Veri analitiği; ham veriyi inceleyerek içgörüler (insight) elde etmeyi, karar destek süreçlerini iyileştirmeyi ve öngörülerde bulunmayı amaçlayan bir süreçtir. Veri toplama ve veri işleme teknolojilerinde kaydedilen gelişmeler sonucu yapılandırılmış verilerin ¹(structured data) yanı sıra metinsel, görsel vb formattaki yapılandırılmamış veriler de işlenebilir olmuştur. Dünyadaki verilerin %80 ila 90'ının yapılandırılmamış verilerden ²(unstructured data) oluştuğu söylenmektedir (Harbert, 2021). Yapılandırılmamış metinsel verilerin; işleme ve analizi, sayısal içerikle

¹ Veritabanı destekli sistemlerden standart bir düzende sunulan veriler (örneğin, bir veritabanından biçimlendirilmiş veri kümeleri) (Svård, 2017, s. 28).

² İçeriklerine doğrudan başvurulamayan ve içerik, düzen ve meta veri ayrımı olmayan her türlü bilgi nesnesi (örneğin, resimler, GIF'ler, videolar, metinler) (Svård, 2017, s. 28).

başı çıkmaktan çok daha zordur (Chen, 2001, s. 18). Yapılandırılmamış metinsel verilerin metin içinden çıkarılmayı bekleyen önemli bilgileri barındırdığı düşünülmektedir ve bu yönüyle önemli bir veri kaynağı olarak kabul edilmektedir. Bu bağlamda metin madenciliği, büyük miktarda yapılandırılmamış metin verisini işleyerek, içindeki anlamlı bilgileri çıkarmayı amaçlayan bir veri analitiği dalı olarak karşımıza çıkmaktadır. Metin madenciliği, doğal dil işleme (NLP - Natural Language Processing), makine öğrenmesi (machine learning) ve istatistiksel yöntemler (statistical methods) kullanarak metinlerden örüntüler (pattern), ilişkiler (relationship) ve içgörüler elde etmeyi amaçlamaktadır.

Literatürde metin madenciliği (Text Mining), geniş anlamda yapılandırılmamış metinlerden anlamlı bilgi ve kalıpları çıkarma genel süreci olarak tanımlanmaktadır. Metin analizi (Text Analysis) bu terimle sıklıkla eş anlamlı kullanılmaktadır. (Pollitt, 6.06.2025). Metin analitiği (Text Analytics) ise, metin verilerinden eğilimleri ve istatistikleri, yani nicel sonuçları belirleme sürecini tanımlamaktadır (Speak Ai, 6.06.2025). Bu terimlerin işaret ettiği kavramlar; metinsel verilerden içgörüler çıkarmak bakımından ilişkili olmakla birlikte farklı süreçleri ifade etmektedir. Metin madenciliği, büyük hacimli yapılandırılmamış metinlerden örüntüler, eğilimler ve bilgi keşfetmek için doğal dil işleme ve makine öğrenme tekniklerinin uygulanmasını içermektedir. Öte yandan metin analitiği, genellikle istatistiksel ve dilbilimsel yöntemler kullanarak metinden anlamlı bilgiler, duygular ve bağlam çıkarmaya odaklanmaktadır. Metin madenciliği gizli örüntüleri ortaya çıkarmayı vurgularken, metin analitiği karar alma için eyleme dönüştürülebilir içgörüler türetmeyi vurgulamaktadır. Her ikisi de yapılandırılmamış metni değerli bilgiye dönüştürmede önemli roller oynamaktadır: metin madenciliği örüntüleri keşfetmekte, metin analitiği yorumlayıcı bağlam sağlamaktadır (GeeksforGeeks, 2024b).

Brecque (2024) tarafından hazırlanan "Text Analytics Vs Text Mining" kaynağına göre; metin madenciliği daha çok "niteliksel sonuçlar" verirken, metin analitiği ise niceliksel sonuçlar (eğilimler ve istatistikler) vermektedir. Çalışma genel anlamda EBYS'lerin kullanım trendlerini anlama ve iyileştirme odaklıdır. Dolayısı ile büyük oranda bilgi erişim sürecini geliştirme gayreti taşıyan bu çalışma, nicelikten çok nitelik ile ilgilenmektedir. Çalışmanın niteliksel sonuçlar üretmesi, gizli bilgi ve örüntüleri keşfetmesi, ilişki çıkarımına odaklanması ve doğrudan bilgi erişim ile ilişkili olması, bu çalışma için en doğru yöntemin metin madenciliği olduğunu göstermektedir.

Günümüzde metinsel verilerin analizini ifade eden ve “veri madenciliğinin bir alt dalı olarak görülen” (Atan, 2020, 224) metin madenciliği, kullandığı birçok teknik bakımından, daha 1950’lerde gerçekleştirilen otomatik dizinleme algoritmalarına dayanmaktadır. Metin formatındaki yapılandırılmamış verilerin sayısal verilere dönüştürülmesi ile başlayan otomatik dizinleme çalışmalarındaki motivasyon, katlanarak artan bilimsel makale sayısı karşısında dokümanların içeriğini belirleyen sözcüklerin metinden otomatik olarak çıkarılması ve bu içerik belirteçleri ile temsil edilen dokümanların erişilebilmesini sağlamak olmuştur; “dizinleme” ve “erişim”, bilgi erişim sistemlerinin temelini oluşturmuştur. Otomatik dizinleme çalışmaları bilgi erişim sorunlarının çözümüne yönelik olarak Luhn’un geliştirmiş olduğu KWIC (Key Word in Context) dizinleme yöntemi ile Ohlman’ın geliştirdiği permütasyon dizinleme sistemine dayanmaktadır. İki sistem/yöntem neredeyse eş zamanlı olarak ortaya çıkmış olmasına ve benzerlikleri bulunmasına rağmen Luhn’un KWIC yöntemi geniş çevrelerce benimsenmiş ve daha yaygın bir kullanım alanı bulmuştur (Williams, 2010).

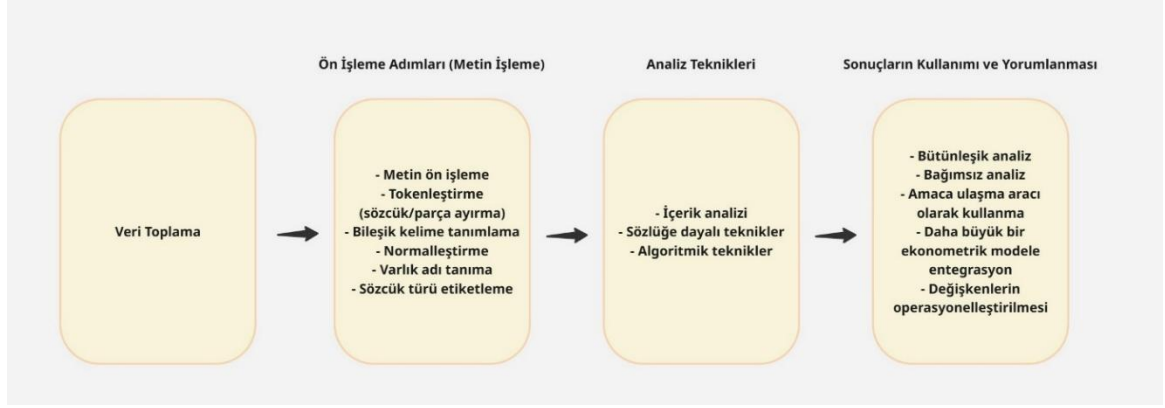
Konu dizinlemesi, temelde bir dokümanın içeriğinin hangi terimlerle temsil edileceğini belirlemektir. Amaç bu erişim uçları ile kullanıcının aradığı bilgiyi içeren dokümanlarla eşleşmenin sağlanması ve ilgili dokümanlara erişebilmektir (Stapleton ve

diğerleri, 2021, s. 794). Otomatik dizinleme bir dizi tekniğin uygulandığı bir süreçtir. Bu sürecin ilk aşaması ön işlemedir. Ön işleme; bir metni noktalama işaretlerinden arındırma, bir karakter dizisi olarak birim sözcükleri elde etme (tokenizasyon), durma listesi (stop-words)nde yer alan sözcüklerin metinden elimine edilmesi, terimlerin görünüm sıklığı (term frequency)nı hesaplama ve gövdeleme (stemming) gibi adımlardan oluşur. Böylelikle doküman(ların) içeriğini temsil etmeye aday terimlerden oluşan bir liste elde edilmiş olur (Willis ve Losee, 2013, s. 1331). Bir sonraki aşamada, edinilen terimler ağırlıklandırma gibi çeşitli tekniklere göre dokümanları temsil etmek üzere belirlenir ve dokümanlara atanır; sorgulama ile etkin bir erişim sağlanmaya çalışılır. Bütün bu süreçte terimlerin dokümanın içeriğini temsil gücünü ve erişimin etkinliğini artırmayı amaçlayan birçok hesaplama yöntemi geliştirilmiştir. Örneğin idf(inverse document frequency) terimin derme içindeki gücünü artırmak; benzerlik hesaplamaları ise sorgu-doküman veya doküman-doküman sınıfları oluşturmak için kullanılmıştır. Böylelikle terim sınıfları veya doküman sınıfları oluşturulabilmiştir. İlgililik ise bilgi erişim sistemlerinin etkinliğini erişim isabeti ve kesin erişim isabeti ile değerlendirmeyi esas alan bir yaklaşımdır (Salton ve diğerleri, 1980, s. 175).

Bu bağlamda başlangıçta otomatik dizinleme, sonrasında ise bilgi erişim kapsamında kullanılan birçok tekniğin günümüzde metin madenciliği alanında kullanıldığı görülmektedir. Geleneksel metin madenciliği sürecini betimleyen Şekil 8'e bakıldığında bu ilişki rahatlıkla görülebilmektedir. Otomatik dizinleme; metinler üzerine ön işlem uygulama, içeriği temsil eden uygun terimleri belirleme ve bu terimlerin temsil gücünü artırmada kullandığı hesaplama yöntemleri ile metinsel verilerin sayısal veriye dönüştürülmesine önemli katkılar sağlamıştır.

Şekil 8.

Metin madenciliği süreci



(GeeksforGeeks, 2024b)

Öte yandan metin madenciliği terimler arasında ilişki kurma, örtük bilgileri bulma, yeni örüntü ve bilgiler ortaya çıkarma gibi yeni yaklaşımlar sunmaktadır (Williams, 2010). Bunun yanı sıra metin madenciliği, sınıflandırma, kümeleme ve ilişki kuralı madenciliği gibi veri madenciliği tekniklerini kullanmaktadır. Metin madenciliği bağımsız bir süreç olarak uygulanabildiği gibi veri madenciliği için bir ön işleme adımı olarak da kullanılabilir.

Tablo 2 metin madenciliğinde kullanılan algoritma ve teknikleri özetlemektedir.

Tablo 2.

Metin madenciliği tekniklerine genel bakış

Metin Madenciliği Süreci Aşaması	Algoritma	Seçilen Soru	Motive	Teknikler
Metin Ön İşlem aşaması	Tokenizasyon	Bir metni kelimelere veya metin formatına nasıl dönüştürebilirim?	Dizeleri tek bir metin belirticine aktarma	Boşluk ayırma
	Bileşik kelime tanımlama	Ortak anlamı olan kelimeleri nasıl belirleyebilirim?	Ortak anlamı kaybolan kelimeleri belirleme	N-gram
İçerik Analizi	Normalleştirme Ve gürültü azaltma	Belge-Terim-Matrisindeki çok fazla değişkenle nasıl başa çıkabilirim?	Belge-Terim-Matrisinin boyutunu azaltma	Kök çıkarma, Lemmatizasyon, Durdurma kelimelerinin silinmesi
	Dilbilimsel analiz	Özel bir anlamı veya gramer işlevi olan kelimeleri nasıl belirleyebilirim?	Kelimelerin etiketlenmesi	Adlandırılmış varlık tanıma, Sözcük türü etiketleme
	Sözlük tabanlı	Gizil sosyolojik veya psikolojik özelliklerin ve durumların doğal dilde nasıl yansıtıldığını nasıl belirleyebilirim?	Bağlamsal, psikolojik, dilbilimsel veya anlamsal kavram ve yapıların ölçülmesi	Önceden tanımlanmış sözlükler ve Özelleştirilmiş sözlükler
İçerik Analizi		Önceden tanımlanmış sınıflara metinleri nasıl atayabilirim?	Metinsel varlıkları önceden tanımlanmış kategorilere sınıflandırma	İkili veya çok sınıflı sınıflandırıcılar gibi denetimli öğrenme teknikleri
	Algoritmik teknikler	Benzer belgeleri nasıl gruplandırabilirim?	Metinsel varlıkların daha önce tanımlanmamış ve bilinmeyen kümeler halinde gruplandırılması	LDA, k-means veya non-negative gibi denetimsiz öğrenme teknikleri

(GeeksforGeeks, 2024b)

Metin madenciliği, metinsel verilerden bilgi, örüntü ve ilişki çıkarmayı amaçlamaktadır ve bu amaçla bazı veri madenciliği teknik ve algoritmalarını kullanmaktadır. Dolayısı ile metin madenciliğini daha iyi anlamak için öncelikle “veri madenciliği nedir, teknikleri nelerdir, hangi algoritmaları kullanır, ne amaçla ve hangi alanlarda kullanılır” soruları üzerinde durmak konuya açıklık getirecektir.

III. 2. Veri Madenciliği: Tanım, Amaç ve Veri Ön İşleme Aşamaları

Veri madenciliği, nicel nitelikteki yapılandırılmış verileri hedefleyen ve bu verilerden güvenilir bilgi toplayarak organizasyonlara yardımcı olan bir süreçtir (Li ve Yanchinda, 2024; Malik ve Gupta, 2022, s. 30). Veri madenciliği ile veri tabanlarında bulunan yapılandırılmış veriler analiz edilerek; daha önce tanımlanamamış olan örüntü ve kuralları ortaya çıkarılmaktadır. Veri madenciliği uygulamaları 1980’lerde araştırma güdümlü olarak başlamıştır; ancak son zamanlarda bu alan istatistikçiler, veri tabanı araştırmacıları, yönetim bilişim sistemleri ve iş çevreleri gibi farklı topluluklar tarafından domine edilmektedir (Mirza ve diğerleri, 2016, s. 437). Belki bu nedenle literatürde veri madenciliğinin farklı isimlerle adlandırıldığı görülmektedir. Bunlardan bazıları; “bilgi çıkarımı (information extraction), bilgi keşfi (knowledge discovery), bilgi hasadı (information harvesting), veri arkeolojisi (data archeology) ve veri örüntüsü işleme (data pattern processing)”dir (Fayyad, et al, 1996; aktaran Jackson, 2002, s. 267).

Mirza ve arkadaşlarının veri madenciliği için “daha önce tanımlanamamış olan önemli örüntüleri ve kuralları ortaya çıkarmak amacıyla devasa veri kümelerinin analizi ve incelenmesi” olarak yaptığı tanıma geri dönecek olursak “daha önce tanımlanamayan bilgi” ifadesinin dikkat çekici olduğu görülmektedir. Bunun anlamı; daha önce kayda alınmamış, ancak veriler arasında ilişkiler kurulması sonucunda ortaya çıkan örtük bilgidir (tacit knowledge) ki bu bizi bilgi keşfine götürmektedir. Bu bağlamda veri madenciliğinin, organizasyonlar için son derece önemli olduğu söylenebilmektedir; veri madenciliği uygulamaları aracılığı ile organizasyonlar rekabet üstünlüğü sağlayabilmekte, karar vericiler ve pazarlamacılar tarafından stratejik bir unsur olarak kullanılabilir. Jackson (2002, s.270) çalışmasında veri madenciliğinin organizasyonların kendi veri tabanlarına ve mevcut verilerine odaklanmalarını sağladığını söylemektedir. Bir organizasyonun elindeki veriler ne kadar çok ise (kaliteli,

çeşitli ve doğru veriler olması koşulu ile) o kadar çok ilişki kurulabilmekte ve bilgi çıkarılabilmektedir. Böylelikle daha doğru ve güçlü analizler gerçekleştirilerek daha etkin bir pazarlama ve karar alma süreci yaşanabilmektedir. Benzer şekilde Lei-da Chen ve Frolick de çalışmalarında (2000) geleneksel yöntemler ile karşılaştırıldığında veri madenciliğinin, bir organizasyonun iç ve dış veri kaynaklarının analiziyle kendisini, müşterilerini ve rekabet ettiği pazarı daha iyi anlayacak şekilde bilgi üretebilme özelliğinden bahsetmektedir. Bu bağlamda organizasyonların sahip olduğu çok miktarda verinin, veri madenciliği ile işlenmesinin organizasyona avantaj sağlayacağı açıktır.

Gelişen teknoloji ile neredeyse her an her yerde veri üretilmektedir. Organizasyonlar adeta veri bombardımanı ile karşı karşıya kalmıştır; bu durumla başa çıkabilmeleri için analitik bakış açısıyla verileri analiz etme ve faydalı bilgiler ortaya çıkarma eğilimi doğmuştur. Analitik bakış açısı; verilerin analizi yoluyla ilişkilerin kurulması, kurulan ilişkiler sonucu yeni bilgiler elde edilerek öngörü sağlanmasıdır. Elde edilen yeni bilgilerin iş süreçlerine uygulanarak organizasyona fayda sağlaması beklenmektedir. İlişkiler kurulması ve yeni bilgi ve örüntülerin elde edilmesi karar almada analitik bakış açısını gerektirmektedir. Örneğin: veri madenciliği kullanan iş profesyonelleri; müşteri davranışları ile ilgili örüntüleri öğrendiklerinde gelecekteki müşteri eğilimlerini tahmin edebilecek ve satışları öngörebilecektir. Bu doğrultuda veri madenciliği organizasyona değerli müşterilerini tanımlama ve değerlendirme bilgisi sağlaması, bunun yanı sıra kendi geleceğini şekillendirme farkındalığı kazandırması ve karını artıracak kararlar almasına yardımcı olacağından organizasyonlar tarafından büyük ilgi görmektedir (Soni, 2015; Aktaran Mirza ve diğerleri, 2016, s. 438).

Lei-da Chen ve Frolick ise bu görüşleri destekler biçimde çalışmalarında (2000) veri madenciliğinin bir pazarlama aracı olarak kullanılmasının sağladığı genel sonuçları şu şekilde sıralamıştır:

- daha keskin bir rekabet avantajı,
- kanıta dayalı bir satış yaklaşımı,
- müşteri odaklı bir pazarlama planı,
- daha kısa satış döngüleri ve daha düşük operasyonel maliyetler.

Malik ve Gupta ise (2022, s. 30) benzer şekilde veri madenciliğinin organizasyonlara kattığı faydaları aşağıdaki gibi sıralamaktadır:

- Şirketlerin güvenilir bilgi toplamasını sağlar,
- Diğer veri uygulamalarıyla karşılaştırıldığında verimli ve uygun maliyetli bir çözüm olarak kabul edilir,
- İşletmenin daha karlı üretim ve diğer operasyonel ayarlamalar yapmasına yardımcı olur,
- Hem eski hem de yeni sistemleri kullanarak genel işletmenin bilinçli kararlar almasına destek olur.

Veri madenciliği, veri tabanlarında bilgi keşfi (Knowledge Discovery in Databases/ KDD) sürecindeki bir adımı oluşturmaktadır. Büyük ölçekli veri tabanlarından; yeni, anlamlı ve yararlı bilgi elde etmek amacıyla; veriler içinde gizli kalmış örüntüler aramakta ve en iyi eylemin belirlenmesini sağlayan içgörü elde etmeyi amaçlamaktadır.

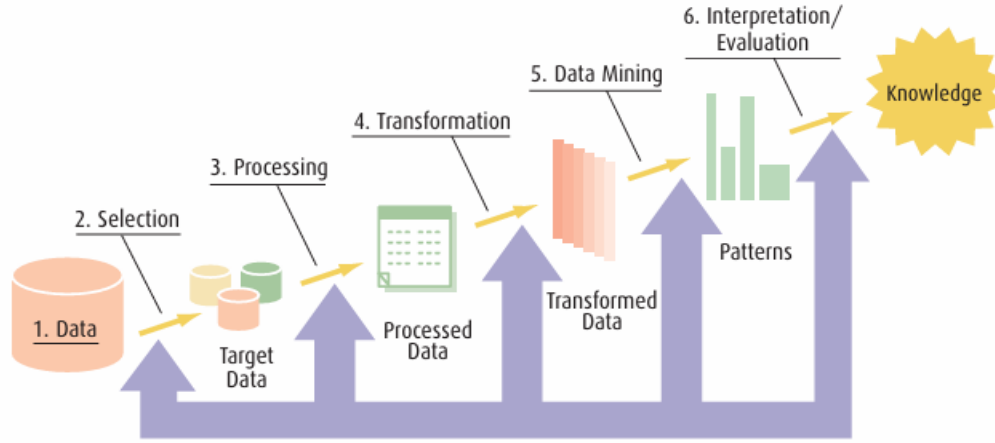
Çalışmalarında büyük veri kümelerini işlemede yaygın olarak kullanılan veri madenciliği algoritmalarını değerlendiren Joseph ve arkadaşları (2016, s. 6806), veri madenciliğinin, veri tabanlarında bilgi keşfinin (KDD) ayrılmaz bir parçası olduğunu söylemektedir; başarı ve iyileştirilmiş karar almanın, içgörü keşfetme hızına bağlı olduğuna dikkat çekmekte ve bu içgörülerin, operasyonel süreçlerde kullanılabilecek daha iyi eylemleri yönlendirmek ve hatta gelecekteki davranışlar ile ilgili öngörü elde

etmek için kullanılabileceğini belirtmektedir. Yazarlar aynı zamanda mevcut soruna bağlı olarak algoritma seçmenin önemli olduğunu, çünkü algoritma seçimine göre farklı sonuçlar elde edilebileceğini vurgulamaktadır. Bu doğrultuda algoritmanın ne olduğunu hatırlamakta yarar vardır. Veri madenciliğinde (veya makine öğreniminde) bir algoritma, verilerden bir model oluşturan bir dizi sezgisel yöntem ve hesaplama adıdır. Algoritma, model oluşturmak için öncelikle edinilen verileri analiz etmekte, belirli türdeki örüntüleri veya eğilimleri aramakta; bu analizi birçok kez yineleyerek oluşturulacak model için en uygun parametreleri bulmaya çalışmaktadır. Daha sonra bu parametreler, eyleme dönüştürülebilir örüntüler ve ayrıntılı istatistikler çıkarmak için tüm veri kümesine uygulanmaktadır (Microsoft, 2023).

Silwattananusarn ve Tuamsuk (2012, s. 13) veri madenciliğinin KDD sürecinin en önemli adımlarından biri olduğunu ve bilgi yönetimi (knowledge management)nin de önemli bir alt alanı olduğunu belirtmektedir. Yazarlara göre KDD, verilerden yararlı bilgi keşfetmenin genel sürecini ifade ederken veri madenciliği, yararlı bilgi çıkarmak için algoritmalara odaklanarak veri tabanlarındaki veri zenginliğinden yeni örüntüler ve modeller keşfetmekle ilgilidir (Han ve Kamber, 2012; Aktaran Silwattananusarn ve Tuamsuk, 2012, s. 15; Fayyad ve diğerleri, 1996; Aktaran Silwattananusarn ve Tuamsuk, 2012, s. 15; Mirza ve diğerleri, 2016, s. 437).

Şekil 9.

Veri tabanından bilgi keşfi süreci



(Chen, 2001, s. 13).

III. 2. 1. Veri Ön İşleme Süreci

Bir veri madenciliği sürecinde verilere, teknik ve algoritmaların uygulanmasından önce verilerin önışlem sürecinden geçmesi önem arz etmektedir. Veri madenciliği sürecinde doğru sonuçların (bilgilerin) elde edilebilmesi; doğru ve kaliteli verilerin işlenmesini gerektirmektedir. Bu bağlamda verilerin; hatalı, eksik, tutarsız, gürültülü veya gereksiz bilgilerden arınmış olması nitelikli sonuçlar üretilebilmesi için önemlidir. Verilerin kaliteli ve kullanışlı hale getirilmesi ise önışleme süreci ile ilişkilidir (García ve diğerleri, 2016, s.2). Oğuzlar'ın (2003, s, 69) çalışmasında belirttiği üzere; veri madenciliği uygulamalarında sürecin %80'i verilerin önışlemden geçirilmesi ve temizlenmesine harcanmaktadır.

Veri önışleme süreci temel olarak üç ana başlık altında sınıflandırılmaktadır. Bunlar: Veri Temizleme (Data Cleaning), Veri Dönüştürme (Data Transformation) ve Veri İndirgeme/Azaltma (Data Reduction)dır (Çetin ve Yıldız, 2022, s. 299).

III. 2. 1. 1. Veri Temizleme (Data Cleaning): Eksik değerlerin doldurulması (örneğin ortalama, medyan, mod ile) ve gürültülü (hatalı) verilerin düzeltilmesi veya kaldırılması (örneğin Binning, Kümeleme veya Regresyon ile) işlemlerini içerir (Yang, 2018; Çetin ve Yıldız, 2022).

III. 2. 1. 2. Veri Dönüştürme (Data Transformation): Verileri analiz için uygun formatlara dönüştürme işlemidir. En yaygın yöntemlerden biri, verileri belirli bir aralığa ölçeklendiren Normalleştirme (Min-Max, Z-skor, Ondalık Ölçekleme). Ayrıca, verilerin özetlenmesi olan Birleştirme (Aggregation) de bu kategoriye girer (Yang, 2018; Çetin ve Yıldız, 2022).

III. 2. 1. 3. Veri İndirgeme (Data Reduction): Veri setinin boyutunu azaltmayı hedeflerken kalitesini korur. Bu, Özellik Seçimi (Feature Selection) ile (gereksiz sütunları azaltma) veya Örnek Seçimi (Instance Selection) ile (gereksiz satırları azaltma) gerçekleştirilebilir. Boyut indirgeme ve veri sıkıştırma da bu kapsama girer (Yang, 2018; Çetin ve Yıldız, 2022).

Özetlemek gerekirse, veri ön işleme, kaliteli veri analizi ve madenciliği sonuçları elde etmek için veriyi hazırlayan kritik bir adımı oluşturmaktadır. Organizasyonlar ise bu verileri işleyerek; içgörü elde etmek, öngörü ve tahminde bulunmak ve veriye dayalı kararlar alabilmek, bir başka deyişle verilerden yararlı bilgi elde edebilmek için veri madenciliği tekniklerini giderek daha fazla kullanmaya başlamışlardır.

III. 2. 2. Veri Madenciliği Teknikleri

Literatür incelendiğinde veri madenciliğinde kullanılan algoritmaların ele alınmasında farklı yaklaşımların benimsendiği görülmektedir. Bu yaklaşımlar soruna,

alana göre deęiřebildięi gibi algoritmaların sistematik gruplanmasına göre de deęiřebilmektedir. Örneęin: Mirza ve arkadaşları (2016, s. 440) alışmalarında veri madencilięinin öngörü (prediction) ve tanımlama (description) olmak üzere iki temel ve öncelikli amacı bulunduęunu belirtmektedir ve algoritmaları buna göre gruplamaktadır.

Öngörü, “gelecekte ne olacak” sorusuna yanıt bulmaya alışır; sorunun yanıtına mevcut veri setlerindeki deęiřkenler üzerinde istatistiksel yöntemlerin uygulanmasıyla ulařılır. Tanımlama ise; bir veri kümesinin tamamında ve/veya alt kümelerinde örüntüleri, eęilimleri ve iliřkileri bulmayla ilgilidir. Öngörü için genellikle sınıflandırma (classification), regresyon (regression) ve anormallik tespiti (anomaly detection) teknikleri kullanılır iken, tanımlamada kümeleme (clustering), birliktelik kuralı öęrenimi (association rule learning) ve özetleme (summarization) teknikleri kullanılmaktadır (Güven, 2007, s. 12; Gorunescu, 2011; Aktaran Silwattananusarn ve Tuamsuk, 2012, s. 16). Bu terimler alışmanın ileriki bölümlerinde daha detaylı açıklanmaktadır.

Silwattananusarn ve Tuamsuk’un alışmalarında bir alan bağlamında uygulanma biimlerine göre algoritmalar gruplanmıřtır. Yazarlar bilgi yönetiminde veri madencilięi uygulamalarını belirlemeye yönelik bir literatür deęerlendirme alışması yapmıřlardır ve incelemiř oldukları literatüre dayanarak bilgi yönetimi alanında uygulanan altı veri madencilięi teknięi saptamıřlardır. Bu teknikler; sınıflandırma, regresyon, kümeleme, bağımlılık modelleme, sapma tespiti ve özetlemedir. (2012, s. 14)

Joseph ve arkadaşları ise, “Data Mining Algorithms: An Overview” bařlıklı alışmalarında algoritmaları denetimli ve denetimsiz öęrenme algoritmaları olarak

kategorize etmiş ve tablo³ halinde özetlemişlerdir. Buna göre denetimli öğrenme algoritmaları: Sınıflandırma(karar öğrenmesi, Naive Bayes Classifiers, K-Nearest Neighbour, Support Vector Machine) ve Regresyon(Linear Regression, Logistic Regression)dur. Denetimsiz öğrenme algoritması ise Kümeleme(k-means, density based, apriori)dir.

Aşağıda genelde alanımız ve özelde EBYS sistemleri için sorunların çözümünde uygulanmasının yararlı olacağı düşünülen bazı tekniklere ilişkin açıklamalara yer verilmiştir;

III. 2. 2. 1. Sınıflandırma (Classification): Makine öğrenmesi üzerine kurulmuş klasik bir veri madenciliği tekniğidir. Sınıflandırmada, sınıflar önceden tanımlanır ve etiketlenir; yeni veri ögeleri bir dizi uygun sınıftan bir ya da birkaçına yerleştirilir. Bir sınıf ortak özelliklere sahip verilerden oluşmaktadır. Sınıflandırmanın amacı; eğitim verilerini analiz ederek ve verilerde bulunan özellikleri kullanarak, verileri önceden belirlenen, birkaç sınıfa ayıran algoritmanın öğrenilmesidir (Joseph ve diğerleri, 2016, s. 6807; Ming-Syan ve diğerleri, 1996; Aktaran Mirza ve diğerleri, 2016, s. 440; Hand vd., 2001; Aktaran Güven, 2007, s. 11). Sınıflandırmaya örnek olarak, sağlık sektöründe hastalara tanı konulması verilebilir. Hastaların göstermiş olduğu semptomlardan yola çıkarak yapılan analizlerle hastalığa en uygun teşhis konulabilmektedir. Böylelikle sağlık alanında sınıflandırma tekniği ile otomatik tanı konulması başarılabılır ve bu, hekimler için büyük kolaylık sağlayabilir. Bunun yanı sıra tespit edilmesi güç ve tecrübe gerektiren karmaşık durumlar için daha hızlı teşhis konulması mümkün olabilir.

³ Tablo için bkz. Joseph, S. R., Hlomani, H., & Letsholo, K. (2016). Data mining algorithms: an overview. *Neuroscience*, 12(3), 719-43.

III. 2. 2. 2. Regresyon(Regression): Regresyon, bir bağımlı değişken ile bağımsız değişkenlere ait değerler arasında bağlantı kuran ve bu sayede bir değişkenin değerine bakarak diğerinin değerini bulan, öngöründe bulunan istatistiksel bir araçtır (Hand vd., 2001; Aktaran Güven, 2007, s. 11; Jackson, 2002, s. 275). Regresyonda öngöründe bulunulan değişken sürekli değişken (belirli bir aralıkta sonsuz sayıda değer alabilen ve sayılamayan rastgele değişken) olabilir. Örneğin; Bir şirket; eğilimleri (trendleri) belirleme, öngöründe bulunma, sapmaları tespit etme ve ilişki kurma gibi amaçlar ile borsa hisselerinin değerini tahmin etmek isterse regresyon kullanılabilir (Joseph ve diğerleri, 2016, s. 6808). Regresyon teknikleri arasında yer alan lojistik regresyon, karar ağaçları veya sinir ağları (neural network) gibi karmaşık teknikler, gelecekteki değerleri tahmin etmek için kullanılabilir. Bu teknikler, daha etkin sonuçlar elde etmek amacıyla farklı kombinasyonlar ile kullanılabilir (Mirza ve diğerleri, s. 440, 2016).

III. 2. 2. 3. Kümeleme (Clustering): Kümeleme, belgelerin bir sınıflama sistemine ihtiyaç duyulmadan içeriklerine göre gruplandırılmasını ifade etmektedir (Schieber ve Kruses, 2014, S. 24). Somut veya soyut nesneleri benzer nesnelerden oluşan sınıflara bölmektir. Kümeleme, veri kümelerini (kayıtlar/demetler/nesneler/örnekler) benzerliklerine göre birkaç gruba (küme) ayırma yöntemidir. Kümelemenin temel amacı, yakınlığa dayalı nesne gruplarını (kümelerini) bulmaktır. Böylece bireysel kümeler içinde birbirlerine büyük benzerlikler olur ve kümeler birbirinden yeterince farklı olur (Gorunescu, 2011; Aktaran Mirza ve diğerleri, 2016, s. 440)

Burada mesafe kavramı benzerlik ya da yakınlık ile ilgilidir ve başlangıç için küme değeri ve küme sayısı rastgele seçilir. Mesafe uzunsa benzerlik azalır, mesafe kısaysa benzerlik artar. Kümelemede veriler genelde etiketsizdir ve kümeleme makine öğrenmesinde bir denetimsiz öğrenme şeklidir. (Joseph ve diğerleri, 2016 s. 6811; Hand ve diğerleri., 2001, s. 31). Dolayısı ile her bir nesnenin etiketi kümeleme algoritması

tarafından önceden bilinmemektedir. Kümelemede benzerlik hesaplanarak nesneler yakınlıklarına göre gruplanmakta/kümelenmektedir.

Kümeleme sayesinde, organizasyonlar için önemli içgörüler elde edilebilmektedir. Örneğin, müşterilerin bir şirketin ürünleriyle ilgili tanımladığı şikâyet veya öneriler hakkında bilgi edinmek mümkün olur (Schieber ve Kruses, 2014, S. 24). Bu yolla veri madenciliği bir ürün ile ilgili; puanlama, zaman bilgisi, müşteri profili gibi bilgilere küme analizi uygulayarak, anlamlı yapılar elde edebilir.

III. 2. 2. 4. Bağımlılık Modelleme/Birliktelik Kuralı Madenciliği (Dependency

Modelling /Association Rule Mining): Denetimsiz veri madenciliğinin en yaygın bilinen tekniği olan birliktelik kuralı madenciliği, büyük bir veri kümesine ait öğeler veya olayların bir veya daha fazla yönü arasında anlamlı ve önemli ilişki ve bağlantıları tanımlayan ve bunları etiketleyen bir model bulmayı amaçlamaktadır (Mirza ve diğerleri, 2016, s. 440; Jackson2002, s. 275). Birliktelik kuralı madenciliği aynı zamanda sınıflandırma ve tahmin ile de yakın ilişkilere sahip olup bilinen bir varlık grubunun öznitelikleri ile bu varlıkların değerini tahmin etmek için de kullanılabilir (Jackson, 2002, ss. 273-275). Belirli ürünleri satın alan müşterileri analiz etmeye ve müşterilerin sıklıkla birlikte almış oldukları ürünlerin kombinasyonları hakkında fikir veren Pazar Sepeti Analizinde bu kural çokça kullanılmaktadır (Mirza ve diğerleri, 2016, s. 440). Böylece pazarlama şirketleri bu bilgileri kullanarak belirli kampanyalar düzenlemekte ve sıkça beraber alınan ürünlerin raflara yan yana dizilmesi gibi stratejiler üretmektedirler. Bu yöntem sadece pazarlama alanında kullanılmamakta olup birçok sektöre uyarlanması da mümkündür. Örneğin kütüphane hizmetleri kapsamında bu teknik; kullanıcılara otomatik kaynak öneri sistemleri geliştirilmesinde veya kullanıcı ilgi alanları tespit edilerek o konuda söyleşi, seminer gibi etkinlikler planlanmasında kullanılabilir.

III. 2. 2. 5. Anormallik Tespiti (Anomaly Detection): Sapma, anomali veya anormallik tespiti, isminden de anlaşılacağı üzere standart davranış ve biçimleri incelendiğinde olağandışı ve en önemli değişikliklerin, örüntülerin veya sapmaların keşfedilmesi ile ilgilenmektedir, sapmaların sayısının artması madencilik algoritmasını olumsuz etkileyerek tahmin gücünü azaltabilmektedir (Mirza ve diğerleri, 2016, s. 440; Rajagopalan ve Krovi, 2002; Aktaran Jackson, 2002, s. 268). Özellikle güvenlik alanında önemli faydalar sağlamaktadır. Sapma tespitine bir örnek olarak; dolandırıcılığın keşfedilmesi noktasında kredi kartındaki olağandışı harcama verilebilir (Jackson, s. 268, 2002) Akıllı saatlerin kalp atış hızını ölçerek anormal bir değer tespit etmesi ve otomatik olarak acil servisi araması sapma tespitine bir başka örneği oluşturmaktadır.

III. 2. 2. 6. Özetleme (Summarization): Özetleme, bir veri madenciliği tekniği olmamakla birlikte aslında tüm bu tekniklerin sonucunu oluşturmaktadır (Mirza ve diğerleri, 2016, s. 440). Veri tabanındaki verilerin tamamı veya bir veri alt kümesi için bir genelleme/tanımlama sağlayarak o veri kümesinin özünü kaybetmeden topluca ifade edebilecek kompakt bir açıklama oluşturmak ile ilgilenmektedir (Mirza ve diğerleri, 2016, s. 440; Hand vd., 2001; Aktaran Güven, 2007, s. 11). Genellikle basit bir tanımlayıcı istatistik ve görselleştirme teknikleri ile gerçekleştirilmektedir. Verilerin tamamı veya veri alt kümesine ait tüm değişkenlerin ortalamalarının ve standart sapmalarının alınarak tablollaştırılması özetlemenin en basit uygulamasını oluşturmaktadır (Hand vd., 2001; Aktaran Güven, 2007, s. 11; Jackson, 2002, s. 275). Bir makale, kitap veya müşteri yorumlarına ait metinlerin anlamını kaybetmeden kısaltılarak kullanıcıya hızlı bir şekilde fikir vermesi bir özetleme örneğini oluşturmaktadır. Bu sayede kullanıcı kaynağın işine yarayıp yaramayacağı ile ilgili hızlı bir şekilde karar verecek, zamandan tasarruf ederek ilgisiz bilgi yığınlarının içerisinde kaybolmayacaktır.

III. 2. 3. Veri Madenciliği Algoritmaları

Veri madenciliğinin bir diğer teknik yanını da algoritmalar oluşturmaktadır. Bu süreçte istatistik ve makine öğrenimi teknikleri önemli yer tutmaktadır (Jackson, 2002, s. 272). Hem istatistik hem de veri madenciliği veriler ile ilgilenen ve buradaki yapıları keşfetmeye çalışan ortak amaca hizmet eden iki alandır, öyle ki birçok kişi tarafından veri madenciliği istatistiğin bir alt alanı olarak tanımlanmaktadır fakat aslında bu doğru bir yaklaşımı barındırmamaktadır. (Jackson, 2002 s. 274). Veri madenciliği interdisipliner bir çalışma alanı olup birçok alandaki araç, teknik ve yöntemlerden faydalanmakta; istatistik alanı ile de bütünüyle örtüşmemektedir (Hand 1999; Aktaran. Jackson, 2002, s. 274). Küme analizi (cluster analysis), korelasyon analizi (correlation analysis), diskriminant analizi (discriminant analysis), faktör analizi (factor analysis), regresyon analizi (regression analysis) veri madenciliğinin kullanmış olduğu istatistiksel algoritmalarından bazılarıdır (Jackson, 2002, ss. 274-275).

Veri madenciliği, makine öğrenmesine dair teknik ve algoritmaları kullanan, bu sayede de en uygun aksiyonun belirlenmesi için öngörü elde eden bir veri analitiği sürecidir. Bu noktada makine öğrenmesi teknik ve algoritmaları, veri madenciliği ve büyük veri konularında önemli bir kesişim noktasını oluşturmaktadır. Makine öğrenimi kısa bir tanım ile eğitim verilerini kullanarak sistemlerin öğrenmesini sağlayan modeller oluşturmaktadır. Sinir Ağları (Neural Networks), Vaka Tabanlı Akıl Yürütme (Case-Based Reasoning), Genetik Algoritmalar (Genetic Algorithms), Karar Ağaçları (Decision Trees), Birliktelik Kuralları (Association Rules) veri madenciliğinde de kullanılan temel makine öğrenimi algoritmalarını (machine learning algorithms) oluşturmaktadır (Jackson, 2002, s. 273).

Algoritmalar ile ilgili olarak iki tür büyük algoritma kategorisi bulunmaktadır. Bunlar: Denetimli Öğrenme (Supervised Learning Algorithms) ve Denetimsiz Öğrenme Algoritmalarıdır (Unsupervised Learning Algorithms) (Joseph ve diğerleri, 2016, s. 6807).

Genel olarak denetimli öğrenme algoritmaları; algoritma çalıştırılmadan önce veri kümesi için sınıf öznitelik değerlerinin bilindiği (etiketli veri/ eğitim verisi) ve bilinen örneklerin kullanılarak bilinmeyen örneklere dair değer otomatik olarak belirlenmesi şeklinde tanımlanabilir (Joseph ve diğerleri, 2016, s. 6807; Gupta ve Lehal, 2009, s. 64). Denetimsiz öğrenmede ise; denetimli öğrenmenin tam tersi şeklinde, veriler genellikle etiketsiz verilerden oluşmaktadır. Verilerin sınıf öznitelikleri önceden bilinmemektedir. Verilerin değerinin bulunması mesafe ölçüsü ile gerçekleşmekte olup birbirleri ile yakınlığa sahip veriler bir arada gruplanmaktadır (Han ve Kamber, 2010; Aktaran Joseph ve diğerleri, 2016, s. 6811).

“Denetimli öğrenme sınıflandırma ve regresyon olarak ikiye ayrılmaktadır. Sınıf özniteliği ayrık olduğunda, buna sınıflandırma; sınıf özniteliği sürekli olduğunda ise buna regresyon denmektedir” (Gundecha ve Liu, 2012; Zafarani ve diğerleri, 2014; Han ve Kamber, 2010; Aktaran Joseph ve diğerleri, 2016, s. 6807). Denetimsiz öğrenme algoritmasına örnek olarak da kümeleme verilebilmektedir. ⁴

⁴ Denetimli ve Denetimsiz Öğrenme türleri ile ilgili örnekler hakkında daha detaylı bilgi için bkz. Joseph, S. R., Hlomani, H., & Letsholo, K. (2016). Data mining algorithms: an overview. *Neuroscience*, 12(3), 719-43.

III. 2. 3. 1. Algoritma seçimini etkileyen temel faktörler

Denetimli öğrenme veya denetimsiz öğrenme algoritmalarından hangilerinin kullanılacağı; projesinin doğasına, eldeki veri setinin özelliklerine, elde edilmek istenen sonuca ve çeşitli pratik faktörlere göre belirlenmektedir (Barut, 2024). Bir sorunun çözümü için (örneğin ev kiralamak için evin değerini otomatik bir şekilde tahmin ettirme) kullanılacak birçok farklı yaklaşım bulunabilmektedir. Evin kira değerinin tahmin edilmesi için farklı teknik ve algoritmalar kullanılabilmektedir. Fakat her algoritma sorunun çözümü için en iyi performansı sağlamamaktadır, hatta bazı algoritmalar birden çok sonuç üretebilmektedir. Bu noktada tüm algoritma türlerinden en iyi sonuç verebileni seçmek önemli ve zor bir iştir (Joseph ve diğerleri, 2016, s. 6806).

Algoritma seçimini etkileyen temel faktörler kısaca aşağıdaki gibidir:

III. 2. 3. 1. 1. Etiketli/Etiketsiz Veri Durumu: Eğer veriler etiketli verilerden oluşuyor ise burada problemin çözümü için denetimli öğrenme algoritmaları seçilmelidir. Verilerin etiketsiz olduğu durumda ise denetimsiz öğrenme algoritmaları seçilmelidir.

III. 2. 3. 1. 2. Veri Boyutu ve Karmaşıklık: Büyük ve karmaşık veri kümelerinin analizi için geleneksel yöntemler yetersiz kalmaktadır (Demircioğlu ve Eşiyok, 2019, s. 370). “Boyut arttıkça veri noktaları arasındaki mesafeler anlamını yitirmektedir” (Tahliz İstatistik, 2024). XGBoost ve LightGBM gibi algoritmalar büyük ve karmaşık veri setlerini analiz etmek için etkili çözümlerden bazılarıdır (Songur, 2025).

III. 2. 3. 1. 3. Doğrusal/Doğrusal Olmayan İlişkiler: Veriler arasındaki ilişkilerin doğrusal olup olmaması algoritma seçimini etkileyen bir diğer unsurdur. Örneğin, daha önce de belirtildiği üzere Lineer Regresyon doğrusal ilişkilerde etkilidir, Karar Ağaçları ve Random Forest gibi algoritmalar ise karmaşık ve doğrusal olmayan ilişkileri işleyebilmektedir (Songur, 2025).

III. 2. 3. 1. 4. Veri Dağılımı ve Şekli: Veri dağılımı ve şekli de algoritma seçimine göre performansı etkileyen bir unsurdur. Örneğin, kümeleme algoritmalarından birisi olan k-means, karmaşık şekilli ve farklı boyutlu kümelerde zayıf performans gösterirken, küremsi (spherical) ve eşit büyüklükteki kümelerde verimli sonuçlar elde edebilmektedir (Songur, 2025). Tam tersi bir duruma örnek olarak ise, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algoritması verilebilir. DBSCAN, k-means algoritmasının aksine düzensiz ve farklı yoğunluktaki kümelerde kullanılacak verimli yaklaşımlardan birisidir (Tahliz İstatistik, 2024).

III. 2. 3. 1. 5. Gürültülü Veri ve Aykırı Değerler: Lineer Regresyon, Lojistik Regresyon, K-Means gibi algoritmalar gürültüye duyarlı olmayıp, yanlış sonuçlar üretilmesine neden olmaktadır. Ridge Regresyon ise özellikle gürültülü verilerde etkili sonuçlar üretmektedir (Songur, 2025). Dolayısı ile analiz edilecek veri setinde gereksiz, hatalı ve anlamsız veriler yoğunluktaysa, algoritma seçiminde gürültüye duyarlılık dikkat edilmesi gereken bir konudur.

III. 2. 3. 1. 6. Dengesiz Veri Setleri: Sınıflar arasında büyük farkların olması dengesiz veri setlerine yol açmaktadır. Bu tür verilerde sınıflar karşılaştırıldığında bazı sınıflarda veya sınıfta eksik veriler gözlemlenmektedir. Bu özellik ise yine algoritma seçimini etkileyen bir unsurdur. Örneğin bir sınıflama algoritması olan Random Forest, dengesiz veri setlerinde, diğer algoritmalar ile karşılaştırıldığında daha iyi sonuçlar elde etmektedir (Kılınç ve diğerleri, 2020, s. 20).

III. 2. 3. 1. 7. Performans Metrikleri: Algoritmanın geliştirilmesinde, doğru metriklerin seçilmesi önem arz eden bir konudur. Doğru performans metriklerinin seçilmesi, geliştirilen projenin (modelin) doğruluğu hakkında geribildirim sağlamaktadır. “Önde gelen performans metrikleri arasında doğruluk, kesinlik, duyarlılık, F1-Skoru, ROC Eğrisi ve AUC, RMSE, MAE ve R-Kare bulunmaktadır. Bu metrikler, hem sınıflandırma

hem de regresyon problemleri için önemlidir ve her biri farklı bilgiler sağlar” (Karakaş, 2024).

Özetle algoritmalar problemin çözümüne yönelik olarak, farklı avantaj ve dezavantajları barındırmaktadır. Bu noktada sorunun çözümü ve projenin başarısı için, algoritmaların seçiminde, veri özellikleri ve performans metrikleri dikkate alınması gereken kriterleri oluşturmaktadır.

