

T.C.
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI
SAYISAL YÖNTEMLER BİLİM DALI

YÜKSEK LİSANS TEZİ

SAĞLIK KONUSUNDA TWİTTER ÜZERİNDEN
YAPILAN PAYLAŞIMLAR VE ÜLKELER
BAZINDA BİR ARAŞTIRMA

BAYSAL SAMET ÇELİK
2501190850

TEZ DANIŞMANI
PROF. DR. ERGÜN EROĞLU

İSTANBUL 2022

ÖZ

SAĞLIK KONUSUNDA TWITTER ÜZERİNDEN YAPILAN PAYLAŞIMLAR VE ÜLKELER BAZINDA BİR ARAŞTIRMA

BAYSAL SAMET ÇELİK

İnsanlar deneyimlerini erişim kolaylığı ve ekosistemi olan sosyal medya araçlarında paylaşabilmekte, böylece karşılıklı etkileşimde buldukları kişiler ve topluluklar ile üstel artan veri yığınlarına neden olunmaktadır. Geçen kısa süre içerisinde küresel olarak yaşanan salgın, sağlık bazında veri yığınlarına neden olmuş, münferit değişkenler ile son kullanıcıların analizlerinin yapılmasını sağlamıştır.

Bu tez çalışmasında seçilen bir sosyal medya aracı üzerinden sağlık bazında atılan ilgili zamandaki paylaşımlar veri gürültüsünden arındırılıp analiz edilmiş ve görselleştirilmiş olup, ülkeler bazında salgın sürecinde sağlık konusundaki paylaşılan düşüncelerin durumu ve yorumlanması sağlanmıştır.

Anahtar Kelimeler: Metin Madenciliği, Veri Analizi, Python, Sağlık, Pandemi

ABSTRACT

THE RESEARCH OF THE TWEETS INCLUDING HEALTH SUBJECT AND BASED ON VARIOUS COUNTRIES

BAYSAL SAMET ÇELİK

People can share their experiences on social media tools with ease of access and ecosystem, thus causing exponentially increasing data stacks with the people and communities they interact with. The global epidemic in a short period of time has led to data stacks on the basis of health, enabling the analysis of individual variables and end users.

In this thesis, the shares at the relevant time on health basis through a selected social media tool were analyzed and visualized after data cleaning, and the status and interpretation of the shared thoughts on health at the time of the epidemic on the basis of countries were provided.

Keywords: Text Mining, Data Analysis, Python, Health, Pandemic

ÖNSÖZ

Çok kıymetli aileme sonsuz destekleri ve emekleri için minnet duyarım.

Tez yazım sürecimde değerli zamanını esirgemeyen ve mentor yaklaşımı ile gelişimime katkıda bulunan Sn. Prof. Dr. Ergün Erođlu'na ve Sn. Prof. Dr. Umman Tuđba Gürsoy Hocalarıma teşekkürlerimi sunarım.

Baysal Samet ÇELİK

İSTANBUL, 2022

İÇİNDEKİLER

| | |
|------------------------|-----|
| ÖZ..... | ii |
| ABSTRACT..... | iii |
| ÖNSÖZ | iv |
| İÇİNDEKİLER..... | v |
| ŞEKİLLER LİSTESİ | ix |
| TABLolar LİSTESİ | x |
| GİRİŞ..... | 1 |

BİRİNCİ BÖLÜM

LİTERATÜR TARAMASI

| | |
|--|----|
| 1.1 Twitter Duygu Sınıflandırması Analizi..... | 4 |
| 1.2. Açık Kaynak İstihbaratı: Twitter Analizi..... | 5 |
| 1.3 Siyasi Bazda Twitter Analizi | 6 |
| 1.4 Twitter’da Aktif Öğrenme Stratejileri Analizi | 7 |
| 1.5 Sosyal Medyada Akan Veri Duygu Analizi..... | 8 |
| 1.6 Türkçe Radyoloji Raporlarında Metin Madenciliği..... | 9 |
| 1.7 Reklam İçerikli Epostaların Metin Madenciliği Yöntemler ile Otomatik Tespiti..... | 10 |
| 1.8 BioBERT: Biyomedikal Metin Madenciliği İçin Önceden Eğitilmiş Bir Biyomedikal Dil Temsil Modeli..... | 10 |
| 1.9 Finans Sektöründe Büyük Veri Analizi için Metin Madenciliği..... | 11 |
| 1.10 Metin Madenciliği ve Siber Suç..... | 11 |
| 1.11 Biyotıpta Metin Madenciliği ve Ontolojiler: Ham Metni Anlamlandırmak | 12 |
| 1.12 Metin Madenciliğinde Benzerlik Ölçümleri Üzerine Bir Araştırma..... | 12 |
| 1.13 İnsan Fenomeninin Bir Metin Madenciliği Analizi | 13 |
| 1.14 Metin Madenciliği ile E-Ticaret Sitelerinin Belirlenmesi..... | 13 |
| 1.15 KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi | 13 |
| 1.16 Metin Madenciliği Teknikleri İle Şirketlerin Vizyon İfadelerinin Analizi | 14 |
| 1.17 Muhasebede Analiz Yöntemi Olarak Metin Madenciliği | 14 |
| 1.18 Örgütsel Araştırmada Metin Madenciliği | 15 |
| 1.19 Metin Madenciliği Teknikleri İle Sosyal Ağlarda Bilgi Keşfi..... | 15 |

| | |
|--|----|
| 1.20 Metin Madenciliği: İmkanlar, Yöntemler ve Kısıtlar | 15 |
| 1.21 Bir Metin Madenciliği Uygulaması: Vosviewer | 16 |
| 1.22 Çağrı Merkezi Metin Madenciliği Yaklaşımı | 16 |
| 1.23 Sosyal Ağlarda Akan Veri Madenciliği | 17 |
| 1.24 Edirne Mektuplarının Metin Madenciliği Analizi..... | 17 |
| 1.25 NoSQL Veri Tabanları Üzerinde Bir Metin Madenciliği Uygulaması | 18 |
| 1.26 Metin Madenciliği Tabanlı Bildirim Takip Sistemi..... | 19 |
| 1.27 Metin Madenciliğinde Anahtar Kelime Seçimi Bir Üniversite Örneği..... | 19 |
| 1.28 Metin Madenciliği Yöntemleri ile Twitter Verilerinden Bilgi Keşfi | 20 |

İKİNCİ BÖLÜM

KAVRAMSAL ÇERÇEVE

| | |
|---|----|
| 2.1 Sosyal Medya..... | 20 |
| 2.1.1 Katılım | 22 |
| 2.1.2 Açıklık..... | 23 |
| 2.1.3 Karşılıklı Konuşma | 23 |
| 2.1.4 Topluluk..... | 23 |
| 2.1.5 Bağlantı | 24 |
| 2.2 Sağlık Nedir? | 29 |
| 2.3 Sağlığı Etkileyen Faktörler Nelerdir? | 30 |
| 2.3.1 Kültür | 30 |
| 2.3.2 Cinsiyet | 30 |
| 2.3.3 Sağlık Hizmetleri | 31 |
| 2.3.4 Biyoloji ve Genetik Faktör..... | 31 |
| 2.3.5 Sağlıklı Çocuk Gelişimi | 31 |
| 2.3.6 Kişisel Sağlık Uygulamaları ve Başa Çıkma Becerileri..... | 32 |
| 2.3.7 Çevresel Faktörler | 32 |
| 2.3.8 Çalışma Koşulları..... | 33 |
| 2.3.9 Eğitim ve Okuryazarlık..... | 34 |
| 2.3.10 Sosyal Destek..... | 34 |
| 2.3.11 Gelir ve Sosyal Durum..... | 34 |
| 2.3.12 Stres ve Sağlık..... | 35 |
| 2.4 Geçmişten Günümüze İnsanlık İçin Yaşanmış Büyük Hastalıklar | 35 |
| 2.4.1 Cüzzam | 36 |

| | |
|---|----|
| 2.4.2 Çiçek Hastalığı..... | 36 |
| 2.4.3 Veba Hastalığı..... | 36 |
| 2.4.4 Kolera Hastalığı | 37 |
| 2.4.5 Koronavirüs (Covid-19)..... | 38 |
| 2.5 Ülkere Göre Sağlık Sistemi Değerlendirilmesi..... | 38 |
| 2.5.1 Çin..... | 38 |
| 2.5.2 Amerika Birleşik Devletleri | 39 |
| 2.5.3 İngiltere | 40 |
| 2.5.4 Türkiye | 41 |

ÜÇÜNCÜ BÖLÜM

VERİ VE YÖNTEM

| | |
|---|----|
| 3.1 Metin Madenciliği Nedir?..... | 41 |
| 3.1.1 Yapısal Veri Nedir? | 47 |
| 3.1.2 Yapısal Olmayan Veri Nedir?..... | 48 |
| 3.1.3 Yarı Yapısal Veri Nedir? | 48 |
| 3.1.4 Süreçler | 49 |
| 3.2 Veri Madenciliği (Data Mining) | 57 |
| 3.3 Bilgi Çıkarımı (Information Extraction) | 58 |
| 3.4 Doğal Dil İşleme Süreçleri (NLP) | 58 |
| 3.5 Bilgi Alma (Information Retrieval) | 59 |
| 3.6 Makine Öğrenmesi..... | 60 |
| 3.6.1 Makine Öğrenmesi ile Çözülebilir Problemler | 65 |
| 3.6.2 Makine Öğrenmesi Algoritma Türleri | 66 |
| 3.7 Yapay Zeka | 69 |
| 3.8 Araştırmada Kullanılan Veriler ve Araçlar | 73 |

DÖRDÜNCÜ BÖLÜM

UYGULAMA

| | |
|--|----|
| 4.1. Veriseti Frekans Analizi, Görselleştirme ve Kalitatif Bazda Yorumlanması..... | 76 |
| 4.1.1 Çin..... | 76 |
| 4.1.2 İngiltere | 77 |
| 4.1.3 Türkiye..... | 78 |
| 4.1.4 Amerika Birleşik Devletleri | 78 |
| 4.2 Duygu Analizi ve Ülke Karşılaştırması | 79 |

| | |
|----------------------|-----------|
| SONUÇ | 82 |
| KAYNAKÇA..... | 84 |



ŞEKİLLER LİSTESİ

| | |
|---|----|
| Şekil 1.1: Kampanya Karşılaştırması | 6 |
| Şekil 1.2: Adayların Kullanım Etkinliği | 7 |
| Şekil 3.1: Metin Madenciliği Akışı | 44 |
| Şekil 3.2: Metin Madenciliği Adımları | 46 |
| Şekil 3.3: Yazılım Dili Karşılaştırması | 54 |
| Şekil 4.1: Çin Görselleştirmesi | 76 |
| Şekil 4.2: İngiltere Görselleştirmesi | 77 |
| Şekil 4.3: Türkiye Görselleştirmesi | 78 |
| Şekil 4.4: Amerika Birleşik Devletleri Görselleştirmesi | 79 |

TABLolar LİSTESİ

| | |
|---|----------|
| Tablo 1.1: Ayata, Model Karşılaştırması | 4 |
| Tablo 1.2: Atalı, Model Karşılaştırması..... | 5 |
| Tablo 1.3: Adayların Kullanım Etkinliği | 7 |
| Tablo 1.4: Akan Veri Model Karşılaştırılması..... | 8 |
| Tablo 1.5: Akan Veri Model Karşılaştırılması 2 | 9 |

GİRİŞ

Dünya üzerinde insanların farklı lokasyonlardaki kümülatif birikimleri, yaşam tarzları ve düşünce yapılarına yansımaktadır. Bu yansıma günlük hayatlarındaki aksiyonları, içerisinde buldukları durumu yansıtmaya şekli, iletişim tercihleri gibi birçok örnek ile ortaya konulabilmektedir. Son kullanıcıya neredeyse herhangi bir ön şart kısıtı koymaksızın doğrudan erişilebilecek ortam sunan sosyal medya araçları ile beraber iletişim hızlı ve küresel bir hale gelmiştir. İnsanlar deneyimledikleri sorunlar için çözüm aramak, belli bir kitle ile iletişime geçmek gibi çeşitli nedenlerle paylaşmak istedikleri mesajları ilgili sosyal medya platformlarında içerik olarak üretilmesine olanak sağlamaktadır.

Bu küresel iletişim tercihi bilgi kaynağı alternatifini olarak insanlar nezdinde kabul görmekte ve lokasyonlar bazında veri yığınlarına, dolayısıyla analizlere olanak vermektedir. Çin'in 31 Aralık 2019 tarihinde gizemli olarak adlandırılan bir solunum yolu rahatsızlığını Dünya Sağlık Örgütü'ne bildirmesiyle rastlanılan Covid-19 salgını Dünya gündemini meşgul eder hale geldi. Kişiler tarafından yaşanan olumsuz deneyimler ve üstel olarak artan vaka sayıları ile beraber gündemin yegane konusu olmaya başladı. Başlarda doğrudan tedavi çözümü olmayan bu salgın için kişiler düşüncelerini paylaşmaya, anında diyebileceğimiz hızla yeni içerikler üretmek ve tüketerek kendileri için enformasyon sağlamaya çalıştılar. Paylaşımlar ülkelerdeki vaka detay bilgileri, tesislerin durumu, koşulları, uygulanan sosyal ve ekonomik politikaları içermektedir. Bu durum uluslararası ve ulusal kurumlara gelen detaylı bilgiler ile aksiyon planlarını şekillendirilmesine ve sorunlara daha çabuk çözüm üretilmesine neden olmaktadır. Tüm bu kümülatif bilgi yığınları karar alıcıları da beslemektedir. Sosyal medya üzerinden yapılan paylaşımlar kurumlar tarafından ürün geri bildirim, potansiyel pazar yoğunluğu, ürün sınıflandırılması ve kampanya planlaması gibi bazı konularda da kullanılmaktadır. Süreç içerisinde yaşanan virüs varyasyonlarından dünyanın diğer ucundaki kişilerin de haberdar olması veya haber kaynaklarına düşmesi gibi nedenlerle kişilerin önlem alması durumun farkındalığını arttırdığı için daha çabuk hale gelmiştir. Bilgi yığınlarındaki doğrudan gelen veriyi

raporlamak son kullanıcılar tarafından anlaşılır olmayacaktır. Bundan dolayı üstel olarak gelişen teknoloji ile beraber doğrudan akan veriyi analiz edecek, üzerine çeşitli algoritmalar veya görselleştirme metotları uygulanarak yorumlanabilir hale getirebilecek araçlar ortaya çıkmıştır.

Oluşan her içerik doğrudan görselleştirme veya analiz sürecine girmemektedir. Özellikle veri analizi proje zamanlamasının oldukça büyük bir kısmını temsil eden veri temizleme adımı önemli bir süreç olarak görünmektedir. Arama motorlarında hız ve doğruluk terimlerine cevap olan etiketleme mantığı çok etkileşimli birçok platform tarafından da kullanılmaktadır. Bu sayede isterlere doğrudan cevap veren içeriklere hızlı ve yüksek doğruluk oranı ile ulaşılmaktadır. Bu gibi isterlere cevap vermek için platformlar geliştiricilere ücretsiz olarak belli bir gizlilik çerçevesinde bilgi paylaşımında bulunmaktadır.

Hızlıca akan veriyi ise tamamen erişilebilir olan Python, R, Julia gibi yazılım dilleriyle kod yazarak ve aynı şekilde tamamen erişilebilir olan, kalabalıkların bilgeliği felsefesini baz alan görselleştirme odaklı veya formüllerin uygulanabilir hali olan kütüphane yapıları ile doğrudan çıktı alınabilir hale getirmiştir.

Tüm bu teknolojik gelişmeler ve bilgi yoğunluğu ihtiyaç doğrultusunda talep edilen verilerin ön işlemlerden geçerek analizlerinin ve durum tespitinin yapılmasını erişilebilir ve hızlı hale getirmiştir. Karar alıcıların bu anahtar göstergeleri baz alarak planlar yapması da verimliliği arttırmıştır.

Sağlık bir toplumu ve dolayısıyla insan figürünü oluşturan önemli yapıtaşlarından biridir. Teknolojinin içerisindeki dinamikliği sayesinde, üstel bir gelişim sürdürmesi içerisinde birden çok evre barındırmasına neden olmuştur. Bulduğumuz teknoloji evresi, çok daha fazla hayatın içerisine müdahil olan, insanların özel yaşamlarında günlük aksiyonlarına veya iş yaşamların günlük operasyonel süreçlerine dahil ettiği çok mühim bir parça haline gelmiştir.

Teknolojinin barındırdığı alternatiflerle insan figürünün etkileşim içerisinde bulunduğu, karşılıklı ihtiyaç analizleri bazında beslenen, edindiği ve üzerine üstel gelişim göstereceği etkileşim çıktılarında yaşamlarını devam ettirebilen iki değişken

olarak hayatlarına devam etmektedirler. Tüm bu etkileşim ve çıktıları teknoloji depolarında tutularak, birden çok disiplin için kullanılacak kaynak haline gelmektedir.

Tüm bu çıktılar içerisinde yapısal ve/veya yapısal olmayan, ilk bakışta son kullanıcının anlamlandırabildiği ve/veya anlamlandıramadığı, kümülatif yığınlar olabilmektedir. Bahsi geçen tüm bu kümülatif yığınlara büyük veri denebilir. Çok fazla gündemde olan, yeni çağın altını ve petrolüne benzetilen veri metası, iş hayatında disiplin farketmeksizin, var olan süreçlerin daha verimli işlemesi, yapılan tahminlemeler için ileriye dönük talep ve ihtiyaçların analiz edilmesi, rekabet avantajı sağlaması, hedeflerin daha tutarlı ve aksiyon alınabilir hale gelmesi için kullanılmaktadır. Sağlık, içerisindeki derinliği ile önemi ve potansiyel veri sağlayıcı yüksek hacmi ile oldukça ön plana çıkan bir disiplindir.

Günümüzde sağlık odaklı veri panelleri, sağlık çalışanlarına verilen veritabanı eğitimleri ve hazırlanan göstergeler bu bazda verilebilecek en iyi örneklerdendir. Bu gibi sebepler nitelikli verinin sıkı bir yoğunlukla miktarının artmasını sağlamaktadır. Tahminler dahilinde artan veri miktarı ve beslenen algoritmalar neticesinde kişisel sağlık asistanları ve dolayısıyla sağlık önlemleri ile beraber yaşam süreleri, yaşam kalitesi gibi bazı kavramların ve verilerin tekrar yorumlanmasına ortam sağlanmaktadır.

Araştırma içerisinde bu denli büyük kapsama sahip sağlık verileri üzerinden hareket edilmiştir. En çok kullanılan sosyal medya araçlarından biri olan Twitter aracılığıyla Türkiye, Çin, Amerika Birleşik Devletleri ve İngiltere lokasyonları ile sağlık etiketine sahip üretilen içerikler baz alınan verisetidir. Veriseti oluşturulurken alternatif sanal özel ağlar ile yapılan paylaşımlar ayıklanmış olup veri gürültüsünden arındırılmıştır. Açık kaynak kodlu Python dili ile veriseti frekans analizi, veri görselleştirme ve kalitatif bazda yorumlama, metin duygu analizi gibi metin madenciliği teknikleri araştırmada kullanılmıştır. Niş bir açığı kapatan araştırma içerisinde farklı odaklara sahip olan ve metin madenciliği teknikleri üzerinden yapılan araştırma örnekleri literatür taraması kısmında yorumlanmıştır.

BİRİNCİ BÖLÜM

LİTERATÜR TARAMASI

1.1 Twitter Duygu Sınıflandırması Analizi

Farklı parametreler içeren veriler üzerinden birçok model ve teknik kullanarak istekler doğrultusunda araştırmalar gerçekleştirilebilmektedir.

Ayata çalışmasında, twitter üzerindeki üretilen içerikleri baz alarak kişilerin sadece anlık duygu ve düşüncelerini değil, olaylara karşı verdikleri tepkileri ölçmek dolayısıyla ilgili haberlerin ve olayların etkilerini ölçmek amacıyla makine öğrenmesi ve doğal dil işleme teknikleri kullanmıştır. Tez içerisindeki çeşitli diller bazındaki duygu çözümlenmelerini, politik tahminlemeleri ve sektör bazlı duygu çözümlenmelerini içermektedir.

Tablo 1.1:Ayata, Model Karşılaştırması

| Classifier | Accuracy % |
|-------------------------|------------|
| SVM | 82.04 |
| Multinomial Naïve Bayes | 76.12 |
| Similarity Based Method | 65.06 |

Ayata, 2000: 46

Uygulama sırasında kullanılan algoritmaların doğruluk oranlarına baktığımızda ise SVM 82.04 doğruluk oranıyla diğer algoritmalara nazaran oldukça iyi bir sonuç vermiş görünüyor. MNB algoritması 76.12, SBM algoritması 65.06 ve tabloda olmayan fakat tez içerisinde belirtilmiş olan LSTM algoritması ise 76.47 gibi bir doğruluk oranıyla SVM kadar iyi olmasa da sıralamada üstte görünen bir doğruluk oranına sahiptir. Araştırmacı analiz sırasında aslında makine öğrenmesi algoritmalarının üzerine

yoğunlaştığını fakat kullanılan boyutlar üzerine yoğunlaşılmasının genel performans için etkili olduğundan bahsetmiştir. Bu oldukça önemli olan süreçten kavramsal çalışmalarda bahsediyoruz.

1.2. Açık Kaynak İstihbaratı: Twitter Analizi

Atalı tezinde içerik üreticilerinin sosyal medya kullanımlarında paylaşılan verilerin hangi amaçlar üçüncü kişilerin eline geçtiğine, bu verilerin ilgili kişiler tarafından nasıl kullanılacağına ve potansiyel oluşabilecek sorunlara karşı bir araştırma gerçekleştirmiştir.

Tablo 1.2: Atalı, Model Karşılaştırması

| Subjektiflik | Sayı | Yüzde |
|--------------|------|---------|
| Objektif | 692 | 70,254% |
| Subjektif | 292 | 29,645% |
| Belirsiz | 1 | 0,102% |

Atalı, 2020: 41

Atalı bir kısım anket ile topladığı veriler üzerinden yaptığı analizleri sonucunda insanların sosyal medya kullanımı konusunda oldukça eksik olduğunu ve eğitime ihtiyaç duydukları kanısı ortaya koymuştur. Bu skalaya biraz daha filtre uygulayacak olursak gerçek hayatta görece olarak daha az deneyimli olan 22 yaş ve altı kişiler bu konuda oldukça içerisinde olmasına rağmen yeterli görünmemektedir. Yaklaşık 287 milyon kullanıcısı olan Twitter üzerinden elde ettiği veri ile yapılan analize bakıldığında insanların yaptıkları objektif yorumlar ile (görselde de yansıtıldığı üzere) ilgili çeyreğin marka satışları arasında doğrusal bir oran söz konusudur.