III. 2. 4. Veri Madenciliği Kullanım Alanları

Veri madenciliği birçok sektörde kurumsal hedeflere ulaşılmasında önemli rol oynamaktadır ve çok çeşitli kullanım alanları bulunmaktadır (Lei-da Chen & Frolick, 2000). Bilgi teknolojilerinin hızlı gelişimi ve yaşanan dönüşümler ile sürekli olarak veri üretimi gerçekleşmektedir. Verinin sürekli üretimi ise, büyük veriyi doğurmaktadır. Veri madenciliği tam bu noktada, büyük verinin analitik olarak işlenmesi ve yararlı sonuçlar elde edilebilmesinde, önemli bir araç olarak karşımıza çıkmaktadır. Bu bağlamda sürekli veri üreten kullanıcılara/müşterilere sahip şirketler, bu araçları kullanmada daha büyük fırsatlara sahip olmaktadır. Veri madenciliğinin özellikle pazarlama sürecinde sağlamış olduğu faydaların yanı sıra sağlık hizmetleri, eğitim, güvenlik, üretim ve finans gibi geniş bir sektör yelpazesinde çeşitli değerler yaratan, kar sağlayan ve/veya maliyetleri düşüren bir araç olarak yaygın şekilde kullanılmaktadır (Jackson, 2002, s. 271).

Aşağıda bazı veri madenciliği kullanım alanları sıralanmaktadır:

III. 2. 4. 1. Pazarlama: Müşterilerin davranışları izlenerek satın alma alışkanlıklarının ortaya konması ve şirketler tarafından çeşitli reklam ve promosyonların yaratılmasında izlenecek politikalar için veri madenciliğinin sağlayacak olduğu bilgi, büyük önem taşımaktadır (Lei-da Chen & Frolick, 2000). Veri madenciliği, müşteri odaklı

pazarlamada önemli bir kullanım alanı bulmaktadır, bağımlılık modelleme başlığında da bahsedilen Pazar sepeti analizi buna önemli bir örnek oluşturmaktadır. Zincir marketler müşterilerinin alışkanlıklarını izleyerek en sık beraber alınan ürünleri tespit edip raflarda yan yana dizerek satın alımını arttırmayı hedeflemektedir. Yine satın alma davranışları izlenerek; kişiselleştirilmiş veya genel kampanya ve reklamlar kar arttırmayı hedefleyen önemli veri madenciliği kullanım alanlarıdır. Bunlar dışında benzer satın alma davranışlarına sahip müşterilere “çapraz satış” ile daha önce almamış oldukları ürünler reklam ve promosyon şeklinde sunulabilmektedir (Jackson, 2002, s. 272). Günümüzde hedefli ürün ve içerik pazarlama ile sıkça karşılaşilmektedir. Bu yolla e-ticaret ve sosyal medya platformlarında “bunlarda ilginizi çekebilir” gibi öneri sistemi ile kişiselleştirilmiş pazarlama stratejileri geliştirilmektedir. Yakın zamanda Uber’in telefonlarının pili azalmış olan müşterilerine daha fazla ücret tahsil etmiş olduğu iddiaları meydana gelmiştir (MailOnline, 2023). Mail Online haberine göre; şarjları sırası ile yüzde 12 ve yüzde 84 olan iki yolcu aynı anda çağırmış oldukları uber tarafından düşük bataryalı müşteriye 17,56 avro, diğer telefona sahip müşteriye ise 16,60 avro teklif edilmiştir. Bu pazarlama stratejisinin etikliği tartışılması gereken bambaşka bir konu iken; kullanılan veri türlerinden çıkarılabilecek bilgi ve yararın uçsuz bucaksız olduğunun da göstergesi denilebilir. Kuruluşların büyük veri kullanarak yapmış olduğu diğer pazarlama stratejilerini ise Jackson (2002, s. 272) çalışmasında şöyle açıklamıştır: Müşterilere ait bir profil çıkarılarak iyi müşterilerin özellikleri belirlenebilmektedir. Böylece kuruluşlar potansiyel müşterilerini hedeflemektedir. Bu yöntem aynı zamanda kaybetme riski bulunan müşterileri geri kazanmak için harekete geçilmesini sağlamaktadır. Çünkü çoğunlukla müşteriye elde tutmak yeni bir müşteri edinmekten çok daha az maliyetli bir iştir.

Bankalar veri madenciliğini en çok kullanan sektörlerden bir tanesidir. Bankaların çok sayıda müşterileri ve ellerinde ciddi miktarda veri bulunmaktadır. Bu verileri yararlı bilgiye geçirmede bankalar büyük avantaj sahibidir. Lei-da Chen ve Frolick (2002) çalışmalarında kredi kartı işinin bilgi işi olduğunu vurgulayarak kredi kartı başvurularını onaylamak, yetkilendirme kararlarını vermek, müşterilerin satın alma davranışlarını analiz etmek gibi noktalarda veri madenciliği kullanımının öneminden bahsetmektedir. Yine aynı çalışmada bankalar ile ilgili olarak Jonathan J. Palmer'ın "bankaların hayatta kalmalarının, detaylı müşteri bilgilerini toplama ve analiz etme, ardından da bu bilgileri pazarlama stratejilerine entegre etme becerilerine bağlı olduğunu" belirttiği sözüne değinmişlerdir.

III. 2. 4. 2. Sağlık Hizmetleri: Canlılar için en önemli konu olan sağlık noktasında veri madenciliği gibi dijital teknolojilerin sağlamış olduğu yarar oldukça anlamlıdır. Veri madenciliği ile hastalıkların teşhisi ve bu süreçte sağlanacak öngörü ve tahminler tıp dünyası için önemli adımları oluşturmaktadır. Dijital teknolojiler mevcutta insan beyni kadar analitik düzeyde olmasa da bir insandan daha hızlı öğrenecek düzeydedir. Doktorların bu teknolojileri teşhis, tanı, hastalıklar arası ilişki kurma (Mirza ve diğerleri, 2016, s. 441) ve tedavi süreçlerinde kullanması sayesinde çok hızlı edinilemeyecek tecrübelerin, çok kısa bir sürede edinilmesi (makinelere öğretmek) ve faydalı hale getirilmesi mümkün olmaktadır. Karmaşık ve işlenmesi zor sağlık verilerinin veri madenciliği ile etkili karar verme ve bilgiye dönüştürme süreçlerinde kullanımı hastalara daha yüksek seviyede hizmet verilmesini sağlamaktadır (Parvathi et.al. 2014; Aktaran Mirza ve diğerleri, 2016, s. 441). Bunlar dışında ilaç firmaları ve tıbbi uygulamalarda da veri madenciliği kullanılmaktadır. Cerrahi prosedürler, tıbbi testler ve ilaçların etkinliği hakkında tahmin elde etmenin yanı sıra hastalıkların tedavisi olabilecek maddeleri

geliştirmek ve keşfetmek için de veri madenciliği kullanılmaktadır (Jackson, 2002, s. 272).

III. 2. 4. 3. Eğitim ve Araştırma: Aslında diğer tüm alanlarda (sağlık, pazarlama, güvenlik, finans vb.) kullanılan veri madenciliği tekniklerinin araştırma alanını da doğrudan etkilediğini söyleyebilir. Her birinin bir bilim dalı veya çalışma alanı olduğu düşünülürse bu alanlardaki gelişmeler bilim dünyasında dolayısı ile de araştırma noktasında gelişmeler sağlayacaktır. Fakat bunlar dışında veri madenciliğinin yeni, yararlı ve gizli; örüntü, bilgi ve ilişkileri ortaya çıkarmasında eğitim ve araştırma alanları da ilerleme içerisinde. Bunun yanı sıra veri madenciliği araştırma alanından ziyade araştırma sürecine de faydalar sağlayabilmektedir. Araştırmacıların işini kolaylaştıracak şekilde çalışma alanlarına uygun kaynak ve araç öneri sistemi araştırma sürecine katkı sağlayan verimli bir uygulama olabilir. Yapılan tüm bu faydalar araştırma ve eğitim sistemindeki bilgi açığının kapatılarak (Mirza ve diğerleri, 2016, s. 441) bilim dünyasına ve insanlığa katkı sağlamaktadır.

III. 2. 4. 4. Finansal Hizmetler: Analistler yatırım stratejileri geliştirmek üzere (alım-satım, risk modelleri) çok büyük finansal verileri analiz etmektedir. Bunun için de veri madenciliği tekniğini yoğun şekilde kullanmaktadırlar (Lei-da Chen & Frolick, 2000). Jackson (2002, s. 272) çalışmasında pazar ve endüstri özelliklerini belirlemenin yanı sıra bireysel şirket ve hisse senedi performansını tahmin etmek için veri madenciliğinin kullanıldığını belirtmektedir. Tüm bunlar dışında döviz kuru, hisse senedi seçimi, kredi puanlaması ve ipotek taraması gibi noktalarda da veri madenciliği kullanılabilmektedir (Lei-da Chen & Frolick, 2000).

III. 2. 4. 5. Güvenlik: Hem sistemlerle hem de insanlarla etkileşimli yapılarda güvenlik önemli bir konudur. Veri madenciliği ile kredi kartı dolandırıcılığı, sigorta sahtekarlıkları,

mağazalarda hileli kasiyer davranışları tespit edilmekte ve güvenlik açıkları saptanmaktadır (Lei-da Chen & Frolick, 2000). Özet olarak şu sonuca varmak mümkündür: Veri madenciliği güvenlik alanında genel olarak olağandışı kalıpları hızla tespit etmek amacı ile kullanılmaktadır.

Bahsedilen veri madenciliği kullanım alanları genel başlıklar halinde işlenmiş olsa da veri madenciliği; sigorta şirketleri, telefon operatörleri, havayolu şirketleri, üreticiler, perakendeciler vb. birçok sektörde kullanılmakta ve bu sektörlerin sürekliliğini sağlayacak olan öngörülü karar verme, strateji geliştirme ve hedefe ulaşma gibi faydaları beraberinde getirmektedir.

Şu ana kadar veri madenciliğinin genel yapısı işlenmiştir. Veri madenciliği istatistik ve makine öğrenimi alanlarına ait teknik ve algoritmaları kullanarak veri ambarları içerisinde bilgi, ilişki ve örüntü çıkaran bir yöntemdir. Metin Madenciliği ise veri madenciliği teknik ve algoritmalarını kullanarak yapılandırılmamış verilerden bilgi çıkaran bir yöntemdir. O halde akla şu soru gelmelidir: Metin madenciliği, veri madenciliği teknik ve algoritmalarını aynı amaç için kullanıyor ise metin madenciliğine neden ihtiyaç vardır?

III. 3. Metin Yoğunluğunda Anlam Arayışı: Metin Madenciliği İhtiyacı

Metin madenciliği ve veri madenciliği, benzer teknik ve algoritmaları kullanarak, verilerden bilgi çıkarma süreçlerini içeriyor olsalar da, ele aldıkları veri türleri, odak noktaları ve ön işleme süreçleri açısından önemli farklılıklara sahiptirler. İki alan arasındaki temel farklardan bir tanesini, veri türleri oluşturmaktadır. Veri madenciliği genel olarak, nicel nitelikteki yapılandırılmamış verileri hedeflemektedir. Metin madenciliği ise, e-posta, web sayfaları, elektronik belgeler gibi yapılandırılmamış veya

yarı yapılandırılmış metinsel veriler ile ilgilenmektedir (Li ve Yanchinda, 2024). Metin işleme ve analiz etme, yapılandırılmış verilere kıyasla çok daha zor bir işlemdir (Chen, 2001, s. 18). İnsan beyni analitik zekası ile bir metindeki bilgiyi kolaylıkla saptayabilmekte iken bilgisayarların bunu yapabilmesi kolay değildir. Fakat insanların da dil becerileri metinsel verileri analitik anlamda yorumlamasına el verse de bilgisayarlar gibi büyük hacimli ve hızlı bir metin işleme altyapısına sahip değildir (Gupta ve Lehal, 2009, s. 60).

Bilgi çıkarma sürecinde veri madenciliğinin kaynağı yapılandırılmış veritabanları iken metin madenciliği bunu yapılandırılmamış metinlerden sağlamaktadır (Agrawal ve Batra, 2013, s. 118). Veri madenciliğinde yapılandırılmış verilerden bilgi çıkarma (information extraction) ve bilgi keşfi (knowledge discovery) gerçekleştirilirken veri ambarları (data warehouse) iyi bir iş çıkarmış olsa da; söz konusu metinsel veriler olduğunda başarısız olmuşlardır (Gaikwad ve diğerleri, 2014, s. 42). Bunun nedeni metinsel verilerin sayısal veriler gibi düzenli bir yapıya sahip olmamasıdır. Düzensiz ve karmaşık yapıda bulunan metinsel verilerin işlenebilmesi için, önışlem süreçlerinden geçmesi ve terim frekansı (TF), ters doküman frekansı (TDF) gibi terim ağırlıklandırma sistemlerinin kullanılması gerekmez. Veri madenciliği ile metin madenciliği arasındaki belki de en belirgin farkı da bu oluşturmaktadır. Aslında her iki alan da önışlem süreçlerini içermektedir ancak metin madenciliğinde önışleme, metinsel verilerin analize hazır hale getirilmesini amaçlamaktadır.

Metin madenciliğinde önışleme süreci; veri temizleme (özel karakterleri, rakamları, noktalama işaretlerini, durak kelimeleri kaldırma) (Vasques ve diğerleri, 2018, s. 493), terim çıkarımı (morfolojik, sözdizimsel ve anlamsal analiz) gibi adımları içermektedir (Schieber ve Kruses, 2014, s. 21). Veri madenciliğinde ise önışlem süreci veri temizleme, veri dönüştürme ve veri indirgeme/azaltma aşamalarını içermekte olup,

metin madenciliğinde olduğu gibi dilbilimsel karmaşıklıkları içermemektedir. Metin madenciliği veri madenciliğinden farklı olarak, metinsel verilerin işlenebilmesi için, dilbilimsel işleme veya doğal dil işleme yetenekleri gerektirmektedir (Chen, 2001, s. 15). Veri madenciliğinin dilbilimsel işleme konusunda bir gereksinimi bulunmamaktadır. Bu iki alanın yetenekleri ele alındığında ise farklı çıkarımlar sağlamaları oldukça olağandır. Dolayısı ile veri madenciliği, daha çok nedensel ilişkileri belirlemeye çalışırken (örneğin, hangi çalışan demografik değişkenlerinin harcama kalıplarını etkilediği) (Chen, 2001, s. 15), metin madenciliği ise, metinsel verilerdeki; gizli kuralları ve kalıpları, anlamlı örüntüleri ve içgörülerini ortaya çıkarmayı hedeflemektedir (örneğin müşteri davranışlarını tahmin etme, doğru kararlar alma ve genel riskleri azaltma gibi iş kararlarının alınmasında) (Malik ve Gupta, 2022, s. 30).

Günümüzde neredeyse tüm organizasyonlar verilerini elektronik ortamda muhafaza etmekte ve yönetmektedir. Söz konusu verilerin yönetilmesinde ise iç ve dış bilgi kaynakları önem arz etmektedir. Yapılandırılmamış veriler resmi bilgi alışverişi (kurumsal yazışmalar) için en yaygın araç olmasının yanı sıra, web üzerinde de verilerin %80'inden fazlası yapılandırılmamıştır. Kuruluşların büyük çoğunluğunda verilerin metinsel formatta kaydedilmiş olması ve geleneksel veri madenciliği tekniğinin metinsel veriler için yetersiz kalması; metin madenciliğinin kuruluşlar için çok daha iyi bir çözüm olarak karşımıza çıktığının göstergesidir (Gharehchopogh ve Khalifelu, 2011; Talib ve diğerleri, 2016, s. 414; Gupta ve Lehal, 2009, s. 60).

Bilgi birçok kaynaktan elde edilebilir fakat yapılandırılmamış metinsel veriler, fazla miktarda bulunmakta ve oldukça da değerli hazır bilgi kaynaklarını oluşturmaktadır. Dijital kütüphanelerde, elektronik arşivlerde, sosyal medyada, e-postalarda, bloglarda vb. platformlarda çok sayıda yapılandırılmamış veri yer almaktadır. Bu noktada veri madenciliğinin, yapılandırılmamış verileri etkili bir şekilde ele alamaması ve

veritabanlarının daha az veriyi depolayabiliyor oluşu, yapılandırılmamış verilerin işlenmesinde eksiklik ortaya çıkartmaktadır (Tandel ve diğerleri, 2019, s. 1022; Birunth ve Sukanya, 2012, s. 271). Metin madenciliği ile bu eksikliğin giderilmesi amaçlanıp; metinsel veriler sınıflandırılmış, belge kümeleri içerisinde bilgi çıkarılıp ilişkiler kurulmuş, desen ve kalıplar çıkarmak için büyük miktardaki veriyi yönetme zorluğu aşılmıştır (Dang ve Ahmad, 2014, s. 25; Gaikwad ve diğerleri, 2014, s. 43). Tüm bu faydaları ile metin madenciliği, veri madenciliği ile farkını ortaya koyarak, veri madenciliğinin hangi noktalarda eksik kaldığı ve metin madenciliğine neden ihtiyaç duyulduğu sorularına açıklık getirmektedir.

III. 3. 1. Metin Madenciliği

Metin madenciliği, doğal dilde oluşturulmuş metinsel verilerin, yapısal metne dönüştürülmesi ile ilgilenen ve bu veriler üzerinden; gizli kuralları ve kalıpları, anlamlı örüntüleri ve yeni içgörülerini ortaya çıkarmayı hedefleyen bir alandır (Malik ve Gupta, 2022, s. 30). Bu özellikleri ile metin madenciliği, doğal dilde yazılmış belgelerde biriken bilginin, yeni keşifler için potansiyel bir kaynak teşkil etmesini sağlamaktadır (Vasques ve diğerleri, 2018, s. 490). Keşfedilen yeni bilgi, ilişki ve örüntüler ile metin madenciliği; iş kararlarının alınmasına yardımcı olur, manuel süreçlerin yarattığı orantısız zaman ve maliyet ihtiyacını azaltır (Schieber ve Kruses, 2014, s. 20), yararlı bilgilerin otomatik bir şekilde tanımlanmasını sağlar.

Literatür incelendiğinde metin madenciliği terimine eş olarak; metin veri madenciliği (Text Data Mining), metinde bilgi keşfi (Knowledge-Discovery in Text), akıllı metin analizi (Intelligent Text Analysis), metinsel veri tabanlarından bilgi keşfi (knowledge discovery from textual databases) gibi çeşitli terimlerin kullanıldığı

görülmektedir (Gupta ve Lehal, 2009, s. 60; Dang ve Ahmad, 2014, S. 22). Metin madenciliği kısaca, veri madenciliği teknik ve algoritmalarını, önışlem sürecinden geçmiş metinsel verilere uygulayan bilgi çıkarımı ve keşfi sürecidir. Tandel ve arkadaşları (2019, s. 1022) çalışmalarında metin madenciliği yöntemini kısaca şu şekilde tanımlamışlardır: “yapılandırılmamış bir bilginin kelime, kelime öbeği ve cümlelerini veri tabanındaki doymuş bilgi (saturated information) ile ilişkilendirerek matematiksel değere dönüştürür ve geleneksel veri madenciliği teknikleri ile analiz eder”. Fakat çok disiplinli bir çalışma alanı olan metin madenciliği, veri madenciliği dışında bilgi erişim, makine öğrenimi, istatistik, hesaplamalı dilbilim, bilgi yönetimi, doğal dil işleme gibi birçok disiplin ile iç içe geçmiş ve bu disiplinler ile etkileşimli bir alandır (Gupta ve Lehal, 2009, s. 73). Metin madenciliği çalışmaları çok uzun yıllardır geliştirilmekte olmasına karşın uygulama açısından genç bir alan olduğu söylenilebilmektedir.

Birunth ve Sukanya (2012, s. 269) çalışmalarında metin madenciliğini; metin belgelerinden ilginç veya erişilemeyen desen veya bilgi çıkarma süreci olarak tanımlamaktadır. Bu tanımda dikkat çekici olan nokta, “erişilemeyen desen veya bilgilerin çıkarılması”dır. Metin madenciliği farklı metinsel kaynakları kullanarak, daha önce bilinmeyen bilgileri keşfedebilmekte/çıkarabilmektedir. Daha önceden bilinmeyen “örtük bilginin” ortaya konulması metin madenciliğinin sunmuş olduğu önemli özelliklerden bir tanesidir. Bu özellik, günümüzde bir proje veya bir iş süreci için, daha genel bir deyişle herhangi bir konuda karar almak için öngörölü bilgiyi sunan paha biçilemez bir değerdir. Gupta ve Lehal (2009, s. 60) çalışmalarında metin madenciliğinin bu özelliği üzerinde durmuş ve metin madenciliği amacını ‘henüz kimsenin bilmediği ve bu nedenle henüz yazılmamış olabilecek “bilinmeyen bilgileri” keşfetmek’ olarak açıklamaktadır. Yine aynı çalışmada metin madenciliğinin, web aramasından farklı olduğunu, metin madenciliğinin keşfetmiş olduğu bu “yeni gerçekler” üzerinden

ayrıştırmaktadır. Web aramasında kullanıcılar zaten var olan ve bir yaratıcı tarafından daha önce oluşturulmuş içeriklere/kaynaklara erişmektedir. Metin madenciliğinde ise var olan bilgilerin yanı sıra daha önce bilinmeyen veya ulaşılamayan gizli/örtük bilginin keşfedilmesi/ortaya çıkarılması söz konusudur. Bu da literatürdeki tanımlarda çokça bahsedilen ilginç, gizli, önemli; örüntü, ilişki, ve bilgilerin çıkarılmasına işaret etmektedir. Agrawal ve Batra (2013, s. 119) metin madenciliğini “geleneksel olmayan bilgi erişim stratejilerinden başka bir şey değildir” şeklinde tanımlamaktadır. Onlara göre geleneksel olmayan stratejiler basit bir bilgi erişimin ötesine geçmesi gereken faydalı bir sistemi temsil etmektedir. Bir başka çalışmada ise, Gaikwad ve arkadaşları da (2014, s. 42) bu bakış açısını eserlerine taşımışlardır. Çalışma, geleneksel bilgi aramasında, kullanıcılar tarafından önceden bilinen ve bir başkası tarafından yazılmış/oluşturulmuş terimlerin taratıldığından söz etmektedir. Buna karşılık metin madenciliği sürecinde ise; veriler analiz edilerek, henüz bilinmeyen ve daha yazıya dökülmemiş bilgilerin keşfi sağlanmaktadır.

Metin madenciliği ile sağlanan bu yeni gerçekler ve keşfedilen ilişkiler, her alanda; fırsatların öne çıkarılması, tehditlerin tespit edilmesi ve risklerin önüne geçilebilmesinde en iyi aksiyonun sağlanmasına yardımcı olabilmektedir. Bilgi teknolojilerinin gelişimi ile veri üretiminin hiç durmadan arttığı bir gerçektir. Daha önce de belirtildiği gibi dünyadaki verilerin %80’inden fazlası yapılandırılmamış veridir. Bu oranın bu kadar yüksek olması ve artış göstermeye devam etmesi, metin madenciliği kullanımının önünü açmaktadır. Yapılan yasal düzenlemeler ile devlet kurumları iş süreçlerini elektronik ortamda yürütmekte ve yönetmektedir. Diğer organizasyon ve endüstrilerde hemen hemen her biri bilgilerini elektronik ortamda saklamaktadır. İş süreçlerinde üretilen bu veriler (büyük oranda yazışmalardan elde edilen) başlık, yazar adı, yayın tarihi, kategori vb. gibi birkaç yapılandırılmış alan barındırır da (Gaikwad ve

diğerleri, 2014, s. 42) metinsel halde bulunan (yazışmalar gibi) yapılandırılmamış veriler çok daha yoğun halde bulunmakta olup içlerinde keşfedilmeyi bekleyen çok önemli bilgiler barındırmaktadır. Bu bilgiler organizasyonlara gelecekleri ve süreklilikleri ile ilgili bilgiler sağlayabilmektedir.

Yapılandırılmamış verilerden bilgi çıkarmak zorlu bir işlemdir. Bilgisayarlar bir metni insanlar gibi algılayıp analiz edememektedir. Metin madenciliği bu verileri işleyerek uyarıcı ve gelişmiş tasarımlar elde etmektedir (Tandel ve diğerleri, 2019, s. 1022). Metin madenciliği bu süreçleri gerçekleştirebilmek için bazı önışlem süreçlerine gereksinim duymaktadır. Böylece yapılandırılmamış halde bulunan veriler yapılandırılarak analize hazır hale gelmektedir. Bu süreçler bir sonraki başlık olan “Metin Madenciliği Süreci”nde işlenmektedir.

III. 3. 1. 1. Metin Madenciliği Süreci: Ön İşleme ve Vektörleştirme

Ertam (2018) çalışmasında metin madenciliği süreci ile ilgili olarak; süreci temelde metin önışleme ve bilgi çıkarma olmak üzere iki adımda özetlemektedir:

- ilk adımda yapılandırılmamış veriler yapılandırılarak sayısal hale getirilir,
- ikinci adımda ise yapılandırılan bu verilere veri madenciliği teknikleri uygulanır.

Çalışmamızın bu başlığında Ertam’ında çalışmasında (2018) belirttiği şekilde iki adım esas alınarak, bu süreçteki adımlar daha detaylı işlenmektedir.

Metin madenciliği süreci yapılandırılmamış verilerin toplanması ile başlamaktadır. Bu veriler çeşitli kaynaklardan (web sayfası, pdf dosyaları vb.) veya veri havuzlarından elde edilebilmektedir (Gaikwad ve diğerleri, 2014, s. 42; Talib ve diğerleri, 2016, s. 414). Veriler elde edildikten sonra metin madenciliği teknik ve algoritmalarının

uygulanabilmesi için, veriler önışleme tabi tutulur. Önışlemede amaç; verilerin temizlenmesi, gürültünün azaltılması ve tutarlılığın sağlanmasıdır (Zhan, 2025, s. 3).

Ön işlemler süreci bazı aşamalardan oluşmaktadır. Bunlar; **Metin Temizleme (Text Cleanup)** (Noktalama işaretlerini ve rakamları filtreleme (Filter punctuations and digits), Normalizasyon (Normalization)), **Tokenizasyon/Parçalara Ayırma (Tokenization)** (Kelime Tokenizasyonu (Word Tokenization), Karakter Tokenizasyonu (Character Tokenization), Cümle Tokenizasyonu (Sentence Tokenization), N-gram Tokenizasyonu (N-gram Tokenization)), **Durdurma Kelimeleri Filtreleme (Filter Stop Words)** ve **Kök Çıkarma (Stemming) ve Lemmatizasyon (Lemmatization)**dur.

Genel olarak yukarıdaki aşamalardan oluşan önışlemler süreci projenin kapsamı, uygunluğu ve ihtiyaçlarına göre farklı bir sıralamayı takip edebilmektedir.

Aşağıda kısaca bu aşamalar açıklanmaktadır:

III. 3. 1. 1. 1. Metin Temizleme: Bu aşamada metnin düzenlenmesi için noktalama işaretleri ve rakamlar filtrelenir. Metinde tutarlılık sağlamak ve gürültü azaltmak için ise metin normalizasyon sürecinden geçer.

III. 3. 1. 1. 1. 1. Noktalama İşaretlerini ve Rakamları Filtreleme: Adından da anlaşılacağı üzere bu adımda belgede yer alan metin tabanlı olmayan (Kılınç ve diğerleri, 2016, s. 90) noktalama işaretleri, sayılar ve özel işaretlerin metinden ayrıştırılması işlemi gerçekleştirilmektedir. Bu adımda da gereksiz veriler temizlenerek sistemin performansı artırılmaktadır.

III. 3. 1. 1. 1. 2. Normalizasyon: Her zaman gerekli olmayan, kullanımı sorunun gereksinimine bağlı olan bir aşama olup, bu aşamada metnin tamamı tek bir büyük harfe veya küçük harfe döndürülür (Kalra ve Aggarwal, 2018, s. 73).

III. 3. 1. 1. 2. Tokenizasyon: Bu işlemde dizeler kelime veya kelimelere ayrıştırılmaktadır. Genelde her iki boşluk arası bir kelime kabul edilerek ayrıştırılan bu yöntemde deyimler, kelimeler veya semboller şeklinde de ayrıştırmak mümkündür (Kalra ve Aggarwal, 2018, s. 72; Noyan, 2019). Bazı tokenizasyon türleri aşağıdaki gibidir (Geeksforgeeks, 2025a):

III. 3. 1. 1. 2. 1. Kelime Tokenizasyonu: Metni ayrı kelimelere böler.

Örnek: "Metin madenciliği metinsel verileri kullanır." cümlesi ['Metin ', madenciliği', 'metinsel ', verileri', 'kullanır ', '.'] olarak ayrılır.

III. 3. 1. 1. 2. 2. Karakter Tokenizasyonu: Metinsel veriyi tek tek karakter dizilerine ayırır.

Örnek: "Metin" kelimesi ['M', 'e', 't', 'i', 'n'] olarak ayrılır.

III. 3. 1. 1. 2. 3. Cümle Tokenizasyonu: Paragrafları veya büyük cümle kümelerini ayrı cümleler halinde böler.

Örnek: "Metin madenciliği süreci iki ana aşamadan oluşmaktadır. İlk aşamada veriler yapılandırılır, ikinci aşamada ise veri madenciliği teknikleri uygulanır." cümlesi ['Metin madenciliği süreci iki ana aşamadan oluşmaktadır.', 'İlk aşamada veriler yapılandırılır, ikinci aşamada ise veri madenciliği teknikleri uygulanır.'] olarak ayrılır.

III. 3. 1. 1. 2. 4. N-gram Tokenizasyonu: Kelimeleri sabit boyutlu (n boyutu) veri parçalarına ayırır.

Örnek: "Metin madenciliği metinsel verileri kullanır." cümlesi n=2 için (bigram); [('Metin', 'madenciliği'), ('madenciliği', 'metinsel'), ('metinsel', 'verileri'), ('verileri', 'kullanır')] şeklinde ayrılır.

III. 3. 1. 1. 3. Durdurma Kelimeleri Filtreleme: Durdurma kelimeleri (Stop-words) metnin anlatımını güçlendiren fakat tek başına bir anlamı olmayan veya bilgi keşfi, çıkarımı veya sınıflamasında bir katkı sunmayacak olan etkisiz kelimelerdir (Kılınç ve diğerleri, 2016, s. 90) Bu kelimeler önceden derlenmiş listelerde yer alan terimlerin kaldırılmasına dayanabilirken (örneğin Türkçede “yani, öyle ki, benzeri vb.”, İngilizcede “the, a, an, am, is, are vb., Almancada ise “Der, die, das vb.” gösterilebilir); bazen de belgede en çok kullanılan ve en az kullanılan kelimelerin ayrıştırılması ile gerçekleşmektedir (Mohan, 2015, s. 10). Örneğin “e-belge” kelimesi, “EBYS” konulu bir belge için önemli bir kelime olabileceken “su arıtma sistemleri” konulu bir belge için etkisiz bir kelime olabilir. Bu noktada terimin önemi, metnin bağlamına dolayısı ile metinde geçen diğer terimler ile yakınlığına bağlıdır. Stop-words’ün ayıklanması metin verilerinin azaltılmasını ve metin madenciliği sürecinde sistemin daha iyi bir performans sağlamasını kolaylaştırmaktadır (Kannan ve Gurusamy, 2014).

III. 3. 1. 1. 4. Lemmatizasyon: “Lemmatizasyon kelimeleri temel veya sözlük formlarına, yani lemma olarak bilinen geçerli bir kelimeye dönüştürme işlemidir. Bu teknik, kelimelerin bağlamını ve anlamını dikkate alır, böylece dönüştürülen temel formun dilin sözlüğünde geçerli bir kelime olmasını sağlar. Örneğin, "running", "ran" ve "runs" kelimelerinin hepsi "run" lemmasına lemmatize edilir” (Geeksforgeeks, 2024a).

Lemmatizasyon birkaç aşamadan oluşur (Geeksforgeeks, 2024a):

III. 3. 1. 1. 4. 1. Kelime Türü (Part-of-Speech (POS)) Etiketleme: Her kelimenin dilbilgisi kategorisini belirleme (örneğin isim, fiil, sıfat⁵),

⁵ Duygu analizi için, sıfatların belirlenmesi önem arz etmektedir.

III. 3. 1. 1. 4. 2. Morfolojik Analiz (Morphological Analysis): Kelimenin kök biçimini anlamak için kelime yapısını analiz etme,

III. 3. 1. 1. 4. 3. Sözlük Arama (Dictionary Lookup): Kelimenin lemmasını bulmak için önceden tanımlanmış bir kelime dağarcığı kullanma.

III. 3. 1. 1. 5. Kök Çıkarma: Özellikle Türkçe gibi eklemeli dillerde bu aşama zorlu ve önemlidir. Bu aşamada sözcükler kök haline indirgenerek eklerden arındırılır. Bunun amacı aynı terime işaret eden iki veya daha fazla kelimenin sanki ayrı iki terimmiş gibi frekansının hesaplanmasının önüne geçmektir. Örnek vermek gerekirse; bir dokümanda geçen “okul” teriminin dokümanda geçme sıklığı yani frekansı 4 olsun ve bu dokümanda 7 kere “okulun,” 3 kere de “okulların” kelimeleri geçiyor olsun. Aslında bu 3 kelime de nihayetinde kelimenin kökü olan “okul” terimine işaret etmekte olup bu ve bunun gibi aynı terimlerin farklı terimler gibi ayrı ayrı frekansının saydırılması sistemi yavaşlatacak ve doğru sonuçlar üretilmesinin de önüne geçecektir. Fakat burada dikkat edilmesi gereken bir nokta vardır. Aslında “okul” kelimesi de “oku” kökünden türemiştir fakat okumak eylemi yapı değiştirerek yeni bir anlam kazanmış ve eğitim-öğretim kurumlarını oluşturan “okul” kelimesine dönüşmüştür. Dolayısı ile stemming işlemi dokümanda anlamı değiştirmeyecek olan çekim eklerine uygulanmalıdır. Yapım eklerine uygulanan stemming işlemi bağlamın kaybolmasına neden olarak yanlış sonuçlar üretilmesine neden olmaktadır. Kök çıkarma, lemmatizasyona göre daha hızlı ve basittir, ancak daha az doğru sonuçlar verebilir ve aşırı veya eksik kök çıkarmaya yol açabilir (Geeksforgeeks, 2024a). Dolayısı ile bu yöntem dilbilimciler ile ortak çalışma gerektirmektedir.

Önişlem aşamasından sonra dokümanın içeriğini oluşturan terimlere bir ağırlık verilmesi gerekmektedir. Bu içeriğin tanımlanabilmesi için önemli bir aşamayı oluşturmaktadır. Bunun için ise metinsel verilerin sayısal hale getirildiği, büyük metin

derleminden her bir elemanın/sıranın ayrı bir terimi, her bir sütunun ise ayrı bir dokümana karşılık geldiği Doküman Terim Matrisi oluşturulmaktadır (Atan ve Çınar, 2019, s. 10; Güven ve diğerleri, 2020, s. 2139). Dokümandaki terimin ne sıklıkla bulunduğunu veren doküman terim matrisi, dokümanda bir terimin yer alıp almadığını sırası ile 1 ve 0 değerleri ile göstermektedir⁶. Eğer ki bir terim dokümanda yer alıyorsa onun değeri sıfırdan farklı işaretlenir, aynı terim birden fazla kez dokümanda yer alıyorsa değeri 1’den büyük olur (Oğul ve Ercan, 2016; Atan ve Çınar, 2019, s. 10; Kılınç ve diğerleri, 2016, s. 90). Dokümanlardaki terimlerin sıklık özelliklerinin hesaplanması terimin önemini (içeriği temsil etme gücünü) hesaplama noktasında önemli olup, içerik analizi ve dizinleme amaçları ile kullanılmaktadır. Özellikle çok yüksek veya çok düşük frekanslı terimlerden ziyade orta frekanslı terimlerin içeriği temsil etme gücü daha yüksektir (Salton ve diğerleri, 1980, s. 175).⁷

⁶ **(Binary Term Occurrence)**: Bir terimin belgede olup olmadığını (1 veya 0) belirtir (Bizer, [Tarih Yok]).

⁷ Yüksek frekanslı terimler, metinde çok fazla geçen kelimeleri tanımlamaktadır. Bir kelimenin metinde çok fazla geçiyor olması, o kelimenin metin içerisindeki önemi azaltan bir unsur olup, belgeyi tanımlama noktasında daha genel kalmaktadır. Örneğin; EBYS’lerde metin madenciliği konulu bir çalışmada, “belge” terimi çok fazla geçebilir. Ancak bu terim konuyu yansıtmada çok genel kalmaktadır. Önişlem sürecinde gerçekleştirilen, terimlerin yüksek frekanslı ve belgeler için ayrıştırıcı olmayacak gereksiz terimlerden arındırılması işlemi, kesin isabet (precision) oranını arttırarak ilgili daha çok belgeye erişimi sağlamaktadır. Çünkü doküman için ayrıştırıcı özelliğe sahip olmayan terimler diğer dokümanlarla da eşleşecek ve benzerlik artacaktır. Bu da erişim (recall) oranını arttırmasına karşın ilgisiz birçok dokümana maruz kalınmasına sebebiyet verecektir.

Kelime Torbası (Bag Of Words (BoW)), terim doküman matrisi oluşturma aşamasındaki temel tekniklerden birisini oluşturmaktadır. BoW ile her doküman terim sıklıklarını (metinde her terimin ne sıklıkla geçtiği) içeren bir vektörle temsil edilmektedir. Bu şekilde metin parçaları makinelerin anlayabileceği ve işleyebileceği yönetilebilir parçalara dönüştürülmektedir. BoW aslında, metin verilerini, bilgisayarlar tarafından erişilebilir hale getiren bir kelime envanteridir (Swarnkar & Thakar, 2023). Diğer doküman terim matrisi oluşturma metotları ise şunlardır: Terim Frekansı (TF), Ters Doküman Frekansı (TDF) ve Terim Frekansı-Ters Doküman Frekansı (TFIDF) (Ertam, 2018; Kılınç ve diğerleri, 2016, s. 91). Temel olarak, TF, her bir terimin metin içerisinde geçme sıklığını yani frekansını vermektedir; TDF, terimin bütün dokümanlar içerisindeki frekansını vermektedir (Oğul ve Ercan, 2016; Kılınç ve diğerleri, 2016, s. 91). Bir terimin frekansının çok yüksek olması onun etkisiz terim olma ihtimalini arttıran bir unsur olup yüksek TF değeri terimin, o dokümandaki önemi ile ilgili bilgi vermektedir (Oğul ve Ercan, 2016). TFIDF ise “terimleri önemlerine göre indekslemek için istatistiksel anlamda oldukça önemli ve sık kullanılan bir yöntemdir” (Borth ve diğerleri, 2013; Aktaran Oğul ve Ercan, 2016). Bu konuyu biraz daha basitleştirmek gerekirse; TF bir terimin bir dokümanda kaç defa geçtiğini verirken, TDF ise; nadir terimlere, tüm doküman koleksiyonu içerisinde, daha çok ağırlık vermektedir. TFIDF hesaplaması ise ikisinin çarpımından oluşmuş olup, nadir kelimelere daha fazla, yaygın kelimelere ise daha az ağırlık vermektedir (Bizer, [Tarih Yok], s. 23). TDF, TF’nin aksine dokümandaki tüm terimleri değil, belirli dokümanlardaki önemli terimleri bulmaktadır. Çok sayıda dokümanda geçen terimlere yüksek ağırlık verilmesinden ziyade TDF, dokümanları birbirlerinden ayırarak (benzerlik yönünden) ilgili belgelere daha yüksek performansta erişilmesini sağlar (Salton ve diğerleri, 1975, s.616). Bu konu şu şekilde somutlaştırılabilir; bir doküman içeriğinde 5 defa “dijitalleşme” terimi geçiyor ve bu

dokümanda toplam 50 tane kelime geçiyor. TF, terim frekansının dokümandaki toplam kelime sayısına oranı şeklinde bulunmaktadır. O halde “dijitalleşme” terimi için TF değeri 0,1’dir. TDF için ise, derlemde 4000 tane doküman olsun. TDF’nin tüm belgeler içerisinde nadir olan terimleri bulduğundan bahsedilmişti. O halde “dijitalleşme” terimi 4000 doküman içerisinde 40 defa geçtiğinde 400 defa geçtiğinden daha nadir olur ve daha yüksek ağırlığa sahip olmaktadır.

Belgeler ne kadar iyi terimler ile temsil edilirse erişim performansı da o kadar iyi olacaktır. Belgelerin iyi terimler ile temsil edilmesi, benzer dokümanların bir arada farklı dokümanların ise birbirlerinden uzak olmasını sağlamaktadır. Böylece kullanıcı, sorgusu ile ilişkili olan dokümanlara erişerek ilgisiz dokümanlar ayıklanmış olmaktadır. Benzerlik, terimlerin birbirine benzemesini değil terimlerin konusal bağlamda birbirine olan yakınlığıdır. Aşağıda Embedding Projector aracı kullanılarak, “library” kelimesine yakın olan kelimeler görselleştirilmiştir. Görselden de anlaşılacağı üzere yakınlık, sözdizimsel benzerlikten ziyade anlamsal benzerliğe sahip kelimeleri tanımlamaktadır. Örneğin görselde “library” kelimesine benzer olarak, “archives, collection, museum...” kelimeleri verilmektedir.

Şekil 10.

'Library' Kelimesinin Anlamsal Komşuları: Vektör Uzayında Konumlandırma



Salton ve arkadaşları (1975) çalışmalarında benzerlikle ilgili şunları söylemektedir: Benzer dizinleme terimlerine sahip iki doküman uzayda birbirlerine çok yakın noktalar ile gösterilmektedir. İki dokümanın benzerlik ölçümü, iki vektörün iç çarpımının alınması veya vektörler arasındaki açı baz alınarak hesaplanabilmektedir. Vektörler için terimler aynı olduğunda açı 0 olur, bu %100 benzerlik demektir. Açı büyüdükçe terimler birbirinden uzaklaşmakta ve benzerlik azalmaktadır.

Yine aynı çalışmaya göre en iyi erişim performansı şu iki koşulu içermektedir:

- Bir küme içerisindeki dokümanlar birbirleri ile maksimum benzerlikle olmalı,
- Farklı kümelerde bulunan dokümanlar birbirleri ile minimum benzerlikte olmalı.

Bu sayede, sorgu sırasında ilgili belgelere bir arada erişim sağlanırken, ilgisizler ayıklanmış olacak ve kesin isabet oranı yüksek bir bilgi erişim süreci yaşanacaktır.

Belgenin içeriğinin tanımlanması oldukça güç bir işlem olup kullanıcı açısından etkililiğini ölçmek de bir o kadar zordur. Çünkü kullanıcının öznel bakış açısını

değerlendirmek mümkün değildir. Bu nedenle bilgi erişim alanında verimi arttıracak ve kullanıcının bilgi isteğine en uygun dokümanlara en hızlı şekilde erişebilmelerini sağlayacak yöntemler geliştirilmiştir. Şu ana kadar yapılan işlemlerde doğal dil dizinleme yöntemi kullanılarak sadece doküman metnini kapsayan terimlerin dizinlendiği ve böylece erişimin dokümandaki terimlerle sınırlandığı görülmektedir. Fakat kullanıcı, dokümana erişmek için sorgusunu farklı terimler ile aratmaktadır. O halde dizinlenen terimlerin yönlendirildiği ve ilişkilendirildiği ontoloji tabanlı bir yapıya ihtiyaç vardır. Bu yapının erişime sunduğu katkılar ve mantıksal çerçevesi, 4. Bölümde işlenmesi nedeniyle burada detaya girilmemektedir.

Bu aşamaya kadar ön işleme sürecinde metin temizlenerek özünü ortaya çıkartmaya çalışılmış, ardından doküman terim matrisi uygulanarak yapılandırılmamış metinsel veriler sayısal hale getirilerek işlenmeye hazırlanmıştır. Bir sonraki aşamada ise elde edilen yapılandırılmış verilerden desen belirlenmektedir (Dang ve Ahmad, 2014, s. 22). Bu aşamada doküman bir metin analizi aşamasından geçmektedir. Gaikwad ve arkadaşları (2014, S. 42) metin analizi aşamasından şu şekilde bahsetmektedir: “Metin analizi, metinden yüksek kaliteli bilgi elde etmek için semantik analizdir. Birçok metin analizi tekniği (sınıflama, kümeleme, bilgi çıkarma vb.) mevcuttur; organizasyonun amacına bağlı olarak farklı kombinasyonlar kullanılabilir. Bazen metin analizi teknikleri bilgi çıkarılana kadar tekrarlanmaktadır. Elde edilen bilgi, bir bilgi yönetim sistemine yerleştirilebilir ve bu sistemin kullanıcısı için bol miktarda bilgi elde edilebilir”. Aslında bu aşamada metin analiz edilerek tasarım ve kalıplar incelenmekte incelenen kalıplar sonucunda ise gerçekler ve ilişkiler tanımlanarak (Agrawal ve Batra, 2013, S. 119) değerli bilgilere ulaşılmaktadır. Analiz etme noktasında farklı metin madenciliği teknikleri kullanılabilir. Bu teknikler metin madenciliği yöntemleri başlığında değinilmesi nedeniyle burada tekrar açıklanmamaktadır. Elde edilen bilgiler bir

veritabanında saklanarak bir bilgi yönetim sisteminde kullanıcıya sunulmaktadır. Kullanıcılar ise bir thesaurus, sözlük veya konu başlığı listesinden yardım alarak sistem üzerinden sorgular gerçekleştirmektedir. Böylece kullanıcılar; yüksek kalitede, yapılandırılmış sonuçlar elde etmektedir (Agrawal ve Batra, 2013, s. 119).

Özetlemek gerekirse metin madenciliği süreci aşağıdaki adımlardan oluşmaktadır:

1. Veri toplanması,
2. Verilerin ön işleme sürecinden geçirilmesi,
3. Doküman-terim matrisi ile verilerin sayısal hale getirilmesi,
4. Metin madenciliği teknikleri kullanılarak verilerin analiz edilmesi,
5. Analiz edilen verilerden desen, örüntü, ilişki ve bilgi çıkarılması,
6. Elde edilen değerli bilgileri bir veritabanında saklanması/bir bilgi yönetim sistemi kullanarak kullanıcıların erişimine sunulmasıdır.

III. 3. 1. 2. Metin Madenciliği Teknikleri/Yöntemleri

Metin madenciliği teknik ve yöntemleri; veri madenciliği, doğal dil işleme, makine öğrenimi, istatistik gibi alanlara ait teknik ve yöntemleri içeren; bilinmeyen kalıpların, ilişkilerin çıkarılmasına ve değerli bilgilerin elde edilmesini sağlayan önemli bir metin madenciliği sürecini oluşturmaktadır. Literatür incelendiğinde farklı araştırmacılar tarafından farklı metin madenciliği tekniklerinin işlenmiş olduğu görülmüştür. Bu başlıkta literatürden derlenen 12 teknik işlenecektir. Bunlar: Bilgi Çıkarma (Information Extraction), Bilgi Erişim (Information Retrieval), Özetleme (Summarization), Kategorizasyon (Categorisation), Kümeleme (Clustering), Trend Analizi (Trend Analysis), Konu Takibi (Topic Tracking), Doğal Dil İşleme (Natural Language Processing), Kavram Bağlantısı (Concept Linkage), Birliktelik Kuralı

Madenciligi/Çıkarımı (Association Rule Mining), Soru Yanıtlama (Question Answering) ve Görselleştirme (Visualisation).

III. 3. 1. 2. 1. Bilgi Çıkarma: Doğal dil metinlerin yapılandırılmamış halde olması nedeniyle madencilik için kullanılamayacak bilgiler barındırmaktadır (Birunth ve Sukanya, 2012, s. 270). Elektronik bilgi, ilişkisel veritabanları gibi yapılandırılmış veritabanlarında bulunması yerine birçok uygulamada doğal dil metinleri halinde bulunmaktadır (Gaikwad ve diğerleri, 2014, s. 44; Gupta ve Lehal, 2009, s. 61). Bu durum ise, uygulamalarda verilerin, karmaşık ve düzensiz bir şekilde (yapılandırılmamış halde) bulunmasına neden olmaktadır. Bilgisayarın, yapılandırılmamış veriyi işleyememesi ise bazı zorluklar yaratmaktadır. Bunlar; veri miktarı nedeniyle depolamada sorun yaşanması, verilerin farklı kaynaklardan gelmesi ve farklı formatla bulunması olarak sıralanabilmektedir. Bu zorlukların üstesinden gelmek için ise bilgi çıkarımı ilk olarak yapılandırılmamış metin dosyalarını yapılandırılmış veritabanlarına dönüştürmektedir böylece metin içerisindeki önemli ve ilginç desen, bilgi ve ilişkiler çıkarılabilir hale gelmektedir (Tandel ve diğerleri, 2019, s. 1023).