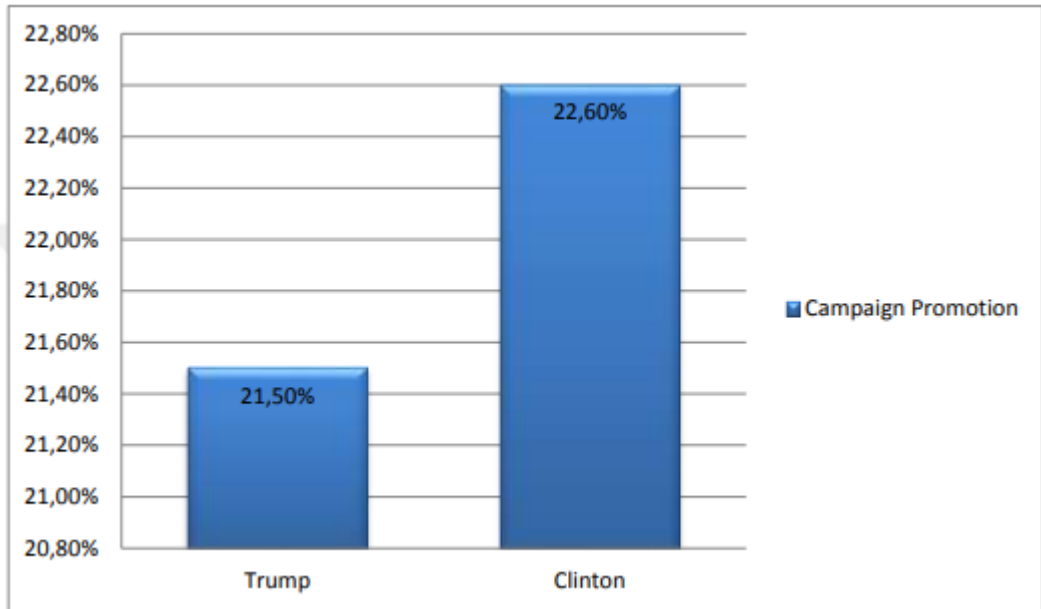
İlgili çeyrek en çok satış yapan markanın aynı zamanda sosyal medya aracı üzerinde en çok yorum yapılan marka olması da aynı şekilde tesadüfi değildir. Tez içerisinde de bahsedildiği gibi en çok veri üretimi aslında marka iletişimi sağlandığında en çok

etkileşim anlamına gelmektedir. Kişilerin yaptığı yorumlara göre ekosistem ve takımlar oluşturulması bu konu bazındaki en güzel örneklerden bir tanesidir. Kullanıcıları ile sosyal medya üzerinden sürekli iletişim halinde olan markalar kendi ürünlerinin eksiklerini ve geleceğine de yön verebilmektedirler. Bu konu hizmet sektörü için markanın ilgili ürün segmentini değiştirmesine kadar ilerleyebilmektedir. Belli dönem sonra sektör değiştiren ve daha az deneyimli olduğu fakat son kullanıcıyla iletişim içerisinde olan girişimler bu konu için verilebilecek önemli örneklerdendir.

1.3 Siyasi Bazda Twitter Analizi

Banavand maliyet ve erişilebilirlik bazında incelendiğinde ve geleneksel medya kadar veya daha fazla tercih edilen ve siyasetler nezdinde de oldukça revaçta olan Twitter üzerinden Trump ve Clinton'ın yaptıkları paylaşımların analizini gerçekleştirmiştir. Yapılan analizler sonucunda seçmenlerin ilgili üretilen içeriklerde siyasetten daha çok karakteristik özellikleri baz alarak değerlendirdiklerini ön plana çıkmıştır.

Şekil 1.1: Kampanya Karşılaştırması



Banavand, 2020: 291

Yapılan paylaşımların analizde kampanya tanıtımları bazındaki karşılaştırma da 22,60 ile Clinton ön plana çıkmaktadır. Trump için ise bu rakam 21,50 şeklindedir.

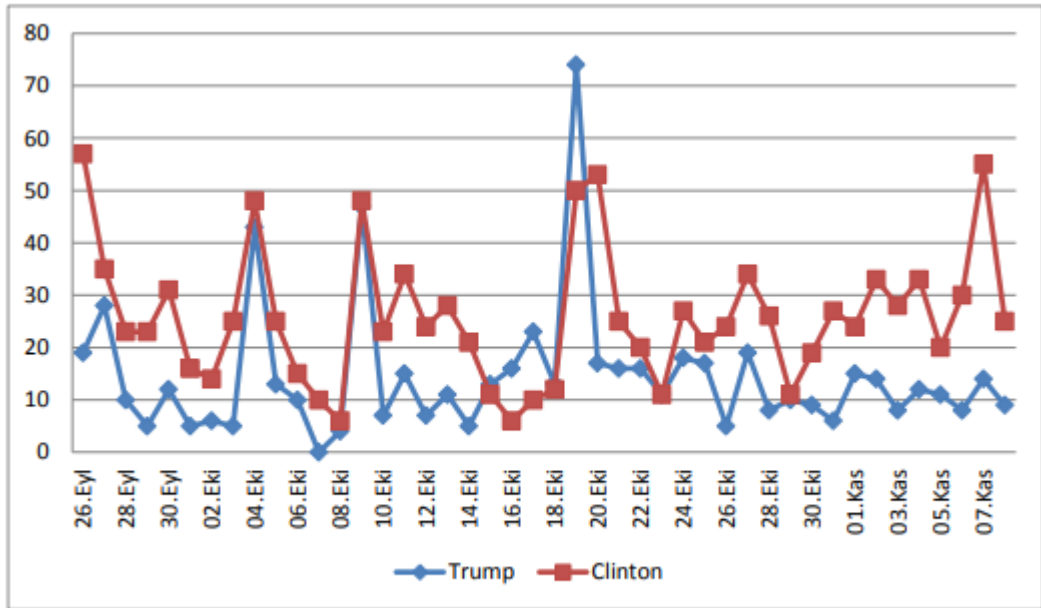
Tablo 1.3: Adayların Kullanım Etkinliği

| Candidate | Frequency | Percentage | Average Daily Post |
|--------------------------------------|-----------|------------|--------------------|
| Trump | 634 | 35.7% | 14.4 |
| Clinton | 1142 | 64.3% | 26 |
| Total | 1776 | 100% | 40.3 |
| $\chi^2 (df = 1) = 145.3, p < .0001$ | | | |

Banavand, 2020: 263

Adayların sosyal medya aracı üzerindeki etkinliği, günlük paylaşım sıklığına bakacak olursak 64,3 yüzdeliği ile Clinton ön plana çıkmaktayken 35,7 gibi bir oran ise Trump tarafından temsil edilmektedir. Günlük yapılan paylaşım sayısı ise Clinton tarafında 26 iken Trump tarafından 14,4 şeklindedir.

Şekil 1.2: Adayların Kullanım Etkinliği



Banavand, 2020: 264

1.4 Twitter'da Aktif Öğrenme Stratejileri Analizi

Aslam, tezinde 2006 yılından itibaren dakikada 350.000 hatta günün önemine göre daha fazla içerik üretimi gerçekleşen son kullanıcı tarafından oldukça rağbet gören ve

erişim kolaylığı neticesinde sürekli fikir paylaşımının gerçekleştiği bu ortamda bir çalışma yapmıştır. Sosyal medya aracı içerisindeki veri yoğunluğu araştırmacılar tarafından oldukça önemli ve üzerine çalışılabilen bir ortam olduğu kadar düzensiz ve formal olmayan yapısı dolayısıyla da çalışılması konusunda çekince ortaya çıkarılabilecek bir ortamdır.

Tez içerisinde bahsettiğimiz gibi hedef odaklı olan ve çok kullanılan denetimli öğrenme metodları bu yapı içerisinde kullanılması zordur. Aslam aktif öğrenme teknikleri ile bu soruna örneklerin etiketlenmesi ve zorluğun daha minimum hale getirilmesi üzerine bir çalışma yapmıştır. Sonuçlar ise belirsiz örnekleme yöntemlerinin sorgulama yöntemlerinden daha verimli olduğu ortaya çıkarılmıştır. (Aslam,2015: 57)

1.5 Sosyal Medyada Akan Veri Duygu Analizi

Gebreyesus, tezinde büyük veri üzerinden gerçekleştirilen analizler noktasındaki açığı kapatmak amacıyla ilgili teknoloji üründen elde edilen sonuçları bir model üzerinden en yüksek doğruluk sonuçlarını verecek şekilde optimal hale getirmekten ziyade büyük veri üzerinden akan gerçek zamanlı ve yüksek hacimli veriler ile bu modellerin karşılaştırılmasını sağlamak amacıyla ilerlemiştir. Bu sayede performans kayıpları da ön plana çıkarılmıştır.

Tablo 1.4: Akan Veri Model Karşılaştırılması

| Model Name | F1 Score |
|------------------------------------|----------|
| SemEval 2014 (best ³³) | 66.23% |
| Experiment #1 (SVM) | 63.11% |
| Experiment #2 (SVM + Spark) | 63.45% |
| Experiment #3 (RNN) | 68.04% |

Gebreyesus, 2018: 48

Büyük veri araçlarının sınıflandırıcılar üzerindeki etkinliğini teoride incelendiğinde aslında çok da büyük bir etkisi olmadığı kanaati ortaya çıkmaktadır araştırmacı ilgili

arařtırmada bununla alakalı bir pürüz potansiyelinden söz ederken son uyguladıđı test ile bu yapının çok büyük bir fark oluřturmadıđından bahsetmektedir. Var olan bir veri üzerinden yapılan temizleme ve üzerine geliřtirilen algoritmalar ve doğrudan akan verinin üzerinden yapılan tahminlemeler aslında oldukça farklı görünmekle beraber veri bilimcisinin en verimli olduđu nokta olarak da düşünölebilir.

Ařađıdaki tabloda bu pürüzün büyük veri aracı ve arařsız şekilde doğruluk yansımamı görebileceđiniz ve her bir iterasyon sonucunu bulabileceđiniz yapı eklenmiřtir. Ortalama ve ortalamadan farklılařmayı gösteren standart sapma deđerleri pürüzün ne kadar fark yaratıp yaratmayacađı konusundaki önemi de göstermektedir.

Tablo 1.5: Akan Veri Model Karřılařtırılması 2

| Experiment | SVM | Spark+SVM |
|-------------------------|------------|------------------|
| Run 01 | 62.87 | 63.03 |
| Run 02 | 62.94 | 63.31 |
| Run 03 | 63.12 | 62.96 |
| Run 04 | 63.14 | 63.17 |
| Run 05 | 63.21 | 63.28 |
| Run 06 | 63.22 | 63.32 |
| Run 07 | 63.23 | 62.98 |
| Run 08 | 63.15 | 63.19 |
| Run 09 | 62.95 | 63.23 |
| Run 10 | 63.16 | 63.27 |
| Mean | 63.10 | 63.17 |
| Std. Dev. | 0.13 | 0.14 |
| t-test (p-value) | 0.22 | |

Gebreyesus, 2018: 44

1.6 Türkçe Radyoloji Raporlarında Metin Madenciliđi

Kocatekin (2013) çalıřmalarında, %94 üzerinde doğruluk ile kategorize edilmiř ve anatomik bölgelere ayrılmıř, kritik pratiklere sahip 230 rapordan oluřan bir veri tabanı üzerinden sözlük tabanlı Türkçe radyoloji dokümanlarının kategorize etmiřtir. Çalıřma eđitim ve uygulama olmak üzere iki kısımdan oluřmaktadır. Öncelikle ayırt etmek olasılıđı daha net olanlar için iřlemi gerçekteřtirip sonrasında yeni gelen

örnekler için kategorize eden bir uygulama ortaya koyulmuştur. Ortalama olarak 98, 99, 98 oranlarında sırasıyla hatırlatma, hassasiyet ve f skor değerlerine ulaşılmıştır. Araştırmacılar gelecek için ses, görsel ve öneri sistemleri ile beraber çok daha büyük bir eğitim seti ve dolayısıyla doğruluk değeri yüksek bir sonuç ortaya çıkartılabileceğine değinmiştir. Bu araştırma ve deneyim ile teknisyenlerin elle yaptıkları bu işi hızlandırmak, otomatik hale getirmek amaçlanmıştır.

1.7 Reklam İçerikli Epostaların Metin Madenciliği Yöntemler ile Otomatik Tespiti

Çalış, Gazdağı ve Yıldız (2013), e-posta üzerinden kişilere erişimin oldukça kolay olması ve bu fiyat performans modelinin iş dünyasında satış tutundurma, satış arttırma gibi isterlerden dolayı yatırıma yer verilmesi e-posta sunucularında yüksek hacimde reklam içerikli e-postaların yer bulmasına aynı zamanda son kullanıcının dönüş yapması ve dolayısıyla ön plana çıkması gereken e-posta iletilerine nazaran yüksek oranda yer kaplamaya başlamasına dair neden ortaya koymuşlardır. Küresel anlamda bu tarz iletilerin engellenmesi ve hem son kullanıcı tarafında hem de donanım tarafında verimliliğin artması konusunda oldukça önemli olduğunu belirtmişlerdir. Araştırmacılar Türkçe içerikli reklam e-postalarının tespiti ve tespit amaçlı yapılan çalışmaların model karmaşıklığı nedeniyle bu araştırmayı ortaya koymuşlardır. Araştırma içerisinde Destek Vektör Makineleri, K En Yakın Komşu ve Naive Bayes sınıflandırma algoritmaları kullanmıştır. K En Yakın Komşu algoritması ile en yüksek skor olan 96 doğruluk oranını elde etmişlerdir.

1.8 BioBERT: Biyomedikal Metin Madenciliği İçin Önceden Eğitilmiş Bir Biyomedikal Dil Temsil Modeli

Lee, Yoon, Kim, So, Kang (2019) araştırmalarında, biyomedikal veri madenciliği artan biyomedikal verilerle beraber oldukça önemli ve araştırmalar arasında popüler hale gelmiştir. Doğal dil işleme ile beraber verimli metin madenciliği modellerinin gelişimi artmıştır fakat doğrudan doğal dil işleme metotlarının veriye uygulanması tatmin edici olmayan sonuçlara yol açmakta olduğunu ortaya koymuşlardır.

Araştırmada daha önce eğitilmiş olan BERT modeli biyomedikal bazda olan veriye adapte edilip sonuçları gözlemlenmiştir. Büyük veri üzerinde ve biyomedikal veriler üzerine geliştirilen BioBERT modeli standart bir BERT modeline göre daha iyi performans göstermekte olduğu ortaya koymuşlardır. BioBERT biyomedikal adlı varlık tanıma, biyomedikal ilişki çıkarımı ve biyomedikal soru cevaplama gibi biyomedikal metin madenciliği vakalarında sırasıyla 0.62,2.80 ve 12.24 oranlarında iyileşme sağlandığı ortaya konmuştur.

1.9 Finans Sektöründe Büyük Veri Analizi için Metin Madenciliği

Bach, Krstić, Seljan, Turulja (2019) araştırmalarında, büyük veri teknolojileri birçok sektörde olduğu gibi finans sektörüne de yön vermekte olduğu ortaya koymuşlardır. Daha çok yapısal veri üzerine odaklanan sektör dinamikleri büyük veri teknolojileri ile beraber yarı-yapısal veya yapısal olmayan verileri de analizlere girdi olarak sağlamıştır. Bu araştırmada: Alanın entelektüel çekirdeği? Hangi teknikler kullanılmaktadır? Hangi veritabanları kullanılmaktadır? Sorularına cevap aranmıştır ve sistematik tarama yapılmıştır. Finans bazında özellikle hisse fiyat tahminleri, finansal hile tespiti, pazar tahminlemesi gibi metin madenciliği temelli analizler izlenmiştir. Analizler büyük çaplı firmaların ve araştırma ekiplerinin analitik bazlı iş çözümleri aramaları üzerine yapıldığı ortaya konulmuştur. Sosyal ağ analizleri, duygular analizleri, tahminlemeler en çok kullanılan teknikler olduğu araştırmada ortaya çıkarılmıştır. Genellikle veri kaynağı olarak e-postalar, finansal bildirgeler ve raporlar, proje raporları kullanıldığı ortaya konulmuştur.

1.10 Metin Madenciliği ve Siber Suç

Kontostathis, Edwards (2010) araştırmalarında, internet avcılığı ve siber zorbalık büyük bir kitle tarafında deneyimlenen ve günlük yaşamda karşılaşılan bir meta haline gelmiş olduğundan bahsetmişlerdir. Özellikle genç kitle arasında karşılaşılan bu durum için araştırmacılar literatür taramasında bulunmuşlar ve sürecin daha yüksek hacimlere ulaşmış kitle ile ilişkisine bakmışlardır. Yeni nesil araçlar ile beraber bu

sayının gittikçe arttığını ve bu detayların farklı alanlara göre veya suç olup olmamasına göre belirlenmesinin yollarını irdemişlerdir. Araştırma sonucunda internet avcılığı ve siber zorbalığın haritalanması için metin madenciliği sınıflandırma tekniklerinin kullanılmasının gelecekte çok önemli olacağını ve proaktif olunması gereken bu alanı oldukça besleyeceğini belirtmektedirler.

1.11 Biyotıpta Metin Madenciliği ve Ontolojiler: Ham Metni Anlamlandırmak

Spasic, Ananiadou, McNaught ve Kumar (2005) araştırmalarında, biyomedikal literatüründeki veri artış hızı ciddi oranlarda artmakta ve sistematik bir şekilde artan veri yığınlarından raporlar, iç görüler ortaya çıkarmakta olduğundan bahsetmişlerdir. Alt dalı olan çeşitli disiplinlerle ilişki ortaya çıkarmak ve hipotezler ortaya koymak için metin madenciliği teknikleri kullanılmakta olduğundan bahsetmişlerdir. Semantik olarak bağlantı kurmak ve haritalamak için ontoloji alanı bu konuda oldukça önemli olduğu hakkında yorumda bulunmuşlardır. Araştırma içerisinde ontoloji ve biyomedikal alanlarını baz alarak çeşitli teknikler ve yaklaşımlar ele alınmıştır. Sonucunda biyomedikal verileri anlamsal olarak açıklamak için ontoloji alanı ile ilişkilendirilmesi gerektiği ortaya konulmuştur.

1.12 Metin Madenciliğinde Benzerlik Ölçümleri Üzerine Bir Araştırma

Vijaymeena, Kavitha (2016) araştırmalarında, artan metinsel veri hacimleri doğrultusunda verilerin toplanması ve organize edilmesi önemli bir hale geldiğini ortaya koymuşlardır. Bu süreçlerde önemli olan detay ise organizasyonun nasıl yapılacağına yöneliktir. Bu süreçlerde ise en çok kümeleme teknikleri kullanılmakta olduğundan bahsetmişlerdir. Farklı kümeleme teknikleri ise farklılığı ortaya koymak için metriklere ihtiyaç duyduğunu ortaya koymuşlardır. Bu benzerlik genel olarak öklit mesafesi ve cosine benzerliği ile bulunmakta olduğundan bahsetmişlerdir. Kümeleme algoritmaları için bu benzerlik ölçümü oldukça önemli olduğunu ve bu araştırmada amaç metin bazında, bilgi bazında, külliyat bazında olmak üzere üç yaklaşım ile benzerlik ölçülmüş olduğundan bahsetmişlerdir. Metin bazında olan algoritmalar

karakter kompozisyonu bazında kullanılırken, külliyat bazlı olanlar büyük hacim üzerine anlamsal benzerlik üzerine kurgulanmakta olduğunu ortaya koymuşlardır. Bilgi bazlı olan ise kelime ve konsept bazlı anlamsal benzerlik için kullanılmıştır.

1.13 İnsan Fenomeninin Bir Metin Madenciliği Analizi

Driel, Bruggeman, Vriend, Brunner, Leunissen (2006) arařtırmalarında, proteinler, genler, hücreler ve çeřit üzerine alıřmalar devam etmekte olduğunu ortaya koymuşlardır. Arařtırma ierisinde ortaya konulan ve biyolojik manada anlamlı sonuçlar elde edilmiş olan fenotipler üzerine bir alıřma gerekleřtirilmiştir. Metin madenciliği tekniđi uygulanan bu alıřmada 5000 fenotip sınıflandırılmaktadır. Fonksiyonel olarak iliřkili olan genler üzerine kurulu fenotipler bulunmuřtur. Aynı zamanda bu fenotipler ierisinde insan salgın genleriyle iliřkili protein türevlerini iermektedir. Bu arařtımayı ilgili adayın genetik olarak salgın potansiyelinin tahminlenmesini sađlamakta olduğunu ortaya koymuşlardır.

1.14 Metin Madenciliği ile E-Ticaret Sitelerinin Belirlenmesi

Kařıkı ve Göken (2021), alıřmalarında internet adres ieriđinin analizini gerekleřtirerek, e-ticaret sitesi olup olmadıđına dair bir sınıflandırma yapmışlardır. Özellikle bu yönde arařtırma yapanlar iin hem dođruluk oranı yüksek hem de hız avantajına sahip bir sonucu hedeflemişlerdir. Baz alınan veri üzerine KNN ve Naive Bayes sınıflandırma algoritmaları uygulanarak deđerlendirme metrikleri üzerinden yorumlar yapılmıştır. Daha yüksek oranda sonuç veren algoritma kullanılarak Java yazılım dili ile masaüstü uygulama gerekleřtirilmiştir.

1.15 KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi

Kılın, Borandađ, Yücalar, Tunalı, řimřek, Özift (2016), arařtırmalarında akademik bilgi paylařımı platformu üzerinden elde edilen makalelerin özetlerini, geliřtirilen yazılım araçları sayesinde sınıflandırmayı hedeflemişlerdir. Sınıflandırma yaparken iki boyutlu tasnif deđerleri ortaya konmuřtur. Bunlar materyal bilimi ve mühendislik,

sosyal bilimler ve insanlık olarak sıralanabilir. Veritabanı olarak Research Gate sistemi kullanılmıştır. Makalelerin özetleri veri tabanını beslemektedir. Veri seti üzerinde yazılım aracı olarak R kullanılmıştır. Açık kaynak kodlu olan R yazılım dili ile K En Yakın Komşu algoritması ile model oluşturulmuştur. Sonuç olarak 96 gibi bir yüzde oranı ile sınıflandırma gerçekleştirilmiştir.

1.16 Metin Madenciliği Teknikleri İle Şirketlerin Vizyon İfadelerinin Analizi

Melek (2012), çalışmasında Capital dergisinde belirlenen En Beğenilen Şirketler üzerine bir araştırma gerçekleştirmiştir. Bu çalışmada listelenen şirketlerinin paylaştıkları vizyon metinleri üzerine metin madenciliği teknikleri uygulanmıştır. Analiz sırasında Statistica programı kullanılmıştır. Araştırma sonucunda şirket vizyonlarında en çok geçen kelimelerin müşteri, ürün, lider, pazar olduğu ortaya çıkmıştır. Vizyon ifadelerinden geçen kelimelerden itibar bazında yapılan faktör analizi sonucunda sırasıyla beklenen imaj, yönetsel performans göstergeleri, konumlandırma gibi boyutlar öne çıkmıştır.

1.17 Muhasebede Analiz Yöntemi Olarak Metin Madenciliği

Yıldız ve Ağdeniz (2018), çalışmalarında işletmelerin paydaşlarına en çok bilgi sağladığı açıklamalar finansal dokümantasyonlar içerisinde yer almakta fakat buna göre tam anlamıyla bilgi edinmek için ek olarak faaliyet, sürdürülebilirlik ve entegre raporlara da bakılması gerekmekte olduğunu ortaya koymuşlardır. Bu raporlar daha çok yapısalıktan uzak olan verileri içeren analiz konusunda okuyucusunda soru işaretleri oluşturmaktadır. Bu ve benzeri sebeplerle araştırmacılar yapısal olmayan bu verilerin analizi ve yapısal veri kadar kolay yorumlanabilmesi adına metin madenciliği yöntemleri ile raporların yorumlanabilmesi amacıyla çalışmayı ortaya koymuşlardır. Bu bazda yayınlanan çalışmaları incelediklerinde daha çok duygu analizi ve sınıflandırma yöntemleri kullanılmakta ve hileli işlem tespiti, işletme performans tahmini, iflas tahmini, sürdürülebilirlik ve kurumsal yönetim gibi bazı başlıklarda yorumlanmışlardır.

1.18 Örgütsel Araştırmada Metin Madenciliği

Kobayashi, Mol, Berkers, Kismiho'k (2018), Hartog araştırmasında, metin verisi oldukça fazla olmasına rağmen organizasyonel araştırma sorularına bu tekniklere görece az uygulanmıştır. Metin madenciliği araştırma sırasında ortaya konan görüşleri destekleyecek verilerde radikal bir artış ortaya koymuştur. Araştırma metin madenciliği üzerine kurulu bir şekilde organizasyonel araştırmacılara farklı bakış açıları sağlayan, içerisinde analitik süreçler barındıran yapıları barındırmıştır. Boyut azaltma, uzaklık ve benzerlik analizleri, kümeleme, modelleme, sınıflandırma gibi analitik teknikler araştırmada gözlemlenmiştir. Teknikleri içerisinde birçok iş ilanı bulunan veritabanı üzerinde gerçekleştirmişlerdir. Böylece metin madenciliği ile beslenmiş çok daha zengin bir araştırma ortaya konmuştur.

1.19 Metin Madenciliği Teknikleri İle Sosyal Ağlarda Bilgi Keşfi

Yüksel ve Tan (2018) araştırmalarında, günümüz sosyal ağ kullanımının popülaritesi nedeniyle sosyal medya üzerindeki mevcut veri akışında gözle görülür bir artış yaşanmaktadır, yaşanan bu durum mevcut veri üzerinden anlam çıkarmanın yanı sıra duygu analizi yapabilme özelliği olan sistemlerin önemini arttırmıştır. Mevcut durum insanlara kullandıkları sosyal ağ üzerinden yemek, içmek, alışveriş yapmak veya ziyaret etmek için planladıkları yer hakkında bilgi veya benzeri yerlerin hakkında yorumlar ve onlara olan uzaklığı/yakınlığı hakkında detaylı bilgiler teknik analizler sonucu ortaya çıkarmış olmayla beraber mevcut durum neticesinde kişiler planladıkları yer hakkında olumlu/olumsuz görüşlere erişebilmekte olup çeşitli sosyal paylaşım siteleri üzerinden onlara karar verme sistemi oluşturmuştur. Araştırmacılar mevcut araştırmalarında bu bilgilere örnek olarak "Foursquare" sosyal paylaşım sitesini kullanmışlardır. Bu veriler ışığında uzmanlar yapmış olduğu araştırmalar sonucu ikili sınıflandırmada %81,97, üçlü sınıflandırmada ise %84,49 başarı oranı elde etmişlerdir.

1.20 Metin Madenciliği: İmkanlar, Yöntemler ve Kısıtlar

Atan (2020) araştırmasında, internetin insan dünyasında kullanımının yaygınlaşması nedeniyle insanların internet üzerinden ürettikleri bilgilerin de bu denli artmasına neden olmuştur. Yapılan araştırmalarda her gün 2,5 GB milyar verinin üretildiğini elde etmişlerdir. Uzmanlar bu verinin yaklaşık %80'inin metin formunda olduğunu söylemişlerdir. İnsanların metin formlarının düz yazı formunda işleme almaları sonucu oluşan devasa metin formlarından anlamlı sonuçlar üretebilmek için metin madenciliğini ön plana çıkarmışlardır. Araştırmacılar metin madenciliği olanaklarını incelemelerine yardımcı olmayı, metin madenciliğinin arka planını inceleyen çalışmalara önem vermeyi hedeflemektedir. Mevcut metinlerin oluşturulan algoritmalar yardımıyla özetlenmesi, sınıflandırılması, etiketlenmesi ve seçilmesinin mümkün olması metin madenciliğinin arka planını ve olanaklarının araştırmacılar tarafından büyük fırsat olarak değerlendirildiği ortaya konmuştur.

1.21 Bir Metin Madenciliği Uygulaması: Vosviewer

Artsın (2020) araştırmasında, veri yığınlarından anlamlı bilgiler üretilme çabası geçmişten günümüze devam etmektedir. Uzmanlar tarafından yapılan araştırmalar sonucunda görselleştirmenin son yıllarda yaygınlaştığı bilgisine ulaşılmaktadır.

Veri madenciliğinde karmaşık ilişkilerin bulunduğu “bibliyometri analizi” çalışmalarında da elde edilen verinin görselliğe dair çeşitli araştırmaları önemli bir konu haline geldiği uzmanlar tarafından belirtilmektedir. Araştırmacılar mevcut araştırmalarında VOSviewer programının bibliyometrik ağların görselleştirilmesi amacıyla yaygın olarak kullanılmakta olduğundan bahsetmektedirler. Mevcut program kullanımı sonrası bibliyometrik ağların görselleştirilmesi ile birçok alandan araştırmacıya bilgilendirici yönlendirmeler bulunmasının yanı sıra arayüzü, kullanımı, altyapısının kullanım açısından detaylı analizler sağlayabilmektedirler.

1.22 Çağrı Merkezi Metin Madenciliği Yaklaşımı

Güvercin, Ferhatosmanoğlu ve Gedik (2017) araştırmasında, çağrı merkezlerindeki görüşme kayıtlarının sesten metne dönüştürülebilmesi metin madenciliğinin

kullanılmasına olanak sağladığını ortaya koymuşlardır. Yapılan arařtırmalarda ele alınan görüřme kayıtları içerisinde görüřmenin ne yönde olduđuna dair sonuçlara varılabileceđini ve müřterinin bu yöndeki tutumunun ölçülmesi amaçlanmıřtır. Metinlerden elde edilen özelliklerden yararlanılarak sınıflandırma ve regresyon yöntemleriyle görüřme kayıtlarının içeriklerinin deđerlendirilmesini sağlayacak tahmin modelleri oluşturulmuřtur. Ayrıca çağrı merkezi alanında geliřtirilmekte olan bu ürün aynı zamanda farklı alanlarda da deđerlendirilmesi için çalıřmalar söz konusudur. Bu durum řirketlerin karar mekanizmaları için büyük önem taşımaktadır. Uzmanlar tarafından bu tahmin modelleri "Türk Telekom" bünyesinde çağrı merkezleri baz olarak gerçekleřmiřtir.

1.23 Sosyal Ağlarda Akan Veri Madenciliđi

řeker (2014) arařtırmasında, son yıllarda kullanımı hızla artan sosyal ağlarda her geçen gün oldukça yüksek veri akmakta olduđunu düşünmektedir. Bu verilerin neredeyse tamamı çeřitli veri merkezlerinde toplanmakta olup veri madenciliđi olarak isimlendirilen yöntemlerle iřlenerek anlamlı istatistiksel deđerler çıkarılmakta olduđu görüřünü belirtmiřtir. Ayrıca bu ağların birçođunun içeriđi internet üzerinden açık olduđundan dolayı internete bađlı herhangi birisi de veri toplayıp analiz yapabilmekte olduđunu ve sosyal ağ üzerinden analizlerin yapılabilmesi için çeřitli adımlar izlememiz gerekmekte olduđunu düşünmektedir. Bunlar; verinin toplanması ve sečilmesi, verinin temizlenmesi, verinin özelliklerinin çıkarılması, makine öğrenmesi algoritmaları kullanılması, anlamlı sonuçların yeniden kullanılabilir hale getirilmesi olduđunu belirtmiřtir. Günümüzde akademik anlamda geliřimi devam eden ve yeni uygulamalar bünyesine katmaya bařlayan veri madenciliđi çalıřmaları ile gün geçtikten verileri analiz edebilmemizde bizlere büyük kolaylık sağlayacak olduđunu düşünmektedir. Dünya üzerinde veri madenciliđi uygulamalarına ihtiyaç her geçen gün artmakta olup, veri madenciliđi ile ilgili çok sayıda yeni iř alanları da ortaya çıkmakta olduđu görüřünü ortaya koymuřtur.

1.24 Edirne Mektuplarının Metin Madenciliđi Analizi

Erduran ve Kayıcı (2018) çalışmalarında, Avrupa, Balkan ve Türk tarihinde Edirne ilimiz her zaman stratejik öneme sahip olduğu belirtmişlerdir. Balkan savaşlarından sonra mevcut dönemin yayınlarında Edirne'nin hali hazırdaki durum değerlendirmesi çeşitli araştırmacı yazarlar tarafından mektuplarla ifade edilmiştir. Bu araştırmaların ilk aşamalarında mevcut yayınlar farklı dillerden Türkçe'ye çevrilmiş olup, istatistiksel yöntemlerle analiz edilmiştir. Mevcut yazılardaki açıklamalar metin madenciliği ile ortaya çıkarılmış ve mevcut yazılar içerisindeki önemli bilgiler değerlendirilmiştir.

1.25 NoSQL Veri Tabanları Üzerinde Bir Metin Madenciliği Uygulaması

Zontul, Aydın (2017) Araştırmalarında, Metin Madenciliği Kısaca, Metin Kaynaklarını Girdi Olarak Kullanıp Bu Metinler Üzerinden Bilgiler Çıkarılması Ve Yapılandırılmamış Verilerin Biçimlendirilmesiyle Beraber Anlamlı Sonuçlar Elde Etmeyi Amaçlayan Sistem Olarak Nitelendirilmekte Ve Metin Madenciliğinde Amaca Hizmet Eden Algoritmalar Seçilerek Veri Tabanı İncelemesi Yapılmakta Olduğunu Belirtmişlerdir. Günümüzde İnternet Kullanımının Artmasından Dolayı İnsanların Da Sosyal Mecralarda Paylaşımları Oldukça Artmıştır. Bu Durum İnternet Dünyasını Metin Madenciliğinin Merkezine Oturtmakta Olduğunu Belirtmişlerdir. Metin Madenciliği 4 Adımdan Oluşturduğu Ve Bunların; 1. Metin Madeni Oluşturma, 2. Metin İşleme, 3. Veri Analizi, 4. Değerlendirme Ve Raporlama Olduğunu Belirtmişlerdir. Bu Projedeki Amaç Müşteri Şikâyetlerini Anlamlandırıp Ölçümlemektir. Veri Madenciliği Kullanılarak Yapılan Bu Araştırmada Değerlendirilen Firmaların Kapsama, İnternet, Fatura Ve Hizmet Bölümleri Değerlendirilmiştir. Turkcell Ve Vodafone 'Nun İşlem Kalitesi Olarak Yakın Oldukları Görünmektedir. Fakat Türk Telekom 'Un Müşterilerine "İnternet" Kategorisinde Diğer Firmalara Göre Daha Çok Sorun Yaşattığı Görülürken "Fatura" Kategorisinde İse Diğer Firmalara Göre Daha Avantajlı Olduğu Ortaya Koymuşlardır. Değerlendirmeler Sonucu Ele Alındığında Tüketiciler Bu Kapsamda Hangi Firmayı Sececeğine Ya Da Seçeneklerinin Neler Olduğuna Daha Kolay Karar Verebilecektir. Yapılan Araştırmada Sosyal Mecraların İnsanlar Tarafından Kullanımının Veri Madenciliğine Sağlamış Olduğu Kolaylıklardan Ve İnsanlara Bu Sayede

Belirleyecekleri Stratejileri Hakkında Karar Verme Süreçlerinde Doğru Bilgilere Ulaşılabilirliği Sağlanabilmekte Olduklarından Bahsetmişlerdir.

1.26 Metin Madenciliği Tabanlı Bildirim Takip Sistemi

Küçük (2019) araştırmasında, yazılım projelerinin önemli bir bileşeni olan bildirim takip sistemlerinin temel amacı kişilerin çalışmalarını olabildiğince basitleştirmek olarak tanımlamıştır. Ayrıca bu sistemin kullanıcılara diğer bir faydası da ekip üyeleri arasında iletişimsel bir fayda sağlamaktır. Mevcut araştırmada Pamukkale Üniversitesi'nde içeriğinde birçok modül barındıran Pusula Bilgi Sistemi için geliştirilen F1 Bildirim Takip Sistemine, Metin Madenciliğinin sağladığı olumlu sonuçlar neticesinde ortaya çıkan başarılı sonuçlardan bahsedilmektedir. Örneğin geliştirilen sistem içerisindeki manuel atama işleminin Metin Madenciliği sistemi sayesinde ortadan kalkmış olması ve bunun da onlara iş ve zaman yönünden fayda sağladığı belirtilmiştir.

1.27 Metin Madenciliğinde Anahtar Kelime Seçimi Bir Üniversite Örneği

Yıldız (2016) araştırmasında, internet üzerinde bulunan metinlerin içerisinde yöneticiler için anahtar kelime çıkarımı oldukça önemli olduğunu belirtmiştir. İnternet kullanımının artması nedeniyle kurumlar ile ilgili öneri, şikayet ve teşekkür gibi mesajlar geçtiğimiz yıllara göre artık daha kolay bildirilmekte olduğundan bahsedilmiştir. Bu durum internet kullanımının bu tip olaylar için mevcut olan iş yükünün azaltması sonucunu bizlere göstermektedir. Yapılan bu araştırmada Yıldız Teknik Üniversitesi'nde iç paydaşlar tarafından kullanılan 7/24 Yıldızlı Hat Yönetim Sisteminde biriken mesajlar anahtar kelime özellik çıkarımı için kullanılmış ve TF-IDF, Ki Kare ve Bilgi Kazancı yöntemleriyle birlikte yapılan analizler sonucunda anahtar kelimeler bulunmuş olduğundan bahsetmiştir. Yapılan bu çalışma sonucu kurum içerisinde raporlama hizmetinde olumlu sonuçların ortaya çıkmasıyla beraber üst yönetime sunulacak olan bilgiler kolay anlaşılacak biçimde olup stratejiler kararların alınması ve uygulanmasına büyük olanak sağlayacak olduğunu tespit etmiştir.

1.28 Metin Madenciliği Yöntemleri ile Twitter Verilerinden Bilgi Keşfi

Beşkirli, Gülbandılar, Dağ (2021) araştırmasında, sosyal medyanın kullanımı bilişim teknolojilerinin gelişmesiyle birlikte her geçen gün daha da artmakta olduğunu belirtmişlerdir. Genellikle bilgi paylaşımının yapıldığı sosyal medya, bilginin artmasına da olanak sağlamakta olduğunu bu dönüşüm için kullanılan yöntemlerden biri de metin madenciliği olduğunu ortaya koymuşlardır. Yapılan araştırmada twitter üzerinden metin madenciliği yöntemleri kullanılarak aşı ile ilgili çeşitli bilgiler elde edilmiştir. Bu veriler aşı öncesi ve aşı sonrası olarak gruplandırılmış olup, kullanılan platformlar, düşünceler, konuşma tipleri, duygu ve cinsiyet gibi çeşitler analizler yapılmıştır. Yapılan araştırmalar sonucunda sosyal medya kullanıcıların Covid 19 aşı çalışmaları esnasında olumlu ve olumsuz sayısı neredeyse eşit olarak görünürken nötr düşünenlerin sayısı ise bir hayli fazla olarak görülmüştür. Fakat Covid19 aşısı için üretime geçilmesi haberleri sonrası sosyal medya üzerinden yapılan araştırmalar sonrası olumlu düşünceler de büyük bir artış görünürken olumsuz düşüncelerde ise büyük bir azalma olduğu sonucu ortaya çıkmıştır.

İKİNCİ BÖLÜM

KAVRAMSAL ÇERÇEVE

2.1 Sosyal Medya

Sosyal medyanın gelişmesi, kişilerin iletişim araçlarını sürekli yanlarında taşıması, teknolojinin daha erişilebilir bir hale gelmesiyle beraber hızlanmıştır. Yapısal ve ön planda olan medyaya alternatif bu araçlar son kullanıcılar arasında oldukça hızlı, kişisel ve dolayısıyla takip edilen ilgi alanları doğrultusunda çok daha fazla içerik tüketimi ve tüketilen içeriğin fikirleri beslemesiyle, çok fazla içerik üretimine neden olmuştur. Tüm bu farklı disiplinlerin Kartezyen çarpımı diyebileceğimiz, veriyi üreten

figür ve ilişkili tüm araç, konu, disiplin paydaşları çok fazla veri üretimine sebebiyet vermiştir.

Veri üretimi konusundaki hacmi tekil olarak sayacak olursak tüm canlı nüfusü ve ilişkili paydaşlarının Kartezyen çarpımı diyebileceğimiz kadar çok bir veri bütününden, bu durumunun zaman boyutunda dolayı kümülatif ve ilişkili sağlık disiplini bazındaki veri havuzu, sağlık konusunun önemi üzerine olan etkisi yadsınamaz bir gerçektir.

Sosyal medya, bireyler ve kuruluşlar için stratejik hedeflere ulaşmayı sağlayan, var olan bilginin dağıtılması ve tarafların bu bilgiyi kendi operasyonlarına, süreçlerine girdi olarak sağlayabilmesi için yayılmasını sağlayan, neredeyse sıfır maliyet ve herhangi bir ön şart koşulu olmadan kişilerin düşüncelerini ifade edebildiği ve münferit birçok taraf için bu bilgileri ulaştırabildiği, içerisinde birçok taraf, araç ve değişken bulunan bir ekosistemdir. Ekosistemin içerisinde yeni dönemde getirdiği ihtiyaçları düşünerek olursak oldukça hızlı bir tüketim söz konusudur. Bundan dolayı yeni haberler ve kaynakları bu tüketime öncülük yapan taraflar olduğunu düşünürsek oldukça reaksiyon alan bir organizma da denilebilmektedir.

Sosyal medya, özellikle yeni dönemde sosyal medyayı kişilerin fikirlerini ortam bağımsız bir şekilde ifade edebildiği platformlar olarak adlandırılmaktadır. Son kullanıcılar kendi özel deneyimleri doğrultusunda oldukça fazla şekilde sosyal medya ile ilgilenmek durumunda kalıyorlar; bunlar aradıkları ürün, araştırdıkları konu, takip ettikleri fikirler, sanal olarak işlerindeki ihtiyaçlar doğrultusunda yaptıkları satın almalar ve tüketimler...

Özellikle firmalar için iş dünyasının son kullanıcıyla iletişim halinde olması gerektiği şu dönemde, ellerindeki en büyük fırsat ve aynı zaman en büyük tehlike de denilebilmektedir. (Akkaya, Akyol, Şimşek, 2018: 363). Bundan dolayı firmaların ihtiyaçları karşılamak için girdikleri yeni organizasyon yapılarında, sosyal medya üzerine uzmanlaşmış ve bunun iletişimini sağlayan oldukça fazla ihtiyaçları bulunmaktadır.

İletişime paralel olarak oldukça gündemde olan diğer bir konu ise veridir. Sosyal medya araçları günümüzde içerdiği birçok iletişim alternatifi ile firmalar ve firmaların içerisindeki karar alıcılar için oldukça elzem bir hale gelmiştir. Oluşturulan misyon ve vizyonları doğrultusunda karar vericiler sosyal medya üzerinden edindikleri verileri kullanmaları gerekmektedir. Bu verileri yeni teknolojiler ve yasal prosedürler çerçevesinde kullanmaları doğrultusunda üretilen çıktılar ile kendilerine iç görü elde edebileceklerdir. (Kaplan & Haenlein, 2010: 61)

Raporlar sosyal medyayı, kullanıcı bazında ortaya çıkarılmış olan içerik ve interaksyonu bir araya getiren dijital teknoloji olarak tanımlamaktadır (Kind, 2009: 1413).

Son zamanlarda ortaya çıkan ve oldukça göz önünde olan diğer bir tanım ise halkla ilişkiler; iletişim, kitle iletişim araçları arasındaki entegrasyon ile gün yüzüne çıkan bir organizma şeklindedir. Sosyal medya ile ilgili farklı bir tanım ise kullanıcıların kendi içlerinde bağlantı kurmasını sağlayan web bazlı algoritmalarının bütünüdür. Bu sayede kullanıcılar birbirleriyle iletişim kurarlar ve etkileşim içerisinde olurlar (Jones, 2012: 152). Sayın Hazar da az önceki örneklerimizi destekleyecek şekilde, sosyal medyayı diğer iletişim araçlarından farkı sorusuna son kullanıcıların ücretsiz bir şekilde herhangi bir konu hakkında diledikleri içeriği oluşturma, paylaşma ve kendi profillerinde ileti olarak göndermek şeklinde iletişim kurabildiklerinden bahsetmiştir (Hazar,2011: 157). Sosyal medya kurumları, fikirleri, organizasyonları örneklerdeki gibi olumlu yönleri sayesinde güç kazanarak yönlendirebilmektedir. Bunun farkında olan kullanıcılar hedef kitlesi ve hedefindeki kişiyle bu vasıta ile iletişim kurabilmektedirler.

Sosyal medya içerisinde birçok dinamik barındırmaktadır. Bunun üzerinde duracak olursak beş temel başlık söz konusudur:

2.1.1 Katılım

Kullanımda olan herkes için doğal bir sentez ile katkıda bulunmak ve konulara geri bildirim vererek beslemek neticesinde paylaşımda bulunan ve paylaşımda bulunan kişiyi takip eden kullanıcılar arasındaki çizgi oldukça azalmıştır. Böylece statü

farketmeksizin aslında paylaşımda bulunan kişiye veya kuruma son kullanıcı görüşlerini belirtebilmektedir. Aradaki çizginin kalkması böylece oldukça fazla verinin ve fikrin üretilmesine sebebiyet vermektedir.

2.1.2 Açıklık

Kullanıcı deneyimi günümüzde oldukça önem kazanmıştır. Sosyal medya araçları içerisindeki dinamiklik sebebiyle aslında son kullanıcıya ücretsiz olarak paylaşımlarda bulunabileceği, değerlendirmelerini karşı tarafa iletebileceği, yapılan paylaşımlara oy kullanabileceği, kendi için hazırladığı özgün bir ara yüz ile karşı tarafa benliğini yansıtabileceği bir yapı sunmaktadır. Son kullanıcı sosyal medya araçlarını kullanırken alışkanlığı ve çağın gereksinimleri nedeniyle oldukça kritik bir gözlem yapmaktadır.

Bu gözlem alternatiflerini sıralayacak olursa en başta, kullandığı sosyal medya aracını içerisinde gezinirken bir engelle karşılaşp karşılaşmaması gelmektedir. Bu deneyim potansiyel kullanıcı kaybı anlamına da gelmektedir çünkü eğer bu kısıtlama deneyim kalitesini arttıran bir yapıda değilse kullanıcı ilk fırsatta alternatif bir araca yoğunlaşmaktadır.

2.1.3 Karşılıklı Konuşma

Eski nesil medya araçları daha çok kullanıcıya baskısını yaptığı, yayımlandığı bilgiyi doğrudan ileterek bu süreci sürdürmektedir. Bu da son kullanıcı tarafından herhangi bir dönüş olma potansiyelini azalmaktadır. Sosyal medya ise daha çok paylaşım üzerine yoğunlaşmaktadır. Bu sayede son kullanıcı ile iletişindedir ve geri dönüşlerini veritabanında tutarak veya sadece izleyerek bunu hizmet sunum şeklinde iyileştirme aşamalarında kullanabilmektedir.

2.1.4 Topluluk

Sosyal medya araçları topluluklar için oldukça önemli bir araçtır. Topluluklara verimli bir iletişim halindeyken rasyonel bir değerlendirme ile düzenli büyüme imkanı sunarken,

belli ilgi alanı etrafındaki bu kişileri topluluğun isterleri doğrultusunda belli bir form yapısında oluşturabilmektedir.

Bu yapı ve büyüme üstel olarak da yapılabilir. Örneğin: Herhangi bir hobi amaçlı bir topluluk oluşturduunuz ve bu konuda çeşitli teknik, yeteneksel paylaşımlarda bulunuyorsunuz. Bu paylaşımları uluslararası olarak kabul edilmiş bir dil ile de yapmaktasınız. Bu doğrultuda büyümeniz sadece fiziki olarak belli bir lokasyonda düzenli organizasyonlar yapan ve topluluğu büyütmeyi hedefleyen yapıya göre oldukça hızlı olacaktır. Bunun en önemli sebeplerinde biri lokasyon bağımsız olması, diğeri ise potansiyel ulaşabileceğiniz son kullanıcısı örnekleminin oldukça büyük olmasıdır.

2.1.5 Bağlantı

Sosyal medya araçları içerdiği link yapısı ve paylaşım ağları ile kaynakları, sayfaları ve insanları oldukça hızlı ve hedef odaklı bir şekilde birbirine bağlayabilmektedir. Bağlantı yapısı genelde platform bağımsızdır ve en kötü ihtimal uygulamanın bağlantısını kopyalayıp herhangi bir araç üzerinden paylaşarak tabiri caizse otomatik hale getirilmemiş olsa bile el ile yapılabilir. (ICrossing, 2008: 5).