Bilgi çıkarma için yapılandırılmış metni analiz etmenin başlangıç noktası denilebilmektedir (Gaikwad ve diğerleri, 2014, s. 43). Bilgi çıkarımı sayesinde devasa metinlerden anlamlı bilgi edinimi sağlanmaktadır (Talib ve diğerleri 2016, s. 415,). Bunun yanı sıra; örüntü işleme ile, metinde önceden tanımlanmış diziler aranarak yapılandırılmış veya yarı yapılandırılmış metinden çıkarılan nesnelerin, anahtar terimlerin ve ifadelerin ilişkileri ve öznitelikleri tanımlanmaktadır (Birunth ve Sukanya, 2012, s. 270; Gupta ve Lehal, 2009, s. 61; Tandel ve diğerleri, 2019, s. 1023). Metinden çıkarılan bu bilgiler (örneğin kişi adı, yeri ve kuruluşu); metnin tamamının okunmasına gerek kalmadan otomatik bir şekilde elde edilmektedir (Gupta ve Lehal, 2009; Aktaran Dang ve Ahmad, 2014, s. 23).

Bilgi Çıkarmanın çeşitli yöntemleri bulunmaktadır. Bunlar: Token çıkarma, varlık çıkarma ve terim ayrıştırma olabilmektedir (Tandel ve diğerleri, S. 1023, 2019). Sıralanan yöntemlerin birisinde “çıkarma” diğerinde ise “ayrıştırma” kullanılmasının nedeni; bilgi çıkarma tekniğinin, metin içerisindeki önemli terimlerin çıkarılarak (belirlenerek) uygulanabileceği gibi, önemsiz terimlerin ayrıştırılması (ayıklanması) ile de gerçekleştirilebileceğini göstermektir. Talib ve arkadaşları çalışmalarında (2016, s. 415) bilgi çıkarma sürecini aşağıdaki şekilde özetlemektedir:

1. Alan uzmanlarının alana göre öznitelik ve ilişkileri belirlemesi,
2. Bilgi çıkarma yazılımlarının dokümanlar içerisinden belirli öznitelik ve varlıkları çıkararak aralarındaki ilişkiyi kurması (Agrawal and Batra, 2013; Aktaran Talib ve diğerleri, 2016, s. 415),
3. Elde edilen derlemin sonraki işlemler için veritabanında saklanmasıdır (Dang and Ahmad, 2015; Aktaran Talib ve diğerleri, 2016, s. 415).

III. 3. 1. 2. 2. Bilgi Erişim: Bilgi erişim en basit haliyle, kullanıcının zihninde oluşan bilgi ihtiyacının hızlı, etkin ve faydalı şekilde, çeşitli bilgi arama davranış, teknik ve yöntemleri ile bir bilgi sistemi veya merkezinden ulaşılması/elde edilmesidir. Metin madenciliği teknik ve yöntemleri ise, sahip olduğu bilgi erişim sistemi geliştirme kabiliyeti ile, bilgi erişimin üst bir seviyeye gelmesini sağlayabilmektedir. Bilgi erişimdeki önemli noktalardan bir tanesi “*kullanıcının bilgi ihtiyacını giderecek bilgi*”yi sağlamaktır. Talib ve arkadaşlarına (2016, s. 416) göre; “bilgi erişim, bir arama motoru üzerinden belirli bir kelime veya kelime grubuna göre, ilgili ve ilişkili bilgilere ulaşmadır”. Benzer şekilde Tandel ve arkadaşları da (2019, s. 2019) kullanıcıların gerçekleştirmiş oldukları tarama işlemi sonucunda eriştikleri bilgilerin ilgili olması gerektiğini vurgulamaktadırlar. Burada bilgi erişim sistemlerinin niteliğini ortaya koyan, “ilgili belgelere erişim” söz konusudur. Peki bu ne demektir? Tonta’nın da çalışmasında

(1995, s. 302) belirttiği üzere bilgi erişim sistemleri, derlemde (corpus) bulunan ilgili dokümanların tümüne erişmeli, ilgisiz olanları ise ayıklamalıdır. Bu bilgi erişim sisteminin temel işlevini oluşturmaktadır. Bilgi erişim sistemlerinin temelini oluşturan bu işlevin ise en verimli sonuç vermesini sağlayacak ve niteliğini arttıracak araçlardan önemli bir tanesini metin madenciliği oluşturmaktadır. Nitekim metin madenciliği ile metinsel verilere erişim arasında da yakın bir ilişki bulunmaktadır (Talib ve diğerleri, 2016, s. 416).

En iyi bilinen bilgi erişim sistemlerinden biri, kullanıcıların önemli ve yararlı bilgilere erişmeleri için taratılan belgelerin işlendiği Google arama motorlarıdır. Google arama motorunun mantığını ise World Wide Web'de belirli bir kelime kümesinin belgelerle ilişkilendirilmesi oluşturmaktadır (Dang ve Ahmad, 2014, s. 23). Google gibi arama motorları kullanıcıların eğilimleri izlemekte ve daha önemli ve ilgili sonuçlara ulaşmak için sorgu tabanlı algoritmalar kullanarak, kullanıcıları tatmin eden sonuçlar çıkarmaktadır (Talib ve diğerleri, 2016, s. 416).

Fakat esasında kusursuz bir bilgi erişim sistemi bulunmamaktadır (Tonta, 1995, s. 302). Metin madenciliği gibi teknolojiler bu sistemleri optimize ederek gelebileceği en iyi noktaya ulaştırmaya çalışmaktadır. Metin madenciliği; metadataları otomatik olarak çıkarmak, bir doküman koleksiyonundaki önemli terimleri indekslemek gibi etkili ve popüler bilgi erişim teknikleri sağlamaktadır (Agrawal ve Batra, 2013, S. 119). İşin temelinde bilgi erişim sistemleri belirli bir konu ile ilgili belgeleri kullanıcıya bir arada sunmakta, ilgisiz olanların ise elenmesini sağlamaktadır. Bu sayede kullanıcı sadece bilgi gereksinimi ile ilgili belgelere erişerek ilgisiz belgerin yaratacağı yükten kurtulmaktadır. Bilgi erişim tekniği ile, sadece ilgili belgelere erişerek, üzerine metin madenciliği uygulanacak belge sayısı azaltılmakta (Tandel ve diğerleri, 2019, s. 1023) böylece daha hızlı ve etkili sonuçlar elde edilebilmektedir. Metin madenciliğinin ön aşaması sayılan

otomatik dizinleme de günümüzde arama motorlarının geliştirilmesi ve bilgi erişimin iyileştirilmesinde ihtiyaç duyulmakta olup, önemli ve eski bir araştırma alanını oluşturmaktadır (Dang ve Ahmad, 2014, s. 23). Metin madenciliğinin bilgi erişimi olumlu yönde nasıl etkileyebileceği detaylı bir şekilde 4. Bölümde EBYS'ler üzerinden anlatılmaktadır.

Aşağıda özet bir bilgi erişim süreci sunulmaktadır:

1. Kullanıcının bilgi gereksinimi duyması,
2. Bilgi gereksinimini karşılayacak anahtar kelimeleri çıkarması,
3. Bir bilgi merkezi veya bilgi erişim sistemi üzerinden sorgulama gerçekleştirilmesi.
(Bu sorgulamayı gerçekleştirirken boolean (ve/veya/değil), joker karakter, coğrafi, tarih vb. sınırlılıklar gibi temel arama seçeneklerinin yanı sıra alaka düzeyine göre sıralanmış doğal dil araması, bulanık arama⁸, kavram araması vb. gibi bazı gelişmiş arama yöntemlerini kullanması (Gupta ve Lehal, 2009, s. 74),
4. Sonuçların değerlendirilmesi ve özümzenmesidir.

III. 3. 1. 2. 3. Özetleme: İki basit özetleme yöntemi bulunmaktadır. Bunlardan birincisi çıkarıcı (extractive) ikincisi ise özetleyici (abstractive) yöntemdir (Tandel ve diğerleri, 2019, s. 1024). İlk yöntemde cümleler istatistiksel olarak ağırlandırılmakta ve metin içerisinden önemli cümleler ayıklanmaktadır (Gupta ve Lehal, 2009, s. 62). İkinci yöntem ise bu başlıkta üzerinde duracağımız olanıdır. Bu yöntemde metin yorumlanarak cümleler

⁸ Bulanık arama (Fuzzy search) ile, sorgu sırasında taratılan terime, sözdizimsel olarak benzer terimler eklenerek sorgunun genişletilmesi sağlanır. Böylece yazım hatalarından kaynaklanan kaçırılmış sonuçlar geri kazanılarak, sorgu genişletilir (IBM, 2025).

kısaltılmaktadır (Gündoğdu ve Duru, 2016). Bir veya birden çok belgeden üretilebilen özetleme, metnin içeriğinin anlamını ve en önemli noktalarını koruyarak belgenin uzunluğunu ve ayrıntılarını azaltmaktadır. Kullanıcıya yararlı bilgi ile metnin sıkıştırılmış özlü temsilini otomatik olarak oluşturma sürecidir (Birunth ve Sukanya, 2012, s. 269; Dang ve Ahmad, 2014, S. 23; Gaikwad ve diğerleri, 2014, s. 44; Tandel ve diğerleri, 2019, s. 1024). Bir örnek vermek gerekirse; “Bilgi erişimde; makine öğrenimi, yapay zekâ, doğal dil işleme vb. teknolojileri kullanmak verim yaratmaktadır” cümlesine bu yöntem uygulandığında “Bilgi erişimde dijital teknolojilerin kullanımı verim yaratmaktadır” şeklinde özetlenebilmektedir.

Metin özetleme, kullanıcıya dokümanın tamamının okunmasına gerek kalmadan bilgi vermektedir. Kullanıcı özeti okuyarak, belgenin bilgi ihtiyacını karşılayıp karşılamayacağı ile ilgili zaman kazanmaktadır. Böylece belgenin tamamının okunmaya değer olup olmadığını saptayabilmektedir (Gupta ve Lehal, 2009, s. 62). Metinlerin özetlenmesi sürecinde; “sonuç olarak, böylece vb.” anahtar kelimelerini takip eden cümleler ile metnin ana noktaları tespit edilebilmektedir. Bunun yanı sıra metinde bulunan başlıklar, giriş bölümü vb. kısımlar metnin kilit noktalarını belirlemede özetleme araçları tarafından kullanılan yöntemleri oluşturmaktır (Gupta ve Lehal, s. 63, 2009). Metin özetleme; sinir ağları, karar ağaçları, anlamsal grafikler, regresyon modelleri ve bulanık mantık⁹(fuzzy logic) gibi çeşitli yöntemleri içermektedir (Dang ve Ahmad, 2014, s. 23).

⁹ Bulanık mantık, karmaşık ve kesin olmayan durumlarda kullanılmaktadır. Klasik mantıkta elemanların kümeye dağılımı kesindir ve 1 (evet) ve 0 (hayır) şeklinde gösterilir. Ancak bazı durumlarda sonuçlar kesin olmamaktadır. Söz konusu durumlar için bulanık mantık, 1 ve 0 arasında bir değer atamaktadır (Ödük, 2025, s.1).

Metin özetleme süreci kısaca aşağıdaki gibidir:

1. “...Öncelikle ham metin alınır ve üzerinde ön işleme ve işleme işlemleri gerçekleştirilir. Ön işlemede tokenizasyon, stemming ve kelime çıkarma yöntemleri olmak üzere üç yöntem uygulanır” (Tandel ve diğerleri, 2019, s. 1024),
2. “İşleme adımında bir algoritma metin yapısını bir özet yapısına dönüştürmelidir” (Gupta ve Lehal, 2009, s. 63),
3. “Son adımda özet yapısından nihai özet elde edilir” (Gaikwad ve diğerleri, 2014, s. 44).

III. 3. 1. 2. 4. Kümeleme: Küme birbirleriyle ilişkili dokümanlardan oluşan bir grubu temsil etmektedir. Kümeleme ise benzer metinleri/dokümanları otomatik bir şekilde gruplara ayıran yöntemdir (Birunth ve Sukanya, 2012, s. 270; Gupta ve Lehal, 2009, s. 74). Her küme bir dizi dokümandan oluşmaktadır ve bir küme içerisindeki dokümanlar birbirlerine benzerken; her bir küme arasında farklılıklar vardır (Birunth ve Sukanya, 2012, s. 270; Gaikwad ve diğerleri, 2014, s. 44). Başka bir deyişle bir küme içerisindeki temalar benzerken, kümeler arası temalar farklı olacak şekilde gruplara ayrılmıştır (Tandel ve diğerleri, 2019, s. 1024). Kümelemenin en önemli özelliklerinden bir tanesi bir denetimsiz öğrenme tekniği olmasıdır. Bu nedenle kümeler, benzer terim ve desenler herhangi bir ön bilgi olmaksızın çeşitli belgelerden çıkarılarak, önceden tanımlanmış konuları kullanmadan o an kümelenmektedir. Dolayısı ile nesnelerle ilişkili etiketlerin tamamı verilerden elde edilmektedir (Gupta ve Lehal, 2009, s. 64; Dang ve Ahmad, 2014, S. 23; Talib ve diğerleri, 2016, s. 416). Bu özelliği ile de kümeleme, kategorizasyondan/sınıflamadan ayrılmaktadır. Kümelemenin amacı bilgideki içsel yapıların ve desenlerin keşfedilmesidir. Bunun için de bir grup nesneyi veya veriyi anlaşılabilir alt gruplara bölmektedir (Dang ve Ahmad, 2014, s. 23; Tandel ve diğerleri, 2019, s. 1024).

Kümelemenin grupları önceden tanımlanmış konular olmaksızın gruplandığını belirtmiştik. Peki kümeleme nasıl çalışmaktadır? Kümeleme, elindeki verileri kullanarak bazı benzerlik ölçütleri ile grupları belirlemektedir (Gupta ve Lehal, 2009, s. 74). Gaikwad ve arkadaşları (2014, s. 44) kümelemenin genel çalışma prensibini şu şekilde açıklamaktadır: Temel bir kümeleme algoritması, her belge için bir konu vektörü oluşturmaktadır ve dokümanın her bir kümeye ne kadar iyi uyduğunu ağırlıklandırarak ölçmektedir.

Kümelemenin bilgi erişim noktasında sunduğu önemli faydaları bulunmaktadır. Kümeleme sayesinde kullanıcılar erişmek istedikleri bir konudaki belgeleri bir arada görüntüleyebilmektedir. Böylece yararlı (ilgili) sonuçların kaçırılması engellenerek kesin isabet oranı artmaktadır. Bunun yanı sıra yine kullanıcı bir belge koleksiyonunu keşfederken, kümeleme sayesinde koleksiyon hakkında genel bir bakış elde etmektedir (Gupta ve Lehal, 2009, s. 74). Bu özellikleri ile kümeleme, özellikle bilgi/belge yönetim sistemleri için önemli bir tekniktir. Bu sistemlerde yönetilmesi gereken çok fazla bilgi ve doküman bulunmaktadır. Dolayısı ile kümeleme, organizasyonlar için büyük önem arz etmektedir. Organizasyonlar dışında kümeleme; “biyoloji, veri madenciliği, desen tanıma, belge alma, görüntü segmentasyonu, desen sınıflandırması, güvenlik, iş zekası ve web araması gibi birçok uygulama alanında faydalıdır” (Dang ve Ahmad, 2014, S. 23).

Kümeleme esasında iki türe ayrılmaktadır. Bunlar; Hiyerarşik ve Hiyerarşik olmayan kümelemedir (Tandel ve diğerleri, 2019, S. 1024). Ancak kümelemenin kullanılan çok daha fazla türü bulunmaktadır. Bunların en çok bilinenlerinden bir tanesi k-means (k-ortalamar)dır.

Gupta ve Lehal (2009, s. 64) çalışmalarında k-means algoritmasının önemi ve sağladığı faydaları aşağıdaki şekilde anlatmaktadır:

K-means kümeleme algoritmasında, metin belgeleri arasındaki benzerlik, yalnızca terim sıklığı istatistikleri algoritmasına dayalı özvektörü dikkate almakla kalmaz, aynı zamanda kelimeler arasındaki ilişki derecesini de birleştirir. Daha sonra anahtar kelimeler arasındaki ilişki dikkate alınır, böylece girdi dizisinin ve sıklığının hassasiyetini azaltır. Bir dereceye kadar anlamsal anlayışı dikkate alır. Küçük metin ve basit cümlelerin benzerliğinin doğruluğunu, metin kümeleme sonucunun kesinliğini ve bilgi erişim oranını etkili bir şekilde artırır.

III. 3. 1. 2. 5. Kategorizasyon: Kategorizasyon, kümeleme tekniğinin aksine bir denetimli öğrenme tekniği örneğidir. Kategorizasyonu aslında bilgi merkezlerinde kullanılan sınıflama sistemleri (dewey, lc vb.) olarak düşünmek çok da yanlış olmaz. Bir nevi bilgi uzmanlarının bilgi kaynaklarına yazar ve konu bilgisini içeren yer numaraları verme işleminin otomatikleştirilmesi denilebilir. Kategorizasyon konuların önceden tanımlandığı bir konu başlığı listesi yapısına dayanan ve yeni belgeleri sınıflamak için girdi-çıkı örneklerini kullanan bir tekniktir (Tandel ve diğerleri, 2019, S. 1024; Gaikwad ve diğerleri, 2014, s. 44). Sınıflandırıcının temel amacı, bilinen örnekler üzerinden eğitilerek bilinmeyen belgelere ait ana tema/konu veya temaların doğru bir şekilde belirlenmesini sağlamaktır. (Birunth ve Sukanya, 2012, s. 270; Gupta ve Lehal, 2009, s. 63; Dang ve Ahmad, 2014, S. 23). Buradan kategorizasyon ile ilgili çıkarılabilecek bir başka sonuç ise bir belgeye bir veya birden fazla konunun atanabileceğidir.

Kategorizasyonun başlarda öngörülen temel kullanımı, kontrollü sözcükler aracılığıyla bilimsel literatürün indekslenmesi iken 1990'lardan itibaren bilgi teknolojilerinin ilerlemesi ve dijitalleşmenin hızlanması ile bu kullanım şu şekilde değişmiştir:

Kategorizasyon artık belgelerin daha kolay kullanımı için erişim amacı ile düzenlenmesi

gereksinimiyle gelişmiştir (Dang ve Ahmad, 2014, S. 23). “Tipik bir metin kategorizasyon süreci; ön işleme, dizinleme, boyutsal indirgeme ve sınıflandırmadan oluşur” (Lam ve diğerleri, 1999; Aktaran Gaikwad ve diğerleri, 2014, s. 44; Lodhi ve diğerleri, 2002; Aktaran Gaikwad ve diğerleri, 2014, s. 44). “Özetlemede olduğu gibi, kategorizasyon da bir belgenin bir konu hakkında bilgi arayan bir kişiyle ilgisini daha fazla belirlemek için konu izleme ile birlikte kullanılabilir. Konu takibinden dönen belgeler içerik ağırlıklarına göre sıralanabilir. Böylece bireyler önce en alakalı belgelere öncelik verebilir.” (Gupta ve Lehal, 2009, s. 63). Bu özellikleri ile kategorizasyonun bilgi erişime sağladığı katkılar çok önemlidir. Öyle ki bilgi uzmanlarının manuel olarak gerçekleştirdiği pek çok işlemin otomatikleştirilmesini kapsamaktadır. Otomatik metin kategorizasyonu; belge ve bilgi erişimi için indeksleme, konu meta verilerin otomatik olarak çıkarılması, bir metnin kapsadığı konuları tespit ederek metnin bağlamı hakkındaki belirsizlikleri giderme, web kaynaklarına ait büyük katalogları düzenleme ve sürdürme gibi birçok pratik uygulamaya sahiptir (Agrawal ve Batra, 2013, s. 119). Bunlar dışında kategorizasyonun başka kullanım alanları da mevcuttur. Örneğin; kamu kurumları veya işletmelerin neredeyse hepsi web sitelerinde müşterilerden gelen sorulara yanıt verebilmek amacıyla bir destek merkezi sağlamaktadırlar, organizasyonlar bu destek merkezlerinde belgeleri konularına göre sınıflandırmak için kategorizasyon şemalarını kullanırlarsa müşteriler veya son kullanıcılar aradıkları bilgilere kolaylıkla erişebileceklerdir (Gupta ve Lehal, 2009, s. 64). Tüm bunlar dışında kategorizasyon; metinleri otomatik veya yarı otomatik indeksleme, kişiselleştirilmiş reklam dağıtımı, spam filtreleme, hiyerarşik kataloglar altında web sayfalarını kategorize etme, otomatik metadata üretme, metin temasını otomatik algılama ve konu takibi ve gibi çeşitli kullanım alanlarına sahiptir (Agrawal and Batra, 2013; Aktaran Dang ve Ahmad, 2014, s. 23). Kategorizasyon günümüzde makine öğrenmesinde kullanılan taze bir konu olup Naive

Bayes, En Yakın Komşu, Karar Ağacı ve Destek Vektör Makineleri gibi istatistiksel sınıflandırma algoritmalarını kullanmaktadır (Agrawal ve Batra, 2013, s. 119; Gaikwad ve diğerleri, 2014, s. 44).

III. 3. 1. 2. 6. Konu Takibi (Topic Tracking): Konu takibi aslında günümüzde çokça yaygınlaşmış ve aktif olarak kullandığımız özelliklerden biridir. Bir konu takibi (konu izleme) sistemi esasında kullanıcının görüntülediği belgelere dayanarak, kullanıcının ilgisini çekecek diğer belgelerin tahmini ile çalışmaktadır (Gupta ve Lehal, 2009, s. 61). Bilgi erişim ve karar almada önemli olan bu teknik, bilgi merkezlerinde sunulan Seçmeli Bilgi Yayımı (Selective Dissemination Of Information) hizmeti ile de benzer olduğunu söylenebilir. Seçmeli Bilgi Yayımı hizmetinde de kullanıcıya ait bir profil çıkarılmakta ve profile uygun şekilde kullanıcının ekstra bir talepte bulunmasına gerek kalmadan, kişiselleştirilmiş bir şekilde en yeni kaynaklar sunulmaktadır. Konu takibinin ise aynı mantıkla, daha gelişmiş ve etki alanının daha yaygın hale getirilmiş versiyonu olduğunu söyleyebiliriz.

Anahtar kelime çıkarımı; bilgi erişim, metin kategorizasyonu, özetleme gibi metin madenciliği uygulamalarının yanı sıra konu tespiti içinde bir temel oluşturmaktadır (Lee and Kim, 2008; Aktaran Gupta ve Lehal, 2009, s. 62). Buna bir örnek olarak “Yahoo, kullanıcıların anahtar kelimeler seçmesine olanak tanıyan ve bu konularla ilgili haberler çıktığında onları bilgilendiren ücretsiz bir konu izleme aracı (www.alerts.yahoo.com) sunmaktadır” (Gupta ve Lehal, 2009, s. 61). Bunun gibi daha birçok örnekte günümüzde kullanılmaktadır. Kullanıcılarına ait bir profil kaydedip sıklıkla incelediği yayınlara benzer kaynakların önerildiği veritabanları ya da araştırılan konuları kaydedip benzer kaynakların mail adreslerine sürekli bir şekilde bildirildiği Academia platformu konu takibi tekniğini kullanan platformlara örnek verilebilir. Konu takibinin kullanıldığı bir diğer teknik de RSS’dir (Really Simple Syndication, ya da Rich Site Summary olarak da

bilinir). RSS ile kullanıcılar; blog, podcast, haber sitesi gibi favori web sitelerine abone olarak, en yeni içeriklere otomatik ve tek bir arayüzden erişebilmektedir. Bu işlem bir RSS okuyucu ile gerçekleşmektedir (Jamali ve diğerleri, 2024). RSS okuyucu kullanıcının abone olduğu tüm web sitelerini düzenli olarak tarar ve yeni bilgileri otomatik bir şekilde çeker. Bu sayede kullanıcı, farklı web sitelerini ziyaret etmesine gerek kalmadan bilgilere hızlı ve kolay bir şekilde erişebilmektedir. Bu özellikleri ile konu takibinin aynı zamanda bir bilgi erişim tekniği olduğu da söylenebilir. Bu tekniğin önemi ise; bir araştırmacı için, güncel bilgilere erişim kolaylığı iken; bir şirket için, rakibinin yer aldığı haberleri görüntülemek ve etkin karar almayı sağlamaktır. Gupta ve Lehal (2009, s. 62) bu örnekleri şu şekilde sıralamışlardır: “...işletmeler kendi şirketleri ve ürünleriyle ilgili haberleri takip etmek isteyebilir. Tıp alanında, hastalıklar için yeni tedaviler arayan ve en son gelişmeleri takip etmek isteyen doktorlar ve diğer kişiler tarafından da kullanılabilir. Eğitim alanındaki bireyler de ilgi alanlarındaki araştırmalar için en son referanslara sahip olduklarından emin olmak için konu takibini kullanabilirler”. Dolayısı ile sadece kişisel ilgi alanları için kullanılabilecek konu takibi aynı zamanda kurumsal amaçlar ile de tüm sektörlerle uyartılabilir.

III. 3. 1. 2. 7. Doğal Dil İşleme: Metin madenciliği ve doğal dil işleme, metinsel verilerden bilgi keşfi yöntemlerine sahip iki teknik olup, doğal dil işleme tekniği metin madenciliğinde ortaya çıkan sorunları azaltarak, açık ve örtük kavramları ve kavramlar arasındaki anlamsal ilişkileri çıkartmaktadır (Gharehchopogh ve Khalifelu, 2011; Talib ve diğerleri, 2016, s. 415; Gupta ve Lehal, 2009, s. 73). Metin madenciliği tekniği genellikle metinsel verilerin işlenip analiz edilmesini sağlayabilmek için Doğal Dil İşleme veya Doğal Dil İşlemeye dayalı olan hesaplamalı algoritmalar kullanmaktadır. Bu sayede metin, bilgisayar tarafından anlamlandırılarak; soru yanıtlama, ilişki ve gerçekleri tanımlama, metni sentezleme gibi özellikler kazanmaktadır (Agrawal ve Batra, 2013, S.

119). Dolayısı ile aslında şunu söylemek yerinde olacaktır: Doğal dil işleme; sentezleme, ilişki kurma, analizleme gibi özellikleri ile insan ve bilgisayar dili arasındaki uçurumu indirgeyerek; bilgisayarların analiz, anlama hatta üretebilme yeteneklerini sağlayan teknolojiler geliştirmiştir (Gupta ve Lehal, 2009, s. 61). 1960’lardan bu yana geliştirilmekte olan Doğal Dil İşleme, yapay zeka ve dilbilimin alt alanı olup amacı, doğal dillerin üretilmesi ve gerçekleştirilmesidir. Başka bir yönünü de metinlerde bulunan kelimelerin işlenmesini içeren istatistiksel teknikler oluşturmaktadır. (Manning and Schutze, 1999; Aktaran Gharehchopogh ve Khalifelu, 2011). Fakat Doğal Dil İşlemede de diğer tekniklerde olduğu gibi karşılaşılan bazı zorluklar bulunmaktadır. Bu zorluklar da genel olarak dilsel problemlerdir. Gharehchopogh ve Khalifelu (2011) bu zorlukları çalışmalarında şu şekilde dile getirmiştir: “NLP sistemlerinin karşı karşıya olduğu bazı sorunlar dilsel çeşitlilik ve belirsizlik içerir. Dilsel çeşitlilik, aynı fikri iletmek için farklı kelimeler veya ifadeler kullanma olasılığı anlamına gelirken, dilsel belirsizlik, bir kelime veya ifadenin birden fazla yoruma izin vermesidir”. Bu tür sorunların tespiti ile bunların tekdüze tanımlanması için ise doğal dil işlemenin bazı kurallar koyması gerekmektedir (Calvillo ve diğerleri, 2013; Aktaran Talib ve diğerleri, 2016, s. 416). Ontoloji, bu sorunları gidermeye yardımcı olacak uygulamalardan bir tanesidir. Ontoloji ile dil ve erişim sorunlarının üstesinden ne şekilde gelindiği 4. Bölümde işlenmektedir. Tüm bu özellikler sayesinde web arama motorlarında, otomatik çeviri araçlarında, özet oluşturucularda ve soru yanıtlama sistemlerinde sıkça başvurulan bir yöntem olan doğal dil işleme, bu alanlarda bilgi erişimi ile ilgili büyük faydalar yaratmaktadır (Baeza-Yates, 2004; Aktaran Gharehchopogh ve Khalifelu, 2011).

III. 3. 1. 2. 8. Soru Yanıtlama: Doğal dil işlemenin bir uygulaması ve aynı zamanda bir metin madenciliği tekniği olarak da kullanılan Soru Yanıtlama; doğal dil sorguları veya belirli bir soruya en iyi cevabın nasıl bulunacağıyla ilgilenmektedir (Gupta ve Lehal,

2009, s. 66). Günümüzde neredeyse bütün organizasyonların kullanıcılarına/müşterilerine sunmuş olduğu bir hizmet olup, aslında bu teknikte birden fazla metin madenciliği tekniği kullanılabilir. Örneğin; kişiler, yerler, olaylar gibi nesneleri/varlıkları ayıklamak için bilgi çıkarma, farklı dillerde yayınlanan web sayfalarını tanımlama ve analiz etme veya soruları bilinen türlere (kim, nerede, ne zaman, nasıl, vb.) atamak için kategorizasyon kullanılabilir” (Gupta ve Lehal, 2009, ss. 66-72). Bir doğal dil uygulaması olarak doğal dil işlemenin karşılaştığı olduğu dilsel zorluklar bu teknikte de mevcuttur. Bir önceki başlıkta bu zorlukların üstesinden gelmenin bir yolu olarak ontoloji uygulaması örnek verilmiştir. Bu sorunların üstesinden ne şekilde gelinebileceği 4. Bölümde Bilgi Tanımlama başlığında daha detaylı işlenmektedir fakat burada da kısaca üzerinde durmak faydalı olacaktır. Ontoloji uygulamaları, erişimi olumlu yönde etkileyen uygulamalar olup bir terimin çok daha farklı seçenekler ile ilişkilendirilmesini ve kullanıcının ilgili ve daha çok sonuca ulaşmasını sağlamaktadır. Ontoloji, bu dil sorunlarını azaltarak hem soru cevaplama sistemlerinde hem de diğer doğal dil uygulama alanlarında kullanıcının bilgi erişim sürecini etkili hale getirmektedir.

III. 3. 1. 2. 9. Trend Analizi: Eğilimlerin tespit edildiği bir diğer metin madenciliği tekniğidir. Pazarlama ve araştırma alanları için önem taşımaktadır. Temel olarak; belirli bir süre boyunca toplanan dokümanlardaki trendleri/eğilimleri belirlemek için kullanılır (Gupta ve Lehal, 2009, s. 74). Literatür incelendiğinde önemli trend analizi çalışmaları görülmektedir; örneğin Sharma ve arkadaşları (2018) metin madenciliği tekniği kullanarak makine öğrenmesi alanında gerçekleştirilen çalışmalara trend analizi uygulamışlar ve makalelerde geçen terimlerin yıllar içindeki eğilimlerini saptamışlardır. Bunun dışında trend analizi bir şirketin veya bir müşterinin ilgi alanlarını bir alandan diğerine kaydırıldığını keşfetmek için kullanılabilir (Gupta ve Lehal, 2009, s. 74).

Bu noktada trend analizinin ilerisi için karar alma mekanizması gerçekleştirme ve buna göre politikalar oluşturmada sağlayacağı faydalardan bahsedilebilmektedir.

III. 3. 1. 2. 10. Kavram Bağlantısı (Concept Linkage): Kavram bağlantısı, bilgi erişim ve bilgi keşfi için özellikle araştırma alanında önem kazanmaktadır. Kavram bağlantısı sayesinde çok büyük bilgi hacmine sahip dokümanlar arasında ortak kavramlar tanımlanarak birbirlerine bağlanmakta ve ilişkilendirilmektedir. Bu sayede kullanıcılar belki de normalde erişemeyecekleri bilgilere erişebilmektedir. Nitekim bir insanın, çok büyük hacme sahip veriler arasında, bunu otomatik bir şekilde yapan sistem kadar iyi performans göstermesi beklenemez (Fan ve diğerleri, 2006, s. 79). Kavram bağlantısının asıl önemi ise şu noktada ortaya çıkmaktadır; bir metin madenciliği çözümü X, Y ve Z konuları arasındaki bağlantıyı keşfedebilmektedir fakat asıl can alıcı kısmı X ve Z arasındaki potansiyel ilişkileri ortaya koyabilmesinde yatmaktadır (Fan ve diğerleri, 2006, s. 80). Bu bilim ve araştırma alanlarının gelişimi için çok değerli bir özelliktir çünkü bazı ilişkilerin kurulması tecrübe ve bilgi birikimine bağlıdır. Örneğin göreve yeni başlamış bir doktorun bir hastaya koyacağı teşhis ve akabinde uygulayacağı tedavi süreci tecrübeli bir doktora göre fark yaratabilmektedir kavram bağlantısının sağlamış olduğu ilişki kurabilme özelliği bir doktorun uzun yıllar sonra elde edebileceği bir tecrübeyi birkaç saniyede uygulamaktadır. Görevine yeni başlamış tecrübesiz bir kataloglama uzmanı veya indeksçi de elindeki kaynağı tanımlamak için bu yöntemi kullanabilir. Bu istisnasız tüm alanlara uyarlanabilecek bilgi keşfi ve erişimi için çok değerli olan metin madenciliği tekniğidir.

III. 3. 1. 2. 11. Birliktelik Kuralı Madenciliği: Birliktelik Kuralı Madenciliğinin ticari karar verme süreçlerinde nasıl kullanıldığı ve faydalarını Veri Madenciliği yöntemlerini açıklarken belirtilmiştir. Bu nedenle bu başlıkta yalnızca metin madenciliğindeki odak noktasına değinilecektir. Birliktelik Kuralının değişkenler arasındaki ilişkilerin

keşfedilmesinde kullanıldığından önceki başlıklarda bahsedilmişti, metin madenciliğinde ise aslında çok farklı değil. “Metin Madenciliği için Birliktelik Kurallarında odak noktası, bir derlemi karakterize etmek için kullanılan konular veya tanımlayıcı kavramlar arasındaki ilişkileri ve çıkarımları incelemektir” (Wong ve diğerleri; Aktaran Gupta ve Lehal, 2009, s. 67). Örneğin bir makalede “Dewey” ve “kütüphane” terimleri geçiyorsa büyük ihtimalle sınıflama sistemleri ile ilgili bilgi verildiği öngörülebilir.

Gupta ve Lehal (2009, s. 67) çalışmalarında metin madenciliğinde birliktelik kuralı sürecini aşağıdaki gibi açıklamaktadır:

1. Anlatı metinlerinden oluşan bir külliyat, konu çıkarımı için bir metin motoruna aktarılır.
2. Madencilik motoru daha sonra metin motorundan konuları okur ve konu ilişkilendirme kuralları oluşturur.
3. Son olarak, ortaya çıkan ilişkilendirme kuralları daha fazla analiz için görselleştirme sistemine gönderilir.

III. 3. 1. 2. 12. Bilgi Görselleştirme: Bilgi görselleştirme, büyük miktardaki metinsel kaynaklara ait eşleşmeleri göstererek kullanıcıların içeriği analizleyebileceği görsel bir hiyerarşi veya harita ortaya koyan araçtır (Birunth ve Sukanya, 2012, s. 270; Gupta ve Lehal, 2009, s. 65). Kullanıcıların bilgiyi daha kolay keşfetmesini sağlamaktadır. Sağlamış olduğu tarama yetenekleri ile belge koleksiyonunun grafiksel bir temsili yaratmaktadır. Bu temsili oluştururken ise özellik çıkarma ve anahtar terim indeksleme yöntemlerini kullanmaktadır (Birunth ve Sukanya, 2012, s. 270; Gupta ve Lehal, 2009, s. 74). Erişim açısından büyük kolaylık sağlayan bilgi görselleştirme; daha iyi ve hızlı anlaşılır bilgiyi sağlama, renkler, ilişkiler, uzaklık vb. ayırt edebilme, haritayla etkileşime geçebilme (yakınlaştırma, ölçeklendirme, alt harita oluşturma vb.), belgenin konumunu ve konularını saptama (Birunth ve Sukanya, 2012, s. 270; Gupta ve Lehal, 2009, ss. 65-

74) gibi özellikleri ile ilgili, örtük ve yeni bilgilerin keşfedilmesini sağlamaktadır. Esasında metin madenciliği açısından değerlendirildiğinde bilgi görselleştirme; dizinleme, vektör uzay modeli vb. yöntemler kullanılmaktadır. Bu yöntemler ile bilginin; yapılandırılmış bir formatta, iki veya üç boyutlu temsilini elde etmektedir (Gupta ve Lehal, 2009, s. 74; Birunth ve Sukanya, 2012, s. 270). Bilgi görselleştirme için dikkat edilmesi gereken bir nokta ise; görselleştirmenin anlamlı olması için belirli görselleştirme özelliklerinin bozulmadan korunmasını sağlamaktır. Buna örnek olarak; bir kelime bulutu (Word Cloud) oluştururken, yüksek frekanslı terimlerin daha büyük puntolar ile düşük frekanslıların ise daha küçük puntolar ile gösterilmesi verilebilir. Bilgi görselleştirme sadece araştırma noktasında kullanıcılara bilgi sağlamanın dışında birçok sektörde faydalar sağlamaktadır; hükümet tarafından terörist ağıları tespit etmek veya daha önce birbiriyle bağlantısı olmadığı düşünülen suçlar hakkında bilgi bulmanın yanı sıra (Gupta ve Lehal, 2009, s. 65), eğitim, sağlık, pazarlama alanlarında da bilgi keşfi ve karar almayı sağlayacak bilgilerin ortaya çıkarılmasında önem arz etmektedir.

III. 3. 1. 3. Metin Madenciliği Kullanım Alanları

Buraya kadar metin madenciliğinin ne olduğu, neden ihtiyaç duyulduğu ve teknikleri üzerinde durulmuştur. Tüm bu başlıklarda aslında metin madenciliğinin nerelerde kullanılabilir olduğu ile ilgili bir fikir elde edilmiştir. Fakat metin madenciliği bunlardan da fazlasını kapsamaktadır. Metin madenciliğinin kullanım alanlarının bir sınırı olmadığını söylenebilir. Tekniklerin birçoğu, birçok alana ve sektöre uyarlanır özelliklere sahiptir. Eğitim ve araştırma alanlarında yeni bilgilerin ortaya konarak bilimin gelişmesinden; kurumların sürekliliklerini sağlayacak öngörülü bilginin keşfedilmesine kadar çok fazla alanda önemli faydalar sağlamaktadır. Bu bölümde metin madenciliği

kullanım alanları toplam yedi başlıkta işlenmiş olup bunlar: Pazar Analizi, Eğitim ve Araştırma, Müşteri İlişkileri Yönetimi, İnsan Kaynakları Yönetimi, Yaşam Bilimleri, Sosyal Medya ve Güvenliktir. Bilgi sistemlerinin geliştirilmesi, bilgi erişim ve organizasyonlara faydalarını içeren yenilikçi metin madenciliği kullanım alanları ise 4. Bölümde ele alınmıştır.

III. 3. 1. 3. 1. Pazar Analizi: Bir organizasyonun büyümesi ve sürekliliğini sağlaması için; rakiplerini, çalışanlarını, satışları, rakip organizasyonların ürünleri vb. gibi pazar koşullarını bilmeye ihtiyacı vardır (Birunth ve Sukanya, 2012, s. 270). Metinsel verilerden bilgi çıkarımının sağlamış olduğu; bilgi keşfi, öngörülü bilgi ve bunların karar almayı sağlaması, pazarı analiz etme ve organizasyonların avantaj sağlamasında büyük öneme sahiptir. Metin madenciliği, organizasyonlarda örtük bilgiyi ortaya çıkararak fırsat ve tehditlere karşı davranış geliştirilmesini sağlamaktadır.

Günümüzde başa çıkılması gereken çok fazla veri vardır bunların içerisinde de çok değerli bilgiler yatmaktadır. Söz konusu verilerin büyük çoğunluğunu metinsel veriler oluşturmaktadır. Bu verilerin insanlar tarafından analiz edilmesi çok zor belki de imkânsız olup yoğun çaba ve emek gerektirmektedir. Verilerin ise hiç durmadan artması, sürekli olarak yenilerinin eklenmesi nedeniyle otomatikleşme kaçınılmazdır. Birunth ve Sukanya (2012, s. 270) çalışmalarında önceden bu işlemlerin manuel bir şekilde ve her organizasyonda ayrı bir bölümde gerçekleştirildiğinden bahsederek, metin madenciliğinin bu süreçte büyük kolaylık sağladığını belirtmektedirler. Pazarlamada bu bilgilerin keşfi noktasında metin madenciliğinin yeri dikkate alınmalıdır. Sadece iç bilgi kaynakları değil, web üzerinde de elde edilecek pazarlama bilgisi, sektörel teknik incelemeler, ticaret dergileri, pazar haberleri, müşteri yorumları gibi çok fazla bilgi kaynağı bulunmaktadır (Gupta ve Lehal, 2009, s. 71).

Organizasyonun bilgi ihtiyacına uygun, organizasyona fayda sağlayacak öngörülü bilgi elde etme, müşteri memnuniyetini artırma ve rakipleri analiz ederek rekabet avantajları sağlamada metin madenciliği kullanımı önemli role sahiptir (Fatima ve diğerleri, 2010; Aktaran Talib ve diğerleri, 2016, s. 417).

Stratejilerini taleplere ve pazarın sunduğu fırsatlara göre düzenleme ve değiştirme ihtiyacı ile organizasyonlar; veriyi toplama, düzenleme, analiz etme, depolama ve kullanmalarını gerektiren bir süreci yönetmeleri gerekmektedir (Gupta ve Lehal, 2009, s. 68).

Organizasyonlara iş zekası, rekabetçi istihbarat ve pazar eğilimlerini analiz etme konularında büyük katkı sağlayarak pazardaki konumlarını, sürekliliği ve ardından gelişimini sağlayacak olan süreç kısaca bahsedilmektedir:

1. İç ve dış bilgi kaynaklarından verilerin toplanması,
2. Toplanan verilere bilgi çıkarımı tekniği uygulayarak organizasyona uygun bilgilerin elde edilmesi,
3. Kategorizasyon ile bilgilerin sınıflandırılması,
4. Bilgilerin bir veritabanına aktarılması,
5. Karar verme ve strateji geliştirme için analiz edilmesi.

III. 3. 1. 3. 2. Eğitim ve Araştırma: Eğitim ve araştırma alanlarındaki metin madenciliği katkısı, genel olarak bilimin ilerlemesine yardımcı olacağı için tüm çalışma alanlarında önem kazanmaktadır. Bunun için en başta bilgi erişim açısından metin madenciliğinin yapacağı katkıları göz önünde bulundurmak gerekmektedir. Nitekim bilgi erişimin performansı ne kadar iyiye ortaya konacak sonuçta o kadar güvenilir ve faydalı olacaktır. Metin madenciliği erişim uçlarının belirlenmesi noktasında; dizinleme, anahtar kelime oluşturma, konu metadataları çıkartma, özetleme gibi bilgiye erişimi etkin hale getirecek

otomatikleşmiş bir süreç ortaya koymaktadır. Dang ve Ahmad de (2014, s. 24) çalışmalarında bahsettikleri gibi; “Büyük bilgi veri tabanlarına sahip yayıncılar için önemli olan konu, erişim için indekslenmeye ihtiyaç duyar. Bu, özellikle oldukça spesifik bilgilerin sıklıkla yazılı metinde bulunduğu bilimsel disiplinlerde geçerlidir”. Bu noktada bilgiye erişimi sağlan ve geliştiren kişiler ve bu hizmetlerin sunulduğu mekanlar olarak; bilgi uzmanları ve bilgi merkezleri için de metin madenciliği önemli bir konudur. Öyle ki kullanıcı için bilgi sistemlerinin ve erişimin geliştirilmesinde bir mimariyi en iyi oluşturabilecek kişiler bilgi uzmanlarıdır. Bu noktada bilgi uzmanlarının dijital teknolojileri çalışmalarına entegre etmeleri, beceri ve yetkinlikler kazanmaları bu alanın gelişmesi için önem taşımaktadır. Makale, bildiri, kitap vb. kaynaklardan kalıplar ve eğilimler oluşturmak için birçok metin madenciliği tekniği vardır (Tandel ve diğerleri, 2019, s. 1024). Trend analizi tekniği kullanılarak farklı konulardaki eğilimleri keşfetmek ve bunların zaman içinde nasıl değiştiğini göstermenin yanında konu izleme için de kullanılabilir (Dang ve Ahmad, 2014, s. 24). Bilgi uzmanlarının hizmetlerine bu teknikleri uygulaması ve kullanıcıya sunması araştırma alanları için son derece değerli bir kaynaktır. Bunlar dışında bilgi uzmanlarının da günümüz gelişmelerine ayak uydurarak kataloglama, sınıflama, dizinleme gibi hizmetler sunarken metin madenciliği araçları gibi otomatikleşmiş sistemleri kullanmaları bilgi ve belge yönetimi alanının gelişimi noktasında da fayda sağlayacaktır.

Günümüzde bilgi teknolojilerinin ilerlemesi ile çoğu hizmet dijital platformlara taşınmıştır. Bu noktada bilgi merkezlerinin, bilgi kaynaklarını dijital platformlarda sunduğu veya direkt olarak dijital platformlarda doğmuş koleksiyonlar (dijital kütüphaneler) ortaya çıkmaktadır. Dijital kütüphaneler, bilgileri düzenlemenin ve çevrimiçi olarak bulunan milyonlarca belgeye erişmenin yenilikçi bir yolunu sağlamaktadır (Tandel ve diğerleri, 2019, S. 1024). Bu platformlarda sunulacak önemli

bilgi erişim hizmetleri bulunmaktadır. Örneğin bilgilerin sınıflandırılmasının yanı sıra arama stratejilerini geliştirebilecek şekilde bilgi çıkarma kullanarak; kişi, yer, kuruluş vb. özellikler ile belgelere erişim sağlanması oldukça faydalı bir uygulama olacaktır. Bu sayede nesneler arası ilişki ve desenleri gösteren şablonlar sunulması, bilgi keşfi ve erişimi için kullanıcılara dikkate değer ölçüde katkı sağlayabilmektedir. (Birunth ve Sukanya, 2012, s. 270). Bunlara ek olarak referans oluşturma, telif hakkı analizi ve özetleme gibi çeşitli işlemler de araştırma alanında önemli metin madenciliği uygulamalarıdır (Witten ve diğerleri, 2004 aktaran Talib ve diğerleri, 2016, s. 416; Dang ve Ahmad, 2014, s. 25). “GATE, Net Owl ve Aylie; dijital kütüphanelerde metin madenciliği için sıklıkla kullanılan araçlardır” (Talib ve diğerleri, 2016, s. 416).

Metin madenciliği bilgi erişim teknik ve sistemlerinin geliştirmesinin yanı sıra eğitim ve öğretim hayatında da faydalar sağlamaktadır. Öğrencilerin bilgileri daha iyi içselleştirebilmelerinin yanında belirli bir bölgedeki eğitim eğilimlerini, öğrencilerin belirli bir alana olan ilgisini ve istihdam oranını analiz etmek için çeşitli metin madenciliği araçları ve teknikleri kullanılmaktadır (Al-Hashemi, 2010; Aktaran Talib ve diğerleri, 2016, s. 416). Bu sayede öğrencilerin eğitim hayatları daha doğru bir şekilde yönlendirilebilmektedir. Dolayısı ile öğretmenlerin öğrencilerine yapacakları danışmanlıklarda da metin madenciliği önemli rol oynamaktadır.

III. 3. 1. 3. 3. Müşteri İlişkileri Yönetimi: Müşteri İlişkileri Yönetimi; çağrı merkezleri, bilgi yönetimi, analitik ve veri madenciliği alanlarındaki yeniliklerin uygulanması için benzersiz bir değere sahiptir (Gupta ve Lehal, 2009, s. 70). Günümüzde organizasyonlar kişiselleştirilmiş müşteri odaklı bir hizmet sunmak için yeni teknolojileri çokça kullanmaktadır. Bu noktada metin madenciliği de bu alana önemli faydalar sağlamaktadır. Gupta ve Lehal (2009, s. 70) çalışmalarında, müşteri ilişkileri yönetiminin daha çok müşteri tarafından yönlendirilen taleplerin otomatik bir şekilde yönlendirildiği veya sıkça

sorulan sorulara otomatik bir şekilde yanıt üretilmesini sağlayan mesaj içerik yönetimi ile ilgili olduğunu belirtmektedirler. Bu noktada müşteri taleplerine otomatik yönlendirme ve sıkça sorulan sorulara otomatik yanıt vermede bilgi çıkarma, kategorizasyon ve soru yanıtlama gibi metin madenciliği tekniklerinin kullanımı söz konusudur. Kullanıcı deneyimini ve memnuniyetini sürdürmek ve geliştirebilmek için müşteri geribildirimlerinin büyük önemi vardır. Kullanıcılar; sosyal medyada, ürün yorumlarında, mesajlarda veya formlarda hizmet veya ürünle ilgili deneyimlerinden bahsetmektedir. Kullanıcıların gerçekleştirmiş olduğu bu geribildirimler, analiz edilmesi gereken önemli metinsel verilerden oluşmaktadır. Organizasyonlar bunları çeşitli metin madenciliği teknikleri ile analiz ederek ürün ve hizmetlerine yönelik karar almalılardır. Dolayısı ile müşteri ilişkilerinde metin madenciliği kullanımı iş zekası geliştirme, öngörülü bilgi elde etme, müşteri odaklı hizmet sunumu ve rakip şirketlerle rekabette etkili yöntemler sunmaktadır.