Sosyal medya yaşam süresi içerisinde; son kullanıcıya sunduğunu hizmet, içerisindeki kullanıcı toplulukları ve bunların etkileşim dinamikleri, kurguladığı misyon ve vizyon çerçevesinde daha çok kitleye ulaşma, son kullanıcı deneyimi, ödül bazlı oluşturulan yarışma alternatifleri, son kullanıcının psikolojisi ve algısı, platforma erişilebilir olması, son kullanıcının deneyim sırasında herhangi bir konuda yapmak istediği iletişime binayen sosyal medya platformunun tarafından oldukça çeşitli bir şekilde alternatifler sunulması (ifade, dinamik görsel, kullanımı kolay klavye, fotoğraf, video...) gibi dinamikleri nezdinde oldukça rekabet oluşturan ve bu rekabet doğrultusunda ön plana çıkmış çok fazla platformlar bulunmaktadır.

Bu platformlar gerek kullanıcı kitlesi, gerek oluşturduğu topluluklar, gerekse altyapısındaki teknoloji nedeniyle birçok kırımla ön plana sunulabilmektedir. Bu kırımlardan birkaç tanesini incelemek gerekirse:

2.1.5.1 Twitter

Tweet adı verilen mesajlar ile kişiler paylaşımlarını yapmaktadır. Twitter bu mesajları belirli bir karakter kısıtlaması ile kullanıcıya sunmaktadır. Kişiler bu mesajlar üzerinden paylaşmak istedikleri düşüncelerini, fotoğraflarını, bağlantı linklerini platform üzerinden diğer kullanıcılara sunabilmektedirler.

Paylaşım kişilerin ana sayfalarına düşmektedir. Bu yapı içerisinde kronolojik olarak paylaşımlar en yeniden en eskiye olacak şekilde son kullanıcının önüne sunulmaktadır. Böylece kullanıcı uygulamaya ilk giriş yaptığında son kaldığı yerde yukarıya doğru gittikçe en yeni yapılan paylaşımı görebilmektedir. Tabi eğer en yeniden eskiye gidecek şekilde bir yapı da son kullanıcı deneyimi bazında ürün özellikleri içerisinde bulunmaktadır ilgili yönlendirmeler ile bu arayüz oluşturulabilir. Twitter içerisinde bulunan en yeniden en eskiye doğru aşağı yönde hareket doğrultusunda ilerleme katedilen kronolojik alt yapısı sayesinde diğer sosyal medya platformlarından kendini ayırt etmektedir. İçerisindeki mesajların sıralanma şekli ve içeriğinin sınırlanması neticesinde kısa olması yapısı sayesinde platformu oldukça dinamik ve güncel tutmaktadır.

Bundan dolayıdır ki haber kaynaklarının bu alana yönelmesi ve son dakika haberlerini bura üzerinden paylaşması tesadüf değildir. Twitter yapısı sayesinde bir marka bazında düşünülecek olursa kısa, açıklayıcı, ilgi çekici bilgiler ve arka plan fotoğrafı ile ve tabi en önemlisi diğer aktif olarak kullandığı ve kullanılan diğer sosyal medya platformları ile bağlantı kurulabilecek şekilde paylaşımları yapmalıdır. Bu değişkenler yapılan paylaşımı, kurum nezdinde tüketicisine verimli bir şekilde ulaşabilmesini sağlamaktadır. Markaları Twitter bu kısıtları ve sunduğu öz ürün özellikleri ile yakalayıcı ve çarpıcı başlıklar bulmaya ve destekleyici görseller ve bağlantılar için zorlamaktadır.

Marka bazında örnek verecek olursak:

Koton, birçok marka gibi Twitter kullanmaktadır, yaptığı paylaşımlar ile tüketicisi nezdinde önemli olan günleri, indirim zamanlarını ve birçok ihtiyaca yönelik iletişimini sağlamaktadır. Kullandığı diğer hesap olan koton destek ile direk mesaj

sayesinde müşteri iletişimini ve dolayısıyla ihtiyaçlarını karşılamaktadır. Diğer birçok markanın aksine kendisini takip eden tüketicilerini geri takip etmektedir. Twitter özellikle dikkat eksikliğinin yoğun olduğu ve birçok enformasyonu oldukça kısa süre içerisinde tüketebilen yeni dönem toplumlarında oldukça revaçta olan bir pazarlama aracı olarak kullanabilmektedir. (Yüzüak, 2018: 11)

2.1.5.2. Facebook

Hepimizin hikayesine aşina olduğumuz şekilde Harvard Üniversitesinin bir yurt odasında okul arkadaşları ile son kullanıcıların kolayca buluşabilmesini sağlayan 2004 yılında kullanılan, gün geçtikçe okul dışında da kullanıma başlanılan ve günümüzde veritabanında oldukça geniş bir çeşitlilikte kullanıcısı bulunan sosyal medya platformudur. (Çiçek and Erdoğan, 2013: 255)

Facebook içerisinde de bağlantıları çekebilmesi için sosyal medya içerisindeki revaçta olan içerik kullanımlarına özen göstermeli ve o tarihi yakalayabilecek nitelikte olması gerekmektedir. Paylaşımın çok nitelikli olması oldukça önemli bir kriter fakat Facebook üzerinden paylaşım yapılacak saat bu kriterin önem seviyesinden bağımsızdır. Sosyal medya platformu üzerindeki en aktif saat, etkileşimin olduğu en doğru zamanla paylaşımın geri dönüşünü arttıracak ve kalitesini belirleyecektir.

Markalar bu gibi kullanım yeteneklerini ve hedef kitlesine ulaşabilecekleri en doğru planlamayı kullanmaları gerekmektedir. Firmaların bünyelerindeki organizasyon ve kadro yapılanmasında bu alanla ilgilenen kişiler bu tarz durumlarda oldukça kritiktir. Facebook daha önce tanım aşamasında oldukça fazla söz ettiğimiz gibi sosyal medya platformlarının içerisindeki kullanıcılar arasındaki etkileşim hacmi ve bu hacmin içerisinde üstel olarak artan veriler ile markalara oldukça kritik iç görüler sunabilecek ve markaların yönetsel ve reklam planlamalarına yön verebilecek analitikler sunmaktadır. Analitikler içerisinde en aktif olan zamanlamayı ve kullanım yeteneklerini geliştirebilecek yapıyı da sunmaktadır. (Yüzüak, 2018: 9)

2.1.5.3 Instagram

Takip ettiğiniz kullanıcıların paylaşımlarını ana sayfanızda, sosyal medya platformlarında daha çok kullanılan diğer bir terim ile ifade edecek olursak zaman çizelgesine gelecek bir yapı ile ürünü sunmaktadır. Diğer birçok platformda olduğu gibi Instagram da içeriğinde bir odak noktası barındırmaktadır. Bu sosyal medya platformumuzda yüksek kalitede olmasa bile daha çok görsel içerik paylaşımı üzerinden hareket edilmektedir.

Masaüstü kullanımı mevcut olsa bile daha çok mobil kullanıma odaklanan Instagram kendi içerisinde paylaşım çeşitliliği sunmaktadır. Son kullanıcı deneyimi doğrultusunda örnekler vermemiz gerekirse; kullanıcılar takip edilebilir, yapılan fotoğraf ve video paylaşımlarına beğeni ve yorum özellikleri ile etkileşimde bulunabilir, bir günlük saat sınırı olan hikaye paylaşımları söz konusudur, genelde kısa diyebileceğimiz belli bir süre kısıtı olacak şekilde yapılan video paylaşım özelliği de son zamanlarda eklenmiştir ve son olarak kullanıcılara izin ve gizlilik kuralları çerçevesinde doğrudan mesaj atabilirsiniz.

Instagram içerisindeki etkileşim oranının artması eklenen yeni özellikler ile de beraber oldukça rekabetçi bir konumda görünmektedir. Güncellemelerini, yeni özellik eklemelerini rekabet edebilir halde kalmak için bu doğrultuda belli bir frekans ile gerçekleştirmektedir. Sosyal medya platformlarında oldukça önemli olan diğer bir dinamik ise kullanıcı deneyimidir. Instagram bu konuda oldukça estetik ve basit bir kullanıcı deneyimine sahiptir. Görsel odaklı bir ürün sunulması bu avantajlardan bir tanesidir. (Yüzüak, 2018: 12)

2.1.5.4 Youtube

Video paylaşımı odaklı bir web sitesidir. Kullanıcılar kendileri için oluşturdukları hesaplar ile içeri aktarma, paylaşım, yorum yapma ve beğenme gibi ürün özelliklerini kullanabilmektedirler. Youtube istatistiklerine bakacak olursak dünya bazında milyonlarca kullanıcı tarafından her dakika 100 saatlik video web sitesine yüklenmektedir. Bu etkileşim kaçınılmaz bir şekilde platform içerisindeki video çeşitliliğini inanılmaz kılmaktadır. Bu etkileşim markalar ve pazarlama sektöründeki

kullanıcılar için oldukça etkili bir iletişim ve pazarlama aracı olarak görünmektedir. İlgili içerik ile bağdaştırıp piyasaya sunulacak potansiyel ürün için bu platform kullanılabilir. Youtube içerisinde oldukça fazla pazarlama için kullanılabilecek çeşitlilik sunmaktadır. Bunlardan bazıları:

- Atlanamaz reklamlar,
- 5 saniye sonrasında atlanabilecek reklamlar,
- Marka sponsorluğunda olan müzik klipleri,
- Meydan okuma içerikleri vb.

Burada oldukça önemli bir rol ise pazarlama uzmanlarına düşmektedir.

- Hedef kitleleri ne tarz içerikler tüketmektedir?
- Hangi ünlü kişileri takip etmektedir?

Bu gibi sorulara cevap bulunup ellerinde pazarlama için kullanılacak olan ürün ile en uygun eşleşmeyi bulup aksiyon almaları gerekmektedir. (Yüzüak, 2018: 13)

Özellikle özgün olan içerikler ürün odaklı bir içerik olmasa bile oldukça yüksek hacimli ve beğenilen bir hale gelebilmektedir. Yeni nesil pazarlama, içerik üretimi ve sosyal medya platformları bu üstel büyüeyebilen, özgün, ve oldukça yüksek etkileşim imkanı olan planlamayı besleyebilmektedir. Bundan dolayı birden fazla disiplinin ürün etrafında bir araya geldiği ve ürün arka planda kalsa bile ihtiyacı oluşturan ve besleyen bir yapıya bürünebilmektedir.

Kurumlar için teknolojinin önemi yadsınamaz. Sosyal medya araçlarının oldukça hızlı bir ivme ile gün yüzüne gelmesi ve birçok kullanıcının hemen alışkanlık haline getirip, günlük hayatlarında düzenli olarak kullandıkları, aksiyon aldıkları ve girdi sağlayabildikleri bir yapıya bürünmesi de bu duruma güzel bir örnektir.

Kurumlar sosyal medya araçlarını müşterileri ile iletişimlerinde kullanırken aynı zamanda sosyal medya odaklı yeni nesil teknolojilerini de kullanmaları gerekmektedir. Sohbet robotları kurumlar ile müşterileri arasındaki sosyal medya platformları üzerinde iletişimi hızlandırmak, sürdürülebilir hale getirmek ve daha çok geri besleme alabilmek için oldukça verimli bir süreç optimizasyonu denilebilmektedir. Bu geri beslemeler kurumlar için kararlarında ve planlamalarında kullanabilecekleri, oldukça

önemli bir süreçtir. Bu sayede belli başlı kalıp sorular için oldukça hızlı bir iletişim sağlanabilir. Sohbet robotları içerisindeki makine öğrenmesi algoritmaları ile süreç ilerledikçe daha çok karar verebilecekler ve anlamlı cevaplar, girdi olarak daha öncesinde verilmese bile, dönüş yapabileceklerdir.

Sohbet robotları Chat ve bot kelimelerinin yan yana gelmesi ile oluşturulmuştur ve insanların konuşmalarına benzemeyi amaçlayan algoritmalar temelli yapılardır. Kavrama, yorumlama ve yanıtlama gibi özellikleri nezdinde insanlara çok benzeyen sohbet robotları gelen sözlü ve yazılı tepkiyi kendi algoritmasına girdi olarak sağlayıp dönüş yapmaktadır.

Sosyal medya kullanıcısı için kişiselleştirilmiş alternatifleri sunan, web bazlı teknolojileri kullanan, kullanıcı kitlesi içerisinde oldukça fazla aşına kişi bulunan bir organizmadır. Kullanıcıları arasındaki aşinalık bu organizmanın doğal yaşamın bir parçası olarak oldukça fazla günlük hayatın içerisine girmesine neden olmaktadır. Universal Mccann Wave 6 (2012) araştırmasına göre bazı istatistikler vermek gerekirse:

- Kullanıcıların %81'lik kısmı aktif bir şekilde bağlantılarını takip etmektedirler.
- Her gün 1,5 milyar tekil kişi kullanıcısı bulunmaktadır.
- Haftalık aktif geçirilen süre 13 saattir.
- Kullanıcıların %47'lik kısmı marka topluluklarına katılmaktadır.
- Kullanıcıların %77'lik kısmı kendi profillerini oluşturmuş ve kullanmaktadır.

2.2 Sağlık Nedir?

Sağlık yalnızca insanların gündelik yaşamlarına hastalık veya sağlık sorunları tesir ettiğinde gündeme gelen bir kavramdır. Genel olarak ele alındığından insanlarda hastalığın olmadığı ve ek olarak psikolojik anlamda kendisini iyi hissetmesi olarak tanımlanmıştır. Bu tanım Dünya Sağlık Örgütü (WHO) tarafından 7 Nisan 1948 yılında yürürlüğe girmiştir. Zaman içerisinde çeşitli hastalıkların ortaya çıkması neticesinde sağlık literatürüne çeşitli tanımlar eklenmiştir.

Sağlık hakkında çoğu insanın ortak görüşü olarak yaşam tarzları oldukça önemlidir. Özellikle gün içerisinde neyi ne kadar yiyip içtikleri, buna istinaden ne kadar spor yaptıkları zararlı alışkanlıkları olup olmadığı, günlük yaşantısında ve iş yaşamında ne kadar strese maruz kaldığı oldukça önemlidir. Bunlar insan sağlığı üzerinde oldukça önemli bir paya sahiptir. Fakat yapılan bu tanımın yanı sıra insanların “Sosyal Refah” düzeyi yaşadıkları hayatı etkilediği gibi sağlıklarına da etki eder.

Kimi zaman toplumun çoğu kesminde görülen hastalık, hastalıktan sayılmayabilir. Örneğin sürekli öksüren birisinin “Acaba gece pencere açık mı kaldı?” düşüncesiyle gelip-geçer bir hastalık olabileceği algısı onu önemli hastalık düşüncesinden uzak tutacaktır. Sağlığı temel olarak ele alacak olursak eğer 2 temel yaklaşım baz alınabilir. Bunlar sosyal yaklaşım ve tıbbi yaklaşımdır. Sosyal yaklaşıma göre ele alacak olursak bireyin mesleği, eğitimi ve eğitim durumu yakalanacağı hastalıkları belirleyen faktörlerdir. Fakat Tıbbi yaklaşıma göre ele alacak olursak, insan sağlığını sadece dışarıdan gelecek olan mevcut etkenler (virus, bakteri benzeri organizmalar) bozmaktadır.

2.3 Sağlığı Etkileyen Faktörler Nelerdir?

2.3.1 Kültür

İnsan sağlığını etkileyen faktörlerden biri de kültürdür. Kültür gerek aile yapısı gerek çevre gerek dini inanç olarak çeşitli kavramlar bakımından insanlığın sağlığa bakışı ve sağlık hakkında genel bilgi konusunda çeşitli farklılıklara yol açmaktadır. Öyle ki günümüz dünyasında ülkeler arası göçlerin ortaya çıkması sonucu kişilerin geldiği ülkenin kültürlerini gittikleri topluma yansıttıkları görülmektedir. Bu durum her zaman geçerli olmamakla birlikte kişisel farklılıklar göstermektedir. Kişi gitmiş olduğu toplumun kültürel değerlerini benimseyerek sağlık sistemini o baz da koşullamaktadır.

2.3.2 Cinsiyet

Cinsiyet başlığı adı altında sağlık konusunda başlıca farklılıklar bulunmaktadır. Bunlar insan bünyesinde bulunan başlıca tıbbi farklılıklardan oluşmaktadır. Temel olarak ele

alındığı zaman kadın ve erkek dahil olmak üzere 2 cinsiyet bulunmaktadır. Anatomik ve hormonal farklılıklar bu temel ayrıma sebep olan başlıca etkenlerdir. Genel olarak ele alacak olursak eğer kişinin sahip olduğu fizyolojik ve biyolojik özellikler cinsiyetini belirler.

Bunun bu cinsiyeti etkileyen diğer etmenlerden biri de kültürdür. Diğer bir deyişle buna toplumsal cinsiyet adı verilmektedir. Toplumsal cinsiyet sosyal olarak kadın ve erkeğin sorumluluklarını ve rollerini ifade etmektedir. (Akın, 2007:1)

2.3.3 Sağlık Hizmetleri

Sağlık hizmetleri faktörü de bireyin sağlığının korunmasında, mevcut hastalığın teşhisinde ve hastalıkların tedavisine yönelik etkisiyle birlikte toplumun sağlık statüsü üzerinde oldukça önemli bir yere sahiptir. Sağlık sistemi başlıca çeşitli kaynakları kullanarak kişide hali hazırda bulunan sağlık problemlerini en aza indirgeyerek kişiyi topluma kazandırmayı ve ek olarak buna toplumun katılımıyla kişiye daha sağlıklı, daha üretken ve daha kaliteli bir hayat sağlamayı hedefler. (Sabuncu, Babadağ, Taşocak ve Atabek, 1996:24)

2.3.4 Biyoloji ve Genetik Faktör

İnsan bünyesinin biyolojisi ve organik yapısı, sağlığın başlıca belirleyicisidir. Dünyada her gün milyonlarca canlı meydana gelmektedir. Bu canlılar ilk geldikleri süre içerisinde dünyaya gelmesine neden olan annesi ve babasının küçük bir örneği olarak görünmektedirler. Buradan yola çıkarak dünyada mevcut her canlıya ait karakteristik bir özelliğin bulunması ve bunun ana-babadan yavruya geçmesidir. Bu durum “soyaçekim“ olarak adlandırılmaktadır. Bazı hastalıklar bu durumda genetik olabilmektedir. Anne ve Baba da bulunan kimi hastalıklar doğan çocukta da ilerleyen yaşlarda görülebilmektedir.

2.3.5 Sağlıklı Çocuk Gelişimi

Bireyin çocukluk döneminde gelişim durumu yetişkinlik dönemine sağlıklı bir şekilde geçiş yapabilmesi bakımından oldukça önemli ve bir o kadar da özen gösterilmesi

gereken bir faktördür. Bu dönemli gerekli beslenme sağlanamamış çocuğun ileri dönemlerde bağışıklık sisteminin bundan olumsuz etkileneceği kaçınılmaz bir gerçektir.

Bireyin ileriki yaşlarda yaşadığı hastalıklar kronik hastalıkların asıl nedeni geçmişte durumu ile doğrudan ilişkilidir. Bununla birlikte kişinin fiziksel gelişimi de bebeklik sürecinde nasıl beslendiği ile ilgili ilişkilidir. (Neyzi, 2000: 1) Bunun yanı sıra çocuğun eğitim olarak da iyi bir şekilde gelişimi buna önemli katkı sağlayacaktır.

2.3.6 Kişisel Sağlık Uygulamaları ve Başa Çıkma Becerileri

Bireylerin hastalıkları önleyebileceği bazı önlemler alması konusunda kişisel gelişim ve kendisine bakım süreci önemlidir. Bu durum kişilerin özgüvenini geliştireceği gibi onların hastalıklar hakkında bilgilendirilmesini de sağlayacaktır. Kişisel sağlık uygulamalarında kişinin yaşadığı ülkenin önemi ve bu ülkenin kimi kültürel değerleri oldukça önemlidir. Öyle ki kimi ülkelerde alternatif tıp oldukça ön plandayken kimi ülkelerde modern tıp oldukça gelişmiş ve bir o kadar önemsenmiştir. Kişilerin yaşamakta olduğu sosyal çevre ve kültürleri bu bazda önemli rol oynamaktadır.

2.3.7 Çevresel Faktörler

İnsanların soludukları hava, içtikleri su, yaşadıkları evler, işe veya okula gidip gelirken kullanmakta oldukları toplu taşıma araçları dahil sağlık konusunda önemli etki oluşturmaktadır. Nitekim kötü bir çevrede yaşayan bireyin uzun ve sağlıklı bir yaşantıdan bir o kadar uzakta olması bu duruma en iyi örnektir.

Geçmişten günümüze yoğun olarak görülen yapılaşmanın sonuçlarından biri olan coğrafi yapıdaki değişimler de insan sağlığına olumsuz şekilde etkilemektedir. Günümüzde insanlar artık yaşadıkları çevrenin daha fazla kirlenmemesi ve bozulmuş olan bu yapılaşmanın daha fazla artmaması için gerekli mücadele vermeye çalışmaktadırlar. Fakat nüfusun hızlı bir şekilde artması da bu tarz yapılaşmanın ortaya çıkmasına zemin hazırlamaktadır.

Her geçen gün hızlı bir şekilde konut ihtiyacı çarpık kentleşme sorununu daha fazla karşılamaya çalışmaktadır. Bu sebepten dolayı insanların doğaya olan özlemleri

artmakla beraber doğadan sağladıkları temiz havanın da bi o kadar onlardan uzaklaşmasına sebep olmaktadır. Bu durum insan sağlığını önemli ölçüde etkilemektedir. Günümüz dünyasında bu duruma örnek verecek olursak eğer nüfusun bu kadar hızla artması ve bu nedenden dolayı iç içe yaşaması yaşanan Covid-19 hastalığının bu denli hızla yayılmasına verilebilecek güncel örneklerden biri olabilir. Bu duruma diğer bir örneğimiz de insan hayatındaki ihtiyaçlardan biri de kişisel ulaşım ihtiyacının doğaya verdiği zarar diyebiliriz. Bu hızda artan nüfus ile birlikte kullanılan araç sayısı da her geçen gün artmakta olup, araçların doğaya vermiş olduğu zararlar da oldukça fazladır. (Güler ve Çobanoğlu, 1994: 35)

2.3.8 Çalışma Koşulları

İş, işyeri ve işgücü koşullarının çalışanların ve toplumun diğer üyelerinin sağlık faktörü açısından etkisi bir hayli fazladır. Günümüz dünyasında insan yaşamak için üretmek zorundadır. Nitekim üretim için de çalışmak gerekmektedir. Öyle ki iş insanlar için hayatlarının bir parçası durumuna gelmiştir. Kişi çalıştığı iş hayatı içerisinde kimi zaman iş kaynaklı kimi zaman ise çalıştıkları iş ortamı kaynaklı sıkıntılardan dolayı sağlık problemleri ile karşı karşıya gelmektedirler.

Bu tip durumlarda insanın sağlık ile ilgili yaşadığı sıkıntılar iş hayatını etkilemekte olup verimlilik anlamında olumsuz bir durumla karşı karşıya kalınmaktadır. İnsanlar kötü çalışma koşullarında riskli işler yaptığı taktirde önlem alınmadığı zamanlarda bu ölümcül sonuçlar doğurmaktadır. İş yaşantısı bireylere sadece fizyolojik tehlikelerle değil aynı zamanda insan bünyesine psikolojik problemlerle de karşı karşıya getirmektedir. Günümüzde bu problemleri en aza indirmek için çeşitli denetlemeler yapılmakta olup “İş Sağlığı ve Güvenliği“ personelleri çeşitli eğitimler verip çalışanların iş hayatında ne gibi durumlarla karşı karşıya kalılabileceğini ve bu tip durumlarla karşı karşıya kalmamak için nasıl önlemler alınması gerektiği hakkında bilgiler verilmektedir. Gerekli denetlemeler sonucunda çalışılan yerlerde çalışanların sağlığının ne kadar düşünüldüğü ve ne gibi koşullar sağlandığı denetimciler tarafından denetlenmekte olup bu neticede olumsuz durumlarla karşılaşıldığında gerekli işlemler başlatılmaktadır. (Vogel, 2007: 22)

2.3.9 Eğitim ve Okuryazarlık

Eğitim ve okuryazarlık faktörü sağlık için iki anahtar kavramdır. Kişinin okuryazarlığının olmaması kötü sağlığın başlıca nedenidir. Eğitim, sağlığı belirleyen önemli faktörlerden biridir. Kişinin hastalık karşısında kendisinde oluşacak özgüvenin başlıca nedenleri arasında yer almaktadır. Öyle ki eğitim düzeyi geliştikçe insanların hayata olan bakış açıları da bir o kadar değişmektedir. Beslenme düzeninden spor a neyin zararlı olup neyin yararlı olduğuna kadar bir çok şeyin farkında olmakla birlikte kişinin kendine olan özgüveninin artmasıyla birlikte topluma olan faydası azımsanamayacak kadar artacaktır. Birey kişisel gelişimini ne kadar arttırırsa sağlıklı bi yaşamın önemini bir o kadar da anlayacak olup, bunu ailesindeki çevresindeki insanlarla paylaşacaktır.

2.3.10 Sosyal Destek

İnsanların kriz ve ihtiyaç zamanlarında daha geniş bir odaklanma ve olumlu bir benlik imajı vermek için aile ve arkadaşlar dahil olmak üzere geniş bir çevreye sahip olma anlamına gelmektedir. Öyle ki kişinin yaşadığını sosyal çevre ile arasındaki olumlu ilişkiler kişinin sağlığına önemli katkı sağlamaktadır. Sosyal desteğin diğer bir faydası da bireyin çevresinden görmüş olduğu olumlu desteğin karşılaştığı psikolojik veya fizyolojik olumsuz durumlar karşısında daha dirençli olması, üstesinden gelmesi veya ona karşı oluşturabilecek olumsuz durumu minimum seviyeye indirgemesini sağlayacaktır. Sosyal destek kişi için özgüven arttırmanın yanı sıra olası bir hastalık karşısında alabileceği desteği bilmesinden ötürü kişi de diğer durumlara göre olabilecek bir rahatsızlığa karşı tampon görevi görür. (Yıldırım, 1997: 81)

2.3.11 Gelir ve Sosyal Durum

Gelir faktörü sağlığın en önemli sosyal belirleyicisidir. Gelir düzeyi insanın genel yaşam koşullarını şekillendirmekle birlikte psikolojik ve fiziksel durumlarda da oldukça önemli bir yapıya sahiptir. Kişinin sosyal durumu süphesiz sağlık sisteminden eğitim durumuna kadar birçok şeyi etkilemektedir. Bir toplumda sosyal durum tabaka halinde sınıflandırılmaktadır. Bunlar alt, orta, üst tabakadır. Bu sınıflandırmanın temel

nedeni ekonomik durum olarak görülmektedir. Kişinin isteklerine ulaşım imkanı sağlığı konusunda da oldukça etkilidir. Bu duruma beslenmeden yaşam standartlarına kadar detaylı örnekler verebiliriz. Örneğin ekonomik durumunun kötü olmasından dolayı kışın soğuk bir evde yaşayan bireyin hastalığa yakalanma oranının ne kadar yüksek olduğu kaçınılmaz bir gerçektir.

2.3.12 Stres ve Sağlık

Günümüz dünyasında gerek çalışma gerek yaşam standartlarında gün içerisinde sürekli karşılaştığımız “stres“in sağlığımızı etkilediği kaçınılmaz bir gerçek. Bu etki kısa veya uzun süreli olabilmektedir. Kısaca ele alacak olursak eğer stres gün içerisinde yaşanan herhangi bir durumun insan vücuduna psikolojik veya fizyolojik tarafından olumsuz sonuçlanmasının yanı sıra yine insan vücudunun psikolojik olarak dengesini bozan bir etkidir. (Dinçer, 1994: 219). Öyle ki stres sadece yıkıcı bir etkiye sahip olmamakla birlikte dozajında olduğu sürece yapıcı etkileri de görülmektedir. (Batlaş ve Batlaş, 2004: 13)

Yukarda belirttiğimiz faktörler sağlık konusunda belirleyici bir öneme sahiptir. Her insanda bu deneyimler yaşamları boyunca bazı değişiklikler gösterebilir. Bu değişimlerde zorluklar ile karşı karşıya kalınması, iş yaşantısında yaşanan bazı aksaklıklar gibi durumlar örnek verilebilir. Bu tip durumların üstesinden gelmek kişiye özgüven aşılacağı gibi hayata dair olumlu bir bakış açısı oluşturacaktır.

2.4 Geçmişten Günümüze İnsanlık İçin Yaşanmış Büyük Hastalıklar

Günümüzde yaşanan ve insan sağlığını büyük ölçüde tehdit eden ve her geçen gün daha da yayılmakta olup, ölümcül sonuçlar ortaya koyan Koronavirüs hastalığı akıllarda geçmişten günümüze yaşanan benzeri hastalıkların yaşanıp yaşanmadığı veya yaşanan hastalıkların ne gibi sonuçlar ortaya koyduğu sorusunu akıllara getiriyor. Bu hastalıklardan bazıları hala günümüzde çok sık olmasa da görülürken kimi hastalıklar da artık görülmediğinden dolayı unutulmuş statüde diyebiliriz.

2.4.1 Cüzzam

Lepra hastalığı olarak da bilinen Cüzzam hastalığı eski çağlardan beri bilinmektedir. (Mycobacterium leprae) adlı mikroorganizmadan dolayı bulaştığı bilinmektedir. Bulaşıcı olmakta birlikte bu oranın oldukça düşüktür. Bu nedenden dolayı insanlar bu hastalığı uzun yıllar boyunca bulaşıcı olmadığı kanısıyla yaşamışlardır. Fakat halk arasında bu hastalık Veba hastalığı kadar korkutucu bir hastalık olarak bilinmektedir. İlgili enfeksiyon her yaşta kapılabildiği gibi mevcut semptomların ortaya çıkması yaklaşık olarak 20 yılı bulabilmektedir.

Cüzzamın bulaşmasının önemli kaynağı sadece insan olarak bilinmektedir. Genç yaştaki bireylerin bu hastalığa yakalanma oranı daha fazla olarak bilindiği gibi hastalığa sahip annelerden çocuklara anne sütü ile geçebilme durumu söz konusudur. Hastalık bulunduğu insan vücudundan dışarıya yara salgılarıyla ve burun salgısıyla beraber çıkar ve etrafa yayılır. Bu durum iç organlara da zarar verebilir. Özellikle ayaklarda ve ellerde sinir kayıplarına yol açmaktadır. Bu tip durumların tedavi edilmemesi gibi durumlar uzuvlarda kısılmalara veya kayıplara yol açmaktadır.

2.4.2 Çiçek Hastalığı

Her yaşta görülebilen bu hastalık ateşli, bulaşıcı, ağır bir hastalıktır. Variola virüsünün neden olduğu bu hastalık daha sık çocuklarda görülmektedir. Özellikler yüz bölgesinde ve kollarda döküntü şeklinde görülür. Hastalık belirtileri öncelikle vucutta ağrılar başlar ve daha sonrasında cilt üzerinde kırmızı lekeler görülür.

Bu kırmızı lekelerin içleri sıvı doludur daha sonrasında bu kırmızı lekeler kabuk bağlayarak çoğalmaya başlar. Hastalığın görüldüğü hasta genellikle 6 hafta karantinaya alınır. Hastaya bu süreçte yaklaşan bireyler aşılanarak denetim altında tutulmaktaydı. Çiçek hastalığı yaklaşık olarak 3000 yıldır varlığını sürdürmüştür. İlgili hastalık zamanının en korkulan hastalıklarından biri olmasının yanı sıra Dünya Sağlık Örgütü'nün de en çok savaştığı hastalıklardan biridir. Mevcut hastalığın 1979 yılından itibaren tamamen ortadan kalkmıştır.

2.4.3 Veba Hastalığı

Antik çağlardan itibaren bilinen *Yersinia pestis* bakterisinin neden olduğu bulaşıcı ve öldürücü bir hastalıktır. Genellikle veba farelerden bulaşır algısı mevcuttur fakat genel olarak ele alırsak eğer küçük memelilerden ve onların pirelerinden bulaşan bir hastalık olarak değerlendirebiliriz. Mevcut hastalık kapıldıktan sonra 1 hafta içerisinde ortaya çıkar. Genellikle nefes darlığı, ateş, vucutta ağrılar, halsizlik, kusma gibi belirtileri başlıca belirtileridir.

Hastalığa yakalandığında vücut içerisinde iç organlarda kanamaların oluşmaya başladığı ve bu kanamalar sonucunda da akan kanın vucutta birikmesiyle ciltte siyah lekelerin oluştuğu görülmektedir. Hastalığın tedavi ile ölüm oranı %10 iken tedavisiz ölüm oranı ise yaklaşık %70 dir. Küresel olarak ele alındığında yıl yaklaşık olarak 600 vaka görülmektedir. İlgili hastalığın en bilindik zamanları 14. yüzyıldır. Yapılan araştırmalar hastalığın sonucunda 50 milyondan fazla ölüm ise sonuçlandığı bilinmektedir.

2.4.4 Kolera Hastalığı

Dönemin en önemli hastalıklarından biri olan “Kolera“ hastalığı *Vibrio cholerae* isimli bir bakterinin neden olduğu mevcut bakterinin bulunduğu yiyecek ve suların yenilmesi/içilmesi nedeniyle ortaya çıkan bağırsak hastalığıdır. Hastalığın belirtileri akut ve şiddetli ishal olarak bilinmektedir. Araştırmalara göre her yıl enfeksiyon nedeniyle dünya çapında yaklaşık 1,3 ila 4,0 milyon kişi kolera hastalığına yakalanırken bunlardan 21.000 ila 143.000 ölüm ile sonuçlandığı bilinmektedir.

Kolera hastalığı çocuklarla birlikte yetişkinleri de etkilemektedir. Özellikle hastalığa yakalanan kişilerin hastalığın belirtilerini yaşaması 12 saat ile 5 gün arası sürmektedir. Fakat bu belirtilere rağmen hala tedavi edilmezse eğer hastalık bireyi öldürebilir. Ölüm riski çok yüksek olan bu hastalık günümüzde de hala binlerce kişiyi etkilemektedir. Hastalık ele alındığında tedavi süreci oldukça basittir. Hastalığa yakalanmamak için öncelikli olarak su kaynaklarının temizliği oldukça önemlidir. Eğer içilen su da şüphe varsa eğer mevcut suyun ısıtılıp daha sonra kullanılması veya mevcut durumla ilgili birimlerin aranması daha sağlıklı olacaktır. Pişmemiş ve çiğ olan gıdalardan uzak durulması bu hastalık için olası önlemler arasında yer almaktadır.

2.4.5 Koronavirüs (Covid-19)

Çağımızın önemli hastalıklarından biri olan şiddetli akut solunum sendromu (SARS-COV-2)'nin neden olduğu bulaşıcı bir hastalıktır. Hastalığın ilk vakası Çin'in Hubei eyaletine bağlı olan Vuhan şehrinde Aralık 2019 tarihinde karşılaşılmıştır. Geçmişte hayvanlarda görülen bu hastalık zamanla mutasyon uğrayarak insanlar içinde tehlikeli bir hale gelmiştir.

Hastalık genellikle solunum yoluyla bulaşmaktadır. İnsanlar arasında temas arttıkça hastalığın bulaşma riski de o kadar artmaktadır. Özellikle kapalı ortamlarda mevcut hastalığın bulaşma riski daha fazladır. İlgili hastalığın ölümcül sonuçlar ortaya çıkarması nedeniyle pandemi süreci başlamıştır. Hastalık ile ilgili semptomlar insanlarda ilgili virüse maruz kalması sonucu 2 ila 14 gün içerisinde ortaya çıkar ve bu kişi bağışıklığına bağlı olarak yaklaşık 10 ila 20 gün bulaşıcı kalmaktadır.

İlgili hastalığın teşhisi laboratuvar da gerçekleştirilir. Hastalığın teşhisi için çeşitli yöntemler geliştirilmiştir. Teşhis için en yaygın kullanılan yöntem sürüntü yardımıyla uygulanan gerçek-zamanlı polimeraz yöntemidir. Mevcut hastalığın önlenmesi için pandemi sürecinde insanlardan genel olarak mümkün oldukça evde kalmaları ve gerekmedikçe evden çıkmamaları istenmiştir. Halka açık yerlerde maske takmaları gerektiği ve mümkün oldukça insanlar ile aralarında mesafeli olmaları, kalabalık olan yerlerden mümkün mertebe kaçınmaları, evlerini sık sık havalandırmalarını ve temizlik konusunda ellerini mümkün oldukça detaylı yıkamaları istenmiştir. Hastalığa yakalanmış bireylerin veya kendisine mevcut belirtilerin (ateş, öksürük, tat ve duyu kaybı, nefes almakta zorlanma) olduğunu bilen kişilerin gerekli sağlık merkezlerini aramaları konusunda bilgilendirme çabaları sürekli olarak desteklenmiştir.

Mevcut hastalık ile alakalı çeşitli aşilar çıkmış olup insanların bağışıklığını hastalık karşısında daha dirençli bir hale getirmek için çeşitli çalışmalar başlamış olup hala devam etmektedir.

2.5 Ülkere Göre Sağlık Sistemi Değerlendirilmesi

2.5.1 Çin

Geçmişten günümüze hızla büyümekte olan ve hemen hemen her sektörde görmekte olduğumuz Çin'i sağlık sektöründe değerlendirecek olursak eğer Çin'in sağlık hizmetleri kamu, özel sağlık kurumları ve sigorta programlarından oluşmaktadır.

Çin halkının yaklaşık % 95'i temel sağlık sigortasına sahip olmasına rağmen kamu sigortası genellikle halkın tıbbi masraflarının sadece yarısını karşılamaktadır. Bu oran daha ciddi hastalıklarda daha da düşmektedir. 2020 yılında Çin bu durumu değiştirmek için olumlu girişimler yapmıştır.

Çin'de hemen hemen tüm büyük şehirde lokasyon bazında büyük hastaneler bulunmaktadır. Mevcut hastanelerin kalite düzeyleri konuma göre değişiklik göstermektedir. Bunları değerlendirecek olursak eğer genel olarak iyi hizmet kamu hastanelerinde verilmektedir. Bu kıyasta diğer sağlık hizmetleri de kliniklerde görülmektedir. Lokasyon olarak büyük şehirlerde çoğu kamu hastanesinde VIP koğuş olarak adlandırılan içerisinde güncel teknolojik hizmete ve tıbbi olarak makul statüde değerlendirilebilecek ayrıca İngilizce bilen kalifiye personele sahip olup, yabancılara da hizmet verilen bölümler bulunmaktadır.

Bu bölümler maddi olarak diğer hizmet alınan sağlık bölümlerinden fazla olmaktadır. Fakat batı ülkeleri bazında kıyaslandığında bu fiyat ucuzdur. Ülkede modern tıbbın yanı sıra Çin tıbbi olarak adlandırılan bir yöntem de bulunmaktadır. Bu yöntem için Çin tarihi ve felsefesi başta olmak üzere Çin kültürü hakkında derin bir bilgi birikimine sahip olmak gerekmektedir. Günümüzde Çin "2030'da Sağlıklı Çin" adı altında çeşitli girişimlerde bulunmaktadır. Hedeflerinden bazılarını gerçekleştiren Çin, farklı ihtiyaçlara hizmet edebilmek için çeşitli sağlık sigortaları oluşturmuş olup bunun yanı sıra şehir hastaneleri, toplum sağlığı merkezleri ve çeşitli kamu sağlığı tesisleri kurmaktadır.

2.5.2 Amerika Birleşik Devletleri

Amerika Birleşik Devletleri, batı yarım kürede Güney Amerika, Kuzey Amerika, Orta Amerika ve bu bölgelere bağlı adalardan oluşan bir ülkedir. Sağlık sektöründe değerlendirecek olursak eğer sağlık hizmetleri bakımından sağlık tesisleri önemli ölçüde özel sektör tarafından yönetilmektedir.

Ülkede bulunan hastanelerin %58'i kar amacı gütmeyen, %21'i devlete ait geri kalan %21 ise kar amaçlı olarak oluşmaktadır. Ayrıca ülkede diğer gelişmiş ülkelerden farklı olarak evrensel bir sağlık programı bulunmamaktadır. Kamu sektöründe çalışan bireyler için sağlık sigortası hükümet tarafından sağlanmaktadır. Ülkede sağlık sigortası oldukça önemlidir.

Sağlık sigortası sistemi eyaletten eyalete göre değişkenlik göstermektedir. Amerika Birleşik Devletleri'nde her yıl sağlık hizmetlerine kişi başı 10.000\$'dan fazla harcama yapılmaktadır. Bu durum gelişmiş ülkeler bazında sıralama yapıldığı takdirde diğer ülkelere göre (İngiltere, Fransa, Kanada) iki katından fazladır. Fakat Amerika Birleşik Devletleri dünyanın en pahalı sağlık sistemlerinden birine sahip olmasına rağmen gelişmiş ülkeler arasında halk arasında düşük yaşam beklentisine sahiptir. Ayrıca doğum oranlarında 2017'de bebek ölüm hızı 1000 doğumda 5,8 ölüm olarak belirtilmiştir.

Amerika Birleşik Devletleri'nde diğer bir tıbbi yöntem de alternatif tıptır. Genellikle tıp fakültelerinde öğretilmeyen veya hastanelerde bulunmayan terapiler olarak tanımlanır. Bunlar genellikle şifalı otlar, masajlar ve enerji şifası olarak adlandırılan hizmetlerle bilinmektedir. Bu tarz alternatif yaklaşımların insanlar tarafından yaygın olarak kullanılmasının asıl nedeni kendi sağlığı üzerinde daha fazla kontrol sahibi olmak ve kronik hastalıklarının nedenlerini hafifletmek olarak ele alınmaktadır.

2.5.3 İngiltere

Sağlık hizmeti güvenlik, sağlık hizmeti ve sosyal hizmetleri kapsamaktadır. Gelir odaklı olmayan tüm popülasyonu kapsayacak bir hizmet söz konusudur. Özellikle güvenlik bazında zor durumda olan kişiler üzerinde durulmaktadır. Hükümet tüm bu potansiyel maddelerden sorumlu olarak bilinmektedir. Yapı mahalli sağlık kurumları ve sosyal yardım kuruluşları üzerinden yürütülmektedir. Sorumlu her ne kadar mahalli sağlık kurumları ve sosyal yardım kuruluşları olarak bilinse de prensip, talimat ve yönlendirme gibi terimlerin sorumlusu hükümet olarak bilinmektedir. Bu yapı

sayesinde İngiltere Nisan 2021 tarihi ile beraber 70 yaş üzeri vaka sayısını sıfıra indirdi.

2.5.4 Türkiye

Sağlık sistemi merkezi bir sistem üzerinden yönetilmektedir. Bu yapı doğrudan Sağlık Bakanlığına bağlıdır. Özellikle 2003 yılında atılan adımla beraber hükümet ayrılan bütçe ve operasyonel koşulları revize ederek kişilerin düşük bütçe ile sağlık hizmetlerinden yararlanılabilir hale gelmesini sağlamıştır. Kişiler toplam sağlık maliyetlerinin yaklaşık 15 gibi bir yüzdesi ile süreçleri deneyimlemektedirler. Binlik sistemde yatak sayısı 2,83 gibi orana denk gelmektedir. Ortalama kişiler için yaşam süresi 73,2 olarak bilinmektedir ve vefatların neden olan hastalıklar:

- Dolaşım sistemi rahatsızlıkları
- Kanser
- Solunum hastalıkları

Önem derecesine göre sırasıyla şeklinde meydana gelmektedir. 2020 yılındaki araştırmada 100 ülke arasından 28. sırada yer almaktadır.

ÜÇÜNCÜ BÖLÜM

VERİ VE YÖNTEM

3.1 Metin Madenciliği Nedir?

Metin madenciliğini bilgisayarlar tarafından yeni olarak nitelendirilen olgunun keşfi diyebiliriz. Yeni olgusu mekan bağımsız ve oldukça çeşitli arka planlardan gelebilmektedir. En önemli özelliklerinden biri ise daha önce bilinmeyen bir durumda olmasıdır.

Metin madenciliği sayesinde çeşitli kaynaklardan gelen enformasyonun otomatik olarak anlamlandırılması, gün yüzüne çıkması ve günlük hayata adapte edilmesi söz konusudur. Özellikle geleneksel deneyler sonucunda ortaya çıkan ve üzerinde soru işareti oluşan yeni tezler ve yeni gerçekler ile bağlantı kurulup bunların çözümüne yönelik enformasyonlar üretilmesi açısından oldukça önemli bir kriteri sağlamaktadır. Metin madenciliği tipik bir internet aramasının oldukça ötesindedir. Kullanıcı internet aramasında elindeki enformasyon doğrultusunda sıradan bir arama ile veya daha önce birisi tarafından yazılmış olan bilgiye ulaşarak kafasındaki soru işaretlerini gidermeyi hedeflemektedir. Alınan bu aksiyon neticesinde aslında son kullanıcı elinde bilgiler doğrultusunda kafasındaki soru işaretlerini giderirken yaptığı araştırmada sadece bu bilgiler ile alakalı olan sonuçlara odaklanmaktadır ki bu durum elindeki bilgiler ile alakalı olmayan ve içerisinden yeni iç görüler sunabilecek, yeni bir bakış açısı ve belki de elindeki enformasyonun niteliği ve niceliğini arttırabilecek olan yeni soru işaretleri ortaya çıkmasını engellemektedir.

Aksine metin madenciliğindeki nihai hedef ise daha önce kişiler tarafından bilinmeyen, henüz bir kaynakta yazılı halde olmayan özetle keşfedilmeyi bekleyen bilgiler üzerine yoğunlaşmaktadır.

Sosyal medya içerisindeki hızlı ve hacimli etkileşim sayesinde oldukça fazla veri üretilmektedir. Bu veriler içerisindeki metin türündeki veri hacmi diğer türlere nazaran oldukça çoğunluktadır. Metin madenciliği, veri madenciliğinin bir varyasyonu olarak nitelendirilmektedir. Söz konusu veri toplulukları yani veri tabanları içerisinde barındığı çeşitli metrikler veya birçok alanla ilişki içerisinde olduğu için makine öğrenmesi gibi diğer konular ile ilişki dahilinde kendi uyarlamalarını yapabilmektedir bu sayede veri tabanlarından ilginç ilişkiler yakalayarak bulduklarını ortaya çıkarma hedefindedir.

Veri madenciliği örnekleri arasında müşteri alışveriş süreci içerisindeki hakkında üretilen ve depolanan yasal verileri baz alarak oluşturulan; alınan kupon için farklı bir kupon önerisinde bulunmak, alınan bir telefon ürünü için batarya veya telefon kılıf önerisinde bulunmak, bankacılık harcamalarındaki olağan üstünde miktarların kişinin

düzenli olarak harcama frekansına göre farklı olması ve bunun anlamlandırılması gibi çeşitli örnek vakalar üzerinden de görülen verilerden anlamlı desenler ortaya çıkarılmaktadır. Bu örnekler tabii kurumlar ve karar alıcılar için planlamalarında önemli bir dinamik olarak yer tutarken diğer tarafta müşteri güvenliği ve kurum marka değeri için de yadsınmaz bir üssel artış anlamına gelmektedir.

Veri madenciliği ve metin madenciliği arasındaki oldukça önemli farklar da bulunmaktadır. Bunlardan biri veri madenciliği yapısal olarak nitelendirilebilecek; boyut ve kayıt sayıları belli, ilk bakışta kolayca yorumlanabilecek, veri tabanları programlar tarafından kolayca okunabilecek düzeyde oluşturulmuş özetle belli bir formatta girdi sağlamaktadır. Metin madenciliği ise daha çok doğal dil metinleri, yapısal olmayan ve kolay okunabilmesi, belli bir format haline gelebilmesi için ön temizlemeden geçmesi gereken bir halde bulunmaktadır.