III. 3. 1. 3. 4. İnsan Kaynakları Yönetimi: Birçok şirkete özellikle de büyük şirketlere her gün çeşitli dillerde, dosya formatında ve farklı biçimlere sahip özgeçmiş gönderilmektedir, metin madenciliği teknikleri bu özgeçmişlerin analizi ve saklanması noktasında daha sistematik ve hızlı bir seçenek sunabilmektedir (Tandel ve diğerleri, 2019, s. 1025). Metin madenciliği bunun dışında personel ve şirket sağlığını analiz etmede de kullanılmaktadır. İnsan kaynaklarını stratejik olarak yönetmek amacıyla personel görüşleri, memnuniyet düzeyi, iç ve dış bilgi kaynakları analizi ile de şirketin durumu ile ilgili bilgi edinerek (Gupta ve Lehal, 2009, s. 70), diğer alanlarda olduğu gibi karar alınır ve organizasyona dair yapılması, değiştirilmesi, kaldırılması veya düzenlenmesi gereken durumlar değerlendirilerek kurumun sürekliliğini sağlayacak adımlar atılır.

III. 3. 1. 3. 5. Yaşam Bilimleri: Yaşam bilimleri, sağlık hizmetleri, biyoenformatik gibi alanlar metin madenciliği için önemli kullanım alanlarına sahip olup insanlığa büyük katkılar sunmaktadır. Hasta bilgileri, hastalıklar, ilaçlar, hastalık belirtileri, tedaviler gibi birçok konuda büyük miktarda metinsel ve sayılar veriler üretilmektedir (Talib ve diğerleri, 2016, s. 416) Bu kadar yüklü miktarda metinsel verilerin bilgisayar tarafından işleme zorluğu daha önce ele alınmış olup bu analizleri manuel olarak gerçekleştirmekte zaman alıcı ve çok yorucu olmaktadır. Tıbbi kayıtların da, bilgi keşfi sürecini çok zorlaştıran, karmaşık, uzun ve teknik kelimeler içeren çeşitli yapılarda olduğunu unutmamak gerekir (Henriksson ve diğerleri, 2016; Aktaran Talib ve diğerleri, 2016, s. 416). Bu noktada sınıflandırma, ilişki çıkarma, özetleme ve soru cevaplama gibi çeşitli üst düzey bilgi erişim görevlerinde metin madenciliği önem kazanmaktadır (Dang ve Ahmad, 2014, s. 24). Bu tekniklerle metinsel veriler analiz edilerek değerli bilgiler elde edilmekte; çeşitli hastalıklar, türler ve genler arasındaki ilişkiler çıkarılmakta; farklı hastalıkları, semptomları ve bunların tedavi süreçlerini karşılaştırarak tıbbi tedavilerin etkinliğinin değerlendirilmesi sağlanmaktadır (Talib ve diğerleri, 2016, s. 417). Tüm bu gelişmelerle hem tıp dünyasına hem de canlılara önemli katkılar yaratılmaktadır. Talib ve arkadaşları çalışmalarında yaşam bilimlerinde metin madenciliği uygulamalarını şu şekilde sıralamaktadır: biyobelirteç keşfi, ilaç endüstrisi, klinik ticaret analizi, klinik öncesi güvenli toksisite çalışmaları, patent rekabet istihbaratı ve çevre düzenlemesi, gen hastalıklarının haritalanması ve çeşitli araçlar kullanılarak hedeflenen tanımlamaların keşfedilmesi (Cohen ve Hersh, 2005; Aktaran Talib ve diğerleri, 2016, s. 417).

III. 3. 1. 3. 6. Sosyal Medya: Metin madenciliği sosyal medya analizinde çok uygulanan bir çalışma alanıdır. Literatür incelendiğinde bununla ilgili birçok araştırma görülmektedir. Metin madenciliğinde sosyal medya analizi daha çok sosyal medya, bloglar, internet haberleri veya e-postalar vb. üzerinden analiz ile gerçekleştirilmektedir

(Talib ve diğeri, 2016, s. 417). Bu tür incelemeler ile takipçi, gönderi ve beğeni sayısı analiz edilerek insanların çeşitli haberlere, gönderilere ve bunların sosyal olarak nasıl yayıldığı hakkında çıkarım sağlamaktadır (Tandel ve diğeri, 2019, s. 1025) ve bu tür analizler sosyolojik ve psikolojik açıdan önem taşımaktadır. Bu tür sosyal medya analizleri sayesinde bir gönderi hakkında benzerlik ve farklılık gösteren görüşlere sahip belirli yaş grubuna veya topluluklara ait kişilerin davranışları ortaya konmaktadır (Talib ve diğeri, 2016, s. 417). Sosyal medya kullanıcılarına yönelik analiz dışında bir diğeri yaygın metin madenciliği uygulamasını da önemsiz e-postaların analizi oluşturmaktadır. Kategorizasyon kullanılarak e-postaların otomatik analizi sayesinde mailler sıklıkla görülen terimlere göre sınıflandırılmakta olup, kullanıcılara faydalı bir kullanım yaratmaktadır (Birunth ve Sukanya, 2012, s. 270).

III. 3. 1. 3. 7. Güvenlik: Güvenlik bir organizasyonun sürekliliği sağlayan önemli bir konudur. Güvenlik sağlanamazsa bilgilerin bütünlük ve doğruluğunu sağlamak mümkün olmayacaktır. “Zaman çok önemli olduğundan ve sorunun ölçeği göz önüne alındığında, e-postaları veya çevrimiçi sohbeti normal şekilde izlemek mümkün değildir” (Dang ve Ahmad, 2014, s. 25). Bilgi çıkarma, sapma tespiti, kategorizasyon gibi teknik veya tekniklerin kullanımı ile bir önceki başlıkta da bahsedilen istenmeyen postaların, saldırgan veya zorbalık içeren materyallerin, (Dang ve Ahmad, 2014, s. 25) kötü amaçla oluşturulmuş blog ve sitelerin önüne geçilmesi sağlanabilmektedir. Bilgi güvenliği ve siber güvenlik kapsamında sistemler üzerinde ne tür önlemler alınabileceği 4. Bölümde Güvenlik başlığında daha detaylı ele alınmaktadır.

IV. BÖLÜM

EBYS'LERDE METİN MADENCİLİĞİ

Bilgi teknolojilerinin gelişimi ile insanlar, her an ve her yerde veri üretilen bir dünyanın içerisinde kendilerini bulmuşlardır. Bu durum kuruluşlar için de söz konusudur. Kuruluşlar günümüz teknolojisinde gündelik faaliyetlerini ağı dayalı elektronik sistemlerde yürütmekte ve yönetmektedir. Bir başka deyişle bu, çeşitli formatlarda ve büyük miktarda veriye sahip oldukları anlamına gelmektedir; söz konusu veriler yapılandırılmış olarak veri tabanlarında tutulabilmekte ve yönetilebilmektedir. Ancak operasyonel düzeyden gelen verilerin yeniden değer katacak biçimde faaliyetlerde değerlendirilmesi veya organizasyonun geleceğine yön verecek şekilde kullanılması, verilerin etkili ve verimli bir biçimde yönetilmesiyle mümkün olabilmektedir. Doğru veriye, doğru formatta, doğru zamanda ulaşabilmek çalışanlar için ama özellikle karar vericiler için kritik bir öneme sahiptir. Bunun yanı sıra kuruluşların rekabet avantajı sağlaması ve içgörü elde edebilmesi de temelde iyi bir veri yönetimine sahip olmalarına bağlıdır.

Öte yandan günümüzde kuruluşlar, faaliyetlerinin sonucu oluşan belgeleri ya da ürettikleri dokümanları, dijital ortama taşıyabilmekte ya da doğrudan dijital ortamda oluşturulabilmektedir. Belgeler, metin verisi olarak ele alındığında bu, kuruluşların büyük miktarda yapılandırılmamış veriye sahip olduğu anlamına gelmektedir. Üstelik yapılandırılmamış veriler belgelerle de sınırlı değildir; ses, görsel, sosyal medya, eposta vb de yapılandırılmamış verilerdendir. Tom Harbert (2021) kuruluşların verilerinin %80 ila 90'ının yapılandırılmamış verilerden oluştuğunu belirtmektedir. Smith ve arkadaşları (2019) anket çalışmalarında, büyük verinin, bu derece etki yaratmasına rağmen işletmelerin çok azının içgörü elde etmede bu kaynakları kullandığına dikkat

çekmektedirler. Bu ankete göre çok az kuruluş (%18) yapılandırılmamış verilerden faydalanmaktadır. Oysa yapılandırılmamış veriler organizasyonlar için büyük potansiyel oluşturmaktadır. Çünkü büyük miktardaki yapılandırılmamış verilerin içerisinde ortaya çıkarılmayı bekleyen önemli bilgiler bulunmaktadır.

Bilgi teknolojilerinin gelişimi ile insanlar, her an ve her yerde veri üretilen bir dünyanın içerisinde yer almıştır. Veri, enformasyonu; enformasyon ise bilgiyi doğurmaktadır. Bilginin kişiler, organizasyonlar ve ülkeler için güç kaynağı olduğu artık yadsınamaz bir gerçektir. Neredeyse her alanda ve her eylemde bilgi odaklı yaklaşımlar süregelmektedir. Dünya artık bilgi uzmanları (daha geniş anlamda bilgi işçileri) ile bilgiye dayanan iş süreçlerinin ön plana çıktığı bir bilgi çağını yaşamaktadır. Bu sayede artık organizasyonların faaliyetleri, geleneksel kaynaklara (insan kaynakları, sermaye ve sabit varlıklar gibi) bağlı değildir (Mukred ve diğerleri, 2022).

Veri/bilgi/belge yönetiminde, kaynak ne olursa olsun amaç aynıdır ve benzer süreçleri içerir. Verinin, bilginin ve belgenin yönetimini kapsayan yaşam döngüsünde; üretim, tanımlama, depolama ve kullanım gibi ortak süreçler yer almaktadır. Veri/bilgi/belge yönetimi en temelde, etkin karar alma süreçlerini geliştirmektedir. Alınan kararlar, organizasyonların iş süreçlerini yüksek verimde sürdürmelerini sağlamaktadır. Dolayısı ile veri/bilgi/belge yönetimi; verileri tanımlayıp erişmenin çok ötesinde olup, veri/bilgi/belgelerin işlenmesini ve değer yaratılmasını amaçlamaktadır (Oracle, 2021).

Günümüzde kritik olan bilgidir. Bilgi doğrudan ulaşılmamaktadır, akıl ve mantık süzgecinden geçirilmesi gerekmektedir yani çıkarımsal ve soyuttur ve iş kararlarını desteklemek için gereklidir (Chen, 2001, s. 2). Organizasyonların iş ve işlem süreçlerinde faaliyetlerini yürütebilmelerinde bilgi, vazgeçilmez bir konuma sahiptir. Bunun için ise doğru veriye/ bilgiye/belgeye zamanında erişebilmek önem taşımaktadır. Konu, belgeler

açısından ele alındığında iki açıdan incelenebilir; belgelerin metin verisi olarak ele alınması ve belge yönetimi. Bir organizasyonun belge akışı ne kadar iyi düzenlenmiş ise yönetim de o derece etkililik göstermektedir. Nitekim karar alma noktasında, belgeler ne kadar hızlı elde edilir, işlenir ve uygulanırsa o ölçüde etki sağlanabilmektedir (Барабан и Колумбет, 2022, s. 296) “Yeterli bilgiye dayanmayan karar alma, organizasyonu riske atabilmektedir” (Mukred ve diğerleri, 2022). Bu kapsamda organizasyonların rekabet avantajı yaratacakları yararlı bilgileri elde edebilmeleri için ellerindeki verileri iyi yönetebilmeleri gerekmektedir. Organizasyonların konumu, verilerini ne derece etkin yönetebildikleri ile belirlenebilmektedir. Rekabet ve bilgi üstünlüğü sağlanabilmesi için verilerin en etkili şekilde işlenmesi ve kurum faaliyetlerine yönelik olarak bir değer yaratabilmesi gerekmektedir. İyi yönetilmiş veriler, organizasyonları etkili karar almaya, hizmetler sunmaya ve ürünler yaratmaya sevk etmektedir. Bu noktada erişim faaliyetleri ne kadar etkili olursa, bilgi varlıkları ne kadar iyi tanımlayabilirse, erişim sistemleri ne kadar iyi geliştirebilirse; o kadar iyi, doğru ve etkili karar alabilir, ilişki çıkarabilir ve içgörüler elde edilebilir. Unutulmamalıdır ki bilgi, bir organizasyonun etkili karar alma ve uygulama gücünü geliştirmek için yönetilen gerekçelendirilmiş inançlar kümesi olarak ele alınmaktadır (Alavi ve Leidner, 2001; Nonaka, 1994; aktaran Khan ve Vorley, 2017, s. 6). Bu güç organizasyonlarda özellikle üretkenlik, etkinlik ve rekabet gücü ile olası faydalar sağlamaktadır (Mukred, 2017; Burkan Hawash et al., 2020; aktaran Mukred ve diğerleri, 2022). Bu noktada bilgi erişim faaliyetlerinin çıkarılabilecek en üst noktada olması gerekmektedir.

Chen çalışmasında (2001, s. 5) metin madenciliği ve veri madenciliğinin bilgi yönetimi ile yakın ilişkisine dikkat çekerek, söz konusu iki tekniğin bilgi yönetimi ile ilişkili iki alt alan olduğunu belirtmektedir. Fakat yine aynı çalışmada organizasyonların metin madenciliği ve veri madenciliğinden ziyade temel bilgi teknolojilerini

benimsediklerini belirtmektedir. Bunun nedenlerinden biri organizasyonların bilgi yönetiminin sağladığı faydaların bilincinde olmaması olabilir.

Benzer şekilde Khan ve Vorley de (2017, s.3) büyük verinin günümüz bilgi çağının hammaddesi olmasına ilgi çekerek bilgi yönetimi alanında yeteri kadar ilgi görmemiş olmamasını şaşırtıcı bulmaktadır. 1990'lardan günümüze bilgi yönetimi, bilgi erişim sisteminden veri madenciliğine, veri madenciliğinden metin madenciliğine kadar birçok uygulamada karşılaşılan popüler bir terim haline gelmiştir (Chen, 2001, s. 2). Amazon gibi bazı büyük şirketler etkin bilgi yönetimini sağlamak, operasyonlarını geliştirmek ve müşteriler ile iletişim kurmak için metin madenciliğini kullanmaktadır (Davenport & Patil, 2012 aktaran Khan ve Vorley, 2017, s. 5). Khan ve Vorley (2017) metin madenciliği kullanımının popüler hale gelmesi sonucunda, metin madenciliğine yönelik akademik eğilimin de arttığını fakat buna rağmen literatürde bilgi yönetiminin sağlayıcısı olarak, metin madenciliğinin rolünü inceleyen çok az çalışma olduğuna değinmektedir.

Çalışmamız metin madenciliğini EBYS süreçlerine entegre ederek; bilgi erişim, bilgi çıkarma, bilgi keşfi, ilişki ve örüntülerin tespiti gibi konularda verimi üst noktaya ulaştırmayı; EBYS uygulamalarındaki sorunları aşmayı böylelikle etkin bir bilgi yönetim süreci gerçekleştirmeyi amaçlayan bir mimari ortaya koymaktadır.

- Bilgi varlıklarının otomatik bir şekilde tanımlandığı,
- Konu metadatalarının otomatik olarak çıkarıldığı,
- ontoloji tabanlı erişim sisteminin geliştirildiği,
- bilgi varlıklarına dair ilişkilerin tanımlandığı EBYSler, organizasyonlara etkili bir yönetim sağlayarak iş zekasını yaratacak dolayısı ile de organizasyonlara rekabet üstünlüğü kazandıracaktır.

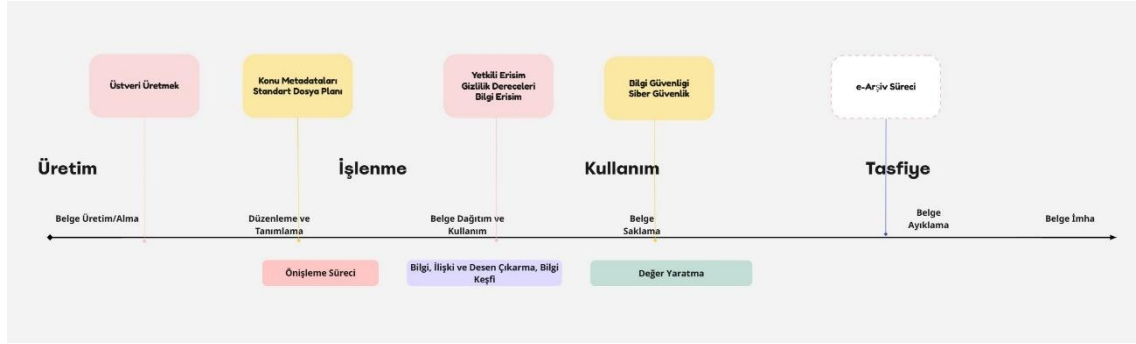
Bu yapının sağlanması için ise, metin madenciliği gibi analitik araçların kullanımı gerekmektedir. Metin madenciliği, yapılandırılmamış metinsel verilerin yorumlanması ve bilgi yönetimini geliştirecek içgörü sağlanması zorluğunu aşarak, organizasyonların erişmesinin ve paylaşmasının oldukça zor olduğu değerli bilgilerin üretilmesinde önemli rol oynamaktadır (Khan ve Vorley, 2017, s. 5).

Çalışmamız söz konusu tekniklerin (metin madenciliği teknik ve algoritmalarının) bu yolda etkili olacağına inanmakta olup, literatürden örneklerle dayanarak, teknikleri bilgi yönetimine uyarlamaktadır. Çalışmamızın 4. Bölümü olan ve çalışmanın varmak istediği asıl fikri ortaya koyan “EBYS’lerde Metin Madenciliği”; metin madenciliği yöntemlerinin EBYS’ler üzerindeki kullanım alanlarını ortaya koymayı amaçlamaktadır. Çalışmamızda EBYS terimi; *tüm organizasyonların faaliyetlerine yönelik olarak kişiselleştirilebilen, süreçlerin otomatikleştirilebildiği ve dahili ve harici veriler kullanılarak organizasyonda değer yaratabilen belge tabanlı bir yönetim sistemi* olarak tasarlanmıştır. Bu noktada çalışma, tek bir EBYS uygulamasını ele almamakta olup, bütünsel ve organizasyon gereklerine yönelik olarak kişiselleştirilebilir bir çerçeve sunmaktadır.

Çalışmanın bu bölümünde metin madenciliği teknik ve yöntemlerinin EBYS’ler üzerindeki kullanım alanlarını ortaya koymak amaçlamaktadır. Bu doğrultuda aşağıda görseli sunulan “belgenin yaşam döngüsü” temel alınmıştır.

Şekil 11.

Belgenin yaşam döngüsünde metin madenciliği kullanım alanları



Belge yaşam döngüsü görüleceği üzere süreç odaklıdır. Amacımız sürecin Belge Tanımlama, Belge Dağıtım ve Kullanım ve Belge Saklama aşamalarında ¹⁰metin madenciliği kullanılabilirliğini sorgulamak ve aşağıdaki sorulara yanıt aramaktır.

- Hangi aşamalarda hangi metin madenciliği tekniği kullanılabilir,
- Ne amaç ile kullanılabilir ve
- Bunların organizasyonlara ve personele sağlayacağı katkı/katkılar neler olabilir?

Belgenin Üretilmesi ve Tanımlanması: Belgenin yaşam döngüsü; üretim, işleme, kullanım ve tasfiye olmak üzere dört ana aşamayı bünyesinde bulundurmaktadır (Aydın, 2023, s. 26). Üretim aşamasında organizasyon, belgeyi kendisi üretebileceği gibi başka bir organizasyon tarafından da edinebilmektedir. Esasında metadata, belgenin yaşam döngüsündeki her aşamada üretilmekte olup, belge üretim ve alma aşamasında ise belgeyi yaratan, belgenin oluşturulduğu yer ve zaman, belgenin başlığı gibi öğelerde kendisini

¹⁰ Belge Ayıklama ve Belge İmha aşamaları, arşiv yönetimine dair süreçlerdir. Gerçekleştirilen çalışma metin madenciliği tekniklerini EBYS süreçlerine uygulamaktadır. Bu nedenle Belge Ayıklama ve Belge İmha süreçleri konu kapsamının dışında tutulmaktadır.

göstermektedir. Bu öğeler daha sonra ilgili belgelere erişilebilmesi için oldukça önemlidir. Fakat belgelere erişilmeden önce belgelerin içeriğinin tanımlanması/nitelenmesi gerekmektedir. Aşağıda, TS 13298 Elektronik Belge ve Arşiv Yönetim Sistemi Standardında, belge tanımlarında yer alması gereken, seçmeli ve zorunlu metadata örnekleri yer almaktadır. Üretim tarihi, gönderim tarihi, üretici gibi metadata alanları genellikle EBYS'lerde yapısal olarak yer almakta ve otomatik olarak işlenebilmektedir. Ancak konu, dosya numarası gibi alanlar belgenin içeriği ile ilgili olup, yapılandırılmamış alanları oluşturmaktadır.



Tablo 3.*Belge tanımları metadata elemanları*

Eleman Adı	Belge adı
Referans Numarası	BE010
Tanım	Belgeye ait ad ve/veya tür bilgisidir.
Zorunluluk	Zorunlu
Notlar	
Eleman Adı	Elektronik dosya adı
Referans Numarası	BE019
Tanım	Elektronik belgenin bilgisayar sisteminde hangi ad ile tutulduğu bilgisidir.
Zorunluluk	Zorunlu
Notlar	
Eleman Adı	Klasör/dosya numarası.
Referans Numarası	KD001
Tanım	Belgenin bağlı bulunduğu klasör veya dosyaya ait referans numarasıdır.
Zorunluluk	Zorunlu
Notlar	
Eleman Adı	Konu
Referans Numarası	BE055
Tanım	Belge içeriğini açıklayıcı bilgisidir.
Zorunluluk	Seçmeli
Notlar	
Eleman Adı	Üretici
Referans Numarası	BE100
Tanım	Belge üreticisini gösteren bilgisidir.
Zorunluluk	Zorunlu
Notlar	
Eleman Adı	Üretim tarihi
Referans Numarası	BE210
Tanım	Belgenin üretildiği tarihtir.
Zorunluluk	Zorunlu
Notlar	
Eleman Adı	Gönderme / alma tarihi
Referans Numarası	BE240
Tanım	Belgenin gönderildiği ya da alındığı tarihtir.
Zorunluluk	Zorunlu
Notlar	
Eleman Adı	Alıcı/Muhatap
Referans Numarası	BE200
Tanım	Belgenin kime gönderildiği bilgisidir.
Zorunluluk	Zorunlu
Notlar	
Eleman Adı	Dil
Referans Numarası	BE350
Tanım	Belgenin hangi dilde üretildiği bilgisidir.
Zorunluluk	Zorunlu
Notlar	
Eleman Adı	Belge türü
Referans Numarası	BE420
Tanım	Belgeye ait tür tanımlamasıdır.
Zorunluluk	Zorunlu
Notlar	Kurumdaki belge türleri (yazışma, rapor, vs.) kullanıcıya liste halinde sunularak çoktan seçmeli bilgi girişi sağlanmalıdır.

Belgelerin içeriğinin tanımlanması işlemi günümüzde manuel olarak gerçekleştirilmektedir. Özne bakış açıları, kayıtların tanımlanmasında; tek biçimlilik ve standart bir temsil sağlamada sorun teşkil edebilmektedir. Nitekim niteleme işlemini yapanlar insandır ve her birinin farklı birikimleri, dolayısı ile de farklı algılayış biçimleri

vardır. Bu nedenle kontrollü bir dizin kullanılsa bile, tanımlama işlemlerine özne bakış açılarının dahil olması ile farklı tanımlamalar yapılması olağandır. Bu duruma örnek olarak, Standart Dosya Planı verilebilir. Standart Dosya Planında, konu başlıkları ve alt başlıklar, sistematik bir şekilde verilmesine karşın; belgenin tanımlanmasında yanlış konu başlıkları seçilebilmektedir. Bu durum ile ilgili örnekler ileriki başlıklarda ele alınacak olup, konu metadalarının girişinde ve sınıflandırma işlemlerinde “neden otomatikleşmeye ihtiyacı var” sorusunun cevabı daha da aydınlanmış olacaktır.

Belgenin Dağıtım ve Kullanımı: Belgenin tanımlanmasından sonraki aşamayı belge dağıtım ve kullanım oluşturmaktadır. Belge dağıtım bir belgenin organizasyon içerisindeki bir birimden başka bir birime veya bir organizasyondan başka bir organizasyona (gerektiği durumlarda belgenin gizlilik, bütünlük ve erişilebilirliğini sağlamak amacıyla gizlilik derecelerinin de kullanılarak) iletilmesidir. Kullanım ise ihtiyaç doğrultusunda ilgili belgeye yetkiler doğrultusunda erişilmesini gerektiren bir süreçtir. EBYS bileşenlerinde de önemle üzerinde durulan yetkili erişim kavramı, dağıtım ve özellikle de kullanım noktasında dikkat edilmesi ve önlemler alınması gereken bir konudur. Belgenin gizlilik ve bütünlüğünün korunması ve kurumsal mahremiyetin sağlanması amacıyla gerekli olan; yetkili erişim ve gizlilik dereceleri, kurumsal süreklilik ve standartlaşmanın sağlanması bakımından otomatikleşmeye ihtiyaç duymaktadır. Fakat belgelere erişmenin tek yönünü yetkili erişim ve gizlilik derecelerinin belirlenmesi (yani erişim hakkına dair süreçler) oluşturmamaktadır bu bilgi erişimin “access” kısmını oluşturmaktadır. Bir de “retrieval” kısmı vardır. Belgelere ve belgelerin içerisindeki bilgilere/bilgi parçacıklarına erişilmesini sağlayan bilgi erişim (information retrieval) konusu, oldukça geniş bir konu olup sistemsal açıdan birçok iyileştirme gerektirmektedir. Bilgi erişim süreçlerinin performansını en üst düzeye getirmeyi de amaçlayan metin madenciliği yaklaşımı, organizasyonların karar alma süreçlerinde kurumsal verimi

önemli derecede arttırması beklenmektedir. Bunu nasıl gerçekleştirebileceği ileriki başlıklarda ele alınmaktadır.

Belgenin Saklanması (Depolanması): Burada son olarak bahsedilecek aşama ise, belge saklamadır. Belge saklama evresinde önem arz eden konu, belgelerin bütünlüğünün ve güvenliğinin tehlikeye atılmadan muhafazasını sağlamak olacaktır. Bu noktada ise bilgi güvenliği ve siber güvenlik kavramları ortaya çıkmaktadır. Metin madenciliği, hassas verilerin tespiti ile bilgi güvenliğinde; sistemdeki anormallikleri tespit ederek de siber güvenlikte önemli faydalar ortaya koyacaktır.

EBYS süreçlerini iyileştirmeye yönelik verilen metin madenciliği adımlarından sonra çalışmamızda yeni bir yapı ortaya koymaya çalışıldı. Ortaya konulan yapı, kurumun fırsat, risk ve tehditlerini saptayabilen; ilişki, desen ve bilgileri keşfedebilen; örtük bilgilerin anlamlandırılarak bir değer/ürün yaratabilen ve organizasyonların karar alma süreçlerinin her aşamasında kullanılabilecek, metin madenciliği tabanlı otomatikleştirilmiş yeni bir belge yönetim sistemidir.

Bilgi çok önemli stratejik bir güçtür ve iyi yönetilmesi gerekmektedir. Küresel rekabetin şiddetini her geçen gün artırdığı günümüzde rekabet üstünlüğünü elde etmek için karar üstünlüğüne, karar üstünlüğünü elde etmek için bilgi üstünlüğüne sahip olmak gerekmekte, bilgi üstünlüğüne sahip olmak için ise iyi bir bilgi yönetimi gerçekleştirmeye ihtiyaç duyulmaktadır. (Önaçan ve diğerleri, 2012, s.2)

IV. 1. EBYS’lerde Metin Madenciliği Kullanım Alanları: Sorunlar ve Çözümler

Bu bölüm EBYS’lerde bilgiye ulaşma süreçlerini iyileştirmek adına, EBYS’lerde metin madenciliği kullanım alanlarını 3 ana başlık altında detaylandırmaktadır. Bunlar;

- Belgelerin tanımlanmasını, konu metadatalarının oluşturulmasını, sınıflandırılmasını, kategorize edilmesini ve ontoloji tabanlı arama motorlarının geliştirilmesini içeren “**Bilgi Tanımlama**”,
- Kurumlarda etkili karar alınmasını sağlayacak olan, bilgi çıkarımı, bilgi keşfi, örtük bilgilerin ve gizli ilişki ve desenlerin ortaya çıkarılmasını sağlayan bilgi haritalarını içeren “**İş Zekası**”,
- Güvenlik açıkları, güvenlik kontrolleri ve veri ihlallerinin tespit edilmesini içeren “Siber Güvenlik” ile hassas bilgilerin tespit edilmesi ile yetkili erişimi sağlayan “Bilgi Güvenliği” konularını içeren “**Güvenlik**” başlıklarıdır.

Çalışmamız bu başlıkları sorun-çözüm sistematğinde ele almaktadır.

IV. 1. 1. Bilgi Tanımlama: Metin Madenciliği ve Ontoloji Tabanlı Yaklaşımlar

IV. 1. 1. 1.Sorun: Organizasyonlar belge yönetim süreçlerinde günlük faaliyetlerini yerine getirirken birçok bilgi ve belge üretmekte ve birçok bilgi ve belgeye de erişimleri gerekmektedir. Ancak incelemeler göstermektedir ki organizasyonlar bir belgeye erişirken çok ciddi vakit kaybı yaşamakta ve bu sorun yıllardır aşılamamaktadır. Borglund ve arkadaşları (2009a, s. 53) küçük ve orta ölçekli kuruluşların büyük bir bölümünün günlerinin önemli kısmını (zamanlarının %25’inden fazlasını) belge ve diğer bilgi kaynaklarını aramak için harcadığını belirtmektedir. Benzer şekilde McKinsey Kuruluşu ise çalışmasında personellerin bilgi erişim sürecinin günlük 2 saatini aldığını ortaya koymaktadır (Chui vd., 2012; aktaran Koyuncu Tunç ve Külcü, 2020, s. 272). Önaçan ve arkadaşları, (2012, s. 12) Gartner tarafından yapılan bir araştırmaya göre belge yönetimini kapsayan günlük rutin eylemlerin personelin zamanının %40 ila %60’ını oluşturduğunu belirtmektedir. Svärd (2017, s. 71) Belçika’da gerçekleştirdiği bir vaka çalışmasında, etkili ve gelişmiş bir belge sisteminin oluşturulmamış olması nedeniyle, çalışanların

çoğunun, önemli zamanlarının büyük kısmını, belge arayarak geçirdiğini belirtmektedir. Personellerin neredeyse yarısı bilgi ve belgelere erişmekte zorlanmakta ve belge yönetim süreçlerinde sorun yaşamaktadırlar (Nintex, 2018; aktaran Koyuncu Tunç ve Külçü, 2020, s. 272).

Bu durum kurumsal verimliliğin önünde önemli bir engel yaratmaktadır. Belgelere zamanında ve kolay erişilememesi, iş akışının (sürekliliğinin) aksamasına, dolayısı ile de zaman, maliyet ve emek noktalarında işlevselliği baltalanmasına neden olabilmektedir. Erişim sorunlarının altında yatan sebeplerden en önemlileri, bilgi ve belgelerin etkin bir şekilde tanımlanamayabileceği ve bilgi/belge erişim sistemlerinin basit tasarımının günümüz teknolojisine kıyasla geleneksel kalması olarak yorumlanabilmektedir.

Bu başlıkta literatürden örnekler ile bilgi erişim noktasında organizasyonların yaşamış oldukları sorunları ve/veya potansiyel sorunların neler olduğu ortaya konulmaktadır. Bilgi erişimin en basit mantığını; iyi tanımlanmış bilgi kaynakları/varlıkları ve bu bilgi kaynaklarına/varlıklarına en hızlı ve kolay eriştiren erişim sistemleri oluşturmaktadır. Bu doğrultuda ilk olarak tanımlama noktasında metadata ve dosya planı ile ilgili sorunlar ele alınacak, ardından erişim sistemleri ile ilgili sorunlar üzerinde durulacaktır.

Yalçınkaya (2016, s. 70) çalışmasında EBYS'lere ilişkin metadata ilkelerini açıklarken; bilginin hangi iş ve işlem süreçleriyle ilişkili olduğunu ve ne maksat ile gerçekleştirildiğini anlamada metadata kullanımının önemine değinmekte olup, metadata kullanımının belgelerin mahiyetinin ve kalitesinin belirlenmesinde önemli rolü olduğunu belirtmektedir. Metadataların kuruluşlardaki önemine karşın Aydın (2023, s. 34)

alışmasında, organizasyonların elle girdikleri metadataların yetersiz olduğunu bunun da belgelere erişmede sorun yarattığını belirtmektedir.

Önceden organizasyonlarda veri giriş işlemlerini belge yöneticileri gerçekleştirmektedir (Borglund ve diğerleri, 2009b, s. 63). Ancak günümüzde dijitalleşmenin de etkisi ile neredeyse tüm personel veri işleyen konumundadır ve metadataların yanlış ve eksik girilmesi belge erişiminde önemli sorunlar doğurmaktadır. Dolayısı ile belge yönetim süreçlerinde, belgelerin içeriğinin tanımlanması noktasında standartlaşma, daha güç bir hale gelmektedir. Standartlaşma, kurumlarda süreklilik sağlamaktadır. Veri işleyenlerin birçoğu farklı disiplinlerden yetişmiş kişilerdir. Dolayısı ile belge yönetimine ait perspektif ve birikime sahip olmadıkları öngörülmektedir. Umut ve Külçü (2014, 119) EBYS sorunları hakkında Kalkınma Bakanlığı üzerinden ele aldıkları çalışmada, metadata ile ilgili karşılaşılan sorunları şu şekilde sıralamaktadırlar; metadata alanlarına eksik girişlerin yapılması, metadata ile ilgili standartlaşma sorunu ve yasal düzenlemelerin eksikliği.

Tüm bu sorunlar bilgi erişimin sağlanmasında kullanıcılara büyük bir engel teşkil etmektedir. Nitekim aynı çalışmada kuruma, söz konusu metadata koşullarının değerlendirildiği bir anket uygulanmıştır ve sonuçta ise personellerin %70'inden fazlasının negatif veya kararsız görüş bildirdiği görülmüştür. Benzer şekilde Alır (2008, s. 70) çalışmasında Adalet Bakanlığını ele almış ve elektronik belgelerde metadata yapısında önemli eksiklikler olduğunu gözlemlemiştir. Metadatalar bilginin etkin tanımlanması ve dolayısı ile de erişiminde önemli faydalar sağlamaktadır. Bu nedenle metadata tanımlamalarının yapılmasının da etkin bir bilgi erişim süreci yaşatması beklenmektedir. Ancak Adalet Bakanlığı üzerinde yapılmış çalışmada görülmektedir ki, personel tarafından, konu metadatasında yapılan veri girişlerinde tutarsızlıklar söz konusudur ve belgeyi içerik bağlamında niteleyen özelliklerden yoksundur (Alır, 2008, s.

70). Bu durum bilgiye erişimin önünde büyük bir problem yaratıp, standartlaşma ile ilgili sorunu gün yüzüne çıkartmaktadır. Standartlaşmada yaşanan sorun ise veri/bilgi/belgeye erişimi zorlaştırabilmekte ve sonuç olarak iş akışlarının aksamasına neden olabilmektedir. İş akışlarının aksaması ise organizasyonun varlığını (sürekliliğini) tehlikeye sokabilmektedir. Standartlaşma ve kurumsal süreklilik sorunu, personele bağlı bir belge yönetim sürecinin geride bırakılması ile aşılabilmektedir. Bunun için ise otomatikleşme gerekmektedir. Yalçinkaya (2016, s. 70) çalışmasında, metadata girişlerinde personelin hata yapma potansiyelinden bahsederek, bu sürecin gerçekleşmesinde insan faktörünün verimsiz ve riskli olabileceğine değinmektedir. Aynı durumun standart dosya planlarında da yaşandığı çalışmalarda öne sürülmektedir.

Standart dosya planı, bir organizasyonun günlük faaliyetleri süresince oluşturdukları belgeleri, konu bağlamında (içerik ilişkileri göz önünde bulundurularak) bir arada bulundurulmasını sağlayan, belgelerin erişilebilirlik, bütünlük ve güvenliğini koruyan sınıflamaya bağlı bir düzenleme sistemidir (T.C. Yükseköğretim Kurulu, 2023). Belge sınıflandırma süreci, belgenin içeriğini en iyi tanımlayacak şekilde, sınıflandırma sisteminden uygun konunun seçilmesi ve belgeye atanmasını içermektedir. Bu sayede birbirleri ile ilişkili tüm belgeler bir arada bulunmakta ve erişilmektedir. Aksi taktirde belgelerin kullanımı ve erişilebilirliği tehlike altına girmektedir (Kiwa ve diğerleri, 2022, s. 145). Eğer belgelere zamanında ve doğru bir şekilde erişim sağlanamazsa, bu durumdan, kurumların karar alma mekanizmaları, olumsuz etkilenecektir. Bu nedenle standart dosya planının veya organizasyonun ihtiyaçlarına yönelik özel olarak oluşturulmuş bir dosyalama sisteminin, etkili bir şekilde kullanılması; kullanılabilirlik, erişilebilirlik ve süreklilik açılarından belirleyicidir. Fakat daha evvel de bahsedildiği üzere, insan faktörü, belgelerin nitelenmesinde hata oluşumuna neden olmaktadır. Hataya açıklık ise, isabetsiz kararlar alınabilmesine neden olabilmektedir.

Solhan (2016) çalışmasında, Devlet Arşivleri Genel Müdürlüğü tarafından gerçekleştirilen EBYS denetiminde; standart dosya planı uygulamasında, personel tarafından yapılan yanlış uygulamaların ortaya çıktığını belirtmektedir. Söz konusu çalışmada, standart dosya planının personel tarafından mahiyetinin anlaşılmadığı ve mevzuat dışı uygulamaların düzeltilmediği belirtilmektedir.

Solhan (2016), çalışmasında bu uygulamalardan aşağıdaki gibi bahsetmektedir:

Şekil 12.

SDP mevzuat dışı uygulama örnekleri



Bu örneklerle benzer ve ek olarak, Menteş ve Solhan, (2019, s. 17) Cumhurbaşkanlığı Bilgi ve Belge Yönetimi Daire Başkanlığının uygulamaları ile kuruluşlardaki EBYS uygulamaları süresince, standart dosya planı ile ilgili yapılmış yanlış uygulamalar ile ilgili şu şekilde bahsetmektedir:

"Belgelerin dosyalanması, saklanması ve tasfiye işlemini sağlayan Standart Dosya Planı (SDP)'nın fiziksel ortama hitap eden yapısının, elektronik ortamda modern belge yönetim süreçlerini sağlayabilmesi için dosyalama yöntemi ve yönetiminin değişmesi gerektiği anlaşılmaktadır. Mevcut SDP yapısının fiziksel ortamda oluşturulan

belgelerin gruplandırılması ve belgenin üretildiği anda ilgili dosyasıyla eşleştirilmesi ile arşivsel süreçte tekrar bir tasnif işlemine ihtiyaç duyulmaması beklenirken; elektronik ortamda detaylı olarak oluşturulan SDP yapısı kullanıcıya tüm kodlar ile aynı anda sunulduğu zaman belge, hiç ilgisi olmayan bir konu kodu ile eşleştirilebilmektedir. Örneğin, Cumhurbaşkanlığı Hükümet Sistemine geçişten sonraki dönemde yürütülen 1. ve 2. “100 Günlük Eylem Planları”nda gerçekleştirilen yazışmalar incelendiğinde, idarelerin birçoğunun 602.05 kodunu kullanarak yazışma yapması gerekirken hiç ilgisi olmayan 951.01.02 kodunu seçtikleri görülmüştür. Bu durumun oluşmasında EBYS’lerin metin üzerinden konu kodu aratmaları da etkili olmakta fakat mevcut SDP yapısının kırılımlı olması, kullanıcıların üst - alt kod bağlantılarını gözden kaçırmalarına sebep olmaktadır” (Menteş ve Solhan, 2019, s. 17).

Şekil 13.

SDP örneği

950 (Güvenlik İşleri, Afet ve Acil Durum Yönetimi ve Topyekün Savunma Sivil Hizmetleri)				600 (Araştırma Planlama İşleri)		
951			Güvenlik İşleri	602		Plan ve Program İşleri
	01		Dış Güvenlik İşleri		05	Eylem Planları
		02	Eylem Planları			

951.01.02 (Dış Güvenlik Eylem Planları)

602.05 (Eylem Planları)

(Menteş ve Solhan, 2019, s. 17)

Örneklerde de görüldüğü üzere, belge erişimin ana zorunluluğu olan, belgelerin içeriğinin tanımlanması işleminde, kuruluşlarda önemli problemler öne çıkmaktadır. Bu problemlerin başlıca kaynağını “insan faktörü” oluşturmaktadır. Her bir kişinin bilgiyi algılayış ve tanımlayış şekilleri farklıdır. Çünkü kimsenin tecrübe ve bilgi birikimi eşit

değildir. Bu nedenle kurumlarda standartlaşmanın sağlanması oldukça önemlidir. Fakat ne kadar standartlaşma sağlanırsa sağlansın, her personelin veri işler pozisyonunda olduğu bir koşulda, belge yönetim süreçlerinin manuel tanımlamalara dayanması süreçleri zora sokacaktır.

Eğitim eksikliği, farkındalık sağlanamaması, bir konu uzmanı ile bilgi ve belge yöneticisinin süreçlerde yer almaması tanımlama aşamalarında, yanlış uygulamaların gerçekleşmesine sebebiyet vermektedir. Bir EBYS üzerinde konu metadatalarına girilen veriler, ne kadar çeşitli şekilde tanımlanırsa bilgi erişim oranları da o kadar verimli olması beklenmektedir. Ancak görülmektedir ki bilgilerin tanımlanmasının, erişim ve diğer süreçler için ne kadar önemli olduğu personel tarafından tam olarak anlaşılamamıştır. Yalçinkaya (2016, s.80) bu durumu şu şekilde özetlemektedir; Belgenin içeriğinde yer alan metinde kelime taraması yapmak “...bazen kullanıcı için tam olarak istediği belge sonuçlarının gelmesine yardımcı olmaz. İçerik aramasından çok metadata üzerinden arama yapmanın önemini kavrayan çalışanlar belgenin üretimi ile ilgili süreçte metadatanın neden çok önemsendiği konusunu da algırlar.”

Bu bağlamda, ne yazık ki kuruluşlarda belge yönetimi süreçlerinin; bilgi yöneticisi perspektifinden uzak olması, eğitim ve farkındalık eksiklikleri, süreçlerin manuel bir şekilde yürütülmesi önemli bilgi erişim problemlerini beraberinde getirmektedir.

Ancak bilgi tanımlama sorunları çözülmüş olsa bile bilgiye erişmek yine kolay olmayacaktır özellikle de sürekli veri/bilgi/belge üreten kuruluşlar için. Veri/bilgi/belge erişimin ana gerekliliği tanımlama olsa da süreç bir tek tanımlamayı içermemektedir. İyi bir bilgi erişim süreci için, mimarisi iyi bir şekilde oluşturulmuş, kullanıcı odaklı bir bilgi erişim sistemi tasarlanması gerekmektedir.

Chen, (2001, s. 27) alışmasında desteklediđi üzere, arama yaparken kullanıcının seçmiş olduđu anahtar kelimeler, belgenin tanımlandığı metadata, dizin terimleri ve konu başlıklarından farklı seçilmektedir, bu nedenle kullanıcılar isabetsiz tarama sonuçları ile karşılaşabilmektedir.

Örnek vermek gerekirse “tarım” konulu bir belgeye erişmek isteyen bir kullanıcı, belge “ziraat” olarak tanımlanırsa bu belgeyi kaırmış olmaktadır. Sistem üzerinde belge hem tarım hem de ziraat olarak tanımlanmışsa bile, bazen kullanıcı aradığı bilgiyi tanımlayamamaktadır. Örneğin kullanıcı aslında “sulama sistemleri” ile ilgili bir belgeye erişmesi gerekiyor ama bunu “tarım” terimi üzerinden aratıyor. Sistemde bu ilişkileri tanımlayan bir thesaurus, kavram haritası veya bir ontoloji yok ise kullanıcının ilgili birçok belgeyi ıskalaması çok olasıdır.

Nitekim, Koyuncu Tun ve Külü (2020, s. 272) Hacettepe Üniversitesi EBYS uygulamasını analiz ettikleri alışmalarında, EBYS uygulamasının, arama sonuçlarında ilişkili belgeleri sunma ve girilen anahtar kelimelerin belgeye ait tüm bilgi alanlarında arama yapma özelliklerinden mahrum kaldığını ortaya koymaktadır. Erođlu ve Külü ise (2014, s. 349) İçişleri Bakanlığının kullanmış olduđu EBYS sistemini (e-içişleri) analiz ettikleri alışmalarında, arama sırasında kullanıcıların boolean operatörlerini kullanamadıklarını ve eşanlamlı sözcükler ile göndermelere ilişkin tanımlamaların eksik olduğunu saptamaktadır. Kutlutürk ve Özdemirci, (2023, s. 17) EBYSlerde, belgeyi tanımlayan unsurlar dışında (konu başlığı, standart dosya planı, gelen-giden birim vb.) metin içerisindeki unsurlardan da belgelere erişimin sağlanması gerektiğini belirtmektedir.

Örneklerde yer aldığı üzere, günümüzde EBYSler üzerinden verimli bir bilgi erişim süreci yaşanmıyor olmasının nedeni, anlamsal ilişkilerin kurulamaması ve anlam

tabanlı bir erişim sistemi sağlanmamasından kaynaklanmaktadır. Arama motorlarının büyük bir çoğunluğu kelime tabanlı sözdizimsel aralamalar yapmaktadır (Lee, 2020, s. 209). Bu arama türlerinde sistem, kullanıcının girdiği anahtar kelimenin bağlamını/anlamını anlamadan yalnızca tam metinde eşleşen belgeleri sonuçlarda verebilmektedir (Sayed ve Muqrishi, 2017, s. 186). Bir diğer deyiş ile “...dizinlenmiş belgeler genellikle o belgeyle ilişkili ham metinden başka bir şey içermemektedir” (Kara ve diğerleri, 2011, s. 301). Bu daha somut bir ifade ile şu anlama gelmektedir; kullanıcı sistemde bir belgeye erişmek istiyorsa, girdisi (anahtar kelimesi) sistem veritabanındaki veriler ile birebir eşleşiyor olmalıdır.

Mevcut sistemlerde standart bir metadata listesi olmadığı için erişim esnasında kullanıcılar, metadata öğelerine girilen verileri tahmin etmek zorunda bırakılmaktadır. Kullanıcı sistemdeki veriyi tahmin edemezse belgeyi ıskalamış olacaktır. Sözdizimsel arama kelimenin bağlamını değil biçimini esas almaktadır. Aşağıda e-devlet platformu üzerinden bu duruma örnek teşkil edebilecek tarama örnekleri verilmektedir:

Şekil 14.

Sözdizimsel arama motoru: e-Devlet örneği

The image displays three sequential screenshots of the e-Devlet search interface, each enclosed in a colored border (blue, purple, and green respectively). Each screenshot shows a search form with two input fields: 'Aranan Kelime' (Search Term) and 'Sıralama' (Sorting). The 'Aranan Kelime' field contains the text 'e- hizmet' in the first two screenshots and 'elektronik hizmet' in the third. The 'Sıralama' dropdown menu is set to 'Sık Kullanıma Göre' (By Frequency of Use). A blue 'Ara' (Search) button is located to the right of the search fields. Below the search fields, the results are displayed. The first screenshot shows 'Toplam 1000 kayıt bulunmuştur.' (Total 1000 records found). The second screenshot shows 'Toplam 1 kayıt bulunmuştur.' (Total 1 record found). The third screenshot shows 'Toplam 55 kayıt bulunmuştur.' (Total 55 records found).