Fakat doğal dil üzerinde gelişen metin veri hacmi ve özellikle verinin önemiyle beraber kurumların bu alana çok rağbet göstermesi metin madenciliği bazında daha kolay yorumlanabilecek ve son kullanıcının önüne oldukça özet bir halde girdi olarak kullanılan metnin bir fotoğrafını gösterebilecek, uygulamalı dil bilimi diğer bir ismiyle doğal dil işleme teknikleri revaçta olmuştur.

Kullanılan teknikler, yazılımlar ve kütüphaneler nezdinde son kullanıcının önüne özet sunulmak üzere girdi olarak kullanılan, görece yapısal olmayan veriler içerisindeki ifadeleri oldukça kolay bir şekilde belli bir formatta, son kullanıcının okuyabileceği şekilde oluşturabilmektedir. Örnek vermek gerekirse tüm metnin içerisindeki en çok geçen isim ve fiilleri yoğunluğuna göre sıralarken diğer bir taraftan son kullanıcının oldukça hakim olduğu ve metnin özetini okurken görmek istemeyeceği bağlaçları ilgili özetten çıkarıp ve bu formatta bir yapı ile analitik bazda ortaya koy şeklinde bir yapı raporlanabilmektedir.

Şekil 3.1: Metin Madenciliği Akışı

Metin Madenciliği Akışı



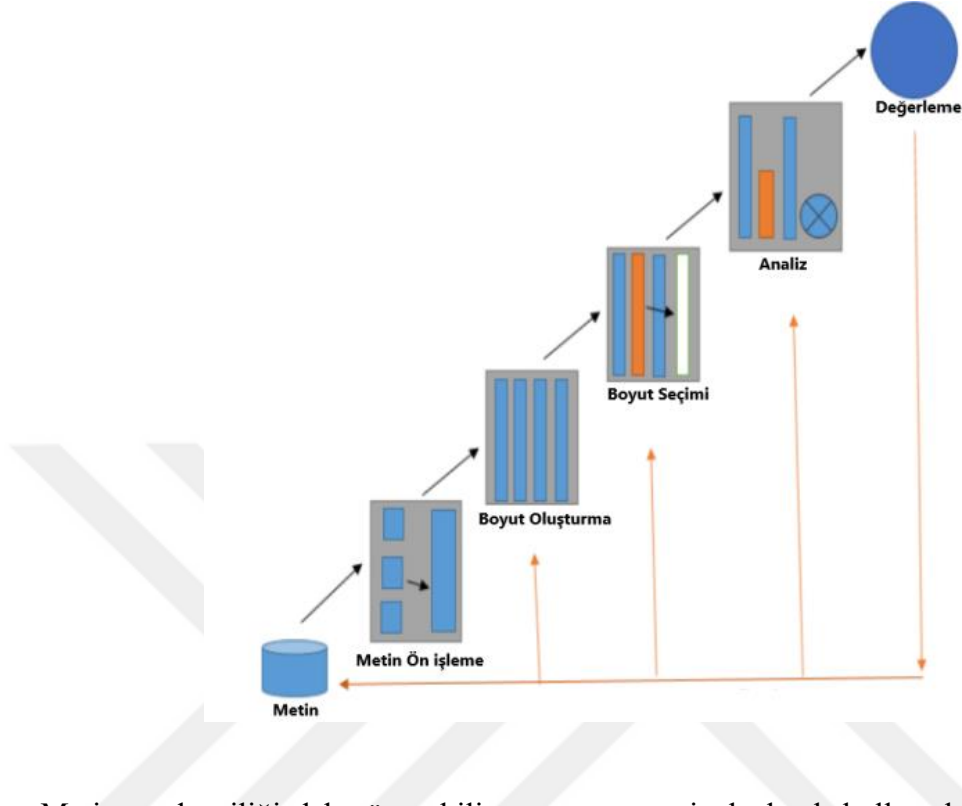
Yüzde 80'lere varacak şekilde çeşitli sektörlerde insan gücünün daha efektif kullanılabilmesi bir görev dağılımı vesilesi ile oldukça önem kazanan bir hal almıştır. Bu durum özellikle insan kaynakları uygulamalarında kullanılan ve sektörü bakımından birçok iş unvanına sahip işe alımlar yapan bir departman için oldukça kritik bir örnek olabilir. Gelen özgeçmiş üzerinden iş yetkinlikleri, oturduğu bölge, bildiği ve belgelerine sahip olduğu teknolojileri analiz ederek oldukça kritik bir doğruluk oranı ile karar alıcıya aday için önemli analitikler sunmaktadır. Tabii bu yapı oldukça hızlı ve ihtiyaç karşılayan bir uygulamaya sahip olsa bile tamamen bir metin madenciliği olarak adlandırılması düşünülmemektedir. Gerçek metin madenciliği var olanı betimlemenin ötesinde aslında bilgi üzerinden yeni görüşler ortaya çıkarmayı ve metin verisi içerisinde yaklaşımları da baz alarak bir akım ortaya koymayı hedeflemektedir.

Bu yapıyı içerisindeki aksiyonlar vesilesi ile detektif rolüne benzetebiliriz. Diğer tarafta ise sadece öne gelen analitikler üzerinden yorumda bulunarak ilerlenmektedir. Ortaya çıkarılan çeşitli analitikler daha önce kurulmuş olan birçok ilişki ile son kullanıcının önüne münferit önemli boyut halinde özet sunabilmektedir. Bunlardan bazıları belki de gerçek hayatta hiç doğrudan süreci etkileyemeyen ve analiz

sonucunda oldukça önemli görünürken, arařtırmacı tarafında en çok aksiyon alınması gereken boyut olarak gün yüzüne gelebilmektedir.

Fakat bu özet halde gelen analitikleri yorumlarken ortaya çıkan olumsuz bir durumdur ve yanlış aksiyon olumsuz sonuçları beraberinde getirirken, aynı zamanda bořa harcanan arařtırma maliyeti anlamına da gelebilmektedir. Metin madenciliğinin doğasında olan ve tabiri caizse veri detektifliğı aslında örnek vaka üzerinden yapılan arařtırmalar, ortaya çıkan soru işaretleri ve olası çözümleri ile alınacak aksiyon tanımlanmadan önce vakanın felsefi ve terminolojisi konusunda oldukça hakim bir arařtırmacı ortaya çıkarmaktadır. Arařtırmacı analitikler içerisinde oldukça ön planda görünen fakat elde ettiğı deneyimler doğrultusunda aslında örnek vakanın iş süreçleriyle hiçbir ilişkisi olmayan bir boyutu dikkate almayacaktır. Böylece alınacak aksiyonu gerçek hayatla ilişkili bir şekilde, en verimli haliyle ortaya çıkaracak ve ortadaki problemi çözerken kaynakları oldukça etkin bir şekilde kullanacaktır. Metin madenciliğinde iki kısıt söz konusudur. Bunlardan bir tanesi var olan veri üzerinde otomatik bir şekilde yorum yapabilecek, anlamlı sonuçlar çıkarabilecek program yazmanın zorluğu ve diğeri bir büyük problem ise kullanılmak istenen verinin tamamen metin halinde tutulmamış olmasıdır. Otomatik halde bir program ile deneyimli bir iş bilgisine sahip arařtırmacının oldukça faydalı metrikler çıkarabilecek fırsat maliyetini göz ardı etmiş ve potansiyel aksiyon alınabilecek boyutları arařtırmacı kaçırmış olabilmektedir. (Hearst, 2003: 3)

Şekil 3.2: Metin Madenciliği Adımları



Metin madenciliği daha önce bilinmeyen, potansiyel olarak kullanışlı, göz önünde olmayan diğer bir deyişle saklı bilginin oldukça yüksek miktardaki metinsel veriden anlamlı sonuçlar çıkarılması işlemidir. (Waegel., 2006: 1).

Oldukça yüksek olan hacim içerisinde aslında yalnızca bir kısım veri modele uygulanırken anlamlı sonuçlar çıkarabilmesi için başarılı olmaktadır. Metin madenciliği araştırmalarında en büyük kısıtlardan biri de bu yapısal olmayan ya da yarı yapısal olarak nitelendirilebilecek veridir.

Yapısal veri kolay bir şekilde organize edilebilir, sabit bir yapıdayken yapısal olmayan veri kompleks ve genellikle kalitatif olarak nitelendirilebilecek verilerdir. Yarı yapısal olan veri ise her iki elementi de bünyesinde taşıyan yapılar için kullanılmaktadır.

3.1.1 Yapısal Veri Nedir?

Organize etmesi ve isterler doğrultusunda aranması oldukça kolay olan veridir. Veri oluşturulan bir formata yüklenmesi sonucunda yapısal hale gelmektedir. Bu format daha önce tanımlamaları yapılmış, analiz yapacak olan kişi için kolay ulaşılabilir ve yorumlanabilir hale gelmiştir.

Günlük hayatta operasyonel süreçlerde en çok kullanılan ve göz önünde olan bir format örneği olarak Excel verilebilmektedir. Yapılan satışların veya işe alınan kişilerin bölge, ürün, tüketici, sektör bazında tutulması ve boyutlara ayrılması en iyi örneklerden biridir. İlişkisel veri tabanları da diğer tarafta en çok kullanılan örneklerden biridir. Çünkü ilişkisel veri tabanları içerisindeki tablo yapıları sayesinde verileri anlamlı bir şekilde gruplamaktadır ve böylece analiz için gönderilen sorgular neticesinde gruplanan niteliğe göre boyutlar bölünmektedir ve analitikleri beslemektedir. Bugünlerde toplam veri içerisindeki payı %20'yi geçmeyecek olan yapısal veriler depolaması, analizi ve erişimi kolay bir yapı dahilindedir.

Yapısal olan veri de makine ve insan tarafından üretilmektedir. Muhasebe ve satış kayıtları bu konuda verilebilecek en iyi örneklerden bir tanesidir. Kurumlardaki oluşturulan kurumsal kaynak planlama stratejileri doğrultusunda isterler doğrultusunda kurumun ilgili iş modeline uygulanan ve yapısal veri üreten bu yapılar oldukça revaçtadır.

Farklı bir örnek ise üretim aşamalarında kullanılan ve insan müdahalesi nadiren gereken makinelerin bakım rutinleri, hata mesajları, üretim aksaklıkları da yazılan ve arka planda depolanan yapısal veriler ile kontrol altında tutulabilmektedir. Tabii sadece fiziki ve belli bir lokasyona bağlı yapılar için bu geçerli değildir. Mobil olarak kullanılan cihazlar için oluşturulan ve kullanıcılarını bu doğrultuda bilgilendiren yapılar da örnek olarak verilebilir.

3.1.2 Yapısal Olmayan Veri Nedir?

Yapısal olmayan veri günümüz dünyasında toplam veri hacmi içerisinde oldukça önemli bir miktarı barındırmaktadır. İçerisinde herhangi bir şekilde veri tabanlarında olduğu gibi satır boyut yapısı ve veri modeli ile ilişkisi bulunmamaktadır.

Bu konuya verilebilecek en iyi örneklerden biri yapısal olmayan ve oldukça önemli potansiyel veriler barındıran, kurumların siber güvenlik uygulamalarında da ön planda tuttuğu e-postalar düşünülebilir. Kurumların önceki evrelerde mümkün olduğu kadar uzak durmak istediği fakat gelişen teknoloji sayesinde yapay zeka uygulamaları ve makine öğrenmesi algoritmaları ile daha kolay yorumlanabilir hale gelen yönetilmesi ve analiz edilmesi oldukça zor olan bu veri yapısıdır.

Veriler bu yapıda belli bir tablo yapısından ziyade NoSQL adı verilen teknoloji ile tutulmaktadır. Bu yapıdaki verileri yapay zeka uygulamaları değerli hale getirmektedir. Yapısal olmayan veri yapısındaki diğer örnekler:

- PDF
- Video
- Fotoğraf
- Sunumlar
- Uydu Görüntüleri
- Çağrı merkezi kayıtları vb.

3.1.3 Yarı Yapısal Veri Nedir?

Yapısal ve yapısal olmayan veri türlerinin ötesinde oldukça çok rastlanılan, yapısal ve yapısal olmayan veri türlerinin karışımı olarak nitelendirilebileceğimiz bu tür yapısal veri kadar tümüyle ilişkilendirilmiş ve yapısal olmayan veri gibi tamamen sınıflandırılmayan bir yapıda görünmemektedir. Örnek üzerinden ilerleyecek olursak telefonunuz ile çektiğiniz bir fotoğraf yapısal olan veri gibi içeriği anlamlandırılmaz

fakat paylaşım esnasındaki coğrafi konum, atanan etiketler, tarih ve zaman verisi bazı detaylar verinin tamamen yapısal olmayan veriden ziyada çeşitli metrikler içermesinden dolayı yarı yapısal veri olarak gruplanmasını sağlamaktadır. Birçok kişinin de bahsettiği gibi sınıflandırılabilir yapısal olmayan veri yarı yapısal veridir.

Metin madenciliği, veri madenciliğinden farklı olarak yarı yapısal ve yapısal olmayan veriden de anlamlı sonuçlar elde etmek istemektedir. Bunlar içerisinde e-postalar, html dosyaları, metin dokümanları bulunmaktadır. Metin madenciliği bu işlemleri ilerletirken elde edilen yapısal olmayan veya yarı yapısal verileri ilk etapta modele sokup anlamlı sonuçlar elde etmemektedir. Veri toplandıktan sonra modele uygun, anlamlı metrikler ve analitikler ortaya çıkarabilmek için ön işleme sürecinden geçirmektedir.

Bu işlem sayesinde veri daha çok yapısal formata benzetilip, çeşitli teknikler sayesinde kümeleme ve sınıflandırma gibi analizler ve durum tespiti açısından doğrudan temiz veriyi görselleştirerek uygulama gerçekleştirilebilir. Metin madenciliği terimi ise doğal dil metinlerinden dilsel kullanışlı ve mümkün olduğunca doğru desenler ortaya çıkarmayı hedeflemektedir.

Metin madenciliğinin anlamlı sonuçlar çıkarmak için birçok alan ile kesişimi mevcuttur. Bunlardan bazılarını sayacak olursak. Veri Madenciliği (Data Mining), Doğal Dil İşleme Süreçleri (NLP), Bilgi Çıkarımı (Information Extraction), Bilgi Alma (Information Retrieval), Makine Öğrenmesi, Yapay Zeka'dır.

3.1.4 Süreçler

Enformasyon bazında analizler içerisinde verimli ve anlamlı bir iç görü ortaya çıkarmak için uygulanması gereken bazı aksiyonlar bulunmaktadır.

3.1.4.1 Metin Önışleme

3.1.4.1.1 Metin Temizleme

Metinler içerisinde istenmeyen birçok veri barındırabilirler. Bunları yapılan analizin doğruluğunun daha anlamlı bir hale gelmesi için temizlememiz gerekmektedir. Bu nedenle metin içerisindeki reklamlar, araştırmanın amacına göre bağlaçları, daha düzgün bir formata gelmesi için temizleme işlemleri yapmak elzemdir.

3.1.4.1.2 Tokenizasyon

Süreç içerisinde modelimizin ihtiyaçları doğrultusunda metine bazı işlemler yapmamız gerekebilmektedir. Bu işlemler arasında daha önce planlanan şekilde boşluklara ve noktalama işaretlerine göre metinleri parçalama işlemi gerçekleştirilebilir. Tokenizasyon işlemin bu ister yerine getirilmektedir.

3.1.4.1.3 Konuşma Etiketlemesi Süreci

Tokenizasyon süreci sonucunda elde ettiğimiz veri yapısı ve formatının girdi olarak kullanılması sayesinde uluslararası manada geçerliliği olan ve araştırmanın isterlerine, kullanılan yazılımın yetkinliğine göre kısıtlanabilecek her bir simgeye yapılan atama işlemlerinin bütünüdür. İşlemler içerisinde kullanılan bazı kural setleri ve kütüphaneler olabilmektedir bu yapılan işe çeşitli araştırma değişkenlerinin bir araya gelmesiyle revize edilebilir. Örnek olarak vermek gerekirse ENGTWOL[8] bunlardan bir tanesidir.

3.1.4.2 Metin Dönüşümü (Boyut Oluşturma)

Bu süreç içerisinde iki teknik oldukça fazla kullanılmaktadır. Bunlardan bir tanesi Kelime Çantası (bag of words) bir diğeri ise Vektör Boşluğu (Vector Space)'dur. Metin içerisinde aksiyonları belirleyen, modelin doğruluğunu arttıracak olan, araştırmanın amacı olan soruya yeni birkaç soru ekleyecek ve bakış açısı kazandıracak olan en büyük etken metin içerisindeki boyutlar yani kelimelerdir. Boyutlar

içerisindeki çeşitlilik ve amaç olarak edinilen problem çözümünün hangi iş modeli içerisindeyse onunla korelasyonu bu konuda oldukça önemlidir. Dolayısıyla veri kaynağı, işlenmesi, yeni boyutların gerekiyorsa oluşturulması, uygulanacak en iyi modelin seçilmesi, temizleme aşamasında potansiyel metrik oluşturabilecek bir veri kaybından kaçınılması gibi birçok faktör devreye girmektedir.

3.1.4.3 Madencilik Süreci

Bu aşamada süre gelen birçok bilgide de aktarıldığı gibi metin madenciliği ile veri madenciliği kesişmekte ve kompleks bir yapıda uygulanmaktadır. Bu tamamen çözülmek istenen problemin isterleri doğrultusunda araştırmacının kullanmak istediği, iş bilgisi ve deneyimi doğrultusunda yorumlamak istediğini model ve kulvara göre çeşitlilik göstermektedir. Bu çeşitlilik birçok yeni alternatif çözümü beraberinde getirirken aynı zamanda çok daha değişken proje, zaman ve maddi maliyet anlamında yorumlanabilmektedir.

3.1.4.4 Değerlendirme

Çıkan sonuçları yorumlanması aşamasında, araştırmacının deneyimi, konuya ne kadar dahil olduğu, teknik manada yeterliliği, iş bilgisi konusundaki yeterliliği, problemin cevabı için oldukça geniş bir bakış açısı ile mi yorumluyor gibi birçok etken devreye girmektedir. Bu aşamada çıktı olarak düşünülen metrikler yorumlanırken aynı zamanda farklı bir araştırma süreci için de girdi olarak kullanılabilir. (Kumar, 2013: 37)

Metin madenciliği doğası gereği birden fazla disiplini bir araya getirmektedir. Bu aynı zamanda birçok noktaya dokunan araştırmacıları ve birçok alanla ilişkisi olduğu için alanların kendi içerisindeki isterleri doğrultusunda yeni metin madenciliği problemlerinin ve aynı zamanda çözümlerinin ve tekniklerinin ortaya çıkmasına neden olmaktadır. Bu da paralel olarak aynı problem için çeşitli algoritmaların uygulanması ve araştırmacılar tarafından aynı probleme birden fazla bakış açısıyla yorum getirilmesi anlamlarına da gelmektedir.

Tüm bu olumlu etkileri doğrultusunda metin madenciliğinin çeşitli alanlarda kullanıma dair örnekler verilebilmektedir: Bunlardan bir tanesi bankacılık üzerine olacaktır. Bankacılık alanında kredi tespit skorlarının veri olarak kullanılması ve son kullanıcıya kredi verilip verilmemesi durumu bir örnektir. Diğer bir örnek ise sağlık alanında forumlarda ve çeşitli web sitelerinde kullanıcıların sağlık bazında sorduğu soruları anlamlandırarak tahmini hastalık verisi ortaya çıkartan çeşitli algoritmalarından söz edilebilir. Diğer bir örnek ise web sitelerinde kullanıcıların deneyimleri doğrultusunda oluşturulan çerezleri girdi olarak alıp, son kullanıcıya reklam gösterme aşamasında bu verileri işledikten sonra potansiyel alabileceği, veriler doğrultusunda ihtiyacı olduğu düşünülen ürünlere yönelik reklamlar gösterilmektedir. Bu reklamlar sadece web site bazında da düşünülmemektedir bu yapılar artık lokasyon bağımlı olmaktan çıkarak tutulan veri türlerinin yapısal, yapısal olmayan, yarı yapısal olacak şekilde depolanıp bu minvalde ilgili teknoloji kullanılıp, süreçlerle veriler işlenip, son kullanıcı için aksiyon alınabilir hale getirilebilmektedir.

3.1.4.5 Metin Madenciliği İçin Kullanılan Bazı Araçlar

3.1.4.5.1 Python

Python son dönemde oldukça fazla kişi tarafından kullanılan, öncelikle yapay zeka ve makine öğrenmesi olmak üzere, veri analizi, internet sitesi arka yapısını geliştirme, oyun geliştirme, masaüstü uygulama geliştirme, kaynak planlama programları geliştirme gibi münferit bazı ihtiyaçları karşılayabilecek düzeye gelmiş bir programlama dilidir. Belli bir spesifik konu için geliştirilmemiş daha çok genel amaç ile kullanılan, açık kaynak kodlu yani yazılım ürünü sahibinin herhangi bir amaç için son kullanıcıya tüm potansiyel yetkileri vermesi ve son kullanıcının herhangi bir maliyete katlanmaması, nesneye yönelik içerisinde sınıf ve fonksiyon yapıları kurgulanabilen, yorumlanabilir, yüksek seviyede yani makine dilinden daha çok kişilerin günlük hayatta kullandığı İngilizce diline yakın ve bu sayede oldukça hızlı ve kolay yorumlanabilen, dinamik bir programlama dilidir.

Guido van Rossum tarafından ilk kez 20 Şubat 1991 ortaya çıkarılmıştır. Kelime anlamı neticesinde daha çok bir yılan türü olarak anılabilecek olan bu dil sanılanın aksine Monty Python's Flyin Circus adında bir komedi programından gelmektedir. Diğer birçok programlama dilinde olduğu gibi bu amaçlar kurulmuş bir kurumsal yapı içerisindeki bir grup mühendis tarafından ticari isterler doğrultusunda geliştirilmiş bir programlama dilinden ziyade tek bir kişi tarafından ortaya çıkarılmış ve kullanıma sunulmuştur. Tabii Python içerisindeki tüm bileşenleri tek bir kişinin geliştirmesinden söz etmiyoruz aslında tek bir kişi tarafından fikrin ortaya konulmasına ek olarak Python dünya etrafında ün kazanmasıyla beraber binlerce kullanıcı, test edici ve mühendis tarafından gerçekleştirilen sürekli çalışma sonucunda bu hale gelmiştir.

1999 yılında Guido van Rossum Python için hedeflerini şu şekilde ortaya koymuştur:

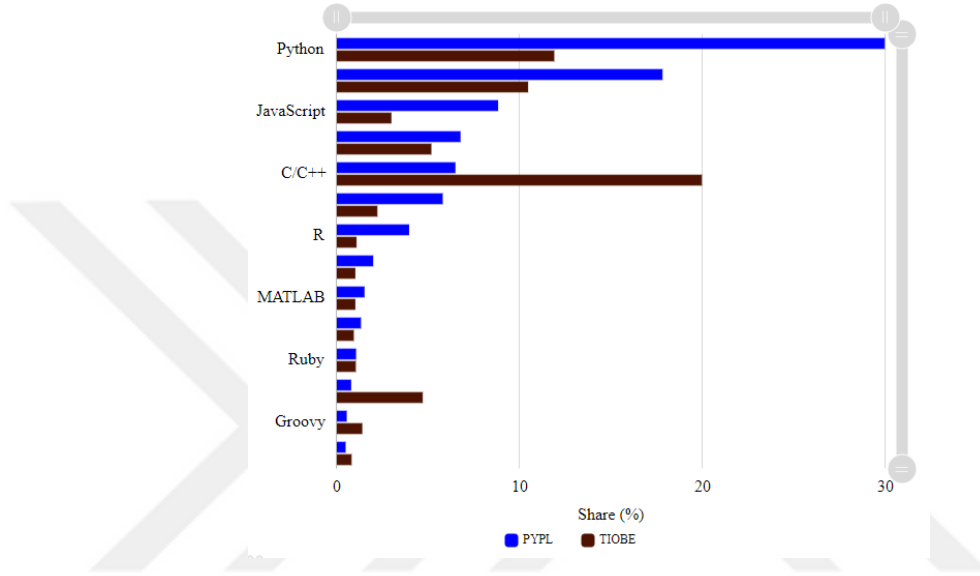
- Büyük rakipleri kadar kolay ve sezgisel bir dil,
- Herhangi bir kişinin katkıda bulunabileceği açık kaynak kodlu bir dil,
- İngilizce dili gibi anlaşılabilir kod yapısı olan bir dil,
- Kısa geliştirme zamanlarına imkan veren günlük hayat ile uyum bir dil,

Hedeflerin gerçekleştirilme oranına bakacak olursak yaklaşık 20 yıl sonra bu isterler uyumlu olarak görüntülenmektedir. Peki son dönemde bu kadar popüler olan bu dilin son kullanıcı tarafında neden bu kadar çok tercih edildiğini sorgulayacak olursak, bu soruya yanıt olması açısından maddeler halinde bu durumun sıralanışı:

- Gerek yüksek seviye bir dil olması gerekse bu amaç doğrultusunda geliştirilmesi Python dili öğrenilmesi oldukça hızlı olan bir dil haline getirmiştir.
- Benzer şekilde öğretmenler tarafından genelde programlama dilleri günlük hayattaki bazı benzetmeler ile akılda kalır hale getirilmektedir, Python günlük konuşma diline yakın bir dil olması nedeniyle öğretilmesi oldukça kolay bir dil olarak yorumlanabilir.
- Kod yazım deneyimi oldukça hızlıdır.
- Farklı bir kodu okumak isteyen son kullanıcı bunun için gerek seviyesi makine diline yakın olan bir dil ile muhattap olmasından dolayı gerekse kullanılan dilin kompleks yapısından dolayı oldukça uzun sürebilmektedir. Bunun yerine Python ile daha hızlı anlaşılır deneyim yaşanılabilir.