“e-hizmet, e- hizmet, elektronik hizmet” terimleri bakıldığında anlamsal farklılıkları bulunmamakta, sadece küçük biçimsel farklılıklara sahiptirler. Fakat bu küçük farklılıklara rağmen e-devlet portalında sistem, kullanıcılara her bir terim için farklı sonuçlar vermektedir.

Lee (2020, s. 209) çalışmasında bu görüşleri destekler biçimde ele almaktadır ki; sözdizimsel kelime tabanlı geleneksel arama motorları, kullanıcılar için endişe verici sonuçlar doğurmaktadır. Bunlardan ikisi şu şekilde ele alınmaktadır:

1. Kullanıcı erişmek istediği belgeye dair anahtar kelimeyi bilemeyebilir ve sorgularını basit doğal dil cümlesi ile oluşturur.

2. Kullanıcı boolean mantığını anlamakta güçlük geçmektedir (Nitekim daha önce bahsedildiği üzere Eroğlu ve Külçü (2014) işleri bakanlığı örneğinde ortaya koymaktadır ki, e-işleri sistemi geleneksel kalan ve kullanıcılar için zorlayıcı olan boole operatörlerinin dahi kullanımına izin vermemektedir)

Şu ana kadar bilgi erişim noktasında edinilen deneyimler göstermektedir ki, kullanıcı eğer sistemde tanımlanan anahtar kelimeyi bulamıyorsa, ilgili belgeye erişmesi de meşakkatli oluyor. Aynı konuya mensup belgeler birbirleri ile ilişkilendirilemiyorsa, kullanıcının sistem üzerinden yaptığı taramalarda, ilgili belgeye/lere erişebilmesi oldukça zor ve vakit alıcı bir süreç yaratmaktadır (Kiwa ve diğerleri, 2022, s. 145).

Geleneksel, kelime tabanlı sistemlerin yaratmış olduğu bağlamsal anlam eksikliğini, ontoloji tabanlı semantik arama motorları çözebilmektedir. Ontoloji gerekli kavramsallaştırma ve bilginin temsili sorunlarını çözerek, (Sayed ve Muqrishi, 2017, s. 182) sorguları anlamsal açıdan zenginleştirmekte, bilgi erişimi verimli bir seviyeye çıkararak bilginin doğru ve etkin kullanımı sağlamaktadır. EBYS'lerin, bağlamı ve ilişkileri ortaya koyan ontoloji tabanlı bir arama motoru yerine kelime tabanlı bir arama motoru kullanmaları, veri/bilgi/belge erişiminde sorunlar yaratmaktadır. Yaratılan bu erişim sorunu ise, organizasyonların gelişimini, süreçlerini ve sürekliliğini olumsuz şekilde etkilemektedir.

IV. 1. 1. 2. Çözüm: Erişim faaliyetleri ne kadar etkili olursa, bilgi varlıkları ne kadar iyi tanımlanabilirse, erişim sistemleri ne kadar iyi geliştirilebilirse; o kadar iyi, doğru ve etkili karar alabilir, ilişki çıkarabilir ve öngörüler elde edilebilmektedir. Bu noktada bilgi erişim faaliyetlerinin çıkarılabilecek en üst noktada olması gerekmektedir. Çalışma, çözüm aşamasında literatürden alınan destekler ile metin madenciliği tabanlı yeni dijital teknolojilerin EBYSlerde bilgi erişim seviyesini iyileştirebileceğini ortaya koymaktadır.

Bilgi organizasyonlar için vazgeçilmez bir kaynak olup bütün iş ve işlem süreçlerinin temelini oluşturmaktadır (Mukred ve diğerleri, 2022). Elektronik belgeler, tanımlandıktan sonra yalnızca ihtiyaç duyulduğunda kullanılan belgeler olmayıp etkin bir yönetim süreci ile erişebilirliği sürdürülmelidir (Tamtürk, 2017, s.857). Organizasyonlar ancak geçmiş belgelere başvurarak etkin planlama ve karar süreçlerini gerçekleştirebilmektedir (Okuonghae ve Bakare-Fatungase, 2023, s. 38). Bunun için ise bilgiye zamanında ve doğru bir şekilde erişim sağlanabilmesi önem taşımaktadır. Organizasyonların pazarda yer edinebilmelerini sağlayan bilgi üstünlüğü, hızlı ve iyi kalitedeki bilgilere erişimi gerektirmektedir (Svård, 2017, s. 69) Bu süreç ne kadar verimli yönetilebilirse, organizasyona sağlayacağı faydalar da o yönde artacaktır. Bunun için ise iyi tasarlanmış belge yönetim sistemlerine ve bilgiyi etkili bir şekilde tanımlayabilmeye ihtiyaç vardır. Çalışmalar ortaya koymaktadır ki, bir EBYS'nin stratejik başarısını, son kullanıcının veri/bilgi/belgelere etkin erişimi ortaya koymaktadır (The Economist Intelligence Unit, 2008 akt.;; Eroğlu ve Külcü 2 , 2014, s. 329).

Çakmak (2017) çalışmasında, bilgi içeriklerinin tanımlanması noktasında, süreçlerin doğrudan metadatalar ile ilişkili olduğunu ve metadataların, bilgi varlıklarına yönelik hizmetlerin, temel yapısını oluşturduğunu belirtmektedir. Bilgiye erişim sürecinde büyük verimlilik sağlayan metadataların, kaynağı en iyi tanımlayacak şekilde, bir standarda bağlı ve çeşitli olması önem taşımaktadır. Organizasyonlarda bilgi/belge erişim sorunları ele alındığında; tanımlama sürecinde insan faktörünün yarattığı yanlışlık ve eksiklikler olduğu görülmektedir. Bilgi varlıklarının doğru yönetilememesi, organizasyonlarda güven eksikliği ve verimsiz karar alma süreçleri yaratan olumsuz sonuçlar doğurmaktadır (Mukred ve diğerleri, 2022). Bu sorunların üstesinden gelebilmek için süreçlerin iyi yapılandırılması ve otomatikleştirilmesi gerekmektedir.

Kullanıcı veritabanındaki bilgileri tahmin edemez bunun için konu metadatasına veri girişinin çok çeşitli tanımlanması gerekir. Genellikle bilgi erişim sistemlerinin tam metin üzerinden anahtar kelime tabanlı teknolojilere dayandığından bahsedilmişti. Anahtar kelime tabanlı sözdizimsel aramalar, belgenin bağlamının kaçırılmasına neden olup verimsiz bir bilgi erişim süreci yaşatmaktadır. Anahtar kelime aramalarında, kullanıcının girdiği terimin; metinde, metaverilerde veya konu başlığında geçen terimler ile aynı olması beklenmektedir (Jurošková ve Závodskás, 2015, s. 459). Geleneksel bilgi erişim tekniklerinin boole operatörleri gibi mantıksal sorgulara dayanması ve sistemlerin sözdizimsel aramaya elverişli olması, deneyimsiz kullanıcıların işini zorlaştırmaktadır (Kara ve diğerleri, 2011, s. 296). Tüm bunlara ilaveten veri depolarının sürekli olarak yeni bilgiler ile dolması, bilgi yığınları içerisinde doğru ve kesin bilgiye erişme zorluğu yaratmaktadır (Sayed ve Muqrishi, 2017, s. 181). Ontoloji tabanlı bilgi erişim sistemleri ve ontolojiye dayanan veri/bilgi/belge tanımlama yöntemleri, söz konusu bilgi erişim sorunlarının üstesinden gelebilmektedir. Çalışmamız ilgili uygulamalara yer vermektedir.

Bilgi, bilgiye erişim için şarttır, bunun için ise EBYSlerin etkin bir bilgi arama ve bilgi erişim mekanizması geliştirmesi gerekir (Hanke ve Thiesse, 2017, s. 2992). Elektronik belgeler, içerisindeki bilgi parçacıklarının yapısını ve anlamını niteleyen metadatalara gereksinim duymaktadır (Барабаш ve Колумбет, 2022, s. 299). Niteleyiciler ne kadar çeşitli olursa bilgi erişimin verimi de o derece artmaktadır. Ontoloji tabanlı bilgi tanımlama ve arama yaklaşımları, bilginin temsilini çeşitlendirerek daha az çaba ile, daha kolay ve daha ilgili daha çok belgeye erişim sağlamaktadır. Felsefe alanından ortaya çıkan bir kelime olan ontoloji, çok güçlü anlamsal modelleme teknolojisidir (Lee, 2020, s. 225). Felsefe alanında ontoloji, “varoluş için sistematik bir açıklama” anlamına gelmektedir (Gruber, 1993; aktaran Tamjid ve diğerleri, 2024, s. 231). İçerisinde bir dizi kavram, taksonomi, ilişki, nitelik, vb. öğeler barındıran ontoloji

(Lee, 2020, s. 225); günümüzde daha çok yapay zeka ve bilgi erişim alanlarında kullanılmakta olup; “bir alanda bilgi paylaşma ihtiyacı duyan araştırmacılar için ortak bir sözlük” olarak tanımlanmaktadır (Noy ve McGuinness, 2001; aktaran Tamjid ve diğerleri, 2024, s. 231). Ontoloji tabanlı yaklaşım bir belgenin hem yapısal hem de içerik yönünden çeşitli şekillerde tanımlanmasını sağlayarak, kullanıcının algılamasını kolaylaştırmaktadır (Барабаш ve Колымбет, 2022, s. 299). Organizasyonlar büyük oranda belgelere sahiptirler. Organizasyonun yapısını ortaya koyan bir ontoloji, bilgi erişim sorunları için önem taşımaktadır. Nitekim Hanke ve Thiesse (2017, s. 2993) çalışmalarında, kamu yönetimi yapısını temsil eden bir ontolojinin oluşturulmasının ve benimsenmesinin, vatandaşlara ve işletmelere, belge içeriklerinin yorumlanması noktasında fayda sağlayacağını belirtmektedir. Belge ontolojisi, belgeyi analiz ederek içeriğe dair kavramsal bir çerçeve oluşturmaktadır. Bu sayede gerekli veriler çıkarılarak, kullanıcıların verilere erişme ve yorumlama performansı artmaktadır (Барабаш ve Колымбет, 2022, s. 299).

EBYS’lerde ontoloji iki amaçta kullanılabilir. İlki, EBYS’lerde ontoloji tabanlı bir bilgi erişim sistemi oluşturmaktır. Bu daha çok içeriğin temsili ile ilgilidir. Şekil 14’teki e-Devlet örneği, ontoloji tabanlı bilgi erişim sistemlerinin önemini ortaya koymakta önemlidir. Ontoloji tabanlı bilgi erişim sistemleri ile, sözdizimsel arama motorlarının yarattığı sorunun üstesinden gelinebilmektedir. Ontoloji ile hem kullanıcının arama motoruna girdiği anahtar kelimeler hem de belgeyi tanımlayan terimler genişletilerek ilgili daha çok belgeye erişim sağlanabilmektedir. Diğer amaç ise hem içerik hem de yapısal temsil ile ilgilidir; burada sadece içerik ilişkileri değil, bir organizasyonun tüm bilgi varlıklarını (organizasyonun sahip olduğu tüm varlıklar; çalışan, müşteri, belgeler, yazılım-donanım sistemleri gibi) kapsayan ilişkilerinin gösterimi söz konusudur. Burada amaç daha çok organizasyonun fonksiyonlarını ortaya

koyabilmek, ilişkileri çıkarabilmek ve karar mekanizmalarını geliştirmektedir. Mevcut sistemlerde organizasyonun fonksiyonlarını ortaya koyan yapı Standart Dosya Planıdır. Ancak Standart Dosya Planı organizasyonun sadece konusal ilişkilerini göstermektedir. Bir ontoloji ise, organizasyonun sahip olduğu tüm bilgi varlıkları arasındaki ilişkileri aktarabilmektedir.

Metin madenciliği ile bir ontoloji oluşturmak için, öncelikle amaç belirlenmelidir, bu amaç doğrultusunda ise ontolojinin oluşturulacağı alan belirlenmiş olur. Bu çalışmada amaç bir organizasyonun bilgi varlıklarına yönelik bir ontoloji oluşturmaktır. Bunun için organizasyona ait eylemlerin analizi yapılmalıdır. Bunun en temel yolu ise organizasyonun sahip olduğu dokümanların toplanmasıdır. Dokümanların toplanmasında dikkat edilecek husus, dokümanların kaliteli, alana özgü ve temsil edici nitelikte olmasıdır. Bu aşamadan sonra dokümanlar içerisinde yer alan metinler ön işlemden geçirilmelidir (Tokenizasyon, durdurma kelimelerini filtreleme, kök çıkarma gibi). Daha sonra ön işlemden geçirilmiş metinlere metin madenciliği tekniklerinden varlık/kavram çıkarma tekniği uygulanmalıdır. Bu sayede aralarında ilişkilerin kurulacağı varlıklar belirlenmektedir (örneğin; kişiler, kuruluşlar, zaman ifadeleri gibi varlıklar) (Bkz. Şekil 17). İlişkiler kurulurken farklı yaklaşımlar tercih edilebilmektedir. Sözdizimsel kalıpların veya alana özgü kalıpların kullanılması bu yaklaşımlardan bazılarıdır (Tamjid ve diğerleri, 2024).

- Örneğin; “X, Y’nin **bir türüdür.** / X’e örnek olarak V, Y ve Z verilebilir. / X **bir dır.**” gibi kalıplar kullanılarak kavramlar arası ilişkiler ortaya çıkarılabilmektedir (Pfeifer ve Peukert, 2013, s. 5) (Bkz. Şekil 16).

Varlıklar arası ilişkiler de kurulduktan sonra son aşama taksonomi eşleştirmedir. Taksonomi eşleştirmede amaç, dokümanlarda aynı anlama gelen fakat farklı ifade edilen

terimlerin birleştirilmesidir. Örneğin dokümanın birinde “kuruluş” kullanılırken diğerinde “şirket” kullanılabilmektedir. Bu aşamada bu terimlerin birbirine eşit olduğu tanımlanabilirken diğer hiyerarşik ilişkiler de tanımlanabilmektedir (daha geniş (Broader Terms), daha dar (Narrower Terms), ilişkili (Related Terms)).

Oluşturulan ontolojinin, alanında uzman dilbilimci ve bilgi yöneticileri ile gözden geçirilmesi önemli olup, ontoloji sürekli olarak geliştirilmeye ve genişlemeye uygun olmalıdır. Aşağıda ontoloji bileşenlerine yer verilmektedir. Genel olarak ontolojiler; sınıflar (classes), örnekler (instances), ilişkiler (relationships), özellikler (properties), işlevler (functions) ve kısıtlamaları (constranints) içermektedir.

Tablo 4.

Ontoloji bileşenleri

Bileşen	Açıklama
Sınıflar (Classes)	Nesneleri tanımlayan kavramlar kümesi
Örnekler (Instances)	Nesnelere ait belirli öğeler
İlişkiler (Relationships)	Bu öğeler arasındaki anlam ilişkileri
Özellikler (Properties)	Bu öğelere ait özellik değerleri
Fonksiyonlar (Functions)	Bu öğeleri tanımlayan işlevler ve süreçler
Kısıtlamalar (Constraints)	Bu öğeleri tanımlayan betimleyici mantık ve kurallar

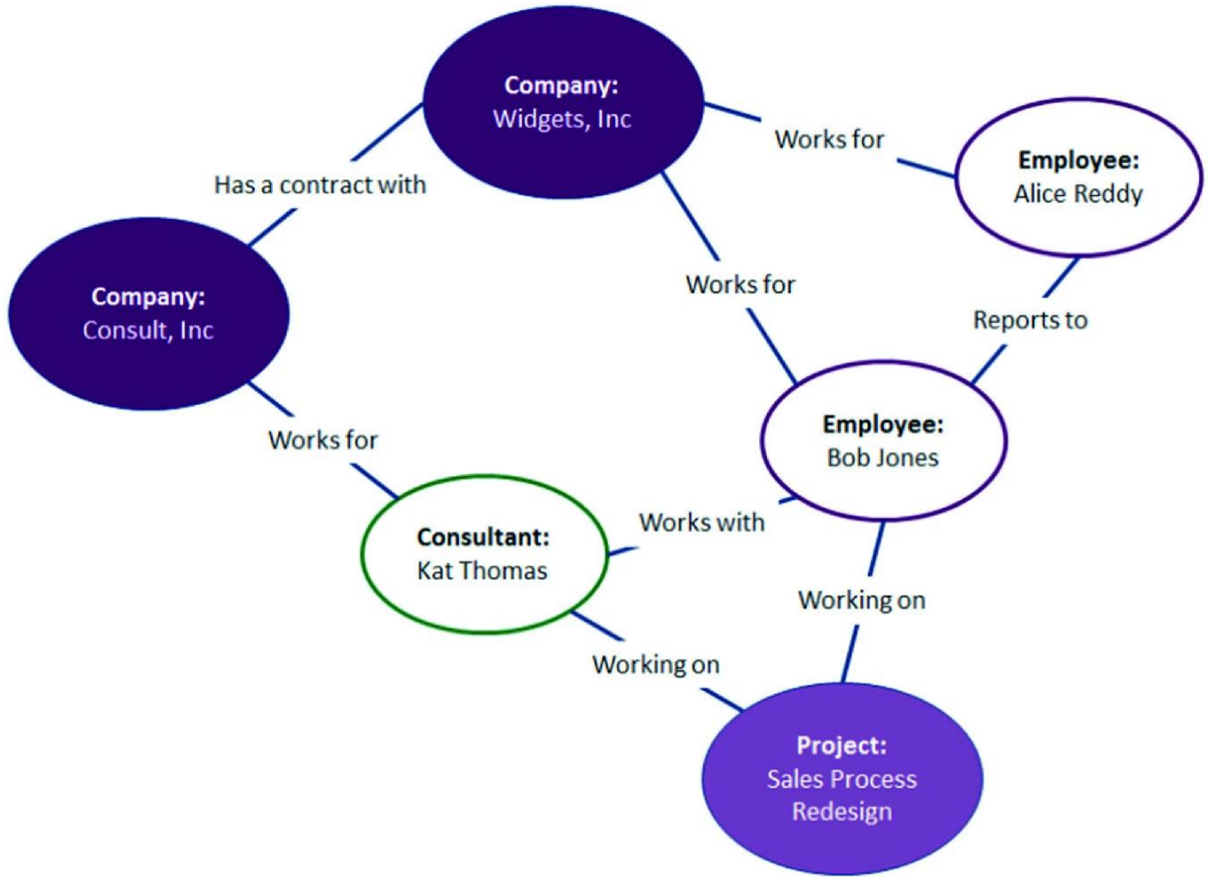
(Lim ve diğerleri, 2011; aktaran Lee, 2020, s. 226)

Yapılmış bazı önemli ontoloji örnekleri bulunmaktadır. Örneğin, HowNet, Çince ve İngilizcedeki kavramsal ilişkilere odaklanırken (Dong ve Dong, 2003, s. 820), Balka-Net'in TR-kelime ağı ve KeNet (76.757 eşanlamlı kümeyi içermektedir) Türkçedeki kavramsal ilişkileri ortaya koymaktadır (Bakay ve diğerleri, 2021). Ontoloji tabanlı arama

motoru geliřtiren alıřmaların, ontoloji olarak genelde WordNet'i kullandığı grlmektedir (Kara ve diğerkleri, 2011; Lee, 2020; Tamjid ve diğerkleri, 2024). WordNet, thesaurus yapısına benzer řekilde alıřan, kelimeler arası anlamsal iliřkileri gsteren, ingilizce bir veritabanıdır (WordNet, Eriřim Tarihi: 04.05.2025). Ařağıda WordNet'teki bir kelime ontolojisi rneğı yer almaktadır.

řekil 15.

WordNet ontoloji rneğı



(Lee, 2020, s. 217)

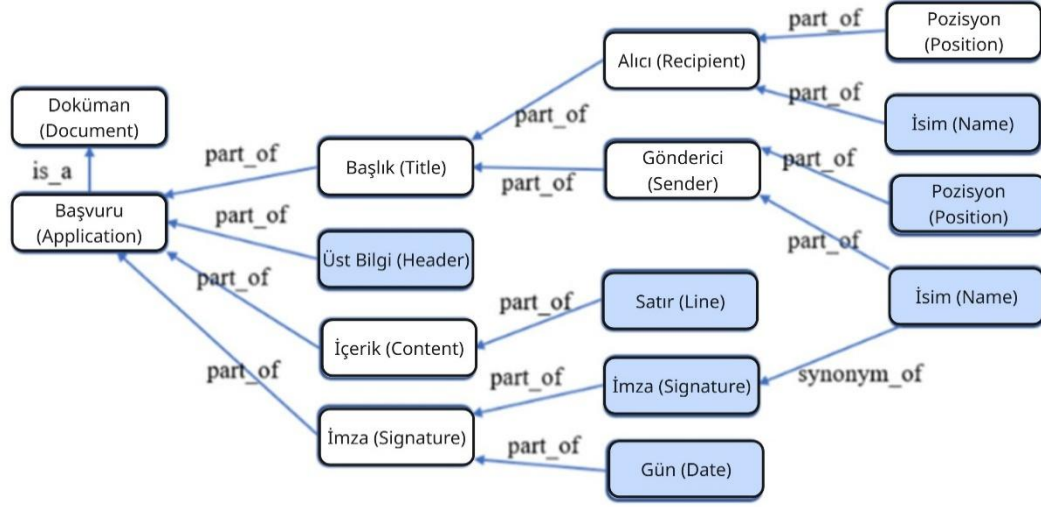
Daha nce de belirtildiğı zere ontoloji EBYS'lerde belge eriřim sreci iin hem belge tanımlamalarında hem de arama sorgularında kullanılabilmektedir. Tanımlama noktasında, ontoloji hazırlandıktan sonra, metin madenciliğı ile metin ierisindeki

terimler TF-IDF matrisi kullanılarak sayısal vektörlere dönüştürülmelidir. Böylece metindeki en önemli terimler sıralanmaktadır. Metin madenciliği ile çıkarılan terimler, ontoloji ile çeşitlendirilmelidir. Bu sayede belgeyi en iyi şekilde tanımlayan ve anlamsal olarak belgeye birçok farklı terimden erişimi sağlayan, konu metadataları elde edilebilmektedir. Bu sayede kullanıcılar tam metin aramanın ötesine geçerek, metinde bulunan terimlerin, farklı anlamsal gösterimleri ile de bilgiye erişebilmektedir. Yapılan bu işlem, belgeyi tanımlama sürecinde personele büyük fayda sağlayacak olup, kesin isabet ve erişim isabeti oranlarını da pozitif olarak etkileyecektir. Fakat ontoloji ile nitelendirici bilgileri genişletmenin yanında, kullanıcının arama motoruna girmiş olduğu anahtar kelimelerin de sistem tarafından ontoloji ile çeşitlendirilmesinin sağlanması daha çok sorgu-belge eşleşmesini yakalayacaktır. Böylece hem tanımlama sürecinde hem de arama sürecinde ontoloji kullanılarak karşılaşılan “kelime farklılıkları” çözülebilecektir (Chen, 2001, s. 27).

Ancak bir belge yalnızca anlamsal bileşenler içermemektedir, aynı zamanda yapısal bileşenler de içermektedir (örneğin imza, gönderici, alıcı, tarih, departman vb.). Belgeyi oluşturan bu bileşenler arasındaki ilişkilerin nitelenmesi; tanımlama, konu analizi ve erişim noktalarında faydalar sağlamaktadır. Барабанш ve Колумбет (2022, s. 298) çalışmalarında bu ilişkileri ortaya koymak için bir ontoloji oluşturmaktadır. Bu ontolojiyi temel ilişki türleri olan “is a”, “part of” ve “synonym of” ile göstermektedir.

Şekil 16.

Başvuru belge sınıfını tanımlayan bir ontoloji örneği



(Барабаш ve Колумбет, 2022, s. 299)

Şekilde yapısal düğümler mavi, anlamsal düğümler ise beyaz olarak gösterilmektedir (Барабаш ve Колумбет, 2022, s. 299). Şekil üzerinden bu ilişki türlerini açıklayacak olursak; “is a” belgenin bir başvuru türü olduğunu, part_of ismin göndericinin bir parçası olduğunu, “synonym_of” ise isim ile imzanın birbirlerine eş olduğunu göstermektedir.

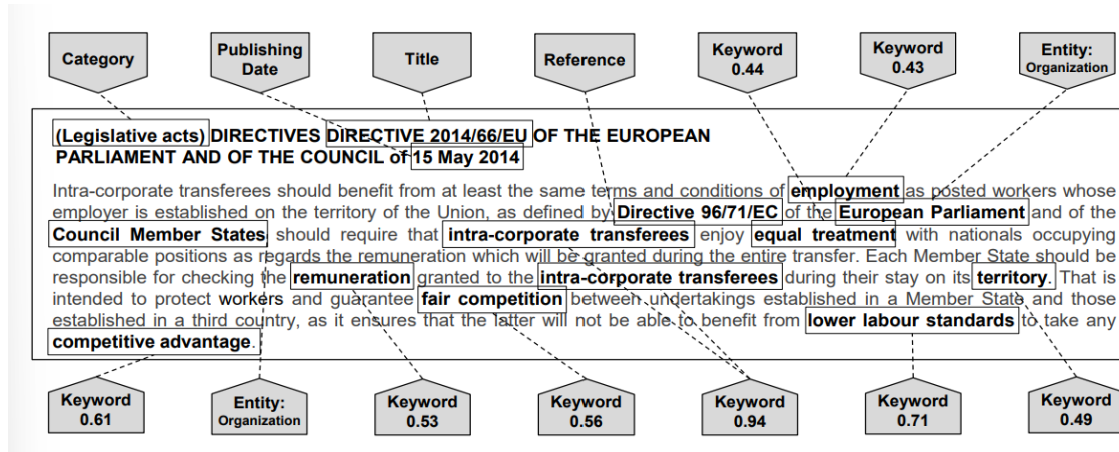
Organizasyonda bu şekilde oluşturulmuş bir ontoloji; belgenin yapılandırılmasında önemlidir. Metin madenciliği ontolojiler sayesinde, bilgi çıkarma tekniği ile bu özellikleri çıkarabilmektedir (örneğin; bu bir başlıktır, bu bir tarihtir vb.). Bu şekilde yapılandırılmış bir belge üzerinden özetleme, sınıflama, kümeleme gibi metin madenciliği teknikleri uygulanabilmekte ve tanımlama, arama ve erişim süreçleri otomatikleştirilerek daha etkili belge yönetimi süreçleri gerçekleştirilebileceği öngörülmektedir.

Önemli bir metin madenciliği tekniği olan bilgi çıkarma için, öncelikli aşama varlık çıkarmadır. Bu tanımlama ve erişimde önem arz etmektedir. Varlık çıkarma ile bir

belgede geçen kurumlar, kişiler, anahtar kelimeler ortaya konabilmektedir. Tüm bunlar metadataların otomatik olarak belirlenmesi için gerekli verilerdir. Hanke ve Thiesse (2017, s. 2995) çalışmalarında IBM AlchemyAPI aracını kullanarak derin öğrenme ile, bir hukuk belgesi üzerinden altı tane bilgi varlığı (Coğrafi Bölgeler, Kişiler, Kuruluşlar, Alan Terminolojileri, Sayısal İfadeler (örneğin tarih), Anahtar Kelimeler) çıkartmayı amaçlamaktadırlar.

Şekil 17.

Belge içerisinden varlık çıkarma örneği



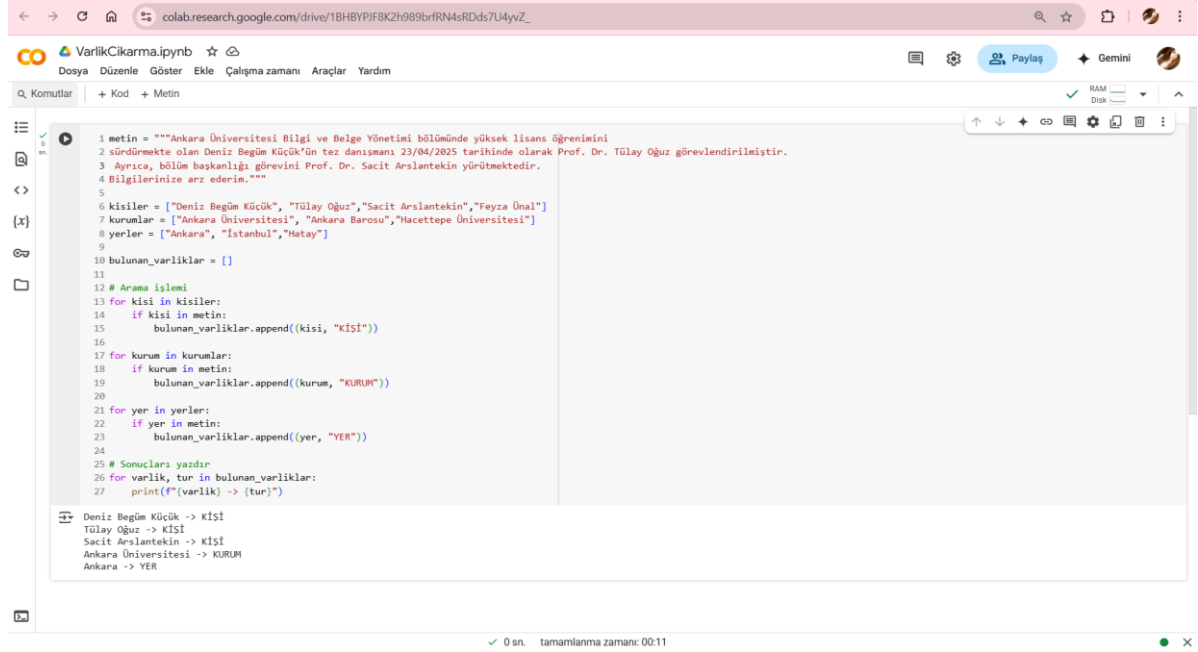
(Hanke ve Thiesse, 2017, s. 2995)

Makine öğreniminin yanı sıra kural tabanlı yöntemler ile de varlıkların çıkarılması mümkündür. Önceden tanımlanmış bir kişi, kurum vb. listeleri oluşturularak, metnin içerisindeki kelimeler ile listelerdeki terimlerin eşleşmesi aratılarak metindeki varlıklar çıkarılabilmektedir. Aşağıda Python ile¹¹ basit bir varlık çıkarma örneği verilmektedir.

¹¹ Metin madenciliği uygulamaları, Python gibi kodlama dilleri ile gerçekleştirilebilmesinin yanı sıra Orange, Rapidminer, Knime gibi yazılımlarla da yapılabilmektedir. Bu çalışmada uygulamalar, Google Colab (Colaboratory) üzerinden Python dili ile gerçekleştirilmiştir.

Şekil 18.

Python ile varlık çıkarma



```
1 metin = """Ankara Üniversitesi Bilgi ve Belge Yönetimi bölümünde yüksek lisans öğrenimini
2 sürdürmekte olan Deniz Begüm Küçük'ün tez danışmanı 23/04/2025 tarihinde olarak Prof. Dr. Tülay Oğuz görevlendirilmiştir.
3 Ayrıca, bölüm başkanlığı görevini Prof. Dr. Sacit Arslantekin yürütmektedir.
4 Bilgilerinize arz ederim."""
5
6 kişiler = ["Deniz Begüm Küçük", "Tülay Oğuz", "Sacit Arslantekin", "Feyza Ünal"]
7 kurumlar = ["Ankara Üniversitesi", "Ankara Barosu", "Hacettepe Üniversitesi"]
8 yerler = ["Ankara", "İstanbul", "Hatay"]
9
10 bulunan_varliklar = []
11
12 # Arama işlemi
13 for kisi in kişiler:
14     if kisi in metin:
15         bulunan_varliklar.append((kisi, "KİŞİ"))
16
17 for kurum in kurumlar:
18     if kurum in metin:
19         bulunan_varliklar.append((kurum, "KURUM"))
20
21 for yer in yerler:
22     if yer in metin:
23         bulunan_varliklar.append((yer, "YER"))
24
25 # Sonuçları yazdır
26 for varlik, tur in bulunan_varliklar:
27     print(f"{varlik} -> {tur}")
```

Deniz Begüm Küçük -> KİŞİ
Tülay Oğuz -> KİŞİ
Sacit Arslantekin -> KİŞİ
Ankara Üniversitesi -> KURUM
Ankara -> YER

Verilen kodda ilk olarak bir belgenin içeriğini oluşturabilecek basit bir metin kullanılmaktadır. Ardından kişiler, kurumlar ve yerleri belirten ayrı ayrı listeler oluşturulmaktadır. Sonraki aşamada ise listeler teker teker metin içerisinde aratılmaktadır. En son aşamada eşleşen varlıkların sonuç içerisinde hangi varlık grubuna ait olduğu yazdırılmaktadır. Bu kodda metin içerisinde geçen “Deniz Begüm Küçük”, “Tülay Oğuz” ve “Sacit Arslantekin”in “KİŞİ”, “Ankara Üniversitesi”nin “KURUM”, “Ankara”nın ise “YER” varlık grubuna ait olduğu çıkarılmaktadır.

Önemli anahtar kelimelerin çıkarılması için ise, TF-IDF matrisini önermekteyiz. TF-IDF bir metindeki en önemli terimleri sıralamayı sağlamaktadır. Ancak önemli terimler bazen tek bir sözcükten değil bir kelime öbeğinden oluşmaktadır. Bu kelime öbeklerini kaçırmamak için ise, n-gram modeli kullanılabilir. Aşağıda bu işlemin Python ile bir örneği kodlanmaktadır.

Şekil 19.

Python ile önemli anahtar kelimelerin çıkarılması

```
1 from sklearn.feature_extraction.text import CountVectorizer
2
3 # Örnek İngilizce metin
4 text = """
5 Text mining, also known as text data mining, is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights. You can use text mining to analyze vast collections of textual materials
6 """
7
8 # TF (term frequency) vektörleştirici - 1'den 2'ye kadar n-gramlar (unigram + bigram)
9 vectorizer = CountVectorizer(stop_words='english', ngram_range=(1, 2))
10 tf_matrix = vectorizer.fit_transform([text])
11
12 # Terimler ve skorları
13 terms = vectorizer.get_feature_names_out()
14 counts = tf_matrix.toarray()[0]
15 term_counts = list(zip(terms, counts))
16 term_counts = sorted(term_counts, key=lambda x: x[1], reverse=True)
17
18 # En sık geçen kelime ve kelime öbeklerini yazdır
19 print("Top keywords and keyphrases (using sliding window / n-grams):")
20 for term, count in term_counts[:15]:
21     print(f"{term}: {count}")
```

Top keywords and keyphrases (using sliding window / n-grams):
text: 4
mining: 3
text mining: 2
analyze: 1
analyze vast: 1
capture: 1
capture key: 1
collections: 1
collections textual: 1
concepts: 1
concepts trends: 1
data: 1
data mining: 1
format: 1
format identify: 1

Yukarıdaki kodda görüleceği üzere, ilk olarak örnek bir text mining konulu metin tanımlanmaktadır. Stop words ile gereksiz terimlerin ayıklaması gerçekleştirilmektedir, ardından n-gram modeli ile 1’li ve 2’li kelime ve kelime grupları tanımlanmaktadır. ngram_range=(1, 3) şeklinde tanımlansaydı 3’lü kelime grupları da dahil olurdu fakat bu kodda en fazla 2’li kelime grupları tercih edilmektedir. Tüm bu işlemler sonucunda ise tüm kelime ve kelime gruplarına TF ile bir ağırlık atanmaktadır. Bu, kelimelerin metin içerisindeki frekansını (geçme sıklığını) göstermektedir. Ardından alınan terim ve skorlar eşleştirilerek en yüksekten en düşüğe ilk 15 tanesi yazdırılmaktadır.

Bu iki örnekte, metin madenciliği ile daha sonra aralarındaki bilgi ve ilişkilerin keşfedilmesine yardımcı olacak, bilgi varlıklarının çıkarılması sağlanmaktadır. Belgeyi tanımlayan ve erişilmesini sağlayan metadataları temsil eden bu bilgi varlıklarının, örneklerde verildiği gibi otomatik bir şekilde çıkarılması, bilgi işleyenlerin hata yapma riskini ortadan kaldırıp, zaman ve emekten tasarruf etmelerine yardımcı olmaktadır. Çıkarılan bu varlıklar özellikle de anahtar kelimeler, erişimi en verimli seviyeye getirebilmek için çeşitlendirilmelidir. Bu ise ontoloji ile gerçekleştirilmektedir. Söz

konusu terimler; konu başlığı listeleri, thesaurus, dizinler vb. kullanılarak oluşturulmuş bir ontoloji ile anlamsal olarak ilişkilendirilmelidir. Bu sayede belgenin verimli bir temsili gerçekleştirilebilmektedir.

Bilgi varlıklarının çıkarılması sadece tanımlama aşamasında değil, arama motorlarında da fayda sağlayabilmektedir. Örneğin bilgi varlıkları sınıflandırılarak kullanıcıların filtreleme yapabilmeleri sağlanabilmektedir. Böylece kullanıcılar anahtar kelime ile aramanın yanında, göndericinin sadece “x kişisi” olduğu belgelere erişmek gibi filtrelemeler uygulayabilmektedir.

Bunun dışında çıkarılan bilgi varlıkları ve anahtar kelimeler, otomatik bir şekilde kümeleme ve sınıflama yapılabilmesi için önemlidir. TF-IDF ile her bir terime atanan puanlar daha sonra belgeler arasındaki benzerliği bulmak için kullanılabilir (Hanke ve Thiesse, 2017, s. 2996). Kümeleme ve sınıflandırma yöntemleri ile otomatik bir şekilde benzer belgeleri, sorguları vb.’leri bir araya getirmek, ileride kurumsal bilgi haritaları oluşturmada (Chen, 2001, s. 24), konu analizlerinde ve karar almada fayda sağlayacaktır. Arıcı ve Kandur (2016, s. 70) çalışmalarında, standart dosya planı ile ilgili olarak, belgeler arasındaki organik bağın kurulması ve korunması ile karar alma süreçlerinde bilgi ve belge erişimine dair verimin arttırılacağını belirtmektedir. Bir organizasyonda, birbiri ile ilişkili belgelerin bir arada bulunması temel gerekliliklerden birisidir. Çünkü neredeyse tüm işleyiş bilgiye bağlıdır. Belgeler arasında kurulmuş ilişkiler, belgenin içeriğinin analizini esas alan, anlamsal anlayış ile tanımlanan, bir düzenleyici görevi görmektedir (Бабабаи ve Колымбет, 2022, s. 297).

3. Bölümde sınıflandırmadan bahsederken, sınıflamanın, verinin önceden tanımlanmış sınıflara otomatik bir şekilde atanması ile gerçekleştiği açıklanmıştı. Bir EBYSde önceden tanımlanmış olan sınıf, standart dosya planı veya organizasyona özel

olarak oluşturulmuş bir sınıflama sistemi olarak ele alınabilmektedir. Veri ise EBYS üzerinden oluşturulmuş veya EBYSye aktarılmış belgeleri temsil etmektedir. Biz bu çalışmada metin madenciliği ile ontoloji tabanlı bir sınıflama gerçekleştirilebileceğine dair bir yaklaşım önermekteyiz. Bu yaklaşımda, standart dosya planındaki veya özel olarak oluşturulmuş bir sınıflama sisteminde yer alan konu başlıkları, sınıflamanın daha doğru gerçekleştirilebilmesi için dizinler, thesauruslar, sözlükler kullanılarak genişletilmektedir. Ardından sınıflandırılmayı bekleyen belgelere metin madenciliği uygulanmaktadır. Belgeler ilk olarak önışlem süreçlerinden geçirilmekte, ardından TF-IDF ile metinler sayısallaştırılmaktadır. Belgeyi temsil eden en önemli kelimeler yine ontoloji ile genişletilerek, konu başlıkları ile terimler arasındaki benzerlik hesaplanmaktadır. Atanan benzerlik puanlarından hangisi daha yüksek ise belge o konu başlığı altında sınıflanmaktadır.

Aşağıda Google Colab (Colaboratory) üzerinden Python ile oluşturulmuş bir sınıflandırma örneği yer almaktadır.¹²Bu örnekte, ilk olarak, konusu tespit edilecek metin için önışlem komutları girilmektedir. Önışlem sürecinde, tüm metin küçük harflere dönüştürülmekte ve işaret ve rakamlar filtrelenmektedir. Ardından Türkçe durdurma kelimeleri tanımlanmakta ve yine metin içerisinden filtrelenmektedir. Filtreleme işleminden sonra, metin tokenlara ayrılmaktadır. Burada n-gram tekniği kullanılmakta olup n=2 olarak belirlenmiştir.

¹² Bu örneğin, tek bir metin üzerine uygulanması nedeniyle, TFIDF yerine TF kullanılmaktadır. Gerçekleştirilen uygulama temel düzeyde olup, önışlem süreçlerinin daha etkili geçirilebilmesi için Colab yerine Visual Studio Code, PyCharm, Python gibi kod editörlerinde zemberek gibi kütüphaneler kullanılmalıdır.

Tüm bu süreçlerin ardından EBYS’ler üzerinde bir belge gövdesini temsil edecek, metin girişi gerçekleştirilmektedir. Bu örnekte, metin kısmına “araştırma görevlisi sınavına itiraz” ile ilgili bir dilekçe girilmiştir. Ardından, SDP’den alınan temsili üç konu başlığı seçilmiştir. Bu konular; “105.02.04 – Doktora”, “907 - Kadro ve Pozisyon İşleri” ve “952 - Afet ve Acil Durum Yönetimi”dir. Her bir konu başlığı için ise, konuları temsil edecek terimler belirlenmiştir. Örneğin “Kadro ve Pozisyon İşleri” için; "araştırma görevlisi", "kadro", "pozisyon"... gibi çeşitli terimler saptanmıştır. Burada amaç, belge metninde yer alan her bir kelimenin, konu başlıklarını temsil eden terim kümeleri ile ne kadar benzer olduğunu tespit etmek ve metne en uygun konu başlığını otomatik atayabilmektir.

Bu uygulamada, frekansı saydırılan terimlerle, konu başlıklarını temsil eden terim kümeleri karşılaştırıldığında; “araştırma görevlisi sınavına itiraz” ile ilgili dilekçe için en yüksek benzerlik; 907 - Kadro ve Pozisyon İşleri **(3)** olarak tespit edilmektedir. Ardından bu sıralamayı, 105.02.04 – Doktora **(1)** ve 952 - Afet ve Acil Durum Yönetimi **(0)** konuları takip etmektedir.

Şekil 20.

Python ile otomatik konu atama

```
import re
from collections import Counter

# Ön işlem fonksiyonu (küçültme ve sadece harfleri bırak)
def preprocess(text):
    text = text.lower()
    text = re.sub(r'[^a-zçışöü ]', '', text)
    return text

# Türkçe stop words
turkish_stop_words = set([
    "acaba", "ama", "aslında", "az", "bazı", "belki", "ben", "bana", "bazen",
    "bazıları", "benim", "beri", "beş", "bile", "bir", "birçok", "biri",
    "birkac", "birşey", "biz", "bu", "buna", "bunu", "da", "daha", "de",
    "defa", "diye", "diğer", "eğer", "en", "gibi", "hem", "hep", "hepsi",
    "her", "hic", "için", "ile", "ise", "keş", "ki", "kim", "mı", "mi",
    "mu", "mü", "nasıl", "ne", "neden", "nerde", "nerede", "nereye", "niçin",
    "o", "sanki", "şey", "siz", "şu", "tüm", "ve", "veya", "ya", "yani"
])

def remove_stopwords(words):
    return [w for w in words if w not in turkish_stop_words]

def get_bigrams(words):
    return [' '.join(pair) for pair in zip(words, words[1:])]

# Metin
text = """
Ben, Deniz Begüm Küçük, Ankara Üniversitesi Bilgi ve Belge Yönetimi Bölümü araştırma görevlisi sınavına katıldım.
Sınav sonuçlarının objektif şekilde değerlendirilmediğini düşünüyorum. Puanlama sürecinde taraflı davranıldığını düşünüyorum.
Bu nedenle değerlendirmelerin yeniden gözden geçirilmesini talep ediyorum.
"""

# Konular ve terimleri
topics = {
    "105.02.04 - Doktora": [
        "doktora", "tez", "savunma", "akademik", "araştırma", "bilimsel", "yayın",
        "yüksek lisans", "tez danışmanı", "doktora programı", "doktora öğrencisi",
        "doktora süreci", "doktora tezi", "doktora komitesi", "doktora başvurusu",
        "akademisyen", "tez konusu", "doktora mezunu", "doktora dersleri", "bilim", "eğitim"
    ],
    "907 - Kadro ve Pozisyon İşleri": [
        "araştırma görevlisi", "kadro", "pozisyon", "atama", "görev", "unvan", "personel", "çalışan", "memur",
        "iş ilanı", "yönetici", "terfi", "tayin", "işe alım", "pozisyon değişikliği",
        "özlük hakları", "maaş", "bordro", "sicil", "disiplin", "performans", "sözleşme"
    ],
    "952 - Afet ve Acil Durum Yönetimi": [
        "afet", "acil durum", "kriz", "yangın", "deprem", "tahliye", "müdahale", "yardım",
        "kurtarma", "risk", "planlama", "yönetim", "seferberlik", "alarm", "sivil savunma",
        "afet raporu", "afet planı", "zarar", "iyileştirme", "güvenlik", "ilk yardım"
    ]
}

# Ön işlemler
processed_text = preprocess(text)
words = processed_text.split()
words = remove_stopwords(words)
bigrams = get_bigrams(words)

# Konu terimlerini de aynı şekilde hazırla (küçült + bigram çıkar)
processed_topics = {}
for topic, terms in topics.items():
    # Terimleri küçült, boşlukları koru, sonra bigramlar da var zaten
    clean_terms = []
    for term in terms:
        pterm = preprocess(term)
        t_words = pterm.split()
        t_words = remove_stopwords(t_words)
        # hem tek kelime hem bigram olarak alalım
        clean_terms.extend(t_words)
        clean_terms.extend(get_bigrams(t_words))
    # tekrar set yapalım fazlalıkları önlemek için
    processed_topics[topic] = set(clean_terms)

# Metindeki kelime + bigram frekanslarını say
metin_terimleri = words + bigrams
metin_counter = Counter(metin_terimleri)

# Her konu için kesişen terimlerin frekans toplamı
scores = {}
for topic, topic_terms in processed_topics.items():
    common_terms = topic_terms.intersection(metin_counter.keys())
    score = sum(metin_counter[t] for t in common_terms)
    scores[topic] = score

# En yüksek skoru alan konu
assigned_topic = max(scores, key=scores.get)

print("Metnin önemli kelimeleri ve bigramları:", metin_terimleri)
print("Ortak kelime ve bigram sayıları:", scores)
print("Metne atanan konu:", assigned_topic)
```

Metnin önemli kelimeleri ve bigramları: {'deniz', 'begüm', 'küçük', 'ankara', 'üniversitesi', 'bilgi', 'belge', 'yönetimi', 'bölümü', 'araştırma', 'görevlisi', 'sınavına', 'sonuçlarının', 'objektif', 'şekilde', 'değerlendirilmediğini', 'düşünüyorum', 'puanlama', 'sürecinde', 'taraflı', 'davranıldığını', 'düşünüyorum', 'bu', 'nedenle', 'değerlendirmelerin', 'yeniden', 'gözden', 'geçirilmesini', 'talep', 'ediyorum'}

Ortak kelime ve bigram sayıları: {'105.02.04 - Doktora': 1, '907 - Kadro ve Pozisyon İşleri': 3, '952 - Afet ve Acil Durum Yönetimi': 0}

Metne atanan konu: 907 - Kadro ve Pozisyon İşleri

Yapılan örnekte, hem metinden çıkarılan terimlerin hem de konu başlıklarını temsil eden terimlerin ontoloji ile genişletilmesi sonucu, doğruluk oranı daha yüksek sınıflama işlemleri gerçekleştirilebileceği öngörülmektedir.

Bu örneklerin haricinde otomatik sınıflama sürecinde, benzerlik kullanılmasının yanı sıra, belgeler önışlem sürecinden geçirildikten ve TF-IDF ile sayısallaştırıldıktan sonra, sınıflama algoritmaları (Karar Ağacı, Naive Bayes, K-En Yakın Komşu vb.) kullanılarak da model eğitilebilmektedir. Böylece yeni bir belge geldiğinde sistem otomatik olarak belgeyi bir sınıfa atamaktadır (Kumbhar, 2012, s. 117).

Kümeleme de ise önceden tanımlanmış sınıflar bulunmamaktadır. Bu özelliği nedeni ile sınıflamadan daha zordur (Lin ve Hsueh, 2006, s. 554). Belgeler direkt olarak birbirleri ile yakınlıkları ölçülerek kümelere atanmaktadır. Burada sınıfları oluşturan etiketlerin tamamı verilerden (belgelerden) elde edilmektedir. Kümeleme, herhangi bir dosya planının bulunmadığı durumlarda veya mevcut dosya planının belgeleri temsil edemediği durumlarda kullanılabilir. Kümelemede de TF-IDF ile sayısallaştırılan ve sıralanan terimleri ontoloji ile geliştirmek, anlamsal olarak daha ilişkili belgelerin bir arada bulunabilmesini sağlayacaktır. Burada sınıflamadan farklı olarak, veriler yani belgeler, konu başlıklarına olan yakınlıklarına göre değil, belgelerin birbirlerine olan yakınlıkları hesaplanarak kümelenmektedir.

Fakat biz EBYSlerde belge erişim sürecinde, iki tekniğin bir arada kullanıldığı arama motorlarının daha iyi verim sağlayacağı inancındayız. Bu şekilde kurulacak bir yapıda, belgeler ilk olarak önceden tanımlanmış sınıflara atanmaktadır. Sınıflanmış belgeler ise kendi içlerinde daha alt/dar konulara kümelenmektedir. Örneğin x üniversitesinde bir personel EBYS üzerinden “tanıma ve denklik işleri” ile ilgili bir belge

arıyor olsun. Fakat bu çok geniş bir anahtar kelime. Dolayısı ile sistem tanıma ve denklik işleri ile ilgili birçok belgeyi getirecektir. Kümeleme “tanıma ve denklik işleri” konulu belgeleri ayrıca kümelediği zaman “yurtdışındaki denklik işleri”, “yurtiçindeki denklik işleri” vb. şeklinde daraltmaktadır. Böylece arama motorlarında daha detaylı ve gelişmiş arama sağlanabilmektedir. Bu özellikle bilgi gereksinimini tanımlamakta zorluk çeken kullanıcılar için önemlidir. Bu sayede sistem kullanıcıyı yönlendirerek kesin isabet oranını arttırmaktadır.

Benzerlik hesaplama aynı zamanda alaka düzeyi sıralamaları ve öneri sistemleri için de önemlidir. Sıralama temel düzeyde, kullanıcının girdiği anahtar kelimenin, derlemde bulunan belgeler içerisinde en çok benzeyenden en az benzeyene doğru sıralanmasını tanımlamaktadır. Sıralama için çeşitli ölçütler bulunmaktadır. Bunlar; terim frekansı, kullanıcıların dokümanla olan etkileşimini gösteren popülerlik skoru veya içerik skoru olabilmektedir. İçerik skoru, kullanıcının girmiş olduğu anahtar kelimenin, dokümanın hangi kısmında geçtiğine göre (başlık, gövde, anahtar kelime vb.) değişkenlik göstermektedir (Lee, 2020, s. 200). Bu ölçütlerin çeşitli kombinasyonları ile sıralama gerçekleştirilebilmektedir.

Kullanıcı odaklı ilgi sıralamalarının ise öneri sistemlerini oluşturduğunu söylemek yanlış olmamaktadır. Çalışanlar arası bilgi aktarımı için önem arz eden öneri sistemleri genel olarak, kullanıcı ile belge arasındaki ilişkileri yansıtmaktadır (Hanke ve Thiesse, 2017, s. 2993).

Hanke ve Thiesse (2017, s. 2997) çalışmalarında öneri sistemleri ile ilgili olarak, içerik tabanlı ve işbirlikçi olmak üzere iki farklı benzerlikten bahsetmektedir. İçerik tabanlı benzerlik; atıflar, anahtar kelimeler, varlıklar ve metadatalar gibi benzer özellikler paylaşılmasını içermektedir. Bu benzerlikte, belgeler arasında doğrudan bir bağlantı

olmasa bile dolaylı bağlantılar sayesinde benzerlikler keşfedilmektedir. Örneğin A ile B belgesinin benzer bir özelliği paylaşması ile B ile D belgesinin dört özelliği paylaşması, dolaylı olarak A ile D belgesini de birbirine bağlamaktadır. İşbirlikçi benzerlikte ise, kullanıcıların aynı belgeleri ilgili veya ilgisiz bulma durumlarına göre benzerlik hesaplanmaktadır. Hanke ve Thiesse (2017) çalışmalarında öneri puanını; içerik tabanlı benzerlik %80, işbirlikçi benzerlik ise %20 olacak şekilde birleştirerek elde etmektedir.

Yapılan çalışmalar ile ontoloji tabanlı uygulamaların bilgi erişim alanda önemli verimler elde ettiği ortaya konulmaktadır. Kara ve arkadaşları (2011, s. 297) çalışmalarında, kelime tabanlı sözdizimsel aramalara göre ontoloji tabanlı aramaların etkinliği test etmişlerdir. Sonuca göre kesin isabet ve erişim isabetinde önemli artışlar elde ederek ontolojinin önemli bir anlamsal anlayışı ortaya koyduğunu belirtmektedirler. Sayed ve Muqrishi (2017, s. 188) benzer şekilde, ontoloji tabanlı bir arama motoru geliştirmişlerdir. Geliştirdikleri arama motorunun ise bilgi erişim açısından bazı avantajlar elde ettiğini gözlemlemektedirler. Çalışmaya göre ontoloji tabanlı arama motorları geleneksel arama motorları ile karşılaştırıldığında daha ilgili sonuçlar elde ederek kesin isabet oranının da artış gözlemlemişlerdir. Ontoloji odaklı bir başka çalışmada da arama motorunun ontoloji sayesinde ilgisiz sonuçları ayıklayarak, ontolojinin, ilgili ve değerli belgeleri elde etme gücünü ortaya koymaktadırlar (Yang ve diğerleri, 2007; aktaran Tamjid ve diğerleri, 2024, s. 231)

Çalışmamız, literatürde başka teknik ve alanlara uygulanan ontoloji tabanlı uygulamaları esas alarak, ontoloji tabanlı metin madenciliği yönteminin EBYS’lerde büyük verim sağlayabileceğini ortaya koymaktadır. Çalışmamızda kurulmuş olan mimari, metinsel belgeleri en iyi şekilde işlemeyi, tanımlamayı ve erişmeyi amaçlamaktadır. Bu noktada, kullanıcılar için bilgiye erişmede zor ve verimsiz sonuçlar ortaya koyan kelime tabanlı arama motoru yerine ontoloji tabanlı arama motoru önerilmektedir. Ontoloji

tabanlı arama motorunun metin madenciliği tekniklerine dayanan yapısı sayesinde ise, veri işleyenlerin hata ve risk payı engellenebilmektedir. Li ve diğerleri (2011), bilgi yönetim süreçlerinde, metin madenciliği tekniklerinin, manuel olarak gerçekleştirilen hataları azaltmaya yardımcı olduğunu belirtmektedir (Hanke ve Thiesse, 2017, s. 2993). Çalışmamızda, ontoloji tabanlı metin madenciliği ile kurulan mimari; verileri otomatik olarak daha iyi tanımlayarak, daha iyi sınıflandırarak, arama motorları sonuçlarını iyileştirerek organizasyonlarda bir standart sağlamayı amaçlamakta ve karar alma süreçlerinde önemli verim sağlayacak bilgi erişim sürecini geliştirmektedir.

IV. 1. 2. İş Zekâsı: Metin Madenciliği ile Bilgi Keşfi ve Öngörü Sağlama

IV. 1. 2. 1. Sorun: Literatür incelendiğinde bilgi yönetiminde metin madenciliği kullanımı ile ilgili sınırlı bilgiye ulaşılabilmektedir. Kaynaklar bilgi yönetiminin rekabet avantajı oluşturmadaki önemine değinmiş olsa da, EBYSlerin değerlendirildiği hiçbir çalışmada; EBYSlerin bilgi keşfi, ilişki çıkarımı gibi özelliklere sahip olduğu belirtilmemiştir. Buna ilaveten, organizasyonların karar destek sistemleri için gereken ve bilgi yönetiminin bir parçası olan öngörünün oluşturulmasında metin madenciliğinin uygulanması ve yararları tam olarak araştırılamamıştır (Khan ve Vorley, 2017, s. 4). Arıcı ve Kandur (2016, s. 65) EBYS kullanılırken, EBYSlerin karar alma süreçlerinde kullanılabileceğinin göz ardı edildiğini ve EBYSlerde karar alma uygulamaları ile ilgili araştırmalara gereksinim duyulduğunu belirtmektedir. Genel olarak bu tarz uygulamalarda, ihtiyaç duyulan verileri bir araya getiren raporlama özelliği bulunmaktadır. Ancak günümüz teknolojisi, büyük veri çağı ve rekabet üstünlüğü sağlama gereksinimi gibi konular göz önüne alındığında raporlama özelliğinin yetersiz kaldığı görülmektedir. Kuruluşlar, strateji geliştirme, performans değerlendirme, faaliyet raporları hazırlama gibi süreçleri gerçekleştirirken, birçok bilgiye ihtiyaç duymaktadır (Arıcı ve Kandur, 2016, s. 68). Raporlama özelliği ancak, işlenmesi ve keşfedilmesi

gereken bilgilerin elde edilmesinde, veri sağlayarak, bir ön aşamayı oluşturabilmektedir. Bunların yanısıra, büyük verilerde metin madenciliği uygulama ve teknikleri ile ilgili de bir anlayış eksikliği söz konusudur. Bu da metin madenciliğinin, bilgi yönetimi dahil olmak üzere başka alanlarda da; veri yakalama, işleme ve analiz etme gibi uygulamaların eksikliğine neden olmaktadır (Khan ve Vorley, 2017, s. 9).

Umut ve Külcü (2014, s. 112) tarafından değerlendirilen, Kalkınma Bakanlığı bünyesindeki elektronik bilgi yönetim sistemlerinde, iletişim süresince çok fazla örtük bilginin kullanıldığı ortaya çıkmıştır. İş ve işlem süreçlerinde bu bilgilerin açık bir şekilde ortaya konması, süreçlerin doğru ve verimli geçmesi için elzemdir. Etkili, belli standartlara bağlı ve sürdürülebilir belge yönetim süreçleri oldukça önem arz etmektedir. Bilgi yönetim sürecinin personele değil, sisteme bağlı olması gerekmektedir. Personele bağlı bir süreç, sürekliliğin sağlanması konusuna zarar verecektir. Bunun içinde sistemlerin analitik teknikler kullanılarak otomatikleştirilmesi gerekmektedir. Bilgi yönetiminin yapılandırılmamış yapısı, işlenmeye gereksinim duymaktadır. Elde edilen kritik ve öngörülü bilgilerin sürdürülebilirlik için organizasyonda tutulması gerekmektedir. Ancak organizasyonlar bunu yapmamakta ve bilginin kaybolmasına izin vermektedir (Svård, 2017, s. 52). Bir organizasyonun başarısızlığının temel nedenleri; organizasyonun etkili belge yönetim süreci gerçekleştirmemesi, belgelerine erişememesi ve yorumlayamaması, eksik pazarlama araştırmaları ve zayıf envanter yönetimidir (Mohd and Chell, 2005; Zheng and Khalid, 2022; aktaran Mukred ve diğerleri, 2022). Chen (2001, s. 5) çalışmasında, bilginin denetimi ve analizi, çözümün tanımlanabilmesi, temel bilgi yönetimi süreçlerinin uygulanabilmesi, bilgilerin insanlar ile ilişkilendirilmesi ve pazar analizinin sağlanması gibi süreçlerin, bilgi yönetimi teknolojileri ile otomatikleştirilmediği sürece sorunlu olacağını belirtmektedir. Literatür incelendiğinde EBYSlerde bilgi yönetimi süreçlerinin genel anlamda manuel olarak, veri işleyen

personeller tarafından gerçekleştirildiği gözükmemektedir. Bu durum insan kaynaklı hatalara neden olmaktadır. Bu hatalar ise süreçte aksaklıklara yol açıp, kuruluştaki standart ve sürekliliğinin sağlanması konularına tehdit oluşturmaktadır. Khan ve Vorley (2017, s.9) bu görüşü destekler şekilde, verinin sürekli olarak artmasına dikkat çekerek, organizasyonların bu veri yığınları içerisinde etkili ve öngörülü bilgi elde edebilmeleri için karar vermeyi destekleyecek araçlara ihtiyaç duyduğunu böylece manuel süreçlere kıyasla zamanında, daha bilinçli ve etkili süreçler geliştirilebileceğini belirtmektedir. Büyük bilgi teknolojisi şirketleri (Oracle, Microsoft gibi) başarılı geçmişlerine rağmen, bilgi haritaları oluşturmak için önemli madencilik yeteneklerinden yoksundurlar. Bu nedenle, az bilgi analizine sahip bu şirketlerin, Bilgi Yönetimi satıcıları yerine, enformasyon yönetimi satıcıları olarak anılması daha yerindedir (Chen, 2001, s. 8). Bilgi haritaları, organizasyonların örtük bilgilerini anlamlandırarak, gizli kalmış ilişki ve kalıpları ortaya çıkarmaktadır. Böylece kuruluşlar, bilginin etkili temsilini sağlayarak, karar alma süreçlerini geliştirmekte, bilgi ve rekabet üstünlüğü elde etmektedir. Ancak sürekli artış gösteren bu veri yığınları içerisinde, bilgi haritalama büyük bir zorluk yaratmaktadır. Bilgi değişikliklerinin hızlı ve kolayca tespiti için ise akıllı bir yaklaşım gerekmektedir (Al Hakim ve diğerleri, 2020, s. 33). Metin madenciliği bu yaklaşımla, organizasyonlara kaliteli ve zengin bilgi üretimini sağlamaktadır. Bilgi tabanlı ekonomide oldukça önemli olan kaliteli ve zengin bilgi üretimi, kuruluşlarda geniş etkiye sahip olup, önemli rekabet avantajları sağlamaktadır (Khan ve Vorley, 2017, s. 19).

IV. 1. 2. 2. Çözüm: Kuruluşlarda etkili bilgi yönetimi, örtük ve açık bilginin en iyi şekilde kullanılmasına dayanmaktadır (Khan ve Vorley, 2017, s. 7). Bir EBYS, iş süreçlerini; organize, verimli, açık ve etkili bir şekilde yönetmelidir. Bunun için, bilgilere en verimli şekilde erişilebilmeli, bilginin değeri en üst düzeye çıkarılmalı ve gereksiz bilgiler ayıklanmalıdır. Bu mimaride kurulmuş bir EBYS, organizasyonun karar alma süreçlerini

geliştirerek, üretkenlik ve etkililiğini arttırmaktadır (Mukred ve diğerleri, 2022). Metin madenciliği tekniği ile EBYSlerde bilgiye dayanan iş süreçlerinin gerçekleştirilmesi, verilerden bilgi derinliği üreterek organizasyonların pazarda kendilerini rakiplerinden sıyırmasını sağlamaktadır (Khan ve Vorley, 2017, s. 8). Bir metin analizi aşaması, belge deposu oluşturma, meta veri deposu oluşturma, özellik çıkarma, kümeleme ve sınıflama, arama motorunun geliştirilmesi ve analizi sağlayacak metin madenciliği araçlarının (birliktelik kuralı keşfi, ilişki çıkarma, trend analizi vb.) kullanılması süreçlerini içermektedir (Li ve Brook, 2006, s. 1413). Bu başlıkla söz konusu süreçlerin EBYSlere entegre edildiğinde, kuruluşlarda etkili karar alma ve iş zekası sağlama yeteneklerinin nasıl artacağı ortaya konulmaktadır.

İş süreçlerinin analizi sayesinde organizasyonların, girdi, çıktı ve müşteri değerleri ortaya çıkarılmaktadır. Buda iş süreçlerinin bütün bir şekilde iyileştirilmesini sağlamaktadır (Svård, 2017, s. 29). Organizasyon ile çevresindeki aktörler işbirliği sonucu bir ürün yaratmaktadır. Bu ürün ise inovasyondur. İnovasyon süreci, fikirlerin üretimi ile başlayıp bir değer yaratılması süresince gerçekleştirilen faaliyetleri içermektedir (Schieber ve Kruse, 2014, s. 3). Bu süreçte, EBYSlerin iyi bir şekilde iş zekası sağlayabilmesi için, dahili bilgi kaynakların yanında harici bilgi kaynaklarını da işleyebilmesi ve analiz edebilmesi gerekmektedir. Tek bir kaynaktan elde edilen veriler kısıtlı ve yetersiz iş zekası sağlamaktadır. İç ve dış bilgi kaynaklarının bütünsel analizi daha verimli sonuçlar elde etmektedir (Li ve Brook, 2006, s. 1412). Örneğin dermesini geliştirmek isteyen bir kütüphaneci, kullanıcıların ödünç alma verilerinin yanında, benzer türde hizmet veren diğer kütüphanelerin sosyal medya hesaplarını ve çevrimiçi kataloglarını analiz ederse, derme koleksiyonunu geliştirmek için daha doğru kaynakları kütüphanesine sağlayacaktır.

Ürün ve hizmetlerin geliştirilmesi noktasında müşteri/kullanıcı bilgisi oldukça önem arz etmektedir. Pazarlama stratejisi direkt olarak müşteri bilgi yönetimi ile belirlenmektedir. Dolayısı ile artık müşteriler/kullanıcılar da fikir tedarikçisi olarak süreçte rol oynamaktadır (Schieber ve Kruse, 2014, s. 3). Fakat günümüzde çok az organizasyon müşteri bilgilerini iyi yönetebilmektedir (Li ve Yanchinda, 2024). Günümüzde neredeyse her organizasyonun ürün ve hizmetleri ile müşteri/kullanıcı etkileşimlerine dair bilgi, web üzerinden elde edilebilmektedir. Organizasyonların müşteri/kullanıcılar ile ilgili bu dış bilgi kaynaklarını EBYS süreçlerine dahil etmesi ve işlemesi karar almada büyük fayda sağlamaktadır. Müşteri yorumlarından yakalanan bilgilerin analizi ile, müşteri davranışlarının nedenleri, ihtiyaç ve tercihleri tespit edilebilmekte ve organizasyonlarda yanlış yürütülen veya geliştirilmesi gereken iş süreçleri saptanabilmektedir (Li ve Yanchinda, 2024)

Çeşitli platformlardaki müşteri yorumlarının analizi, karar alma süreci için çok önemli olsa da, sürekli artan, yönetilmesi güç hatta imkansız, veri yığınları arasında, değerli bilgileri gereksiz olanlardan ayıklamak zor bir süreçtir (Schieber ve Kruse, 2014, s. 3). Bu verileri en iyi şekilde yöneterek bilgiye değer katan ve ticarileştiren organizasyonlar, rakiplerine karşı büyük güç elde etmektedir. Metin madenciliği ile çıkarılan ilişki ve bilgiler, müşteri/kullanıcı odaklı hizmet ve ürünlerin geliştirilmesine olanak sağlamaktadır. Malik ve Gupta, (2022, s. 28) müşteri/kullanıcı beklentisinin tespitinde kullanılan metin madenciliğinin, organizasyonun niteliğinde fark yarattığını belirtmektedir. Bir metin madenciliği tekniği olan kümeleme, yalnızca kurumdaki bilgi varlıklarını gruplandırmakla kalmamakta, aynı zamanda karar mekanizmaları için büyük önem sağlayan içgörüler de elde etmektedir (Khan ve Vorley, 2017, s. 20). Müşterilerin yorumlarından elde edilen veriler, benzerliklerine göre kümelenerek, niteleyici terimler ile özetlenebilmektedir. Burada sınıflama yerine kümele yönteminin kullanılması,

önceden tanımlanmamış yeni fikirlerin keşfedilmesini sağlayabilmektedir. Kümelerin teker teker analiz edilmesi, karar alma süreçlerinde önemli içgörüler sağlamaktadır. Kümelemenin yanı sıra, yine müşteri yorumları analizi ile, kelimeler arasında ilişkiler çıkararak, metinlerin niyet, konu ve duygularına göre sınıflandırılmasını sağlanabilmektedir (Malik ve Gupta, 2022, s. 31). Elde edilen bilgiler ile, hedef pazarın analizi büyük oranda kolaylaşabilmektedir. Potansiyel ve mevcut müşterileri elde tutmak için gerekli olan bu bilgiler (ürünler/hizmetler hakkındaki düşünce ve şikayetler) organizasyona bir fikirler listesi sağlamaktadır. Bu liste organizasyonun ar-ge departmanında analiz edilip değerlendirildikten sonra bulguların aktarılması ile üst düzey inovasyon süreci gerçekleştirilmektedir (Schieber ve Kruse, 2014, S. 24).

Kümeleme ve sınıflama gibi tekniklerin yanında birliktelik analizine dayalı bir metin madenciliği tekniği de, karar alma süreçleri için büyük faydalara sahiptir. Bu teknik kullanılarak, büyük veri yığınları içerisindeki nedensellik ilişkileri ve örtük ilişkiler ortaya konulabilmektedir (Vasques ve diğerleri, 2018, s. 493). Bilgilerin keşfedilmesinde, birliktelik analizinde önemli yeri olan ilişki kuralları çıkarımı ön plana çıkmaktadır. İlişki kuralları, organizasyonun dahili veya harici bilgi kaynaklarında yer alan bir yazışma, rapor gibi metinsel veriler üzerinden, kelimeler arası ilişkiler analiz edilerek, çıkarılabilmektedir. İlişki kuralları çıkarımına göre $A \rightarrow B$ olmaktadır. Bu, bir A terimi geçtiğinde B teriminin de geçme olasılığını ifade etmektedir. (Vasques ve diğerleri, 2018, s. 493)

Örneğin, bir organizasyon, müşteri yorumlarının yer aldığı bir rapora ilişki kuralları çıkarımını uygulamak istiyor olsun. Bunun için ilk olarak rapora önışlem süreci gerçekleştirilmelidir. Önışlemden geçirilen belgelere n-gram gibi teknikler uygulanarak anlamlı sözcük grupları elde edilebilmektedir. Ardından ilişki kuralları çıkarılmaktadır. Bunun için öğeler (kelimeler) arasındaki ilişkileri keşfetmek ve kurallar oluşturmak için

Apriori algoritması gibi klasik teknikler uygulanabilmektedir (GeeksforGeeks, 2025b). Böylece işlem sonucunda, Kargo → Gecikme, Teslimat → Gecikme vb. gibi kurallar elde edilebilmektedir. Çıkarılan kurallar sayesinde organizasyon, kargo şirketini değiştirmek gibi önlemler alarak, müşteri memnuniyetini sağlayabilmektedir.

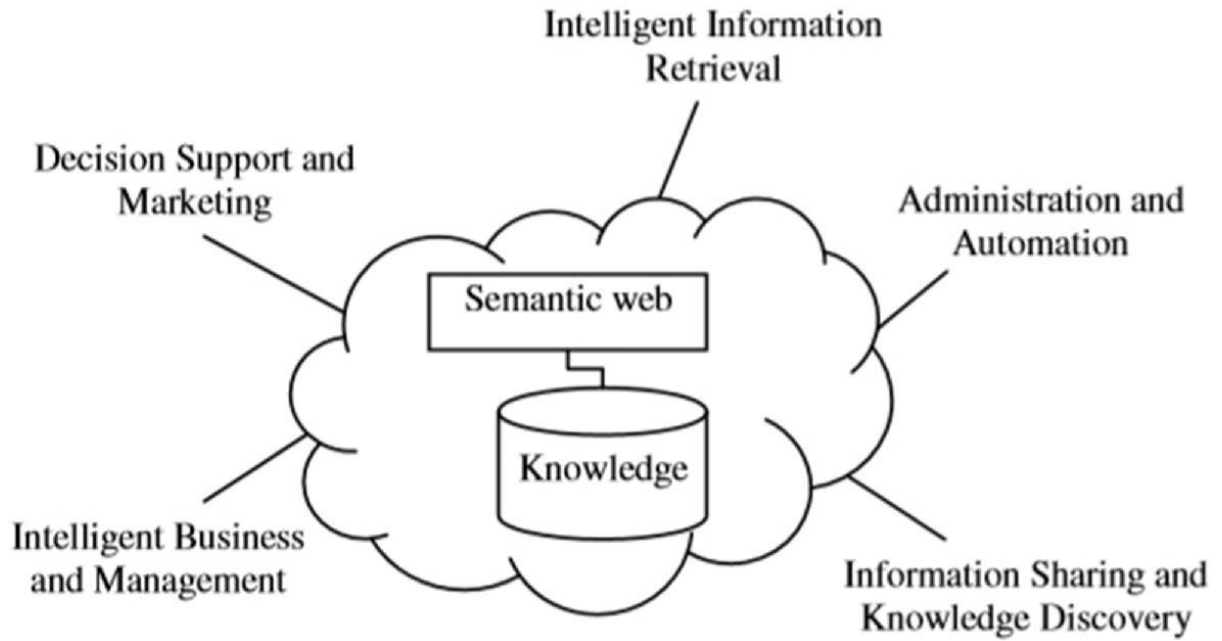
Birliktelik kuralı keşfi, bir metindeki nesneler arasındaki ilişkileri bulmak için etkili bir yöntemdir. Keşfedilen ilişkiler sayesinde kurumsal ontoloji geliştirmek mümkündür. Keşfedilen kurallar, kurumun ontolojisine aktarılarak, bilginin temsili ve ifade gücü zenginleştirilebilmektedir. Böylece organizasyon karar alma süreçlerinde işlerine yarayacak kaliteli bilgi üretimini arttırabilmektedir (Vasques ve diğerleri, 2018, s. 491). Bir terimin bağlamının, diğer terimler ile ilişkilerinin (dar, geniş, eşanlamlı, ilişkili gibi) bilgi temsili dolayısı ile de bilgi erişimi önemli derecede geliştirdiğinden daha önce söz edilmektedir. Bu süreç aslında organizasyonda karar almanın da temelini oluşturmaktadır. X üniversitesinde öğrenci değişim programları ile ilgili bir analiz yapılmak istendiğini varsayalım. Kurulan ilişkiler sayesinde sistem, içerisinde “öğrenci değişim programı” geçen terimlerin yanında diğer ilgili metinsel verilerin (Erasmus, Mevlana, Farabi vb.) analizini de sağlayacaktır. Bu sayede ilgili daha çok belgeye analiz yapılarak daha etkin bir karar verme süreci gerçekleşecektir.

Anlamsal web’in yapı taşı olan ontoloji, anlamsal web içeriğini oluşturmada önemli bir unsurdur (Lee, 2020, s. 221). Kuruluşlara karar vermede yardımcı olacak; ilişki çıkarma, varlık çıkarma, duygu analizi gibi metin madenciliği teknikleri, anlamsal web ile geliştirilirse ilişkilerden çıkarılan bilgilerde bağlam daha iyi ele alınabilmektedir. Anlamsal web, önemli akıl yürütme yeteneklerine sahiptir. Bu özelliği ile, geleneksel iş uygulamaları ile karşılaştırıldığında, statik veriler sağlamak yerine (ürün bilgisi, müşteri bilgisi, işlem kayırları vb.) potansiyel müşterilerin satış departmanları ile ilişkilendirilmesini sağlamaktadır (Lee, 2020, s. 222). Böylece daha doğru ve daha etkili

karar verme süreçleri gerçekleştirilebilmektedir. Örnek vermek gerekirse daha önce bilgi varlığı çıkarmak için Python ile bir örnek kod oluşturmuştuk. Metin içerisinde çıkarılan kişi, kurum, zaman gibi varlıklar, anlamsal web sayesinde, metinde geçen eylem ve nesnelerle ilişkilendirilebilmektedir. Kurulan bu ilişkiler, ürün analizi, müşteri analizi, duygu analizi, otomatik raporlama ve soru-cevap sistemleri oluşturmada kullanılarak organizasyonların etkili ve doğru karar almalarını sağlayabileceği öngörülmektedir.

Şekil 21.

Semantik web ve bilgi



(Sheth 2013; aktaran Lee, 2020, s. 221)

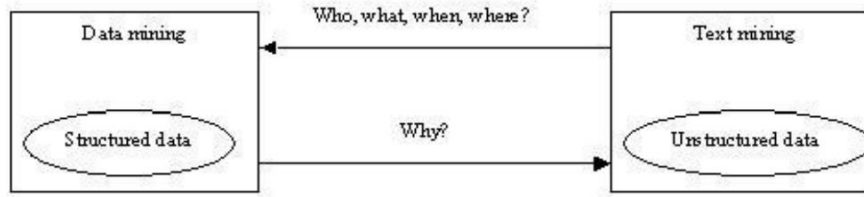
Metin madenciliği teknikleri işlenirken bahsedilen, trend analizi de organizasyonların öngörülü bilgi elde etmesi için önemli bir tekniktir. Metin madenciliği sayesinde, kümeleme ve sınıflama gibi yöntemler ile birbirleri ile ilişkilendirilmiş belgelerin trendleri/eğilimleri analiz edilebilmektedir. Bu yöntem ile konu alanındaki dönemsel değişiklikler tespit edilerek fırsat, tehdit ve riskler hakkında etkili kararlar alınabilmektedir. Örneğin, bir x üniversitesinin dilekçelerine trend analizi uygulansın.

Bunun için ilk olarak dilekçeler önışlemden geçirilip TF-IDF ile sayısallaştırılır. Sayısallaştırılan belgeler ontoloji tabanlı metin madenciliğı tekniğı ile sınıflanır veya kümelenir. Sınıfların veya kümelerin her biri bir zaman çizelgesinde (veya farklı görselleştirme yöntemleri ile) gösterime sunulur. Değışiklik gösteren sınıflar/kümeler ve bu sınıfların/kümelerin temaları tespit edilir. Elde edilen sonuçlara göre üniversite bir aksiyon alır. Trend analizi ile benzerlik gösteren olay keşfi (Event Discovery) tekniğı ise, bir organizasyondaki yeni olayları keşfetmektedir. Organizasyonlar, rakiplerinin reklam ve ürün lansmanlarını analiz ederek, terim sıklık dağılımları veya kelime dağarcığındaki bir kayma ile, rakiplerinin strateji değışiklikleri hakkında bilgi sahibi olabilmektedir (Li ve Brook, 2006, s. 1413).

Şu ana kadar bahsedilen tüm analiz işlemleri ile yeni dijital teknolojilerin bir arada kullanımı daha etkili öngörülerin elde edilmesi açısından iş zekasını arttırmaktadır. Nitekim tekniklerin kullandığı yöntem ve algoritmalar incelendiğinde de her birinin birbirlerinden etkilendiğı ortaya çıkmaktadır. Bu durum göz önüne alındığında metinsel verilerin analizini sağlayan metin madenciliğı tekniklerinin yanına (ilişki çıkarma, trend analizi vb.), yapılandırılmış sayısal verilerinde analizi de eklendiğinde (veri madenciliğı) daha ayrıntılı ve etkili keşifler elde edilebilmektedir. Bu iki yöntemin beraber kullanıldığı sistemlere ise “bilgi madenciliğı” adı verilmektedir (Li ve Brook, 2006, s. 1411).

Şekil 22.

İş zekâsı için veri ve metin madenciliğinin gücü



(Li ve Brook, 2006, s. 1411)

Bilgi yönetimini iyileştirmeye yönelik önemli bir uygulama olan verilerin görselleştirilmesi (Khan ve Vorley, 2017, s. 1), bilgilerin analizi ve anlamlandırılmasında kuruluşlara kolaylık sağlamaktadır. Sınıflama, kümeleme, ilişki çıkarma, birliktelik analizi, trend analizi gibi metin madenciliği tekniklerinden elde edilen bilgilerin görselleştirilmesi, bilginin daha açık, anlaşılır ve keşfedilir olmasını sağlamaktadır. Moon ve arkadaşları (2018, s. 4792) çalışmalarında, inşaat sektöründe artan veri miktarını yönetmek için metin madenciliği tabanlı bir belge yönetim sistemi geliştirmektedirler. Geliştirdikleri sistem (UNİ (User Needed Information)-Tacit) ile büyük veri yığınlarından yararlı bilgiler çıkararak iş stratejileri geliştirmeyi amaçlamaktadırlar. Bu sistem; güncel belgelere erişim için otomatik veri toplama, toplanan verilere metin madenciliği uygulayarak anahtar kelimeler çıkarma, son olarak ise örtük bilginin ortaya çıkarmak için kelime bulutları oluşturma aşamalarını içermektedir. Kelime bulutu ile bilgiler görsel şekilde özetlenerek, anahtar kelimelerin pratik bir gösterimi elde edilebilmektedir. Bilgi yönetimi sürecinde örtük bilginin anlamlandırılmasını sağlayan bir diğer görselleştirme yöntemi bilgi haritalarıdır. Bilgi haritaları kurduğu ilişkiler ile, kelime bulutlarından çok daha kapsamlı bir yapı sunmaktadır. Bilgi haritaları; karar alma, eğitim ve bilgi erişim için önemli faydalar sunmaktadır. Böylece organizasyonlarda, çeşitli faktörle ilgili

olasılıkları ilişkilendirerek karar verme sürecini iyileştirmektedir (Lin ve Hsueh, 2006, s. 552). Bunun yanı sıra, kuruluşlarda karmaşık veya gizli kalmış bilgilerin açığa çıkarılmasında temel bir işleve sahiptir. Bu yöntem ile, yapılandırılmamış metinsel veriler, anlamlandırılarak, açık, anlaşılır, erişilebilir bilgilere dönüştürülmektedir. Bilgi haritalarının bir diğer özelliği, nesnelerin konumu ve diğer nesneler arasındaki ilişkilerin açık bir gösterimini sağlamasıdır, bu yolla kullanıcıların bilgiye erişme ve anlama hızı önemli derecede azalmaktadır (Tsai ve Ji, 2018, s. 382). Bilgi haritası oluşturmak güç bir iştir bununla birlikte bu haritaları oluşturacak uzmanları bulmakta zor ve maliyetlidir (Lee ve Segev, 2012, s. 353). Bu noktada metin madenciliği otomatik bir bilgi haritası oluşturmada etkili bir tekniktir. Otomatik olarak metin madenciliği teknikleri ile oluşturulmuş bilgi haritaları, manuel tekniklere göre daha kaliteli ve doğru sonuçlar çıkarmaktadır (Chen, 2001, s. 27).

Kısaca özetlemek gerekirse; bilgi haritaları bilginin anlaşılabilirliğini sağlamada önemli bir materyaldir. Bilgi haritaları sayesinde organizasyonlar, iç ve dış bilgi kaynaklarını daha verimli şekilde temsil edebilmektedir. Bu haritalar, önemli terimler ve bu terimler arasındaki ilişkileri açık bir şekilde ortaya koymaktadır. Böylece kullanıcılar tarafından anlaşılması zor ve uzun metinleri hızlı ve kolay bir şekilde öğrenilebilmektedir.

Lee ve Segev (2012) çalışmalarında, öğrenme materyali oluşturma amacı ile metin madenciliği tekniklerini kullanarak bir bilgi haritası oluşturmuş ve oluşturulan bilgi haritasının belge öğrenimi ile karşılaştırıldığında bilgi haritalarının çok daha önemli fikirleri tanımladığını kanıtlamışlardır.

Chen ve arkadaşları (2008) makalelerden otomatik kavram haritası oluşturdukları çalışmalarında, anahtar kelimeler arasındaki ilişkileri dört varsayım ile tanımlamaktadırlar: (Lee ve Segev, 2012, s. 354)

- 1- Belgede listelenen her anahtar kelime, önemli bir terimi temsil etmektedir.
- 2- Belgede iki anahtar kelimenin yer alması, aralarında bir ilişki olduğunu ortaya koymaktadır.
- 3- İki anahtar kelimenin bir cümlede geçme sıklığı ne kadar fazlaysa, aralarındaki ilişki de o derecede fazladır.
- 4- Bir cümlede iki anahtar kelimenin arasındaki mesafe ne kadar az ise aralarındaki ilişki o kadar fazladır.

Lee ve Segev, (2012, s. 358) çalışmalarında bu varsayımların ilk üçünü esas alan bir bilgi haritası oluşturmaktadırlar. Uygulamalarında bir bilgi haritası oluşturma sürecini şu şekilde gerçekleştirmektedirler:

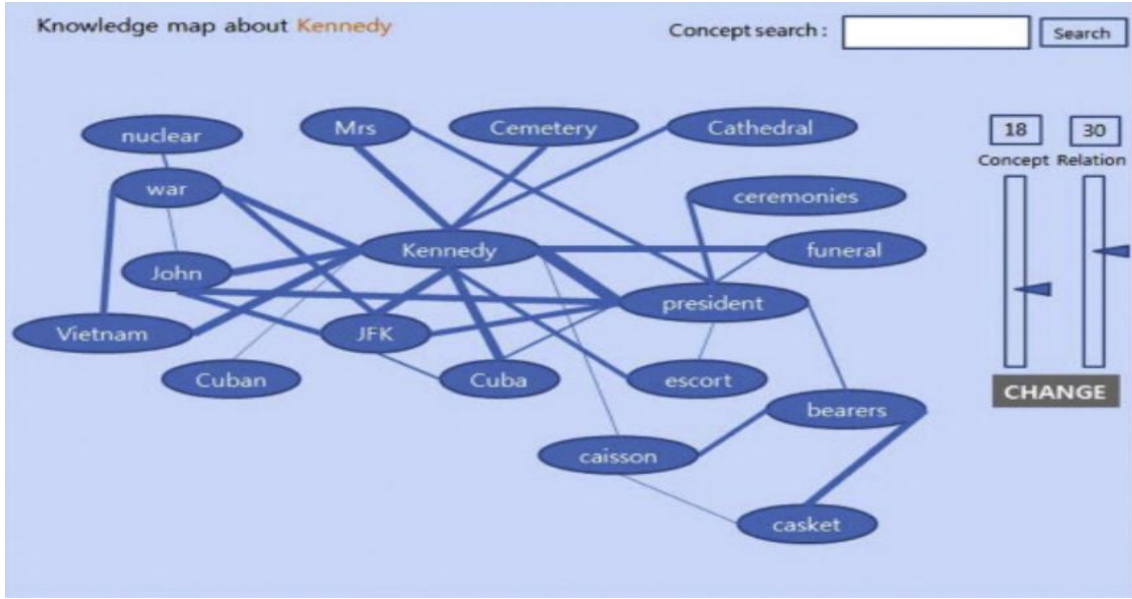
İlk olarak belgeler önişlem sürecinden geçirilir. Önişlem sürecinden geçilen belgelere TF-IDF uygulanarak terimler ağırlıklandırılır ve önemli anahtar kelimeler çıkarılır. Ağırlıklandırma her belge için ayrı ayrı gerçekleştirilmektedir. Bu nedenle aynı kelime farklı dokümanlarda farklı ağırlıklara sahip olabilmektedir. Ağırlıklandırma aşamasından sonra ilişkiler çıkarılmaktadır. Bunun için iki anahtar kelimenin bir cümlede beraber geçme sıklığı ve iki anahtar kelimenin cümledeki önemini gösteren ağırlık puanı hesaplanmaktadır. Bir cümledeki kelime sayısının artması her kelimenin ağırlığını azaltan bir faktördür, dolayısı ile az kelimedenden oluşan bir cümledeki ilişki puanı, çok kelimedenden oluşan ilişki puanından daha fazladır (Lee ve Segev, 2012, s. 354). Bu işlem sonucunda ise ağırlık puanları, ilişkileri etiketlemek için kullanılmaktadır. Yüksek ağırlığa sahip kelimeler kalın çizgiler ile düşük ağırlığa sahip çizgiler ise ince çizgiler ile kelimelere bağlanmaktadır. Ancak ilişkilere tıklandığında kullanıcının karşılaşacağı çok fazla cümle ortaya çıkabilmektedir. Bu sorunu çözmek için ise sonuçları daraltacak bir yöntem uygulanmaktadır. İki anahtar kelime arasındaki ilişkileri ortaya koyan yeni anahtar

kelimeler çıkarılmaktadır. Çıkarılan anahtar kelimeler, iki anahtar kelimenin geçtiği cümleleri kümelemektedir. Böylece kullanıcı aradığı bilgi ile ilgili daha kesin ve spesifik sonuçlar elde edebilmektedir.

İlgili sürecin uygulaması aşağıdaki gibidir:

Şekil 23.

John F. Kennedy'nin bilgi haritası

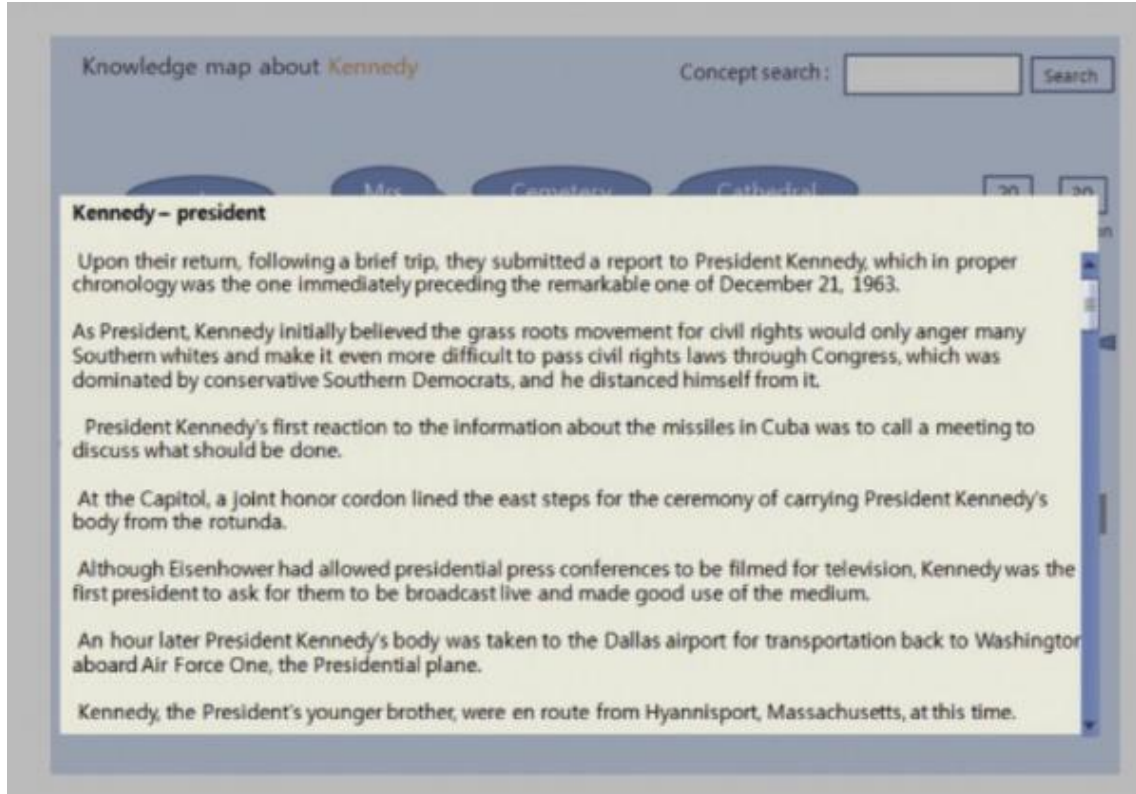


(Lee ve Segev, 2012, s. 355)

Oluşturdukları harita ile bir konu hakkındaki, anahtar kelimeler ve aralarındaki ilişkiler birbirlerine bağlanmaktadır. Haritaya göre yüksek ilişkiye sahip bağlantılar daha kalın, az ilişkiye sahip bağlantılar daha ince şekilde gösterilmektedir. Yukarıda John F. Kennedy hakkında oluşturmuş oldukları bir bilgi haritası örneği yer almaktadır (Lee ve Segev, 2012, s. 355).

Şekil 24.

'Kennedy' ve 'president' kelimelerini içeren cümleler

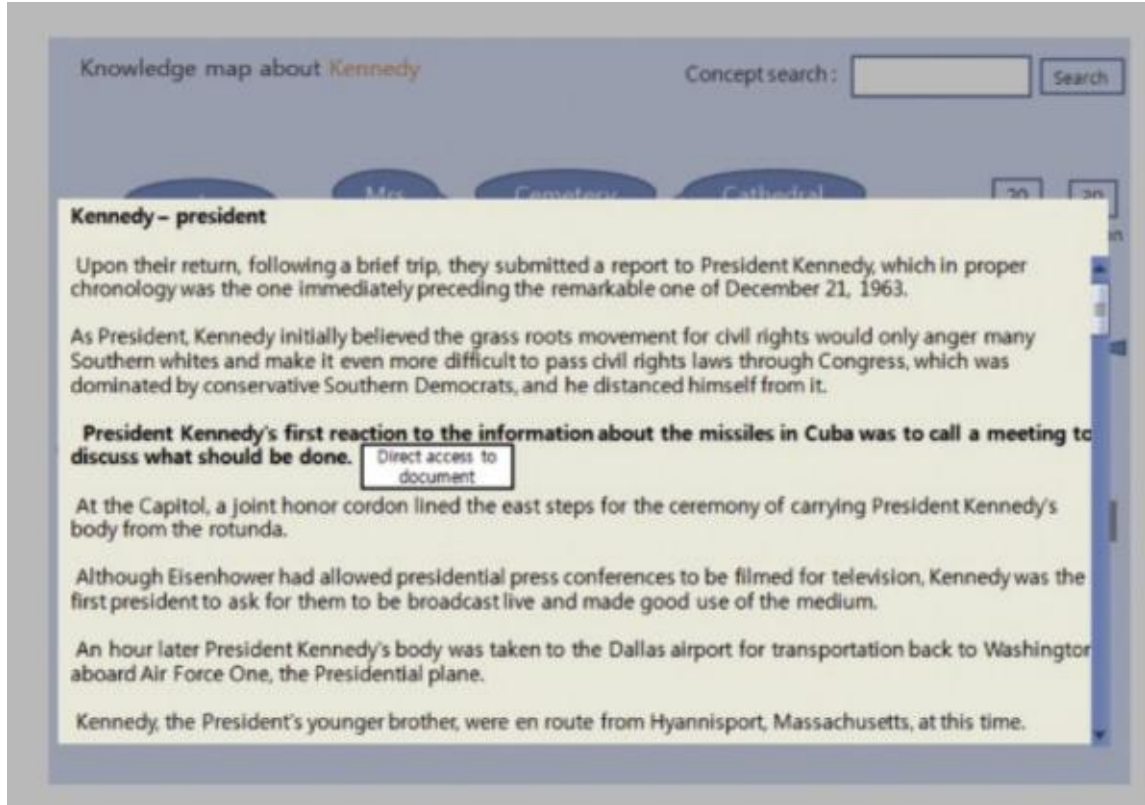


(Lee ve Segev, 2012, s. 356)

Haritada, kelimeler arasındaki bağlantılara tıklandığında, bağlantının iki ucunda yer alan anahtar kelimelerin geçmiş olduğu cümleler sıralanmaktadır. Cümlelerden herhangi birine tıklandığında ise, cümlenin yer aldığı belgenin tam metni görüntülenmektedir (Lee ve Segev, 2012, s. 356).

Şekil 25.

Bir cümle aracılığıyla bir belgeye doğrudan erişim



(Lee ve Segev, 2012, s. 356)

Bunun yanı sıra bilgi haritaları ile arama daha da daraltılabilmektedir. Örneğin kurulan sistem ilişkinin iki ucundaki anahtar kelimelerin arasındaki ilişkiyi gösteren farklı anahtar kelimeler de sıralamaktadır. Seçilen anahtar kelimelerden birisine tıklandığında ise, üç anahtar kelimenin de beraber geçtiği cümleler sıralanmaktadır (Lee ve Segev, 2012, s. 355). Bilgi haritalarının sağlamış olduğu bu yapı, mevcut arama motorları ile kaçırılan, anlaşılmayan ve vakit alan sorunları çözebilmektedir (Lee ve Segev, 2012, s. 355).

Lee ve Segev'in e-öğrenme amacıyla gerçekleştirmiş ve önemli verimler elde etmiş olduğu bu örnek, EBYS ile entegre edilmesi durumunda; kuruluşlarda örtük bilgilerin açığa çıkarılmasında, bilgi keşfinde, bilgi erişiminde ve diğer önemli belge

yönetim süreçlerinde hızlı, etkili ve doğru kararların alınmasını sağlayarak iş zekasını geliştirecek bir yapı sağlayacağı açıktır.

Buraya kadar bahsedilen tüm örnekler göstermektedir ki, metin madenciliği teknikleri; kurumsal hafızanın geliştirilmesi, iş süreçlerinin iyileştirilmesi, etkili karar almanın sağlanması gibi iş zekasını etkileyen noktalarda sağlayacağı yenilik ve iyileştirmeler ile kuruluşlara büyük katkı sağlayacaktır.