- Yazılım dilleri içlerinde erişilebilir olma durumuna göre, son kullanıcı tarafından maliyet katlanması gerekebilecek gibi bazı kısıtlar yüzünden dezavantajlı olabiliyorlar, Python ücretsiz ve kolay erişimi sayesinde bu konuda avantajlı konumdadır.

Şekil 3.3: Yazılım Dili Karşılaştırması



3.1.4.5.2 R Dili

Günümüzde akademik ve özel sektör bazında en çok kullanılan bir diğer dil ise R dilidir. R Python aksine çok daha fazla yapay zeka, makine öğrenmesi ve veri analizi odaklı geliştirilmektedir ve ek olarak oluşturulan model ve analiz çıktılarını belli bir formatta internet sitesi üzerinde de gösterilmesine imkan sağlamaktadır.

Python çok daha geniş yelpazede kütüphaneler çıkarmışken R bu konuda daha odaklı bir ilerleyiş halindedir. Deneyimlerim doğrultusunda R çok daha fazla veri bilimcileri ve analitikçiler tarafından görsel odaklı kütüphaneler ile ön plana geçmekte ve bu bazda olan veri üzerinde yapılacak görselleştirmeler için kullanılmaktadır. R dili Auckland Üniversitesi'nden fakülte arkadaşı olan Robert Gentleman ve Ross Ihaka tarafından 1990'larda uygulanmaya başlanmıştır. R dili, John Chambers, Rick Becker, Trevor Hastie, Allan Wilks ve Bell laboratuvarındaki diğer kişiler tarafından ortaya çıkarılan, 1970 ortalarında geliştirilen ve 1980 yıllarında piyasa sunulan S diline yakın bir şekilde modellenmiştir.

S dili istatistiksel hesaplamalar yapmak için geliştirilmiş bir araçtır. Robert Gentleman ve Ross Ihaka tarafından 1995 yılında açık kaynak olarak kurgulanmış ve 1997 yılından bu yana da R projesi R Core Grup bünyesinde bulunmaktadır. 2000 yılının ilk Şubat ayında piyasaya sürülmüştür. R dili açık kaynak kodlu bir model üzerine oturtulmuştur. Bundan dolayı R dili de Python dilinde olduğu gibi son kullanıcı ve diğer adıyla geliştiriciler tarafından ellerindeki problemleri çözmek amacıyla oluşturulan fonksiyon ve fonksiyon grupları üzerinden kütüphaneler oluşturularak dilin kullanıcı hacminin artmasına ve iş modelindeki bu katmanın daha fazla genişlemesine olanak sağlamaktadır. R dili kullanıcılar tarafından oluşturulmuş 10 binden fazla kütüphaneye sahiptir. Bu durumda R dilinin işlevselliğine oldukça katkıda bulunmuştur. R dili görselleştirme tarafının daha güçlü olduğu varsayılırsa uzmanlara verileri üzerinden oluşturulmuş görseller ile verilerini daha kolay yorumlama ve verileriyle iletişim kurma imkanı sunmaktadır.

R son kullanıcı tarafında oldukça ilgi gören dillerden biridir. Bu tarz durumlarda son kullanıcının deneyimi önem kazanmaktadır. R dil seviyesi bakımından yüksek seviyeye sahip bir dildir. Bu konu da yazım farklılıkları olsa da Python diline benzemektedir. Yüksek seviye bir dil olması R dilinin günlük konuşma diline yakın olması ve kolay yazım deneyimine son kullanıcı tarafından sahip olunması anlamına gelmektedir. R dili deneyimi için daha çok karşılıklı interaktif bir yapı sunulmasından bahsedilmektedir. Bu interaktif yapı sayesinde çok daha fazla deneyimsel, keşifsel ve veri ile oynanabilecek bir platform söz konusu olmaktadır. Son kullanıcı tarafından oluşturulan ve münferit birçok problem için çözüm üreten kütüphaneler sayesinde sorun çözümüne yardım bulabileceksiniz. Daha çok kütüphane olarak kullanılan fakat R dili içerisinde fonksiyonların koleksiyonu olarak adlandırabileceğimiz terimin adı R paketleridir. Gerek terminal üzerinden gerekse arayüz üzerinden paket yüklemeleri kolay bir deneyimle ve ücretsiz bir şekilde oturuma yüklenebilmektedir.

R dili tarihçesi

- 1992: Robert Gentleman and Ross Ihaka tarafından araştırma projesi olarak uygulanmaya başlanmıştır.
- 1993: Statlib içerisinde ilk ikili sayı sistemi olarak uygulanmıştır.
- 1995: GLP2 (Genel Halka Açık Lisans) lisans altında ilk açık kaynak kodlu proje olarak yayınlanmıştır.
- 1997: R Core grup bünyesindeki.
- 1997: CRAN (Kapsamlı R Arşiv Ağı) kurulmuştur.
- 1999: r-project.org kurulmuştur.
- 1999: Viyenede ilk R takımı ile buluşma gerçekleştirilmiştir.
- 2000: R 1.0.0 piyasaya sunulmuştur.
- 2000: John Chambers R takımına katılmıştır.
- 2001: R Haberleri bulunmuştur daha sonrasında R dergisi olarak düzeltilmiştir.
- 2003: R Foundation kurulmuştur.
- 2004: UseR! Konferansı Viyenede gerçekleşmiştir.
- 2004: R 2.0.0 piyasaya sunulmuştur.
- 2009: R dergisinin ilk sayısı yayınlanmıştır.
- 2013: R 3.0.0 piyasaya sürülmüştür.
- 2015: R Konsorsiyumu kuruldu.
- 2016: Yeni R logosu oluşturuldu.
- 2017: CRAN (Kapsamlı R Arşiv Ağı) 10.000 aşkın kullanıcıya ulaştı.
- 2020: R 4.0.0 piyasaya sürüldü.

3.1.4.5.3 Julia Dili

Julia diğer dillere göre çalışmaları daha yakın bir tarihe dayanan ve kullanım istatistiği ve duyum frekansı artan bir dildir. Jeff Bezanson, Stefan Karpinski, Viral B. Shah, ve Alan Edelman tarafından 2009 yılında çalışma başlanmıştır. Bu çalışma başlatılırken esas amaçları yüksek seviyede konuşma diline yakın, hızlı ve açık kaynak kodlu bir dil ortaya koymaktır. Takım 2012 yılında internet sitesi üzerinden yaptığı paylaşımlar aracılığıyla potansiyel kullanıcılarıyla iletişim kurmuştur. 2012 yılından bu yana piyasaya sunulan Julia 10.000'den fazla şirket tarafından kullanım, 29 milyondan fazla indirme, 1.500'den fazla üniversite tarafından indirilme gibi istatistiklere sahiptir.

Julia takımı bu dili bilimsel hesaplamalar, makine öğrenmesi, veri madenciliği, büyük ölçekte veri içeren lineer cebir problemlerinde, dağıtık ve paralel hesaplama alanlarını baz alan işlerde kullanılmak üzere çalışmaktadır.

3.1.4.5.4 Matlab Programı

Veri analizi için kullanılan bir diğer araç ise Matlab'tir. 1970 sonları ve 1980 başlarında Yeni Meksika Üniversitesinde görev yapan bir öğretmen tarafından, lineer cebir ve sayısal analiz dersleri verirken daha verimli bir deneyim sunmak ihtiyacıyla geliştirilmiştir. Öğretmenin ismi Cleve Moler'dır. Moler Fortran programı üzerinden öğrencilerin dersi işleyebilmesi için LINPACK ve EISPACK erişimlerini hiçbir kod yazmadan gerçekleştirmek istiyordu. Daha sonrasında ortaya çıkan bu ihtiyacı karşılamak için Moler neler yapabileceğini araştırmaya başladı ve karşısına Niklaus Wirth'in Algoritmalar + Veri Yapıları = Programlar kitabı çıktı ve bu kitaba çalışarak programlama dillerini nasıl bir yapıya sahip olduğunu ve bu yol ile nasıl program yazabileceğini öğrendi. Bu bilgileri baz alarak öğrencilerine daha iyi bir deneyim sunabilmek için Fortran üzerinde ilk MATLAB'ı yazmış oldu.

Daha sonrasında öğrencilerin katkıları ve dönemde ortaya çıkan ihtiyaçlar doğrultusundaki geliştirme süreçleriyle bugünkü lineer cebir, optimizasyon, nümerik analiz, istatistik, iki ve üç boyutlu grafik çizimi gibi çok daha geniş yelpazede analiz yetenekleri olan MATLAB ortaya çıkmış oldu.

3.2 Veri Madenciliği (Data Mining)

Veri madenciliğini veri içerisinden anlamlı iç görüler çıkarma süreci diyebiliriz. Veri madenciliği veri içerisinden daha önce bilinmeyen, belli bir kaynakta yazılı olmayan, kullanışlı iç görüler çıkarma süreci de denilebilmektedir. Bilgiyi baz alarak bu süreçte son kullanıcı veri madenciliği araçlarını kullanarak durum analizleri, gelecek tahminlemeleri, eğilim çıkarımı, davranış modellemesi gibi çeşitli öngörüler ile söz konusu iş modelindeki karar alıcılar ve modeldeki doğruluk oranını oldukça besleyen bir süreç sunmaktadır. Özellikle kullanılan araçlar sayesinde daha önce belli bir

yapıdaki veriden anlamlı sonuçlar çıkarma süreci geleneksel olarak oldukça zaman kaybedici olmaktadır.

Günümüzde olduğu gibi geniş çaplı ve detaylı bir yapıda bu süreçleri ilerletebilmek çok mümkün olmadığı için araştırmacı oldukça büyük bir potansiyel kullanışlı metrikler çıkarılabilecek anlamlı bir veriden düşünülen kadar verimli bir sonuç çıkartabiliyordu. Artık verinin önem kazanması ve üstüne rekabet stratejisinin üstel olarak geliri artışa geçirebilmesi için veri bazlı kararlar çok önemli hale gelmiştir.

Veri bazlı kararların bu denli önemi kurumların müşterileriyle daha çok etkileşime geçmesi ve bu etkileşimi birçok aşamada veri olarak yasal düzeyde elde ederek kullanabilmesiyle bu hale gelmiştir. Ellerinde veriler sayesinde kullanılan veri madenciliği araçları ile kurumlar hem günümüzdeki koşullarını ölçümlerken hem de araçların önlerine sunduğu gelecek için ön görüşleriyle daha tutarlı ve potansiyel getirisi daha fazla olan planlamalar gerçekleştirebilmektedirler.

3.3 Bilgi Çıkarımı (Information Extraction)

Bilgi çıkarımı yapısal olmayan veya yarı yapısal veri üzerinden otomatik bir şekilde anlamlı veri çıkarım süreci olarak düşünebiliriz. Doğal dil işleme sonucunda anlamlandırılan insan dili metinleri üzerinden daha çok süreçler ilerletilmektedir ve oldukça önemli bir diğer konu ise Google Search Engine bu metodolojiyi kullanmaktadır.

3.4 Doğal Dil İşleme Süreçleri (NLP)

Doğal Dil işleme çok eski ve temel yapay zeka problemlerinden biri olarak düşünebiliriz. Çok fazla kırılımda, çeşitte dil, yaşam biçimi, lokal olarak kullanılan deyimler ve uluslararası anlamda bir şey ifade ettiği için yaşam biçimi bazında oldukça farklılaşan kelimeler olduğu ve bunları herhangi bir yapısal formatta girdi olarak kullanılmadığı düşünülürse, bir bilgisayarın insanların iletişim kurarken kendi aralarında gerçekleştirdiği gibi karşılıklı dil ve cümleleri anlamak, yorumlamak

gerekirse buna cevap vermek şeklinde olan dönüşleri durumu oldukça zor bir problem olarak gün yüzüne çıkarmaktadır. Süreç içerisinde belli başlı algoritmalar ile bu yapı doğruya yakın ve anlamlılığı olacak bir şekilde sonuç verebilmektedir.

Bunlar: İnsanların nasıl bu iletişimi birbirleriyle gerçekleştirebildiği, iletişim esnasında anlamlı bir süreç ortaya çıkması için ne gibi yönlendirici unsurlar bulunmaktadır, kullanılan kelimeler arasında belli başlı bir korelasyon, desen mevcut mudur, anlam bakımından set çekilmiş bazı yapılar mevcut mudur şeklinde olan belirsiz durumlar için sorular yönelmekte ve bunun takibini yapmaktadır. Doğal dil işleme süreçleri metin madenciliği süreçlerine girdi haline getirdiği verileri sunmaktadır. Bu veriler üzerinden işlemler metin madenciliği araştırmacısı tarafından gerçekleştirilmektedir.

3.5 Bilgi Alma (Information Retrieval)

Dokümanların içerisindeki kullanıcı tarafından aranan bilgilerin enformasyonlar bazında yoğunlaştırılmış ve işlenmiş doküman uzantıları olarak düşünülebilir. Kullanıcı tarafından oluşturulan ve geri dönüş için ileriye sürülen sorgular için dokümanlardaki verilerin özet aşaması diyebiliriz. Sistem ulaşılmak istenen ve soru işaretlerinin ortadan kaldırılması hedeflenen konu hakkındaki dokümantasyon çeşitliliğini daraltmaktadır. Özetle birçok potansiyel veri kaynağı için yapılan araştırmadaki uygulanacak olan algoritmaların, ilişki dokümanlara uygulanması için çeşitliliği anlamlı bir şekilde daraltmaktadır. (Kumar ve Bhatia, 2013: 36)

Metin madenciliği, doğal dil metinlerinden anlamlı verilerin edinme sürecidir. (Witten, 2003: 1). Bu bilgi çıkarma sürecini belli bir amaç güdümlü şekilde ilerletebilirsiniz. Tabi bu süreç veritabanlardaki gibi yapısal ve düzenli bir veri topluluğu üzerinden ilerletilmemektedir. Mümkün olan en doğru yapıyla veriden anlam çıkarma işlem otomatik hale getirilmek istenmektedir.

İlk olarak 1980'lerde metin madenciliği yapısı ortaya çıkmıştır. (Scholtes, 2009: 2)

Tabi o zaman şimdiki gibi metin madenciliği araçları değil süreç daha çok el yordamıyla ilerletilmektedir. Araştırma ise daha çok araştırmacı üzerinden yoğunlaşarak çok daha maliyetli bir haldedir.

El yordamıyla sürecin ilerletilmesi araştırmanın niteliğini arttırmak için nitelik, harcanan zaman ile doğru orantılı şekilde ilerletilmektedir. Süreç geçtikçe teknolojinin gelişmesi ve çeşitlik matematik tekniklerinin, istatistik tekniklerinin, dil bilimi bazında ve desen yakalama amaçlı ortaya çıkan teknikleri vesilesi ile oldukça hızlı ve üstel bir ilerleyiş yolu kat edilmiştir. Bu sayede çok daha az maliyet ve zaman içerisinde çok daha anlamlı bütün verinin yapısını, ilişkisini ortaya koyan modeller ortaya çıkarılmıştır. Böylece daha araştırılabilir ve kolay ulaşılabilir metrikler sektörlere yön vermiştir.

3.6 Makine Öğrenmesi

Makine öğrenmesi, deneyimleri kullanarak hesaplamalı bir şekilde performansı arttırmak ve doğru tahminlemeyi gerçekleştirmek için kullanılan bir metottur. Deneyimden kasıt ise aslında kurulan model içerisindeki öğrenciye verilen geçmişte yaşanmış olan verilerin bütünüdür. Elektronik halde elde edilen ve belli bir yapıda depolanan bu veriler analizler için uygun halde öğrenciye sunulmaktadır. Genelde bu veri setleri ise çevreden etkileşim sonucu elde edilmiş olan veya insan tarafından etiketlenen yapılar olarak gün yüzüne gelmektedir. Tüm bu konuların çatısı olarak düşünebileceğimiz bu durum algoritma içerisindeki öğrenci tarafından oldukça önemlidir. Çünkü öğrencinin yaptığı tahminlerin başarısını toplanan verilerin kalitesinde önemli bir rol oynamaktadır. Makine öğrenmesi içerisinde tümüyle dizayn edilmiş verimli ve doğru tahminleme algoritmalarını içermektedir. Fakat çoğu bilimde de olduğu gibi özellikle bilgisayar tabanlı olanlarda, bu algoritmalarının kritik göstergeleri çalışma zamanları ve kompleks alanlarıdır.

Makine öğrenmesi içerisinde ise kavram ailesini değerlendirebilmek için örneklem karmaşıklığı kavramı da kullanılmaktadır ki daha çok teorik öğrenme bunları

garantilemektedir. Günümüzde bilim dallarının gelişen teknoloji, fiziki hayattaki durumlar ve insan ihtiyaçları doğrultusunda birden fazla disiplin iletişim halinde olması ve problemlere kendi nezdinde metotları ile çözümler üretmesi talebini ortaya koydu. Bu konuya verilebilecek en güzel örneklerden biri finans ve psikoloji bilimlerinin birleşimi ve ortaya koydukları davranışsal finans kavramıdır.

Davranışsal finans, sadece rakamlar üzerinden yeteri kadar yorumlamanın yapılamayacağını ve bundan dolayı insan psikolojisinin de devreye girdiğini savunarak bu yeni bir kavram ortaya konulmasını ve bu bazda yazılan kitabın Nobel ödülü almasını sağlamıştır. Bunun gibi birçok alan da kendi içerisindeki kesişimleri doğrultusunda yeni yorumlar ortaya koymaktadırlar. Veri bilimi, yapay zeka ve makine öğrenmesi kavramları da bu konu oldukça revaçta olan gerek tıp gerekse finans vb. birçok alan kesişim halinde bulunan alanlardır. Öğrenme teknikleri daha çok veri üzerinden ilerlemektedir. Oluşan bu veri bazlı teknikler makine öğrenmesini de beslemektedir.

Makine öğrenmesi içerisinde istatistik, bilgisayar bilimi, olasılık ve optimizasyon konularını içermektedir. Tabi bu konuya ek olarak ilgili modelin uygulandığı disiplin, eğer bu bir kurumsal firma içerisindeyse bunu bir departman ve iş modeli olarak da düşünebiliriz, doğrultusunda yorumlanabilecek ve böylece verimli bir besleme sağlayabilecek yapı olarak düşünebiliriz. Veri madenciliği içerisinde ortaya çıkan birçok problem için araştırmacı süreç içerisinde analiz hataları yapabilmektedir ve bu tarz durumlara çok kez de karşılaşılmıştır. Özellikle birden çok boyut için iş bilgisi doğrultusunda tamamen sorunu anlayıp, yorumlamadan en verimli şekilde ilişki kurması gereken aşamada süreç boyutlar arasında anlamsız bir ilişki kurulması gibi olumsuz bir sonuç ile devam edebilmektedir.

Bu durumda belli olan problemin çözümü için bir alternatif bulunamaması anlamına gelmektedir. Fakat makine öğrenmesi bu tarz durumlar için oldukça proaktif ve verimli bir süreç ortaya çıkmasını sağlayan mentaliteye sahiptir. Bu tarz problemlere makine öğrenmesi algoritmalarının uygulaması süreci otomatik hale getirmeyi ve verimliliğini arttırmayı sağlamaktadır. Makine öğrenmesi içerisinde birçok problem için benzer

değişkenleri kullanmakta ve bu değişkenler üzerine modeller kurmaktadır. Bu özellikler sürekli, kategorik ve ikili sayı sistemi denilen alternatifler olarak genel görünmektedir.

Eğer veriler etiketler atanarak ve çıktı ile konuşabiliyorsa bu tarz makine öğrenmesi içerisindeki algoritmalara denetimli öğrenme algoritmaları denmektedir. Eğer veriler etiket atanmadıysa ve belli bir çıktı değeri ile iletişim halinde değilseler bu tarz makine öğrenmesi algoritmalarına denetimsiz öğrenme algoritmaları denmektedir.

Denetimsiz öğrenme algoritmaları araştırmacılar tarafında daha çok bilinmeyen, faydalı, sınıflandırma nesnelere keşfetmeyi amaçlamaktadır. Diğer bir makine öğrenmesi algoritmalarındaki öğrenme türü ise pekiştirmeli öğrenmedir. Bu öğrenme türü ile araştırmacı kullandığı model içerisinde sürecin performansı ölçen yapılar kurmuştur ve performansın olumlu veya olumsuz olmasına göre derecelendirmektedir ve böylece hedefe ulaşma doğruluk değerini arttırmayı sağlamaktadır. Makine öğrenmesi algoritmalarında kullanılan modellerin ve bu modeller içerisindeki yazılan koda girdi sağlayan, araştırmacı tarafından değiştirilebilen ve bu değişiklik sayesinde doğruluk katsayısının arttıran değişimler oldukça önemlidir.

Bu önem kadar modele girdi sağlayacak olan veri setinin ne kadar hazır olduğu, ön temizleme işleminden geçip geçmediği veya geçiş sırasında temizlik işlemini başarısı da oldukça önemlidir.

Veri hakkında neler hatalı olabilir? Birkaç başlık ve açıklamalar silsilesi ile bu konuyu ele alalım:

- İmkansız ve muhtemel olmayan değerlerin girdi olarak alınması,
- Değer olmayan girdiler,
- Alakasız boyutların elde bulunması,

İmkansız ve muhtemel olmayan değerler için literatürde oldukça fazla kullanılan kelime gürültülü veri kavramıdır. Gürültülü veri aynı boyut üzerinden diğer değerleri anımsatan fakat veri giriş veya dönüşüm sırasında yapılan hata neticesinde anlamlı

olmayan verilere denmektedir. Gürültülü arařtırmacı tarafından veriye ilk bakıř olarak sylenebilecek zet analitikler incelenirken diđer metriklere gre farklılařtıđı iin hemen gz nne gelebilmektedir. Bu farkına varma aynı zamanda yapılan alıřmanın veri kalitesini arttırmak iin yapılabilecek veri dzeltmesi anlamına da gelmektedir.