IV. 1. 3. Güvenlik: Metin Madenciliği ile Siber Güvenlik ve Bilgi Güvenliği Uygulamaları

IV. 1. 3. 1. Sorun: Bir EBYS'nin güvenliğinin sağlanabilmesi, elektronik belgelerin bütünlük, gizlilik ve erişilebilirliğini doğrudan etkileyen bir faktördür. Bu nedenle organizasyonlarda tüm iş ve işlemlerin güvenceye alınması ve süreçlerin sürdürülebilirliğinin sağlanmasının temel şartı güvenlidir. Güvenliğin sağlanabilmesi için ise bazı önlemler alınması gerekmektedir. Bu önlemlerden en önemlisi yetkili erişimdir. Yetkili erişim ile kurumsal mahremiyet sağlanarak, yetkisi olmayanların, erişim hakkının bulunmadığı belgelere erişimi engellenmektedir. Bu da gizliliğin yalnızca kurumlararası değil kurum içerisinde de uygulanması gerektiğini göstermektedir.

Nitekim güvenlik açıklıklarının büyük bir bölümünü insan hatası yaratmaktadır (Önaçan ve diğerleri, 2012, s.16). 13298 Elektronik Belge ve Arşiv Yönetim Sistemi standardı incelendiğinde, EBYSlerde erişim haklarının düzenlenmesi ile ilgili önlemler, oldukça detaylı ve net şekilde ele alınmaktadır. Standart, erişim hakları ile ilgili şu konuların üzerinde önemle durmaktadır: EBYS'lerde erişim hakları tanımlanabilir olmalıdır (çok gizli, gizli, hizmete özel), her bir EBYS kullanıcısı için sistem bir fonksiyon tanımlamalıdır (yönetici, kullanıcı vb.), tanımlanan fonksiyonlara erişim haklarını düzenleyecek nitelikte roller atanmalıdır, tanımlanan erişim hakları kullanıcı

profillerine işlenmelidir, kullanıcılar arama motoru üzerinden yaptıkları taramalarda erişim haklarını sağlayan belge ve metadatalara erişim sağlayabilmelidir.

13298 standardında ve diğer ilgili mevzuat ve rehberlerde güvenlik ile ilgili alınabilecek önlemler detaylı bir şekilde yer almasına karşın organizasyonlarda yanlış uygulama ve eksiklere rastlanmaktadır. Umut ve Külçü'nün (2014, s. 111) Kalkınma Bakanlığı bünyesinde bulunan elektronik belge ve arşiv yönetim sistemlerini değerlendirmiş oldukları çalışmada; Bakanlığın KBTarama adlı sisteminde, sınırlamalar ve yetkili erişimin sağlanması konularında yetersizlik tespit edilmiştir. Benzer şekilde Menteş ve Solhan'da (2019, s. 21) çalışmalarında, idarelerin EBYSler üzerinden gerçekleştirdiği, gizlilik dereceleri uygulamalarının, yanlış bir şekilde gerçekleştirildiğini belirtmektedir. Güvenlik önlemleri kapsamında, kuruluşlarda gerçekleştirilecek bilinçli veya bilinçsiz her türlü yanlışlık veya dikkatsizlik, alınan önlemleri geçersiz kılmaktadır (Aydın ve Saydam, 2019, s. 4). Bu noktada bilgi ve belgelerin paylaşımında gizlilik derecelerinin belirlenmesi ve yetkili erişimin sağlanması, insan yönetimine bırakılmaması gereken bir işlemdir.

EBYS'lerde güvenliği etkili bir şekilde sağlayabilmek için risklerin en az seviyeye indirilmesi gerekmektedir (Aydın ve Saydam, 2019, s. 5). Gizlilik derecesine sahip bilgiler, karar alma süreçlerinde kritik ve stratejik öneme sahiptir. Bu nedenle bu bilgilerin, erişime hakkı bulunmayan kişilerin eline geçmesi, organizasyonlar için oldukça risklidir (Arıcı ve Kandur, 2016, s. 73). Bilgi güvenliğinde en kritik tehdit ise, kurum personeli tarafından sızdırılan verilerdir. Bu veriler yaygın olarak bilgilerin elektronik ortamda paylaşımı sırasında gerçekleşmektedir (Katz ve diğerleri, 2014, s.137). Dış bilgi kaynaklarının organizasyonlar tarafından elde edilmesi iç bilgi kaynaklarına kıyasla daha maliyetli ve zordur. İç bilgi kaynakları kurumsal istihbarat için önem taşımaktadır. Bu nedenle bu bilgiler yetkisiz bir kullanıcının veya rakiplerin eline

hiçbir şekilde geçmemelidir (Mukred ve diğerleri, 2022). İnsan faktörü güvenliğin sağlanmasında sorun yaratmaktadır, bu nedenle şifreleme çözümleri, bilgi güvenliğinin sağlanmasında yeterli değildir (Gonzalez-Compean ve diğerleri, 2019, s. 432).

İşin siber güvenlik kısmında ise, veri hacmi her geçen gün büyük bir artıştır. Bu durum, siber tehditlerin insan analizi yolu ile tespitini imkansızlaştırmaktadır (Ignaczak ve diğerleri, 2021, 140:28). Bir EBYS'nin, sisteme dair hata ve sorunlarını kayıt altına alır yapıda olması gerekmektedir (Eroğlu ve Külcü, 2014, s. 339). Hata sezme ve düzeltme işlemi (Aydın ve Özdemirci, 2011, s. 110), bir EBYSde güvenliğin sağlanması için önemli bir konudur.

Bu bilgilerin güvenliği için alınacak önlemlerde ilk aşama, bilgi varlıklarının tanımlanmasıdır (Aydın ve Saydam, 2019, s. 6). İkinci aşamada tanımlanan varlıkların sınıflandırılması gerekmektedir (Personel varlık sınıfı, ağ ve sistem varlık sınıfı gibi). Ardından varlık sınıflarına dair bir güvenlik derecesi atılmalıdır. Bu derecelerin atanması için bir anket sistemi geliştirilip, puanlandırılabilir (Bilgi ve İletişim Güvenliği Rehberi, 2020). Bu şekilde kurumdaki her varlığa bir risk değeri atanır ve daha etkili güvenlik önlemleri gerçekleştirilir.

Tüm bu süreçlerde, insan kaynaklı hataların önüne geçebilmek ve daha etkili güvenlik önlemleri alabilmek için, süreçlerin otomatikleştirilmesi ve iyi yapılandırılması gerektiğine inanmaktayız. Bu çalışmamızda güvenlik verimini arttırabilmek için madenciliği kullanımı önermekteyiz.

IV. 1. 3. 2. Çözüm: Günümüzde veriler çoğalarak artmaktadır, artan verilerin önemli bir kısmını yapılandırılmamış veriler oluşturmaktadır. Siber güvenlik sistemlerinde söz konusu yapılandırılmamış veriler önemli bir artış içerisindedir. Bu noktada

yapılandırılmamış metinsel veriler içerisindeki anormallikleri, tehditleri ve saldırıları tespit etmede metin madenciliği göz ardı edilemez bir öneme sahiptir (Sharma ve Arjunan, 2023, s. 1). O halde bu siber güvenlik ortamında metin madenciliği, güvenlik faaliyetlerine karşı, verimliliği arttıracak önemli bir yöntem olarak karşımıza çıkmaktadır. Şu ana kadar metin madenciliği ile tespiti sağlanan bazı tehdit örnekleri; siber zorbalık, bot ağları, yanlış bilgilendirme ve dolandırıcılıktır (Ignaczak ve diğerleri, 2021, s. 140:7).

Siber güvenlik ve bilgi güvenliği birbirleri ile karıştırılması muhtemel iki terimdir. 7545 sayılı Siber Güvenlik Kanunu siber güvenlik ile ilgili; siber uzayda kayıt altına alınan verilerin gizlilik, bütünlük ve erişilebilirliğinin sağlanmasının oldukça kritik olduğu; gizlilik, bütünlük ve erişilebilirliği tehdit edecek siber saldırılar karşısında uyarıcı tepki ve alarm mekanizmalarının harekete geçirilmesi gerektiğini belirtmektedir. Siber güvenliği ise, bir siber saldırı sonrasında durumun saldırı öncesindeki haline döndürülmesini sağlayan süreçler bütünü olarak tanımlamaktadır. Von Solms ve Van Niekerk (2013, s. 101) çalışmalarında, bilgi güvenliği ve siber güvenliğin ayrımını şu şekilde vermektedir: Bilgi güvenliği; bilgi varlıklarını ve bilgi özelliklerini tehditlerden korumayı içermektedir. Siber güvenlik ise yalnızca siber uzayın kendisini değil, siber uzayda faaliyet gösteren ve siber uzay ile erişilen varlıkların (ağlar, ağı kullanan insanlar, yazılımlar vb.) korunmasını da içermektedir. Bilgi güvenliğinde tehditler genel olarak insan faktörü ile ilişkilidir dolayısı ile güvenlik tedbirleri daha çok insanlara yöneliktir. Siber güvenlik ise bunun ötesinde insanların da dahil olduğu, siber uzay ile etkileşimli tüm varlıkların güvenliğini hedef almaktadır.

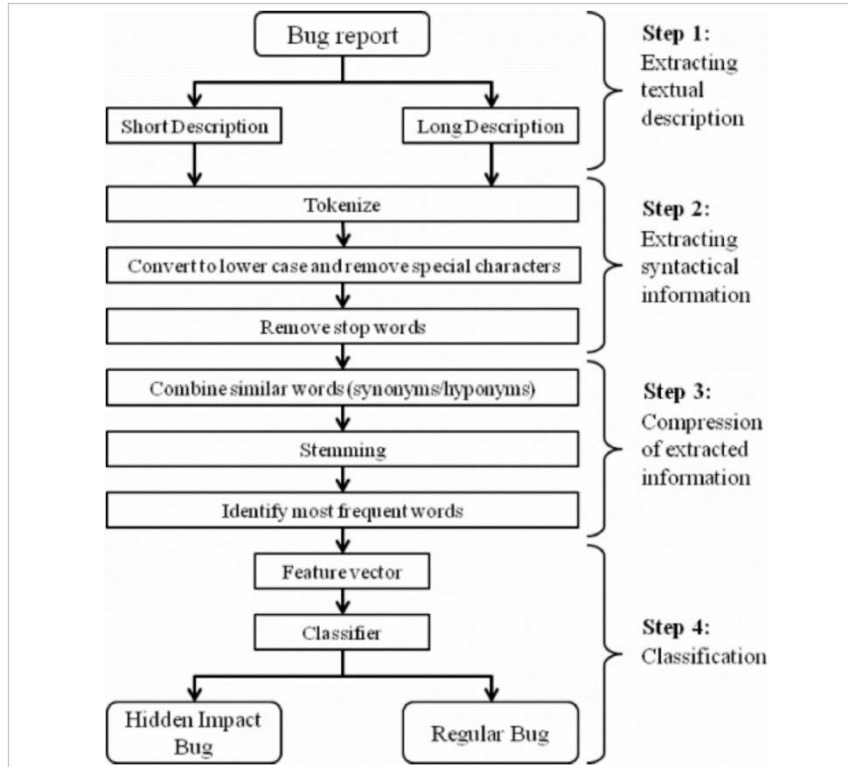
Siber suçluların güvenlik, erişilebilirlik ve bütünlüğün engellenmesine karşı kötü amaçları bulunmaktadır. Riggs ve arkadaşları (2023, s.2) bu amaçları şu şekilde sıralamaktadır: “...bilgi engelleme, siber güvenlik önlemlerini baltalama, karar alma sürecini geciktirme, kamu hizmetleri sağlamayı reddetme, kamu güveninin azalması,

kurban ülkenin itibarının zedelenmesi ve yasal bir çıkarın yok edilmesi...” Metin madenciliği, zafiyet yönetimini sağlayabilmek için, kaynak kodlarını analiz ederek güvenlik açıklarını tespit edebilmekte veya hata veritabanlarından bilgi çıkararak yeni zayıflıkları keşfedebilmektedir (Ignaczak ve diğerleri, 2021, s. 140:6). Hata veritabanları; yazılımcılar tarafından tanımlanmamış yazılım hatalarını tanımlamayı ve ileride karşılaşılabilecek hataları veritabanında aramayı sağlamaktadır (Wijayasekara ve diğerleri, 2014, s. 3613).

Wijayasekara ve arkadaşları, (2014, s. 3612) metin madenciliği ile hata veritabanlarından potansiyel ve mevcut güvenlik zafiyetlerini tespit ettikleri bir çalışma gerçekleştirmektedir. Linux üzerindeki güvenlik zafiyetlerini tespit ettikleri çalışmada, yalnızca metinsel verileri kullanarak %28 ila %88 oranında verim elde etmişlerdir. Kaynak olarak hata raporlarının metinsel açıklamalarından yararlanılan çalışmada, uyguladıkları yöntem üç aşamadan oluşmaktadır. İlk adımda metin madenciliği teknikleri ile metin içerisindeki sözdizimsel bilgiler çıkartılmaktadır. Ardından yine metin madenciliği ile metinleri önişlemeden geçirerek daha kolay işlenilmesi sağlanmaktadır. Son olarak ise özellik vektörü oluşturularak veriler sayısallaştırılmaktadır. Bu işlemler verilerin sınıflandırılması için önem arz etmektedir. Sistemin verileri sınıflandırması ile de kural tabanlı veya öğrenme yöntemleri kullanılarak, yazılımda bulunan açıklar, henüz tanımlanmadan tespit edebilmektedir.

Şekil 26

Hata raporlarını kullanarak güvenlik açıklarının belirlenmesi metodolojisi



(Wijayasekara ve diğerleri, 2014, s. 3613)

Hataların tespit edilmesinin dışında, metin madenciliği teknikleri ile gizli ve hassas bilgilere karşı önlemler alınabilmektedir. Yetkisiz kullanıcıların kurumsal mahremiyeti oluşturan gizli ve hassas bilgilere erişimi, erişim politikalarının otomatikleştirilmesi ile engellenebilmektedir. Hassas bilginin tanımı; finansal bilgiler, fikri mülkiyet ve müşteri bilgileri gibi kritik bilgileri içeren geniş bir kapsayıcılığa sahiptir (Katz ve diğerleri, 2014, s.137)

13298 standardında da belirtildiği üzere, erişim haklarının tanımlanabilirliği EBYSler için büyük bir önem taşımaktadır. Ancak çalışmalar, gizlilik derecelerinin yanlış uygulandığını ortaya koymaktadır. Bilgilerin gizliliklerine göre sınıflandırılması; bilgilere yetkisiz erişimin sağlanması, bilgilerin değiştirilmesi ve kaybedilmesi gibi durumlardan korumaktadır (Svård, 2017, s. 18). Gizlilik Dereceli Belgelerde

Uygulanacak Usul ve Esaslar Hakkında Yönetmelik ve bu yönetmelik esas alınarak hazırlanan Gizlilik Kılavuzu, derecelerin nasıl uygulanması gerektiği konusunda detaylı şekilde rehberlik etse de insan hatalarını önlemede yeterli değildir. Bu noktada metin madenciliği ile bu kaynaklara dayanarak oluşturulmuş otomatik bir sistem belgelerin güvenliğini korumada önemlidir.

Veri sızıntısı, hassas bilgilerin bilinçli veya bilinçsiz şekilde yetkisi bulunmayan üçüncü taraflara aktarılması olarak tanımlanmaktadır (Ojoawo ve diğerleri, 2014, s. 14). Katz ve arkadaşları (2014) bilgi sızıntısı ile başa çıkmak için içerik tabanlı yaklaşım ile (kural ve sınıflandırıcı tabanlı) otomatik bir şekilde belgelere gizlilik puanı atan ve belgelerin gizlilik düzeyini belirleyen bir uygulama gerçekleştirmektedir. Uygulamada kural tabanlı yaklaşım ile metinde görülebilecek terimler ile ilgili kurallar tanımlanmaktadır. Bu sayede belirli terimlerin metinde gözükmeye sıklığı hesaplanarak bir gizlilik düzeyi belirlenmektedir (Katz ve diğerleri, 2014, s.138). Hesaplanan puanlar gizlilik derecelerinin tespit edilmesinin yanında risk envanteri oluşturmak için de kullanılabilir. Böylece bilgilerin sızdırılmasının, kuruluş için ne kadar riskli olacağı da otomatik bir şekilde ortaya konmaktadır.

Sorunlar kısmında, veri sızıntısının sıklıkla bilgi aktarımı sırasında gerçekleştiğinden bahsedilmişti. Kurum içerisinde verilerin sızdırılmasını önlemek için kurumun erişim politikaları ile uyumlu, elektronik posta sistemleri ile entegre otomatikleştirilmiş bir yapı kurulması gerekmektedir. Gonzalez-Compean ve arkadaşları (2019, s. 430), hassas bilgileri tespit ederek risk seviyesi belirleyen SecFilter adında bir güvenlik filtresi geliştirmişlerdir. Bu filtre, gizliliğin etkili bir şekilde korunmasını sağlayarak, dört aşamalı bir görev tanımlamaktadır. Bu görevler şunlardır: Filtre ilk olarak şifrelenmeden buluta gönderilen belgelerin gönderilmesini engeller, ikinci olarak madencilik teknikleri ile hassas bilgileri tespit eder, üçüncü aşamada her belge için bir

risk seviyesi hesaplar, son aşama olan dördüncü aşamada ise belgeye, riske göre bir güvenlik seviyesi atar. Benzer şekilde Elmahdy ve arkadaşları da, (2022) veri çıkarma yaklaşımı ile metinleri sınıflandırarak hassas bilgileri tespit ettikleri bir çalışma gerçekleştirmişlerdir.

Hassas bilgilerin tespit edilebilmesi sistemin gizlilik ihlalini fark etmesini sağlayan önemli bir aşamayı oluşturmaktadır. Bu tekniğin EBYS'ye uyarlanması şu şekilde gerçekleştirilebilmektedir: sistemdeki tüm belgeler içeriklerine göre kümelenmelidir. Ardından erişim hakları tanımlanmış olan kullanıcılar gönderici ve alıcı olarak kümelere atanmalıdır. Bu şekilde kullanıcı başka bir kullanıcıya ileti gönderdiği sırada sistem, eğer kullanıcının yetkisi yoksa sızıntıyı tespit edebilmektedir (Katz ve diğerleri, 2014, s.153). Böylece belgelere yetkisiz erişim, kullanıcının inisiyatifine kalmadan engellenmiş olmaktadır.

E-Postalara harici kaynaklardan gelen kötü niyetli saldırı girişimleri de bir metin madenciliği tekniği olan sapma tespiti kullanılarak tespit edilebilmektedir. e-Postalar metin madenciliği ile düzenli bir şekilde gözden geçirildiği ve kümelendi taktirde, yeni gelen posta da en benzer kümeye dahil edilir. Benzerlik hesaplama ile iletişim düzenindeki sapmalar (bağlam, sıklık, ilgili taraflar vb.) analiz edilebilmektedir böylece bir anormallik tespit edildiğinde ileti riskli olarak işaretlenebilecektir (Ojoawo ve diğerleri, 2014, s. 20).

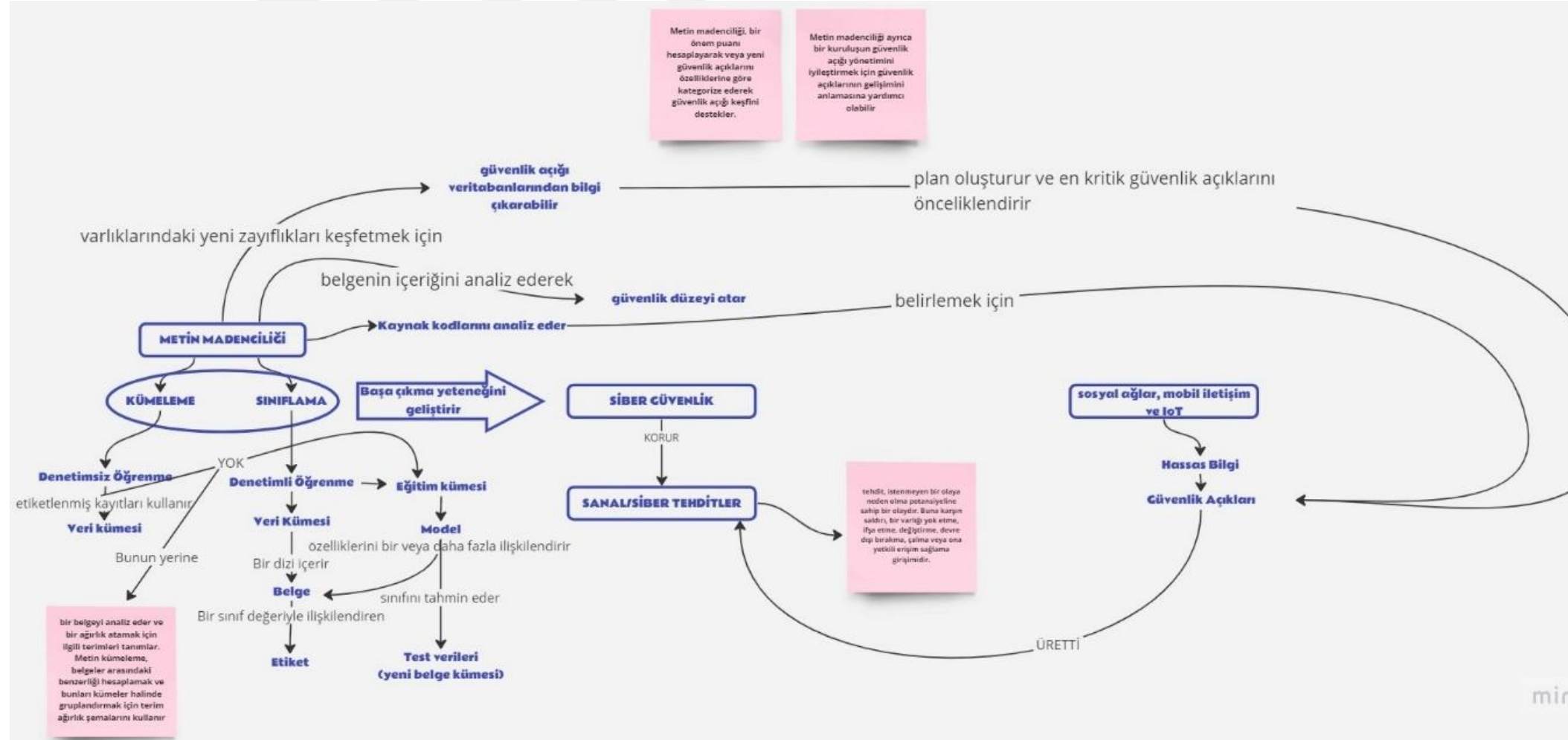
Siber güvenlikte yapılandırılmamış verilerden hata ve anormallik tespiti; bilgi güvenliğinde ise hassas verilerin tespiti ile insan kaynaklı güvenlik zafiyetlerinin önüne geçmede, metin madenciliği önemli faydalar ortaya koymaktadır. Güvenlik açıkları organizasyonların bütünlüğü ve sürekliliğini sağlamada önem arz etmektedir. Literatürde insan kaynaklı sorunların, kuruluşlarda güvenlik açıkları yarattığı görülmektedir. Ne

kadar kapsayıcı ve açıklayıcı mevzuat, standart ve rehberler yayımlansa da süreçler otomatikleşmediği sürece bu sorunların üstesinden tam anlamıyla gelinmesi mümkün gözükmemektedir. Metin madenciliği hem bilgi güvenliği hem de siber güvenlik süreçlerini otomatikleştirerek tehditlerin, saldırıların ve olayların tespitinde ve engellenmesinde kuruluşlara büyük destek sağlayabilmektedir (Ignaczak ve diğerleri, 2021, 140:28).



Şekil 27.

Metin madenciliği ve güvenlik bileşenleri arasındaki ilişki şeması



V. BÖLÜM

SONUÇ VE ÖNERİLER

Çalışmamız organizasyonlarda, yapılandırılmamış metinsel veriyi işleyerek; değer/bilgi/ürün yaratmayı, verimliliği arttırmayı ve rekabet gücünde artış yaratmayı hedefleyen bir dijital dönüşüm projesidir. Bu noktada organizasyonların iş süreçlerini yönetmesi ve içerisinde büyük miktarda yapılandırılmamış metinsel veri bulundurması ile EBYS'ler, bu çalışma için son derece önemli bir nitelik taşımaktadır.

Bilgi ve belge yönetimi perspektifi ile oluşturulmuş metin madenciliği tabanlı bu mimari, EBYSlerdeki veri/bilgi/belge yönetimini en üst seviyeye getirmeyi ve bu veri/bilgi/belgeyi de işleyerek organizasyonlara değer katacak öngörülü bilgiyi elde etmeyi amaçlamaktadır. Bu bağlamda çalışma; literatürde başka alanlara uygulanan metin madenciliği uygulamalarını analiz ederek (eğitim, inşaat, bilgisayar mühendisliği ve diğerleri), metin madenciliği ile, EBYSlerdeki mevcut belge yönetim süreçlerini iyileştiren (üretim, tanımlama, dağıtım, depolama, erişim) ve süreci geliştirecek yeni aşamaları ekleyen (iş zekâsı yaratma) bir yapı ortaya koymaktadır.

Veri/bilgi/belge yönetimi süreçlerini geliştirmenin yanında çalışmanın ortaya koymuş olduğu önemli bir işlevi daha bulunmaktadır. Bu işlev, bilgi ve belge yöneticilerinin bilgi kaynaklarını, kullanıcıların erişimine sunmasının yanında; bilgi erişime dair bilgi ve becerilerini kullanarak, sistemlerin geliştirilmesindeki önemli rolünün ortaya konulmasıdır. Bir bilgi ve belge yöneticisi tarafından hazırlanan bu çalışmanın da ortaya koyduğu üzere, bilgi ve belge yöneticileri yeni dijital teknolojileri kullanmanın ve/veya erişime sunmanın ötesinde bu sistemleri bilgi ve becerileri ile de geliştirebilecek kişilerdir.

Çalışmada ilk olarak, EBYS'nin ne olduğu, özellikleri ve Türkiye'deki mevzuat, standart ve rehberlerde hangi açılardan ele alındığı kronolojik olarak ortaya konmaktadır. EBYS'ler e-devletin dolayısı ile de dijital dönüşümün oluşmasında kritik bir adımı oluşturmaktadır. Çalışmada; dünyada ve Türkiye'de gelişimi, olumlu ve olumsuz etkileri ve organizasyonlardaki işlevleri ile açıklanan dijital dönüşüm kavramı kısaca; iş süreçlerine yeni dijital teknolojileri uygulayarak değer ve köklü değişiklikler yaratmaktır. Bu kavram çalışmamız için önem arz etmektedir. Çünkü “dijital dönüşüm” aslında, çalışmamızın varmak istediği son noktayı oluşturmaktadır.

Bilgi yönetimi, organizasyonlarda bulunan çeşitli bilgi kaynaklarına yönelik, bilgi çıkarma ve arama süreçlerinin düzenlenmesi ile ilgilenmektedir. Bu sürecin en iyi şekilde ilerleyebilmesi için ise yeni dijital teknolojilere ihtiyaç duyulmaktadır (Krishnapuram, 2003, s. 1494). Çalışmamız yeni dijital teknolojilerden metin madenciliğine ağırlık vererek, metin madenciliğini EBYSlerdeki bilgi yönetimi süreçlerine uygulamaktadır. Üçüncü bölümde metin madenciliği; veri madenciliğine dayanan teknik ve algoritmalarıyla açıklanmakta, veri madenciliğinden hangi yönleri ile ayrıştığı ve neden metin madenciliğine gereksinim duyulduğu ortaya konmaktadır. Açıklanan teknik ve algoritmalar, dördüncü bölüm olan “EBYS’lerde metin madenciliği” başlığı için bir temel oluşturmaktadır. Dördüncü bölümde, söz konusu teknik ve algoritmalar, EBYSlerde belge yönetim süreçlerine uygulandığında, organizasyonlar için;

- mevcut ve potansiyel sorunlara nasıl çözüm sağlayacağı,
- iş süreçlerini ne şekilde optimize edeceği
- ve değer/ürün/hizmet yaratılmasını sağlayarak nasıl rekabet avantajı elde edileceğini açıklamaktadır.

Çalışma EBYS'lerde metin madenciliğini üç ana başlıkta ele almaktadır. Bunlar; Bilgi Erişim, İş Zekası ve Güvenliktir. EBYS'ler kapsamında, tek tek geniş bir şekilde ele alınan üç başlığın da ortak bir özelliği bulunmaktadır. Bilgi Erişim, İş Zekası ve Güvenlik başlıklarında ele alınan sorunlarda görülmektedir ki, sorunların büyük bir çoğunluğu insan kaynaklı ortaya çıkmaktadır. Hem insan kaynaklı hataların önlenmesinde hem de veri işlemenin güç olduğu büyük veri çağında otomatikleşme, kurumların sürekliliği için büyük önem arz etmektedir.

Bu noktada çalışma metin madenciliğini

Bilgi Tanımlama süreçlerine uyarlayarak;

- Hatalardan ve yanlışlıklardan arınmış, otomatikleşmiş bilgi tanımlama (metadaların oluşturulması) sınıflama ve kümeleme süreçleri sağlanmakta,
- Kesin isabet ve erişim isabeti oranlarını iyileştiren ontoloji tabanlı bilgi erişim sistemleri oluşturmaktadır.

İş Zekasına uyarlayarak;

- İnsanlar tarafından analiz edilmesi imkânsız olan büyük metinsel verileri, bilgi haritası gibi yöntemlerle anlamlandırarak; organizasyonların bilgi ve rekabet üstünlüğü sağlayacağı bilgi ve desenleri keşfetmekte ve örtük bilgileri ortaya çıkartmadır.

Güvenliğe uygulayarak ise;

- Güvenlik açıklarını, güvenlik kontrollerini ve veri ihlallerini tespit ederek Siber Güvenliği sağlamakta,
- Hassas bilgilerin tespiti ile bilgilerin sızdırılmasını ve yetkisiz kişilerin erişimini engelleyerek Bilgi Güvenliğini sağlamaktadır.

Günümüzde büyük verinin ve bu verileri işleyerek elde edilen bilginin, çok büyük bir güç yarattığı; aynı zamanda söz konusu verilerin işlenmesinde kullanılan araçların ne kadar önemli olduğu, her alanda güncelliğini koruyan bir konudur. Bu konunun güncelliğine ve popülerliğine rağmen, literatürdeki çalışmalar göstermektedir ki, organizasyonlar için büyük verim sağlayacak metinsel verilerden bilgi keşfi, bilgi yönetimi alanında neredeyse hiç değinilmemiştir. Bu noktada çalışmamız, EBYSlerde metin madenciliği uygulamalarını bütünsel bir şekilde ele alarak (erişim, tanımlama, arama, güvenlik, iş zekası oluşturma gibi çeşitli konulara değinerek) hem literatürdeki boşluğu doldurmakta hem de erişilebilen kaynaklar çerçevesinde ilk olma niteliği taşımaktadır.

EBYS’lerde metin madenciliğinin aşağıdaki yararları sağlaması beklenmektedir;

- Bilginin verimli kullanımı sağlanarak değer elde edilmektedir,
- Etkili bilgi erişim süreçleri sağlamaktadır (kesin isabet oranını arttırmaktadır),
- Kelime tabanlı arama motorları yerine ontoloji tabanlı arama motorları ile sorgu ve nitelemeler genişletilmektedir,
- Veri işleyenden kaynaklanan hatalı nitelemelerin önüne geçmektedir,
- Yanlış dosya planı uygulamaları engellemektedir,
- Ontolojiler sayesinde veri/bilgi/belgelerin çeşitli şekilde tanımlamaları yapılmaktadır,
- Kullanıcıların bilgi ihtiyaçlarını tanımlayamadığı noktada, ontolojiler sayesinde bilgi arama süreçleri yönlendirilmektedir,
- Kümeleme sayesinde, kullanıcı sorgusuyla paralel olarak, arama sonuçları daha spesifik hale getirilmektedir,

- Kullanıcının bilgi ihtiyacını karşılayan belgeye, benzer diğer belgeler önerilmektedir,
- Bilgilere daha hızlı ve etkili erişim sağlanmaktadır,
- Otomatikleşme sayesinde standartlaşma sağlanmaktadır,
- Organizasyondaki metinsel verilerden varlıklar çıkarılarak, kurumsal bir ontoloji oluşturulmaktadır,
- İnsanlar tarafından analiz edilmesi güç hatta imkânsız olan büyük metinsel veriler analiz edilmektedir,
- Belgelere metin madenciliği teknikleri uygulanarak; örtük bilgiler anlamlandırılmakta, bilgi, desen ve ilişkilerin çıkarılmakta ve keşfedilmektedir,
- Keşfedilen bilgiler ve çıkarılan ilişkiler sayesinde öngörülü bilgiler elde edilmekte ve etkili ve zamanında karar alma süreçleri gerçekleştirilmektedir,
- İç ve dış bilgi kaynaklarına dayalı metinsel veriler bütünsel şekilde analiz edilerek etkili karar mekanizmaları oluşturulmaktadır,
- Kullanıcı/Müşteri odaklı ürün/hizmetlerin gelişimi sağlanmaktadır,
- Bilgi üstünlüğü sağlamak ve sürdürmektedir,
- Elde edilen bilgi üstünlüğü sayesinde organizasyonlar, rakipleri arasında sıyrılmakta ve öne çıkmaktadır,
- Kuruluşlarda etkin bir bilgi kültürü yaratmaktadır,
- Metinsel veriler içerisindeki anormallikler, tehditler ve saldırılar tespit edilerek güvenlik açıkları belirlenmektedir,
- Hata veritabanlarından güvenlik zafiyetlerini tespit etmektedir,
- Çalışanlar tarafından gerçekleşen, veri ve bilgi sızıntılarını tespit etmekte ve engellemektedir,
- Hassas bilgileri tespit ederek belgelerin risk seviyelerini puanlamaktadır,

- Yetkili erişimi sağlamaktadır,
- Organizasyonların fırsat, tehdit ve risklerini daha açık hale getirmektedir,
- Tüm bu faydalar ile etkili bilgi yönetim süreçleri sağlayarak organizasyonun varlığını korumakta ve değer yaratmaktadır.

Bir EBYS, fiziksel ortamda gerçekleştirilen belge yönetim süreçlerinin, elektronik ortama taşınmasının ötesinde; dijital teknolojileri bu süreçlere entegre ederek, dijital dönüşümü gerçekleştirecek bir yapıya sahip olmalıdır. Bu görüşü destekler şekilde çalışmamız metin madenciliği ile, organizasyonlarda veri ve enformasyonla kurulan ilişkiyi yeniden şekillendirerek iş süreçlerinde önemli bir dönüşüm yaratmaktadır (Khan ve Vorley, 2017, s. 5).

Organizasyonların belge yönetim süreçlerini işleyiş şekilleri ve bu süreçlere olan bakış açıları ile standart, mevzuat ve rehberler göz önüne alındığında, aşağıda sıralanan önerilerin veri/bilgi/belge yönetim süreçlerini iyileştireceği öngörülmektedir:

- EBYS’lerde iş süreçlerinin sürekliliğini ve güvenliğini tehdit eden en önemli unsur insan faktörüdür. Belgelerin yanlış tanımlanması ve tasnifi, ihtiyaç durumunda belgelere erişim sağlanmasında sorun yaratacaktır. Erişilemeyen belgeler ise organizasyonların karar alma süreçlerini olumsuz etkileyecek, iş süreçleri aksatacak ve maliyetleri arttıracaktır. Bu durumun önüne geçilmesindeki en kesin çözüm süreçlerin otomatikleştirilmesidir. Ontoloji tabanlı metin madenciliği tekniği kullanılarak, bilgiler hem otomatik bir şekilde tanımlanmakta ve sınıflanmakta hem de ontolojinin yapısında barındırdığı ilişkiler ile erişimin verimini arttıran çeşitli erişim uçları tanımlanmaktadır.
- Belgelerin tasnifinde yapılan hataların önemli bir nedeni de standart dosya planının çok kırılgan ve karışık yapısı olabileceği tahmin edilmektedir. Her

organizasyonun farklı iş süreçleri ve fonksiyonları bulunmaktadır. Bu noktada standart dosya planının EBYS'lere uygun, esnek, basit ve uyarlanabilir yapıda olması önemlidir. Bu şartları sağlayan bütünsel bir dosya planının oluşturulması belge yönetim süreçleri için daha verim sağlayabileceği düşünülmektedir.

- Kurumsal ilişkiler ile iş ve fonksiyonların ortaya konmasında ontolojiler oldukça önem arz etmektedir. Her kuruluşun kendi iş süreçlerini yansıtan bir ontoloji oluşturması; kurum içinde tutarlılık, standartlaşma, bilgi tanımlama, erişim ve paylaşma ve kurumsal hafıza gibi konularda fayda sağlayacaktır.
- TS 13298 standardında (2015) EBYSler üzerinden arama işlemlerinin mantıksal operatörler kullanmaya izin veren bir yapıda olması gerektiği belirtilmektedir. Ancak mantıksal operatörlerin kullanımı kullanıcılar açısından anlaşılması zor bir süreci doğurmaktadır. Halbuki bir bilgi erişim sisteminin; basit, anlaşılır, hızlı ve etkili sonuç üretmesi beklenmektedir. EBYS'lerde bilgi arama süreçlerinde, mantıksal operatörler gibi hem geleneksel hem de anlaşılması zor teknolojiler yerine, çalışmamızın ortaya koymuş olduğu ontoloji tabanlı bir bilgi erişim sistemi bilgilere daha kolay ve etkili şekilde erişilmesini sağlayacaktır.
- Yine TS 13298 standardında, EBYSlerde kayıt ve tasnif işlemleri gerçekleştirilirken, belgelere ait metadataların ve dosya planında temsil edilen elemanların (dosya adı, klasör gibi) manuel olarak işlendiği belirtilmektedir. Literatür incelendiğinde kuruluşlarda manuel olarak oluşturulan bu süreçlerin hatalı ve verimsiz uygulandığı ortaya konulmaktadır. Bu noktada metadata oluşturma ve sınıflama süreçlerinin sağlam bir ontolojiye dayanan otomatik sistemler ile gerçekleştirilmesi gerekmektedir.
- EBYS, sistem içerisindeki eylemleri ve istatistikleri içeren bir raporlama modülüne sahiptir ancak bu modül organizasyonların karar alma süreçleri için

yeterli değildir. Kurumlara karar aldırarak olan, bu raporları işleyecek, analiz edecek ve öngörüler sağlayacak teknolojilerdir. Bu bağlamda hem bu raporların analizinde hem de kurumun elde ettiği metinsel bilgi kaynaklarında, iş zekasını oluşturacak metin madenciliği teknikleri, EBYS'lere entegre edilmesi gereken olmazsa olmaz bir teknolojidir.

- EBYS'lerdeki belgeler için yapılandırılmış bilgi haritaları; belgelerin içerisindeki bilgi ve/veya bilgi parçacıklarına erişmede, örtük bilgilerin anlamlandırılmasında ve ilişkilerin ortaya çıkarılmasında kurumlar için büyük fırsatlar sunacaktır. Bu noktada organizasyonlar; fırsat, risk ve tehditlerini tespit edebilmek ve etkili kararlar alabilmek için bu teknolojiyi sistemlerine entegre etmesi gerekmektedir.

- TS 13298, TS 15489-1, Elektronik Doküman ve Belge Yönetim Sistemi Koruma Profili, Gizlilik Kılavuzu, Bilgi ve İletişim Güvenliği Rehberi, Gizlilik Dereceli Belgelerde Uygulanacak Usul ve Esaslar Hakkında Yönetmelik gibi EBYS'leri yakından ilgilendiren mevzuat, rehber ve standartlarda; hassas bilgilere yetkisiz erişimin engellenmesi ve belgelerin gizlilik, bütünlük ve erişilebilirliğinin sağlanması üzerine önemli bilgiler ve tedbirler işlenmektedir. Ancak söz konusu kaynaklarda görülmektedir ki; insan faktörünün, güvenliği sağlamada en büyük tehditlerden birisi olduğu kabul edilmesine rağmen, süreçler yine insanlar tarafından manuel olarak yürütülmektedir. EBYSlerde yetkilendirme işlemlerinin gerçekleştirilmesi için sistem yöneticisinin önemli yetkileri bulunmaktadır. Bu kadar önemli yetkiler yanlış uygulandığı takdirde belgelerin orijinalliği, bütünlüğü ve doğruluğu tehlike altına girmektedir. Bu ise karar alma süreçlerini ve kurumun sürekliliğini zora sokmaktadır. Bu gibi risklerin önüne geçilmesi için organizasyonlarda, yetkili erişim süreçleri otomatik bir

şekilde gerçekleştirilmelidir. Bu noktada, sistemlerin, yanlışlıkla gerçekleştirilen bilgi ve veri sızıntılarını, otomatik bir şekilde engeller bir yapıya sahip olması önem arz etmektedir.

- Yetkilendirme işlemleri ile benzer şekilde, gizlilik kılavuzunda, farklı gizlilik derecelerinin üst yönetici, birim yöneticisi ve alt birim yöneticisi tarafından manuel olarak gerçekleştirildiği belirtilmektedir. Gizlilik derecelerinin yanlış uygulanması kuruluşlarda güvenlik ve hak ihlali yaratmaktadır. Yanlış uygulanan gizlilik dereceleri, belgelere yetkisiz kişilerin erişmesine ve/veya belgelerin gereksiz şekilde erişiminin ve kullanımının kısıtlanmasına neden olmaktadır. Organizasyonların bu yanlışlıkların önüne geçebilmesi için gizlilik dereceleri, hassas bilgiler esas alınarak otomatik puan ataması ile gerçekleşmelidir.

- Organizasyonlarda bilgi varlıklarına yönelik oluşturulan risk analizinde de otomatik puanlama sistemi kullanılarak metinsel bilgi varlıklarına yönelik risk değeri atanması mümkündür. Organizasyonlar daha tutarlı, hızlı ve ekonomik süreçler sağlayabilmek için bu uygulamaları tercih etmelidir.

- EBYSler ile ilgili mevzuat, standart ve rehberlerde, yeni dijital teknolojilerin politikalarda yeterli düzeyde yer almadığı görülmektedir. Dijital dönüşümün sağlanması, köklü değişimleri gerektirmektedir. Dijital devlet için atılan adımlarda öncelikli amaç, politikaların da bu bakış açısı ile güncellenmesi olmalıdır. Teknolojinin sağladığı faydalar mevcut sistemin iyileştirilmesi ve sorunların çözümü için kullanılmalıdır.

- Alanımızda, dijital teknolojiler ile üretilen ürünlerin kullanılmasının ötesinde bu teknolojilerin geliştirilmesi ve iş süreçlerine entegre edilmesi

gerekmektedir. Bilgi ve belge yöneticileri, uzmanlıklarını bu teknolojilerde ne şekilde kullanacağına yönelik daha çok çalışma ortaya koymalıdır.

- Bu çalışmanın da ortaya koyduğu üzere, bilgi ve belge yöneticilerinin; niteleme, sınıflama, dizinleme gibi alanlardaki uzmanlıkları, bir bilgi erişim mimarisi oluşturulması için önem taşımaktadır. Bu noktada bilgi sistemlerinin geliştirilmesi ve sürdürülmesinde bilgi ve belge yöneticilerinin sürece dahil olması, organizasyonların gelişimi ve etkililiği için son derece gereklidir. Bu noktada hem ülke çapında veya kurumsal ölçüde oluşturulacak politikalarda hem de kuruluşların iş alım süreçlerinde bu hususu göz önüne almaları gerekmektedir.
- Dış bilgi kaynaklarının analiz edilmesi ve yeniden kullanımı organizasyonların değer yaratmasında ve karar süreçlerinde büyük önem arz etmektedir. Ancak organizasyonların bu verilere erişmesi oldukça maliyetlidir. Ülkenin, kamusal verilerin açık erişimli bir şekilde paylaşımını sağlayacak politikalar geliştirmesi gerekmektedir. Geliştirilen bu politikalar sayesinde organizasyonlar daha kolay ve etkili ürünler/hizmetler sunarak, hem kurumsal çapta hem de ülke çapında büyüme ve kalkınma sağlayacaktır.
- Veri/bilgi/belge yönetim süreçlerinin organizasyonların sürekliliği için ne kadar büyük katkı sunacağı üst yönetime aşılması gereken bir konudur. Ancak üst yönetim desteği alındığı takdirde işin uzmanları tarafından etkili teknolojiler kullanılarak süreçler iyileştirilebilecektir. Bu noktada alan uzmanı olan bilgi ve belge yöneticilerine büyük sorumluluk düşmektedir. Bilgi ve belge yöneticileri çağın gerektirdiği teknolojik becerileri ne kadar iyi edinir ve iş süreçlerine uyarlarsa, sahip olduğu potansiyel gücü o kadar iyi ortaya koyacak ve bu güçle alanı temsil edecektir.

Rekabet avantajı elde etmek ve çağın gereğine uygun ürün ve hizmetler sunabilmek için, organizasyonlar kendilerini sürekli olarak yenilemeli ve geliştirmelidir. Mevcut iş sürecinde bir aksaklık yaşanmıyorsa bile, bilgi avantajı elde etmek ve süreçleri daha verimli hale getirmek, her zaman birincil hedef olmalıdır. Bu bağlamda organizasyonlar için, yeni dijital teknolojilerin iş süreçlerini iyileştirme ve sorunları gidermede nasıl rol oynayacağı, araştırılması gereken bir konudur. EBYS’ler şu anki işleyişleriyle de, organizasyonların belgelere erişimini hızlandıran, güvenliği ve hesap verilebilirliği sağlayan ve karar vermeyi güçlendiren sistemlerdir. Ancak rekabet avantajı elde edebilmek, maliyetleri arttırmak, bilgi üstünlüğü sağlamak, zaman ve emekten tasarruf etmek için iş süreçlerinin en verimli noktaya ulaşması sağlanmalıdır. Bu noktada çalışmamız dijital teknolojilerden metin madenciliğini kullanarak, bilgi ve belge yönetimi perspektifiyle, EBYSlerde belge yönetim süreçlerini en üst seviyeye getirmeyi ve mevcut sorunlara çözüm bulmayı hedefleyen bir mimariyi ortaya koymaktadır.

Bu noktada çalışmamızın başında Temel hipotezimiz şu şekilde ortaya konmuştur:

H0: EBYS’lerde insan kaynaklı hataların giderilmesinde metin madenciliği tekniği kullanılabilir

Bu temel hipoteze bağlı olarak alt hipotezimiz de,

H1: EBYS’lerde metin madenciliği tekniği kullanılarak; sistemin gelişimi, bilgi güvenliği, siber güvenlik ve kurumsal politikaların oluşturulmasında yarar sağlanabilir.

olarak oluşturulmuştur.

Çalışmamızda, literatürde daha önce EBYS’lere yönelik olarak ortaya konmuş sorunlar tespit edilmiş ve bu sorunların büyük çoğunluğunun insan kaynaklı hatalar

olduğu belirtilmiştir. Bu bağlamda, hipotezimiz bir varsayımın ötesinde, somut olarak ortaya konulmuş bir gereksinimi de ifade etmektedir. Öte yandan çalışmamızda; insan kaynaklı bu hataların çözümü için, daha önce başka alanlara başarılı bir şekilde uygulanan; otomatikleşmiş, metin madenciliği tabanlı, teknolojiler önerilmektedir. Bu noktada çalışmamız, söz konusu sorunların çözümüne yönelik olarak, metin madenciliği ile ilgili başka alanlara uygulanan yöntemleri belge yönetimi alanına teorik düzeyde uyarlamaktadır. Metin madenciliği uygulamalarının diğer alanlarda göstermiş olduğu başarılı sonuçlar, ebylerdeki insan kaynaklı sorunların da çözülme potansiyelini ortaya koymaktadır. Literatür incelendiğinde EBYS uygulamalarında; belgelerin tanımlanması, tasnifi, gizlilik derecelerinin uygulanması ve bilgi ihlali konularında insan kaynaklı hatalar tespit edilmiştir. Bu hataların çözümüne yönelik olarak; metin madenciliği ile ilgili uygulamalar incelenmiş ve metin madenciliğinin başarılı bir şekilde;

- metinsel verileri analiz ederek metni en iyi tanımlayan kelimelerin otomatik olarak çıkarttığı,
- sınıflama tekniğini kullanarak belgeleri önceden etiketlenmiş sınıflara atadığı,
- hassas bilgileri tespit ederek metinlere güvenlik skorlarının atandığı,
- güvenlik skorlarının atandığı belge kümelerinin kullanıcılarla (erişim hakları esas alınarak) ilişkilendirildiği uygulamalar ortaya konulmaktadır.

Bu örnekler ve çalışmada bahsedilen diğer tüm bulgular, metin madenciliğinin başarısını ortaya koymakta ve EBYSlerde insan kaynaklı sorunların çözümünü sağlamada güçlü bir dayanak oluşturmaktadır.

Buna ilaveten çalışmamız, EBYSdeki bilgi erişim, bilgi güvenliği, siber güvenlik ve karar alma alanlarındaki mevcut ve potansiyel sorunları ortaya koymakta ve metin madenciliğinin gerçekleştirmiş olduğu uygulama örnekleri ile bu sorunlara ne şekilde

özüm saęlanabileceęini açıklamaktadır. Sorunlar incelendięinde, problemlerin büyük bir çoęunluęunun insan kaynaklı olduęu, geri kalanının ise veri hacminden kaynaklandığı görölmektedir. EBYS'ler ile ilgili mevzuat, standart ve rehberler incelendięinde ise, süreçlerin yeni dijital teknolojilerin kullanılarak otomatikleştirilmedięi, bazı uygulamaların ise geleneksel kaldığı saptanmaktadır. Bu noktada, metin madencilięi kullanarak; bilgi erişim, bilgi güvenlięi, siber güvenlik ve karar alma konularında gerçekleştirilen başarılı örnekler, EBYSlere entegre edildięinde, metin madencilięinin; alt hipotezimizi destekler şekilde; sistemin gelişiminde, bilgi güvenlięinde, siber güvenlikte ve kurumsal politikaların oluşturulması konusunda yüksek bir potansiyele sahip olduęu ortaya konulmaktadır.

Organizasyonların saęlıklı kararlar alabilmesi, varlığını koruyabilmesi ve etkili hizmetler sunabilmesi önemli ölçüde verilerin ve içeriğin iyi yönetilmesine baęlıdır (Svård, 2017, s.15). Çalışmamız, metin madencilięinin analiz, keşif ve ilişki kurma gücünü kullanarak, EBYSlerdeki sorunların nasıl üstesinden gelinebileceęini ve sistemin veriminin nasıl arttırılacaęını ortaya koymaktadır. Bilginin gücü sürekli artmasına raęmen organizasyonlar yaygın olarak, bilgi gücü elde etmede bilgi yönetiminin etkililięinin farkında deęildir (Mukred ve dięerleri, 2022). Bu bağlamda çalışmamız, zaten organizasyonların verimlilięi için önemli olan belge yönetimi konusunu, dijital çağın gereklerine göre düzenleyerek, organizasyonlar için üst düzey verim saęlanacaęına inanmaktadır. Bir organizasyonun iyi bir bilgi kültürüne sahip olması, iş başarılarını da olumlu şekilde etkilemektedir (Svård, 2017, s. 30). Bu yaklaşımdan hareketle bilgi ve belge yöneticilerinin, teknik ve becerilerini kullanarak, bilginin potansiyel gücünü ortaya koyması, organizasyonlarda bilgi kültürünün oluşmasında ve buna baęlı olarak iş süreçlerinin iyileştirilmesinde büyük fayda saęlayacaktır. Bir organizasyonun başarısının temeli, sahip olduęu yetkinliklerdir ve bu yetkinlikler, ortaya konulan ürün ve/veya

hizmetlerden daha dirençlidir (Wissensmanagement Forum, 2003, s. 8). Bu noktada organizasyonlarda verimli iş süreçlerinin geliştirilmesi, bilgi ve belge yönetiminin uzman katkısı olmadan düşünülmemelidir.

V. 1. Gelecek Çalışmalar İçin Öneriler

Çalışmamız, bilginin güç olduğu ve ancak bu gücü elinde tutabilenlerin kendini ön plana çıkarabildiği çağda, bilgi ve belge yönetiminin uzman bakış açısını kullanarak, metin madenciliği teknolojisini, bilgi yönetim süreçlerine entegre eden ve süreçleri iyileştiren bir çerçeve sunmaktadır. Bilgi istisnasız her alanda kullanılmakta olup bilgi ve belge yönetiminin çeşitli uygulama alanları bulunmaktadır. Çalışmamız EBYS'ler için organizasyonlara yönelik bir bilgi yönetim sürecini ele almaktadır. Ancak söz konusu teknikler diğer sistem ve bilgi merkezlerine de uygulanabilir yapıdadır. Metin madenciliği teknikleri kütüphanelerde, otomasyon sistemleri ve çevrimiçi kataloglara uygulanması ile etkili bilgi erişim süreçleri sağlanabilir ve kullanıcı odaklı bilgi hizmetleri sunulabilir. Yine veritabanlarında bilgi kaynaklarının daha etkili tanımlanması, aranması ve erişilmesi için kullanılabilir. Yayıncılık alanında ise, dergi içerisinde yer alan makalelerin, bir standarda oturmuş, daha çeşitli ve etkili anahtar kelimeleri tanımlamasına olanak sağlayabilir. Bunların yanı sıra, bilgi ve belge yönetimi alanına trend analizi, konu takibi, kavram bağlantısı gibi önemli metin madenciliği teknikleri uygulanarak; literatürdeki araştırma boşlukları, keşfedilmemiş konu ve ilişkiler kolaylıkla saptanabilir ve ihtiyaca yönelik, yenilikçi bilimsel çalışmalar ortaya konulabilir.

Metin madenciliği birçok disiplin ile de iç içe geçmiş ve farklı teknolojilere ait teknik ve algoritmaları kullanan bir yöntemdir. Bir uygulamadan daha yüksek verim elde etmek, disiplinler arası yaklaşımları gerektirmektedir. Bu çalışma metinsel verilere

odaklandığı için metin madenciliğinin ağırlıklı olduğu uygulama örnekleri ortaya konulmaktadır. Ancak daha bütünsel çalışmalar için (görsel, sesli vb. verilerin de analiz edildiği), metin madenciliğinin yanı sıra; derin öğrenme, görüntü işleme, makine öğrenmesi, yapay zeka vb. yöntemler kullanılması süreçlerin daha da verimli ele alınabilmesini sağlayacaktır.

Bilgi ve belge yönetimi alanına mensup kişiler; bilgi erişim, bilgi nitelendirme ve bilgi yönetimi alanlarında uzmanlığa sahip kişilerdir. Günümüz bilgi çağının hammaddesi “bilgi”dir. Bilginin en yüksek verimi nasıl elde edeceği yetkinliği ise bilgi ve belge yöneticilerinde mevcuttur. Bu noktada, bu bilgi ve becerilerin dijital teknolojiler ile nasıl buluşturulacağı, “organizasyonların iş süreçlerinde nasıl verim elde edeceği” sorununun da cevabını oluşturmaktadır. Gerçekleştirdiğimiz çalışma, iş süreçlerini iyileştirmenin yanında “bilgi ve belge yönetimi” alanının perspektifini sunması ile önem arz etmektedir. Çalışmamızın ortaya koyduğu bu perspektifin ilerideki çalışmalar ile beraber daha da gelişmesi ve bilginin bulunduğu her alanda uygulanması en büyük gayemizdir.

KAYNAKÇA

Agrawal, R., & Batra, M. (2013). A detailed study on text mining techniques. *International Journal of Soft Computing and Engineering*, 2(6), 118-121.

Akça, M. F. (2020). Metin Madenciliğinde Veri Ön İşleme. 27.12.2024 tarihinde <https://medium.com/deep-learning-turkiye/metin-madencili%C4%9Finde-veri-%C3%B6n-i%C3%9C%87%C5%9Fleme-3e3e7ade4290> adresinden erişilmiştir.

Al Hakim, S., Sensuse, D. I., Budi, I., Prima, P., & Safitri, N. (2020). Exploring an Intelligent Approach in Knowledge Mapping with Ontology and Text Mining: Systematic Literature Review. *IICST*, 33-40.

Alir, G. (2008). E-Türkiye uygulamaları: Elektronik belge yönetimi ve üst veri.

Ankara Üniversitesi Sosyal Bilimler Enstitüsü. (2025, Mart). *Tez yazım yönergesi* [PDF]. Ankara Üniversitesi Sosyal Bilimler Enstitüsü. Erişim adresi: https://sosbilens.ankara.edu.tr/wp-content/uploads/sites/271/2025/03/tezyazim.yeni1_.pdf

Arıcı, G., & Kandur, H. (2016). Elektronik Belge Yönetim Sistemleri (EBYS) Yazılımlarının Geliştirilmesinin Kurumsal Karar Destek Sistemleri (KDS) İçin Önemi. *Kurumsal Belleklerin Geleceği*, 65.

Atan, S. (2020). Metin Madenciliği: İmkânlar, Yöntemler ve Kisitlar. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (31), 220-239.

Atan, S., & Çınar, Y. (2019). Borsa İstanbul'da finansal haberler ile piyasa değeri ilişkisinin metin madenciliği ve duygu (sentiment) analizi ile incelenmesi. *Ankara Üniversitesi SBF Dergisi*, 74(1), 1-34.

Aydın, C. (2005). Bilgi teknolojilerinin belge yönetimine etkisi ve elektronik belge yönetimi. *Bilgi Dünyası*, 6(1), 89-97.

Aydın, A. (2023). 1 — Text Preprocessing Techniques for NLP. 27.12.2024 tarihinde <https://ayselaydin.medium.com/1-text-preprocessing-techniques-for-nlp-37544483c007> adresinden erişilmiştir.

Aydın, C. (2023). Dijital dönüşüm bağlamında belgenin yaşam döngüsü süreçlerinin yönetiminde yapay zekâ kullanımı. *Bilgi Yönetimi*, 6(1), 21-42.

Aydın, C., & Özdemirci, F. (2011). Elektronik belgelerin arşivlenmesinde gerçekliğin ve bütünlüğün korunması. *Bilgi Dünyası*, 12(1), 105-127.

Aydın, F., & Saydam, V. (2019). Elektronik Belge Yönetimi Sistemlerinde Bilgi Varlıklarının Kullanımı. *Bilgi Yönetimi*, 2(1), 1-8.

Aydoğdu, Ü. R., Karamustafaoğlu, O., & Bülbül, M. Ş. (2017). Akademik Araştırmalarda Araştırma Yöntemleri ile Örneklem İlişkisi: Doğrulamalı Doküman Analizi Örneği. *Dicle University Journal of Ziya Gokalp Education Faculty*, (30).

Bakay, Ö., Ergelen, Ö., Sarmış, E., Yıldırım, S., Arıcan, B. N., Kocabalcıoğlu, A., ... & Yıldız, O. T. (2021, January). Turkish wordnet kenet. In *Proceedings of the 11th global wordnet conference* (pp. 166-174).

Balcı, A. (2016). Sosyal Bilimlerde Araştırma Yöntem, Teknik ve İlkeler. Ankara: Pegem Akademi.

Best Edit & Proof. (2021). *Kusursuz bir literatür taraması için 7 teknik*. 9.06.2025 tarihinde <https://besteditproof.com/tr/academy/kusursuz-bir-literatur-taramasi-icin-7-teknik> adresinden erişilmiştir.

Bilgi ve İletişim Güvenliği Tedbirleri (2019/12 Sayılı) Genelgesi. (2020) T.C. Resmi Gazete, Sayı: 30823, Temmuz 2019. Erişim adresi: <https://www.mevzuat.gov.tr/MevzuatMetin/CumhurbaskanligiGenelgeleri/20190706-12.pdf>

Bingöl Üniversitesi. [Tarih Yok]. *Araştırma yöntem ve tekniklerinin seçimi* (Bölüm 9). <https://www.bingol.edu.tr/media/205521/sayt-bolum9-Arastirma-Yontem-ve-Tekniklerinin-Secimi.pdf>

Bizer, C. [Tarih Yok]. *Data Mining: Text-Mining* [PowerPoint slaytları]. Universität Mannheim. 7.06.2025 tarihinde https://www.uni-mannheim.de/media/Einrichtungen/dws/Files_Teaching/Data_Mining/Slides/DM05-Text-Mining.pdf adresinden erişilmiştir.

Borglund, E. A., Anderson, K., & Samuelsson, G. (2009a). How requirements of record managers change after implementing new electronic records management systems. In *3rd European Conference on Information Management and Evaluation, Sep 17-18, 2009, Gothenburg, Sweden* (pp. 59-66). Academic Conferences Limited.

Borglund, E. A., Anderson, K., Sirkemaa, S., Wahlberg, O., & Sandberg, K. W. (2009b). Business values of electronic records management in SMEs. In *Proceedings of the 3rd European Conference on Information Management and Evaluation* (pp. 51-58). UK: Academic Conferences Limited.

Boyalı, H. (2023). Türkiye’de Vatandaş Odaklı İdare: E-Devlet Ötesi Dijitalleşen Kamu. *Bucak İşletme Fakültesi Dergisi*, 6(2), 172-190.

Brecque, C. (2024, 10 Mart). *Text analytics vs text mining*. TextMine. 6.06.2025 tarihinde <https://textmine.com/post/text-analytics-vs-text-mining> adresinden erişilmiştir.

Chen, H. (2001). Knowledge management systems: a text mining perspective. Knowledge Computing Corporation.

Chieh-Yuan, T., & Wei-Zhong, J. (2018, July). Knowledge map construction using text mining and artificial bee colony algorithm. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 1, pp. 382-387). IEEE.

Çakmak, T. (2017). Bilgi merkezlerinde metadata paylaşımı ve entegrasyonu. Hüseyin Odabaş ve Mehmet Ali Akkaya (Editör). *Bilişim Teknolojilerinin Bilgi Merkezlerine ve Hizmetlerine Etkileri içinde* (s.49-65). İstanbul : Hiperyayın. ISBN: 978-605-201-526-1

Çetin, V., & Yıldız, O. (2022). A comprehensive review on data preprocessing techniques in data analysis. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 28(2), 299-312.

Dang, S., & Ahmad, P. H. (2014). Text mining: Techniques and its application. *International Journal of Engineering & Technology Innovations*, 1(4), 22-25.

Demircioğlu, M., & Eşiyok, S. (2020). COVID-19 Salgını ile Mücadelede Kümeleme Analizi ile Ülkelerin Sınıflandırılması. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, 19(37), 369-389.

Demirtel, H., & Bayram, Ö. G. (2014). Elektronik belge yönetim sistemlerinin verimliliğe katkısı: Kalkınma Bakanlığı örneği. *Bilgi Dünyası*, 15(1), 91-101.

Digital Regulation Platform. (2023). National digital transformation strategy – mapping the digital journey. Erişim Adresi: <https://digitalregulation.org/national-digital-transformation-strategy-mapping-the-digital-journey/>

Ding, Y., Shi, Z., Xi, R., Diao, Y., & Hu, Y. (2024). Digital transformation, productive services agglomeration and innovation performance. *Heliyon*, 10(3).

Dong, Z., & Dong, Q. (2003, October). HowNet-a hybrid language and knowledge resource. In *International conference on natural language processing and knowledge engineering, 2003. Proceedings. 2003* (pp. 820-824). IEEE.

Elektronik İmza Kanunu. (2004). T.C. Resmi Gazete, Sayı: 25355, 23 Ocak 2004. Erişim adresi: <https://www.resmigazete.gov.tr/eskiler/2004/01/20040123.htm>

Elmahdy, A., Inan, H. A., & Sim, R. (2022). Privacy leakage in text classification: A data extraction approach. *arXiv preprint arXiv:2206.04591*.

Eren, A. & Kaya, M. D. (2019). Elektronik Belge Yönetim Sisteminin Algılanan Kullanılabilirliğinin Değerlendirilmesi. *Ergonomi*, 2(2), 88-100.

Eroğlu, Ş., & Külcü, Ö. (2014). TS 13298 çerçevesinde kurumsal bilgi sistemleri ve elektronik belge yönetimi standartlarının değerlendirilmesi: İçişleri Bakanlığı örneği. *Bilgi Dünyası*, 15(2), 327-352.

Ertam, F. (2018, September). Deep learning based text classification with Web Scraping methods. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)* (pp. 1-4). IEEE.

e-Yazışma Projesi Konulu (2017/21) Genelge. 2017. T.C. Resmi Gazete. Erişim adresi: <https://www.resmigazete.gov.tr/eskiler/2017/10/20171014-11.pdf>

Eymen, E. (2007). Tez Önerisi Nasıl Hazırlanır.

Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9), 76-82.

Gaikwad, S. V., Chaugule, A., & Patil, P. (2014). Text mining methods and techniques. *International Journal of Computer Applications*, 85(17).

García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. *Big data analytics*, 1, 1-22.

GeeksforGeeks. (2024a). *Lemmatization vs. Stemming: A Deep Dive into NLP's Text Normalization Techniques*. <https://www.geeksforgeeks.org/lemmatization-vs-stemming-a-deep-dive-into-nlps-text-normalization-techniques/#what-is-lemmatization>

GeeksforGeeks. (2024b). *Text mining in data mining*. <https://www.geeksforgeeks.org/text-mining-in-data-mining/>

GeeksforGeeks. (2025a). *NLP – How tokenizing text, sentence, words works?*
GeeksforGeeks. <https://www.geeksforgeeks.org/nlp-how-tokenizing-text-sentence-words-works/>

GeeksforGeeks. (2025b). *Apriori Algorithm*. Erişim adresi: <https://www.geeksforgeeks.org/apriori-algorithm/>

Gharehchopogh, F. S., & Khalifelu, Z. A. (2011, October). Analysis and evaluation of unstructured data: text mining versus natural language processing. In *2011 5th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1-4). IEEE.

Gizlilik Dereceli Belgelerde Uygulanacak Usul ve Esaslar Hakkında Yönetmelik. (2022). Resmi Gazete, Sayı: 31821, 26 Nisan 2022. Erişim adresi: <https://www.resmigazete.gov.tr/eskiler/2022/04/20220426-11.pdf>

Gonzalez-Compean, J. L., Telles, O., Lopez-Arevalo, I., Morales-Sandoval, M., Sosa-Sosa, V. J., & Carretero, J. (2019). A policy-based containerized filter for secure

information sharing in organizational environments. *Future Generation Computer Systems*, 95, 430-444.

Government Records Office Archives of Manitoba. (2023) Glossary of Records and Information Management Terms. 11.12.2023 tarihinde <https://www.gov.mb.ca/chc/archives/gro/recordkeeping/docs/glossary.pdf> adresinden erişilmiştir.

Griffin, A., Kaekopa, S., Mansfield, W., Millar, L., & Podolksy, L. N. (2009). Glossary of Terms. London: International Records Management Trust. 11.12.2023 tarihinde https://www.irmt.org/documents/educ_training/term%20modules/IRMT%20TERM%20Glossary%20of%20Terms.pdf adresinden erişilmiştir.

Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, 1(1), 60-76.

Gündoğdu, Ö. E., & Duru, N. (2016). Türkçe Metin Özetlemede Kullanılan Yöntemler. *18. Akademik Bilişim Konferansı-AB'16*.

Gürbüz, S., & Şahin, F. (2017). Sosyal Bilimlerde Araştırma Yöntemleri. Ankara: Seçkin Yayıncılık.

Güven, A. (2007). Türkçe belgelerin anlam tabanlı yöntemlerle madenciliği. *FBE, Yıldız Teknik Üniversitesi*.

Güven, Z., Diri, B., & Çakaloğlu, T. (2020). Comparison of n-stage Latent Dirichlet Allocation versus other topic modeling methods for emotion analysis. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35(4).

Hai, T. N., Van, Q. N., & Thi Tuyet, M. N. (2021). Digital transformation: Opportunities and challenges for leaders in the emerging countries in response to COVID-19 pandemic. *Emerging Science Journal*, 5(1), 21-36.

Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. Bradford Book.

Hanke, J., & Thiesse, F. (2017). Leveraging text mining for the design of a legal knowledge management system.

Harbert, T. (2021). Tapping the power of unstructured data. Erişim adresi: <https://mitsloan.mit.edu/ideas-made-to-matter/tapping-power-unstructured-data>

Hobbs, J. R. (2002). Information extraction from biomedical text. *Journal of biomedical informatics*, 35(4), 260-264.

Hsiao, M.-H. (2024), "Resource integration and firm performance through organizational capabilities for digital transformation", *Digital Transformation and Society*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/DTS-07-2023-0050>

IBM. (2024). *Methods for fuzzy search*. 23.06.2025 tarihinde <https://www.ibm.com/docs/en/db2/11.5.x?topic=methods-fuzzy-search> adresinden erişilmiştir.

Ignaczak, L., Goldschmidt, G., Costa, C. A. D., & Righi, R. D. R. (2021). Text mining in cybersecurity: A systematic literature review. *ACM Computing Surveys (CSUR)*, 54(7), 1-36.

ISO/IEC 27001 Bilgi Güvenliği Yönetim Sistemi Standardı. (2013). Ankara: Türk

https://tr.wikipedia.org/wiki/%C4%B0kili_say%C4%B1_sistemi#:~:text=%C4%B0kili%20say%C4%B1lar%20say%C4%B1lar%C4%B1n%202%20taban%C4%B1nda,g%C3%BCn%C3%BCm%C3%BCz%20bilgisayarlar%C4%B1n%C4%B1n%20neredeyse%20tamam%C4%B1nda%20kullan%C4%B1l%C4%B1rlar adresinden erişildi.

Jackson, J. (2002). Data mining; a conceptual overview. *Communications of the Association for Information Systems*, 8(1), 19.

Jamali M., Ali N.I., Maree M., Soomro, A.H. Soomro, S., & Memon, A. G. (2024). RSS Feeds Filtering from multiple sources using automated techniques of Natural Language Processing.

Joseph, S. R., Hlomani, H., & Letsholo, K. (2016). Data mining algorithms: an overview. *Neuroscience*, 12(3), 719-43.

Jurošková, Z. J., & Závodská, A. (2015). Using of Text Mining in Online Public Access Catalog. In *Knowledge Management in Organizations: 10th International Conference, KMO 2015, Maribor, Slovenia, August 24-28, 2015, Proceedings 10* (pp. 451-462). Springer International Publishing.

Kalra, V., & Aggarwal, R. (2018). Importance of Text Data Preprocessing & Implementation in RapidMiner. *ICITKM*, 14, 71-75.

Kandur, H. (2006). Elektronik belge yönetimi sistem kriterleri referans modeli (2. bs). Ankara: Devlet Arşivleri Genel Müdürlüğü.

Kannan, S., & Gurusamy, V. (2014). Preprocessing techniques for text mining. *International Journal of Computer Science & Communication Networks*, 5(1), 7-16.

Kara, S., Alan, Ö., Sabuncu, O., Akpınar, S., Cicekli, N. K., & Alpaslan, F. N. (2012). An ontology-based retrieval system using semantic indexing. *Information Systems*, 37(4), 294-305.

Karakaş, S. (2024). *Makine öğrenmesinde başarıyı tartmak: Temel metrikler ve anlamları*. İLGE. 17.06.2025 tarihinde <https://ilge.com.tr/makine-ogrenmesinde-basariyi-tartmak--temel-metrikler-ve-anlamlari> adresinden erişilmiştir.

Karasoy, H., & Babaoğlu, P. (2020). Türkiye’de elektronik devletten dijital devlete doğru. *Karadeniz Sosyal Bilimler Dergisi*, 12(23), 397-416.

Katz, G., Elovici, Y., & Shapira, B. (2014). CoBAn: A context based model for data leakage prevention. *Information sciences*, 262, 137-158.

Kayıtlı Elektronik Posta Sistemine İlişkin Usul ve Esaslar Hakkında Yönetmelik. (2011) T.C. Resmi Gazete, Sayı: 28036, 25.08.2011. Erişim adresi: https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=15224&MevzuatTur=7&MevzuatTe_rtip=5

Khan, Z., & Vorley, T. (2017). Big data text analytics: an enabler of knowledge management. *Journal of Knowledge Management*, 21(1), 18-34.

Kılınç, M., Tarhan, Ç., & Aydın, C. (2020). Kitle Fonlaması Projelerinin Karar Ağacı ve Rastgele Orman Algoritmalarıyla Sınıflandırılması. *Journal of Information Systems and Management Research*, 2(2), 16-25.

Kilinc, D., Borandag, E., Yucalar, F., Tunalı, V., Simsek, M., & Ozcift, A. (2016). Classification of scientific articles using text mining with kNN algorithm and R language. *Marmara Journal of Pure and Applied Sciences*, 3, 89-94.

Kiral, B. (2020). Nitel bir veri analizi yöntemi olarak doküman analizi. *Siirt Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 8(15), 170-189.

Kiwa, F. J., Chinofunga, S., Shangwa, M., & Chikondowa, T. (2022). Review on an integrated archives electronic document and records management systems. *Information Technologist (The)*, 19(2), 141-154.

Koyuncu Tunç, S., & Külcü, Ö. (2020). Elektronik belge yönetim sistemlerinin sezgisel değerlendirme yöntemi ile kullanılabilirlik açısından değerlendirilmesi. *Bilgi Dünyası*, 21(2), 269-297.

Kumbhar, R. (2012). *Library classification trends in the 21st century*. Elsevier.

Kutlutürk, L., & Özdemirci, F. (2023). Elektronik Belge Yönetim Sistemleri ve Kullanılabilirlik. *Türk Kütüphaneciliği*, 37(1), 4-28. <https://doi.org/10.24146/tk.1139012>

Lamba, M., & Madhusudhan, M. (2022). Text Mining for Information Professionals. Springer. <https://doi.org/10.1007/978-3-030-85085-2>.

Lee, J. H., & Segev, A. (2012). Knowledge maps for e-learning. *Computers & Education*, 59(2), 353-364.

Lee, R. S., & Lee, R. S. (2020). Ontological-Based Search Engine. *Artificial Intelligence in Daily Life*, 193-241.

Lei-Da Chen, Toru Sakaguchi & Mark N. Frolick (2000) Data Mining Methods, Applications, and Tools, *Information Systems Management*, 17:1, 65-70, DOI: 10.1201/1078/43190.17.1.20000101/31216.9

Li, P., & Yanchinda, J. (2024, June). The Customer Requirements About Thai SMEs Product Based on Customer Knowledge Management by Using Text Mining.

In 2024 5th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON) (pp. 1-5). IEEE.

Li, Q., & Brook, Y. F. (2006). Information mining: Integrating data mining and text mining for business intelligence. AMCIS 2006 Proceedings, 182.

Lin, F. R., & Hsueh, C. M. (2006). Knowledge map creation and maintenance for virtual communities of practice. *Information processing & management*, 42(2), 551-568.

Liu, M., Li, H., Li, C. and Yan, Z. (2023), "Digital transformation, financing constraints and enterprise performance", *European Journal of Innovation Management*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/EJIM-05-2023-0349>

Magnusson, J., Elliot, V. and Hagberg, J. (2022), "Digital transformation: why companies resist what they need for sustained performance", *Journal of Business Strategy*, Vol. 43 No. 5, pp. 316-322. <https://doi.org/10.1108/JBS-02-2021-0018>

MailOnline. (2023) Uber is charging users more if they have low phone battery, investigation claims. 27.12.2024 tarihinde <https://www.dailymail.co.uk/sciencetech/article-12010459/Uber-charging-users-low-phone-battery-investigation-claims.html> erişim tarihi 27.12.2024 adresinden erişilmiştir.

Malik, S., & Gupta, D. S. K. (2022). The importance of text mining for services management. *TTIDMKD*, 2(4), 28-33.

Margiono, A. (2021), "Digital transformation: setting the pace", *Journal of Business Strategy*, Vol. 42 No. 5, pp. 315-322. <https://doi.org/10.1108/JBS-11-2019-0215>

Menteş, S. ve Solhan, S. (2019) “Cumhurbaşkanlığı Hükümet Sistemi ile Birlikte Kamu’da Bilgi ve Belge Yönetimi” Bilgi Yönetimi ve Bilgi Güvenliği, 2019, ss.11-23. 11.12.2023 tarihinde <http://2019.ebeyas.org/wp->

[content/uploads/2020/01/BILGI_YONETIMI_VE-](content/uploads/2020/01/BILGI_YONETIMI_VE-)

[BILGI_GUVENLIGI_KITAP_23122019.pdf](#) adresinden erişilmiştir.

Microsoft. (2023). *Data mining algorithms (Analysis Services - Data Mining)*.

Microsoft Learn. 6.06.2025 tarihinde <https://learn.microsoft.com/en-us/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=asallproducts-allversions> adresinden erişilmiştir.

Mirza, S., Mittal, S., & Zaman, M. (2016). A review of data mining literature. *International Journal of Computer Science and Information Security (IJCSIS)*, 14(11), 437-442.

Moon, S., Shin, Y., Hwang, B. G., & Chi, S. (2018). Document management system using text mining for information acquisition of international construction. *KSCE Journal of Civil Engineering*, 22(12), 4791-4798.

Mukred, M., Hawash, B., Nofal, M. I., Mokhtar, U. A., & Yusof, Z. M. (2022, October). Big data and electronic records management systems for public sector organizations competency. In *2022 International Conference on Cyber Resilience (ICCR)* (pp. 1-7). IEEE.

Nadeem, K., Wong, S. I., Za, S., & Venditti, M. (2024). Digital transformation and industry 4.0 employees: Empirical evidence from top digital nations. *Technology in Society*, 76, 102434.

Noyan, M. (2019). Doğal Dil İşleme (Natural Language Processing). 27.12.2024 tarihinde <https://merveenoyan.medium.com/do%C4%9Fal-dil-i%C5%9Flleme-natural-language-processing-2d7c72daf245> adresinden erişilmiştir.

Oğul, B. B., & Ercan, G. (2016, May). Sentiment classification on Turkish hotel reviews. In *2016 24th Signal Processing and Communication Application Conference (SIU)* (pp. 497-500). IEEE.

Oğuzlar, A. (2003). Veri ön işleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (21).

Ojoawo, A. O., Fagbolu, O. O., Olaniyan, A. S., & Sonubi, T. A. (2014). Data leak protection using text mining and social network analysis. *Int J Eng Res Dev*, 10(12), 14-22.

Okuonghae, N., & Bakare-Fatungase, O. D. (2023). Cloud Computing: a tool for effective records and information management system in Nigerian academic libraries. *Folia Toruniensia*, 23, 35-52.

Oracle. (2021). What is business intelligence? 12.06.2025 tarihinde <https://www.oracle.com/business-analytics/what-is-business-intelligence/> adresinden erişilmiştir.

Ödük, M. N. (2019). *Bulanık mantık yöntemi ve uygulamaları*. İksad Yayınevi. 23.06.2025 tarihinde <https://iksadyayinevi.com/wp-content/uploads/2020/02/BULANIK-MANTIK-Y%C3%96NTEM%C4%B0-VE-UYGULAMALARI.pdf> adresinden erişilmiştir.

Önaçan, M. B. K., Medeni, T. D., & Özkanlı, Ö. (2012). Elektronik Belge Yönetim Sistemi EBYS'nin Faydaları ve Kurum Bünyesinde EBYS Yapılandırmaya Yönelik Bir Yol Haritası. *Sayıştay Dergisi*, (85), 1-26.

Özdemirci, F. (2019). “Kurumlar İçin EBYS ve E-Arşiv Sistemi İdari Yapılanma ve Yönetim Süreci: Bileşenler ve Entegrasyonlar” Bilgi Yönetimi ve Bilgi Güvenliği,

2019, ss.11-23. 11.12.2023 tarihinde http://2019.ebeyas.org/wp-content/uploads/2020/01/BILGI_YONETIMI_VE-BILGI_GUVENLIGI_KITAP_23122019.pdf adresinden erişilmiştir.

Özdemirci, F., & Torunlar, M. (2018). Bilgi-değişim-siber güvenlik-bağımsızlık. *Bilgi Yönetimi*, 1(1), 78-83.

Özdemirci, F., Gökkurt, Ö., Torunlar, M., Saraç, S., & Yalçinkaya, B. (2013). Elektronik Belge Yönetimi ve Arşivleme Sistemi Geçiş Süreci ve Uygulama Yönetimi.

Özdemirci, F., Savaş, A., & Akdoğan, Z. (2017). EBYS Uygulaması e-Arşiv midir? TÜRK SAT-Ankara Üniversitesi BEYAS Koordinatörlüğü e-Arşiv Deneyimi ile Yeni Yaklaşımlar. *Bilgi Sistemleri ve Bilişim Yönetimi: Beklentiler ve Yeni Yaklaşımlar*, 35-46.

Özkan, U. B. (2023). *Eğitim bilimleri araştırmaları için doküman inceleme yöntemi* (6. baskı). Pegem Akademi. <https://depo.pegem.net/9786052417232.pdf>

Özkol, İ. *Ulusal bilgi politikaları bağlamında belge ve arşiv yönetimi stratejilerinin geliştirilmesi* (Doctoral dissertation, Ankara Üniversitesi).

Parrish Jr, J. L., & Courtney, J. F. (2007). Electronic records management in local government agencies: The case of the clerk of courts office in Lake County Florida. *Information Systems Management*, 24(3), 223-229.

Pfeifer, K., & Peukert, E. (2013). Mapping Text Mining Taxonomies. In *Kdir/kmis* (pp. 5-16).

Pollitt, E. (Tarih Yok). *Text mining vs. text analysis vs. text analytics: Understanding the differences*. Displayr. 6.06.2025 tarihinde <https://www.displayr.com/text-mining-vs-text-analysis-vs-text-analytics-understanding-the-differences/> adresinden erişilmiştir.

Public Record Office. (1999) Functional Requirements for Electronic Records Management Systems. 11.12.2023 tarihinde <https://cdn.nationalarchives.gov.uk/documents/requirements.pdf> adresinden erişilmiştir.

QuestionPro. [Tarih Yok]. *Betimleyici araştırma: Tanım, özellikler, yöntemler, örnekler ve avantajlar.* 9.06.2025 tarihinde <https://www.questionpro.com/blog/tr/betimleyici-arastirma-tanim-ozellikler-yontemler-ornekler-ve-avantajlar/> adresinden erişilmiştir.

Resmi Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik. (2015). Resmi Gazete, Sayı: 29255, 2 Şubat 2015. Erişim adresi: <https://www.resmigazete.gov.tr/eskiler/2015/02/20150202.pdf>

Resmi Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik. (2020). Resmi Gazete, Sayı: 31151, 10 Haziran 2020. Erişim adresi: <https://www.mevzuat.gov.tr/mevzuatmetin/21.5.2646.pdf>

Riggs, H., Tufail, S., Parvez, I., Tariq, M., Khan, M. A., Amir, A., ... & Sarwat, A. I. (2023). Impact, vulnerabilities, and mitigation strategies for cyber-secure critical infrastructure. *Sensors*, 23(8), 4060.

Sak, R., Şahin Sak, İ. T., Öneren Şendil, Ç., Nas, E. (2021). Bir araştırma yöntemi olarak doküman analizi. *Kocaeli Üniversitesi Eğitim Dergisi*, 4(1), 227-256. <https://doi.org/10.33400/kuje.843306>

Salomão, A. (2023, Aralık 20). *Betimleyici çalışma nedir ve araştırmada neden önemlidir?* Mind the Graph. <https://mindthegraph.com/blog/tr/betimleyici-calisma-nedir/>

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620.

Salton, G., Wu, H., & Yu, C. T. (1981). The measurement of term importance in automatic indexing. *Journal of the American Society for Information Science*, 32(3), 175-186.

Sayed, A., & Al Muqrishi, A. (2017). IBRI-CASANTO: Ontology-based semantic search engine. *Egyptian Informatics Journal*, 18(3), 181-192.

Schieber, A. (2014). Idea Mining: Wissensmanagement und Text Mining im Innovationsprozess.

Sharma, D., Kumar, B., & Chand, S. (2018, October). Trend analysis in machine learning research using text mining. In *2018 international conference on advances in computing, communication control and networking (ICACCCN)* (pp. 136-141). IEEE.

Sharma, S., & Arjunan, T. (2023). Natural language processing for detecting anomalies and intrusions in unstructured cybersecurity data. *International Journal of Information and Cybersecurity*, 7(12), 1-24.

Siber Güvenlik Kanunu (32846 S.Y.)". Resmî Gazete, 7545, Mart 2025.

Silwattananusarn, T., & Tuamsuk, K. (2012). Data mining and its applications for knowledge management: a literature review from 2007 to 2012. *arXiv preprint arXiv:1210.2872*.

Smith, T., Stiller, B., Guszczka, J., & Davenport, T. (2019). Analytics and AI-driven enterprises thrive in the Age of With. *Deloitte Insights*, 16.

Solhan, S. (2016). Fizikselden Elektroniğe; Belge Yönetim ve Arşivleme Sürecinin Sürdürülebilirliği. *KURUMSAL BELLEKLERİN GELECEĞİ*, 49.

Songur, M. (2025). *Makine öğrenimi algoritmaları rehberi: Hangi algoritma ne işe yarar?* Mehmet Songur. 17.06.2025 tarihinde <https://www.songurmehmet.com/post/makine-%C3%B6%C4%9Frenimi-algoritmalar%C4%B1-rehberi-hangi-algoritma-ne-i-%C5%9Fe-yarar> adresinden erişilmiştir.

Speak Ai. (Tarih Yok). *Ücretsiz metin analiz aracı*. 6.06.2025 tarihinde <https://speakai.co/tr/araclar/metin-analiz-araci/> adresinden erişilmiştir.

Standart Dosya Planı Konulu (2005/7 Sayılı) Genelge. (2005). T.C. Resmi Gazete. Erişim adresi: <https://www.resmigazete.gov.tr/eskiler/2005/03/20050325-10.htm>

Stapleton, S. C., Dinsmore, C. S., Van Kleeck, D., & Ma, X. (2021). Computer-assisted indexing complements manual selection of subject terms for metadata in specialized collections. *College & Research Libraries*, 82(6), 792.

Subaşıoğlu, F. (2022) Araştırma Yöntemleri Ders Notları.

Sukanya, M., & Biruntha, S. (2012, August). Techniques on text mining. In *2012 iee international conference on advanced communication control and computing technologies (icaccct)* (pp. 269-271). IEEE.

Svård, P. (2017). Enterprise content management, records management and information culture amidst e-government development. Chandos Publishing.

Swarnkar, N., & Thakar, C. (2023). *Bag of Words: Approach, Python Code, Limitations*. QuantInsti. <https://blog.quantinsti.com/bag-of-words/>

Şen, S. (2015). *Bilimsel araştırma yöntemleri*. <https://sedatsen.com/wp-content/uploads/2015/02/bilim4.pdf>

T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (2020). Bilgi ve İletişim Güvenliği Rehberi. Erişim adresi: https://cbddo.gov.tr/SharedFolderServer/Genel/File/bg_rehber.pdf

T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (2020). e-Yazışma Teknik Rehberi 2.0. Erişim Adresi: https://cbddo.gov.tr/SharedFolderServer/Projeler/File/EYP_2.0/EYP2.0_teknik-rehberi.pdf

T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (2021). Bilgi ve İletişim Güvenliği Denetim Rehberi. Ankara: T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi.

T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. Erişim Adresi: <https://cbddo.gov.tr/>

T.C. Cumhurbaşkanlığı İdari İşler Başkanlığı Destek ve Mali Hizmetler Genel Müdürlüğü Bilgi ve Belge Yönetimi Daire Başkanlığı. (2022). Gizlilik Dereceli Belgelerde Uygulanacak Usul Ve Esaslar Hakkında Yönetmelik Kılavuzu. Erişim Adresi: <https://www.tccb.gov.tr/assets/dosya/resmiyazisma/dosyalar/gizlilik-kilavuzu.pdf>

T.C. Cumhurbaşkanlığı İletişim Başkanlığı. Elektronik Belge Yönetim Sistemi Uygulama Yönergesi. 11.12.2023 tarihinde <https://www.iletisim.gov.tr/uploads/docs/%C4%B0leti%C5%9Fim%20Ba%C5%9Fkanl%C4%B1%C4%9F%C4%B1%20Elektronik%20Belge%20Y%C3%B6netim%20Sistemi%20Uygulama%20Y%C3%B6nergesi.pdf> adresinden erişilmiştir.

T.C. Kalkınma Bakanlığı. (2012). Birlikte Çalışabilirlik Esasları Rehberi Sürüm

2.1. Erişim adresi: http://www.bilgitoplumu.gov.tr/wp-content/uploads/2014/04/Birlikte_Calisabilirlik_Esaslari_Rehberi_2.1.pdf

Tahliz İstatistik. (2024, 29 Ağustos). *Makine öğrenmesinde kümeleme analizi: Teknikler, algoritmalar, uygulama alanları ve zorluklar*. 17.06.2025 tarihinde <https://www.tahliz.com/makine-ogrenmesinde-kumeleme-analizi/> adresinden erişilmiştir.

Talib, R., Hanif, M. K., Ayesha, S., & Fatima, F. (2016). Text mining: techniques, applications and issues. *International journal of advanced computer science and applications*, 7(11), 414-418.

Tamjid, S., Nooshinfard, F., Hosseini Beheshti, M. S., Hariri, N., & Babalhavaeji, F. (2024). Semi-automated ontology development scheme via text mining of scientific records. *The Electronic Library*, 42(2), 230-254.

Tamtürk, E. (2017). Kamu yönetiminde elektronik belge yönetim sistemi. *Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi*, 5(3), 851-863.

Tandel, S. S., Jamadar, A., & Dudugu, S. (2019, March). A survey on text mining techniques. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)* (pp. 1022-1026). IEEE.

The National Archives and Records Administration (NARA) içinde Electronic Records Management [ERM]. 11.12.2023 tarihinde <https://www.archives.gov/records-mgmt/initiatives/context-for-erm.html> adresinden erişilmiştir.

Tonta, Y. (1995). Bilgi erişim sistemleri. *Türk Kütüphaneciliği*, 9(3), 302-314.

TS 13298. (2009). Elektronik belge yönetimi. Ankara: Bilgi Teknolojileri ve İletişim İhtisas Grubu Türk Standartları Enstitüsü.

TS 13298. (2015) Elektronik Belge ve Arşiv Yönetim Sistemi standardı (TS 13298). Ankara: Türk Standartları Enstitüsü.

Tunç, S. K., & Külçü, Ö. (2020). Elektronik belge yönetim sistemlerinin sezgisel değerlendirme yöntemi ile kullanılabilirlik açısından değerlendirilmesi. *Bilgi Dünyası*, 21(2), 269-297.

Umut, G., & Külçü, Ö. (2014). Elektronik belge yönetimi uygulamalarında karşılaşılan sorunların analizi ve çözüm önerileri: Kalkınma Bakanlığı örneği. *Bilgi Dünyası*, 15(1), 102-124.

Uysal, Y., Kurban, S., & Çığman, M. Z. (2023). Cumhurbaşkanlığı Dijital Dönüşüm Ofisi ve E-Yönetişim. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (78), 211-231.

Vasques, D. G., Rezende, S. O., & Martins, P. S. (2018, October). A semantic approach to uncovering implicit relationships in textual databases. In *2018 XLIV Latin American Computer Conference (CLEI)* (pp. 490-499). IEEE.

Von Solms, R., & Van Niekerk, J. (2013). From information security to cyber security. *computers & security*, 38, 97-102.

Wang, H., Jiao, S., Bu, K., Wang, Y., & Wang, Y. (2023). Digital transformation and manufacturing companies' ESG responsibility performance. *Finance Research Letters*, 58, 104370.

Wijayasekara, D., Manic, M., & McQueen, M. (2014, October). Vulnerability identification and classification via text mining bug databases. In *IECON 2014-40th Annual Conference of the IEEE Industrial Electronics Society* (pp. 3612-3618). IEEE.

Williams, R. V. (2010). Hans Peter Luhn and Herbert M. Ohlman: Their roles in the origins of keyword-in-context/permutation automatic indexing. *Journal of the american society for information science and technology*, 61(4), 835-849.

Willis, C., & Losee, R. M. (2013). A random walk on an ontology: Using thesaurus structure for automatic subject indexing. *Journal of the American Society for Information Science and Technology*, 64(7), 1330-1344.

Yalçinkaya, B. (2016). Elektronik belge yönetim sistemi EBYS uygulamalarında başarı faktörü ve fayda analizi. *AJIT-e: Academic Journal of Information Technology*, 7(23), 67-96.

Yang, H. (2018). Data preprocessing. *Pennsylvania State University: Citeseer*.

Yazıcı, S. ve Özdemirci, F. Bilgi Yönetim Sistemlerinin Birlikte Çalışabilirlik Gereksinimleri ve Elektronik Belge Yönetim Sistemi Birlikte Çalışabilirlik Olgunluk Modeli. *Bilgi Yönetimi*, 2(2), 84-104.

YILDIZ, A. (2022). Bir Araştırma Metodolojisi Olarak Sistematik Literatür Taramasına Genel Bakış. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 22(Özel Sayı 2), 367-386.

Yüksek Öğretim Kurulu. (2023). Yükseköğretim Üst Kuruluşları ve Yükseköğretim Kurumları Saklama Süreli Standart Dosya Planı V.4. *Erişim adresi: https://www.yok.gov.tr/Documents/Universiteler/Standart_Dosya_Planı.pdf*

Zhan, Z. (2025). Comparative Analysis of TF-IDF and Word2Vec in Sentiment Analysis: A Case of Food Reviews. In *ITM Web of Conferences* (Vol. 70, p. 02013). EDP Sciences.

Zhu, F., Patumcharoenpol, P., Zhang, C., Yang, Y., Chan, J., Meechai, A., ... & Shen, B. (2013). Biomedical text mining and its applications in cancer research. *Journal of biomedical informatics*, 46(2), 200-211.

Барабаш, О. В., & Колумбет, В. П. (2022). Інтелектуальний аналіз та обробка текстових документів на основі мультиагентного підходу. *Інформатика та математичні методи в моделюванні*, 12(4), 296-305.



ÖZET

Bu çalışma, Elektronik Belge Yönetim Sistemleri (EBYS) içindeki yapılandırılmamış metinsel veri karmaşası ve insan kaynaklı hataların yol açtığı bilgi yönetimi sorunlarını ele almaktadır. Amacı, bilgi ve belge yönetimi perspektifiyle geliştirilen metin madenciliği tabanlı yenilikçi bir mimari ile EBYS'lerde dijital dönüşümü sağlamaktır.

Tez, metin madenciliği tekniklerini EBYS süreçlerine entegre ederek üç ana alanda köklü iyileştirmeler ve otomasyon sunar:

Bilgi Tanımlama: Manuel metadata girişlerindeki tutarsızlıkları ve hatalı sınıflamaları gidermek için ontoloji tabanlı metin madenciliğiyle otomatik metadata oluşturma, sınıflama ve kümeleme yaparak belgelere kesin isabetli ve etkili erişim sağlar.

İş Zekâsı: EBYS'lerdeki metinsel verilerden öngörülü bilgi ve iş zekâsı elde edilmesini sağlayarak karar alma süreçlerini güçlendirir; bilgi haritaları aracılığıyla gizli ilişki ve desenleri ortaya çıkarır.

Güvenlik: İnsan hatasından kaynaklanan güvenlik açıklarına karşı otomatik tedbirler geliştirir; kaynak kodlarındaki zafiyetleri ve hassas bilgileri otomatik tespit ederek veri sızıntılarını ve yetkisiz erişimi engeller.

Bu çalışma, EBYS'lerin sadece bir belge depolama aracı olmaktan öte, akıllı, öngörülü ve güvenli bir bilgi merkezine dönüşmesinin yol haritasını sunmaktadır

Anahtar Sözcükler

Elektronik Belge Yönetim Sistemi (EBYS), Metin Madenciliği, Yapılandırılmamış Metinsel Veriler, Ontoloji, Metadata, Bilgi Erişim, Bilgi Keşfi, Belge Yönetimi, Dijital Dönüşüm.

ABSTRACT

This study addresses the information management problems caused by the large volume of unstructured textual data within Electronic Records Management Systems (ERMS) and human errors resulting from manual processes. Its aim is to enable digital transformation in ERMS through an innovative text mining-based architecture developed from a knowledge and document management perspective.

The thesis offers radical improvements and automation in three main areas by integrating text mining techniques into ERMS processes:

Information Identification: To eliminate human-induced issues like inconsistencies in metadata entries and erroneous classifications, it provides automatic metadata creation, classification, and clustering using ontology-based text mining techniques, ensuring precise and effective access to documents.

Business Intelligence: It strengthens decision-making processes by enabling the extraction of predictive information from textual data in ERMS; it reveals hidden relationships and patterns through knowledge maps.

Security: It develops automatic measures against security vulnerabilities arising from human error; it aims to prevent data leaks and unauthorized access by automatically detecting vulnerabilities in source codes, sensitive information, and anomalies in emails.

This study presents a roadmap for ERMS to transform from merely a document storage tool into an intelligent, predictive, and secure information hub.

Keywords

Electronic Records Management System (ERMS), Text Mining, Unstructured Textual Data, Ontology, Metadata, Information Retrieval, Knowledge Discovery, Records Management, Digital Transformation.