Grltl veriler kullanılan yazılım aracı sayesinde kolayca fark edilip arařtırmacı tarafından veya veri sađlayan kiři, ara tarafından yeniden girmesi sađlanabilir. Bu dnřm veri kalitesi arttırabilir rneđi dođrudur fakat eđer geređi kadar zen gsterilmezse veri kalitesi arttırma sreci potansiyel veri kaybı ve hatalı veri giriři anlamlarına da gelebilmektedir. Bu kategoriye geiř sonucunda veriden kaldırma kaliteyi arttırmak iin olduka nemli bir ařama olarak deđerlendirilebilir. Bu tarz aykırı (outlier) denilebilecek deđerler gnmzde retilen veri hacmi baz alındıđında olduka n plana ıkmaktadır. Bu ihtiya dođrultusunda birok aykırı deđer tespitine ynelik algoritmalar ortaya ıkarılmıřtır. zellikle gerek dnya deneyiminde tamamlanmamıř diyebileceđimiz veriler yadsınamaz bir gerektir. Bu tarz deđerleri ele alırken bilinmezliđine ynelik birok tekniđi gz nne almak gerekmektedir. Bunlar verinin belirsizlik durumu, arařtırmacı ve arařtırma modeli tarafından nemsenmeyen bir deđer olabilmektedir, verilen rnek iin gerekleřmemiř olabilir tm bu alternatif deđerkenler bazında kayıp verinin niteliđi, modele girdi sađlayıp sađlamama kararı alınabilmektedir.

Bu neticede yapılan gruplamalar, verinin daha az deđerřkene indirgenme iřlemleri ve kayıp veri temizleme, dzeltme iřlemleri alakasız ve artık olarak nitelendirilebilecek olan grltnn veri ierisinden kalkmasına neden olmaktadır. Bu indirgeme iřlemleri kullanılacak olan algoritmaya gre deđerřkenlik gstermekle beraber modelin daha hızlı iřlenmesi ve daha dođru sonular ortaya ıkarmasına fayda sađlamaktadır. Model ierisindeki bu artıřı ve performansı girdi olarak kullanılan boyutlar etkilemektedir ve bu yapı her bir boyutun bir diđerine olan bađı model performansı ve dođruluk katsayısı ile olduka iyi bir dzeyde iliřki ierisindedir. Bu problem ise genellikle dođruluk katsayısının artması ile ařırı đrenmeye neden olabilecek potansiyele sahip, baz olarak alınan veri ierisinden yeni boyutların ortaya ıkarılmasını nedeniyle gerekleřmektedir.

Hangi algoritmanın kullanılacağına verilecek olan karar da bu konuda kritik bir nedendir. Bu konuda ise arařtırmacının deneyimi ve yapılacak olan ön eleme testi oldukça tatmin edici olacaktır. Genellikle rutin kullanımlar sırasında ön plana çıkan sınıflandırıcılara bu konuda oldukça rastlanmaktadır. Bu sınıflandırıcıları seçerken oldukça önemli olan tahmin performans değerlendirmesidir. Sınıflandırıcının doğruluğunu belirlerken genelde kullan birkaç teknik bulunmaktadır. Bunlardan biri eğitim setinin parçalara bölünmesi ve bölünen parçalardan iki tanesi eğitim için girdi sağlarken diđer bir tanesi ise hedeflenen kırımla örtüşüp örtüşmediğini veya hedefe yaklaşma derecesini teyit eden performans ölçümü için kullanılabilir.

Bir diđer teknik ise çapraz doğrulama tekniğidir eşit olarak alt kümeler halinde parçalanmış veri seti ve bu alt kümelerin bir tanesi doğrulayacak şekilde her bir alt kümenin modele girdi olarak sağlanarak performansın test edilmesine dayanmaktadır. Her bir işlem sonrasındaki ortaya çıkan olması gereken değerlerle aradaki farkı gösteren hata oranı ise performans metriği olarak değerlendirilebilmektedir.

Diđer bir değerlendirme işlemi ise birini dışarıda bırak şeklinde isimlendirilen algoritmadır. Bu algoritma diđerlerine göre donanım, zaman ve maliyet açısından oldukça gideri olan bir algoritmadır. Tüm alt kümelerin sadece birer örnek ile adreslendiği ve bu yapı üzerinden performans değerinin ölçümlendiği bu algoritma diđer alternatiflerine göre kısıtları da göz önüne alırsak daha az kullanılmaktadır. Maliyetine nazaran yüksek doğruluk katsayısı istenen örnekler için oldukça faydalı bir algoritma diyebiliriz. Maliyeti ise daha çok her bir örnek için adreslenen alt kümelerin kendi araların bir döngüye girmesi ve döngünün iřetilmesi için gereken hesaplama maliyeti diyebiliriz. Özetle makine öğrenmesi modellerini birçok tür altında gruptandırabilmek mümkündür. Bunlar denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme şeklindedir. Daha sonrasında bölümlerde oldukça detaylı açıklayacağımız bu alternatifler için günümüz dünyasında hem akademik manada hem de özel sektör bazında iş zekası ve veri analizi iş modelleri içerisinde kullanım frekansı en yüksek olan tür denetimli öğrenmedir. Durum böyle olunca denetimli öğrenme teknikleri üzerinden yapılan araştırma ve üretim sayısı diđerlerine nazaran çok daha fazla görünmektedir. Fakat yeni dönemde özellikle nesnelerin interneti olarak adlandırılan fiziki nesnelerin internet bazında birbiriyle iletişim halinde olmaları ve bu

iletiřim neticesinde üstel olarak artış gösterecek hacime sahip olan veri artışını da baz alırsak denetimli öğrenme teknikleri yerini daha çok diđer alternatiflerine bırakacak gibi durmaktadır. (Maglogiannis, 2007: 4)

Makine öğrenmesi, makinelere ayrıntılı ve açık bir şekilde programlanmadan öğrenme imkanı sunmaktadır. Diđer bir ifade ile makine öğrenmesi insanların yaptıklarını belli başlı pratikler ile makinelerin yapabilmesine hazırlık yapmaktır. Makine öğrenmesi, yapay zeka olarak bilinen çok daha geniş olan bu disiplinin aslında bir alanı olarak değerlendirilebilir. Belli bir problemi çözmek veya var olan bir veri seti üzerinde potansiyel çözümlere ya da potansiyel araştırma konularını ortaya çıkarmak amacıyla kullanılmaktadır. Sürekli olarak veri seti üzerinden hedeflenen ve bu bazda ayrıştırılan veriler yeni yapılar ve desenler keşfetmek amacıyla algoritma tarafından öğrenilmektedir.

Bu sayede veri bazında ve niteliksel deneyime göre kurumlar veya kişiler için baz alınan veri seti içerisindeki iç tehdit unsurlarını, kişilerin güvenli olmalarına engel olan unsurların tespiti ve aksiyon bildirisi gibi amaçlarla kullanılabilir. Bu doğrultusu verinin belli bir lokasyonda, bulut teknolojisiyle, kurulu sunucular üzerinden, kurumsal kaynak ve lisanslı ürün veya açık kaynak kodlu bir veritabanı ile tutuluyor olmasının herhangi bir farklılığı yoktur tabi bu konuda güvenlik gereklilikleri yerine getirilmiş olmalıdır.

3.6.1 Makine Öğrenmesi ile Çözölebilen Problemler

- Metin veya doküman sınıflandırması, bu sayede istenmeyen metinler e-posta adreslerimize düşmemekte ve kurulan yapı sayesinde istenilen sınıflandırmaya göre dosyalara parçalanmaktadır.
- Doğal dil işleme ile elde edilen kelimelere etiket ataması yapılabilmektedir, Bu doğrultuda iletiřim halinde olunan herhangi bir örnek için tüm metni anlamlı bir özet şeklinde son kullanıcıya sunmaktadır.

- Konuşma işleme uygulamaları, kullanıcının talimatları yerine getirmek için öncelikle sesini tanması ve doğrulaması gerekmektedir bu özellik sayesinde bu aksiyon alınabilir. Kullanıcı tarafından gelen talimatlar tekrar bu özellik kullanılarak algılanmakta ve cihaz tarafından yerine getirilmektedir.
- Bilgisayar görüşü uygulamaları; nesne tanıma, nesne doğrulama ve nesne üzerinden verilen kısıta göre olması veya olmaması gereken niteliklere göre, uyarı sistemlerini aktif hale getirmesi veya son kullanıcıya analitikler sunması örneklerdendir.
- Hesaplamalı biyoloji uygulamaları, protein fonksiyon tahminlemeleri, gen ve protein ağı analizleri verilebilecek örneklerdir.
- Finans sektöründe sahtekarlık tespiti,
- Yardımcı olmadan sürülen araçlar,
- Öneri sistemleri,
- Arama motoru algoritmaları,
- Öğretilmiş oyun oynayan algoritmalar,

Bu konuda gündemde olan makine öğrenmesi vakalarıdır. (Mohri, Rostamizade, 2018: 2)

3.6.2 Makine Öğrenmesi Algoritma Türleri

Makine öğrenmesi dönemin ihtiyaçları ve akademik bazdaki isterleri de göz önüne aldığımız oldukça çeşitli algoritmalarına sahiptir. Özellikle makine öğrenmesi için kullanılan yazılımların açık kaynak kodlu olması (yazılım ürünü sahibinin herhangi bir amaç için son kullanıcıya tüm potansiyel yetkileri vermesi ve son kullanıcının herhangi bir maliyete katlanmaması), kişilerin oluşturdukları fonksiyonları topluluklarını kütüphane şeklinde diğer kişilerle paylaşım kolaylığı sayesinde akademik platformlar üzerinden paylaşılan bir makale ile hemen koda dökülebilmekte ve oluşturulan fonksiyon veya fonksiyon toplulukları ile diğer kişilerin karşılaştıkları problemlere hemen uygulayabilmelerine olanak sağlanabilmektedir. Hatta var olan bir

kütüphane üzerine geliştirme aşamasındaki deneyimi kolaylaştırmak için bir yenisini de çıkarabilecek kadar verimli bir ortam söz konusudur.

3.6.2.1 Denetimli Öğrenme

Belli bir hedef değer çerçevesinde tasarlanan öğrenme şeklidir. Bu bazdaki makine öğrenmesi modellerinde ise hedef değer denilen etiketler üzerine kurulmakta ve eğitilmektedir. Denetimli öğrenme de kendi içerisinde kısıtlımlara ayrılmaktadır. Bunlardan bir tanesi sınıflandırma modelidir.

Sınıflandırma modelinde hedef değer “Olumlu”, “Olumsuz”, “Tutarlı”, “Tutarsız” gibi görünmektedir. En çok kullanılan sınıflandırma modeli ise yapay sinir ağlarıdır (ANN). Yapay Sinir Ağları doğrusal olmayan problemler için özellikle baz alınabilecek bir yapıya sahiptir. Birkaç boyuta sahip ve yeterli veri miktarına sahip etiket atanmış veri setleri üzerinde kullanılmaktadır ki algoritmanın çalışma mantığını bir yapıya benzetecek olursak içerisindeki milyonlarca simüle edilmiş ve hedef doğrultusunda çalışan beyin hücresi sayesinde adete insan beyni gibi emirleri alan, uygulayan ve iletken bir sistematik ile çalışmaktadır.

Sinir ağları genellikle birbirleriyle iletişim halinde olan milyonlarca zihin hücresi gibi davranan yazılım simülasyonları tarafından oluşmaktadır. Bu yapı karmaşıklıkların çözümü, yargılara varım gibi aksiyonları vesilesiyle insan zihni gibi işleyen bir süreçten söz edilmesine mahal vermektedir. Özellikle orta kısımda görünmez olarak bulunan ve eklenmiş verileri bir dizi veri seti içerisinde gelen boyut ile destekleyen aktivasyon katmanından oluşur. Diğer bir kısıtlım ise regresyon modelidir. Regresyon modeli gerek sektörde gerekse araştırmalarda oldukça fazla gözlemlenmektedir. Regresyon modelinde hedeflenen değer ise dinamik bir şekilde olan miktar, satış tutarı gibi örnekler verebileceğimiz değerlerdir.

Denetimli öğrenme makine öğrenmesi algoritmaları içerisindeki belirlenmiş datalar çerçevesinde geliştirilen öğrenme türüdür. Pratikte kullanılan veriler ise sembol halinde kullanılacak olursa eğer değişkenlerin kombinasyonu şeklinde düşünülebilir. Baz aldığımız veride kullanılan değişkenleri “x” ve “y” olarak adresleyecek olursak eğer “x” modele girdi sağlayan değişkenleri temsil ederken “y” modelin ürettiği ve “x”

değerlerini baz alarak belli başlı katsayı oranları ile düzenlenmiş üretilen değeri temsil etmektedir. Bu sayede model girdi ve çıktı değişkenleri ile taslak bir çizim uygulamaktadır.

3.6.2.2 Denetimsiz Öğrenme

Denetimsiz öğrenmede ise denetimli öğrenmede olduğu gibi veri içerisinde hedef olarak kullanılabilir ve modelleri bu bazda kurabileceğimiz bir yapı bulunmamaktadır. Bundan dolayı denetimsiz öğrenme modeli kullanılacak olan veri seti diğer öğrenme modeline göre oldukça farklılaşmaktadır. Veri seti kendi içerisinde belli bir dağılım gösteren, farkedilebilecek bir yapıya sahip olması elzem bir hal almıştır. İçerisinde modele boyut olarak girdi sağlayacak olan özellik değişkenleri üzerine kurulacak olan bu öğrenme modelinin başarısında önemli ve aktif bir rol oynamaktadır.

En çok rastlanılan türlerinden biri ise kümeleme problemleridir. Çeşitli anlamlı gruplamalar ile bu problem çözülebilmektedir. Model performansını ölçümleyebileceğimiz bir hedef değer yapısı bulunmadığı için denetimsiz öğrenme metodlarının problem çözümünde kullandığı en önemli özellikler sıfırdan oluşturmaya çalıştığı bağlantılar ve desenlerdir. Bu konu da satış gelirleri için düşünce olursak yine aynı amaca hizmet eden satışların iş boyutlandırmasını sağlayıp, sosyal bir araştırma sayesinde daha son kullanıcı ile iletişim halinde bir yapı sunmaktadır. Bu sayede doğru yatırımlar ve pazarlama planları bu boyutlandırma ile daha verimli hale gelebilmektedir.

Boyut azaltma ve anomali tespiti denetimsiz öğrenme modelleri içerisinde kullanılan diğer örnekleridir. Boyutsal azaltma metodu veri içerisindeki rastgele olan ve anlam düzeyi düşük olan boyutları azaltma aksiyonu olarak model çıktısı vermektedir. Bunu yaparken algoritması gereği bir dizi temel değişken ele almakta ve amacını uyarlamaktadır. Veri içerisinde analizin efektif olmasını sağlayan özellik değişkenlerini boyut azaltma metodu ile daha anlamlı ve efektif hale getirebilmekteyiz. Anormallik algılama tekniği ise oldukça büyük ve çeşitli dağılımlara sahip olan veri içerisinde diğer verilerle belli bir anlamlılığı bulunmayan,

belli bir deęişkenlik içermeyen verileri edindięi varsayım kısıtı ile tespit ederek projenin isterleri doęrultusunda ortadan kaldırılması konusunda destek olmaktadır.

Kalan örnekler içerisinde birimleri gözlemleyerek normal olduęu varsayımı ile bu teyit gerçekleştirilmiş olmaktadır. Bu tespit sayesinde aksiyon alınmak için analitikler öne sürülecek olan projede aslında ister nezdinde hiçbir anlamı bulunmayan, ihtiyaçları karşılamayan özelliklerin ortadan kaldırılmasını sağlamaktadır.

3.6.2.3 Pekiştirmeli Öğrenme

Belli bir yapı içerisinde kurulan algoritma akış şemasına sahiptir. Akış çevresel faktörlere, veriye, model kısıtlarına, gözlemlere ve deneyime baęlı olarak saęlanan girdiler ile şekillenmektedir. Akış içerisindeki en verimli noktaları ve en verimsiz noktaları belirlerken bunları her bir kısım için örneklere göre cezalandıran ve ödüllendiren algoritmalara pekiştirmeli öğrenme algoritmaları denmektedir. Süreç içerisindeki iş akışı esasında hareket temelden başlamaktadır ve hareket edilen noktada sorgulama tamamlanırsa an için ödül ve ceza sonucu ortaya çıkmakta ve bu yapı sayesinde hedeflenen optimum sonuca ulaşmak için fayda katsayısı düzenlenmektedir. Bu süreçte akış devam ettięi, hedeflenen sonuca veya başarıya potansiyel tüm aşamalar denenerek ulaşılmak istenmektedir ve yenilenmektedir. (Gençalp, 2020: 18)

3.7 Yapay Zeka

Teknoloji kendi içerisindeki dinamiklerle beraber oldukça deęişen hatta tek bir temel üzerine kurulu olduęunu varsayarsak sadece deęişim ile var olan bir terimdir. Süreç içerisinde günlük hayatta karşılaşılan bir çok problem için çözüm üretilmiştir. Bu durum ise teknolojinin birçok araç ve sistem ile meydana çıkmasına neden olmuştur. Bu cihazlar ve sistemler ise genellikle belli birçok insan beyninin karşılaştığı, var olan bir işleyişi daha verimli ve hızlı hale getirmek ile veya daha az maliyetli hale getirmekle ilgili olan problem ile uğraşmıştır. Tüm bu uğraşlar ilgili olduęu var olan

sistemlerin ve kurumlar dijital dönüşümü veya endüstri köklü değişimleri gibi yadsınamaz bir iyileştirmeye mahal vermiştir.

Tüm bu iyileştirmeler toplumlardan insanlara, kurumlardan çalışanlara ve hayata tüm tarafları etkileyecek şekilde gerçekleşmiştir. (Makridakis, 2017: 59) 1950 yılları civarında literatürel bazda kaynakları bulunan fakat işlevsel olarak çok daha yakın bir dönemde revaçta olan ve günlük hayata giren yapay zeka teorisi teknoloji yeniliklerinin tarihine baktığımızda alternatiflerine göre oldukça yenilikçi ve çağ atlatabilecek düzeyde ve altyapıda bir yeniliktir. Gerçi içerisindeki potansiyeli düşündüğümüzde aslında hızı konusunda hala yavaş olduğuna dair soru işaretleri olan iş modelleri içerisinde bakıldığında gerek yapay zekaya uygun olan algoritmalar ile uyum sağlama gerekse yapay zeka bazında güçlendirilmiş yapılarla uyum konularında halan gereken aksiyon almamış ve gerekli yatırımları yapmamış firmalar görünmektedir.

Bu da potansiyel gelişim hızını olumsuz etkilemektedir. Yapay zeka iş modellerine baktığımızda sağlık, bankacılık, pazarlama ve ekonomi ve çok daha çeşitlendirilebilecek alternatifleriyle uyumlu olduğu görülmektedir. Bundan dolayı kurumlarda ve araştırma yapan tüm potansiyel karar alıcıların yapay zekanın ne olduğunu, nasıl ortaya çıktığını ne gibi isterlere cevap verdiğini anlamak oldukça hayati önem taşımaktadır. 1950 yıllarının başları ve öncelerinde elektronik araçlar ve zekalarına dair birçok makale ortaya çıkmıştır fakat bu makaleler sayısal zekanın varlığına dair net bir karar ortaya koymamışlardır. Fakat ilk defa etkili bir sonuç ortaya çıkaran makale “Bilgi İşlem Makineleri ve Zeka” ismiyle Alan Turing tarafından ortaya çıkarılan makaledir. Makineler düşünebilir mi? Sorusuyla 1950 yılında elektronik bir makinenin zekice davranabilmesi için programlanabileceği sorusunu araştıran Turing oldukça etki yaratmıştır. Bu makaleye ek olarak Alan Turing bir makinenin zeki olup olmadığını belirleyebilecek olan Turing Test’i modelini geliştirdi. Turing testi bir makinenin aynı zaman insan ile potansiyel olarak sorgulayıcı bir kişi tarafından test edilip düşünüp düşünemediğinin araştırılmasına dayanmaktadır. Mind adı verilen bir dergide ilk olarak yayınlanmıştır. Matematikçi ve Bilgisayar Bilimcisi Alan Turing tarafından ortaya konulan Turing testi ve Bilgi İşlem Makineleri ve Zeka makalesi kısa bir süreç sonrasında 1956 yılında Dartmouth konferansında Yapay Zeka

bir alanın başlığı olarak dile getirilmiştir. (Turing, 1950: 460) Tüm bu gelişmelerle beraber sonrası yapılan bilgisayar bilimleri üzerine olan konferanslarda bu alan yer almaya başlamıştır. Yıllara göre yapay zeka alanında ortaya konuş bazı gelişmeleri örnek olarak gösterecek olursak.

- 1990 Sinir Ağı cihazı el yazısı olan rakamları okumuştur.
- 1993 Robot Polly görüş teknolojisini kullanarak geziniyor.
- 1997 Deep Blue dünya satranç şampiyonunu yeniyor.
- 1998 Robot Furby konuşmayı öğreniyor.
- 2002 Roomba adı verilen otonom ev süpürgesi piyasaya sürüldü.
- 2005 Robot ASIMO restaurant müşterilerine servis yapıyor.
- 2008 Google iPhone telefonlarında ilk konuşma algılayan teknoloji kullanılmıştır.
- 2009 Google'ın ilk kendi kendini yöneten arabası.
- 2010 Belli bir düzende dans eden Nao robotları fuarda gösteri sergiledi.
- 2011 Watson bilgisayarı Jeopardy yendi.
- 2014 Eugene Goostman 64 yıl sonra Turing testini geçti ve Google Skype firmaları bu konuda aksiyon aldı.
- 2016 AlphaGo Go şampiyonunu yendi.

Yıllar içerisinde yapay zekanın kapasitesini de gösterebilecek olan sıralama içerisinde ses üzerine analizler yapan algoritmadan, satranç oyunu için kurulmuş model ile dünya şampiyonunu yenen algoritmaya kadar, insana özgü olan ve duygusal baz yaklaşımlar içerebileceğini düşünebileceğimiz dans etme eyleminden tutun, görüş ile etraf gezintisine kadar ve el yazısı ile fiziki olarak var olan rakamları okuyarak dijital bir halde dönüştüren teknolojiye kadar süreç içerisinde birçok gelişme meydana gelmiştir. Bu durum günlük yaşamı etkileyebilecek olan potansiyelini oldukça yüksektir. (Doğan, 2018: 6)

Süreç içerisinde birçok yapay zeka tanımı gündeme gelmiştir. Oxford görsel algı, konuşma tanıma, karar verme ve diller arasında çeviri gibi normalde insan gerektiren görevlerin bilgisayar bilimi ve gelişmesidir şeklinde yorumlamıştır.

1956 yılında gerçekleştirilen bu Dartmouth Konferansında, ki az önce bahsettiğimiz McCarthy tarafından yapay zeka konseptinin ilk kez kullanıldığı (Smith, 2006: 4) tarafından yürütülen ve konsept oluşturulan bu konferans makinelerdeki zeka davranışını bazında temel konferans olarak varsayılmaktadır. (Boden, 1996: 142). Konsepti daha iyi anlamak ve farklı tanımları değerlendirme ve kollektif bir gelişim sağlamak araştırma ve geliştirme gereken konseptlerde oldukça önemlidir. Bundan dolayıdır ki yeni nesil kavramlar üzerine toplulukların çalışması ve bu çalışmalarını açık kaynak denilen bir paradigma ile diğer kişilerle paylaşmaları ve bu paylaşımın toplulukların bilgeliği sayesinde çok daha iyi bir yere gelmesi tesadüfi değildir.

Yapay zeka üzerinde yapılan bir diğer tanım Barr ve Feigenbaum tarafında ortaya çıkarılmıştır. Barr ve Feigenbaum'a Yapay Zeka akıllı bilgisayar sistemleri ortaya koyan bilgisayar biliminin bir alt dalı olarak düşünülebilir. Bu yüzden bahsetmişlerdir. Ek olarak bu sistemler insan zekasına benzeyen bazı karakteristiklere sahip olduğunu ve ileri düzey bilgisayar bilimi terminoloji için kullanılan birçok disiplini kanatları altına aldığı için şemsiye olarak simgeleştirilen bu terimden bahsederler. (Jacob, Moore, Whinston, 1988: 24). Yapay zeka için olan bir diğer tanım ile insanlar tarafından gerçekleştirilen bilgi temelli işlevleri bilgisayarlar tarafından otomatik hale getirilmesi şeklinde yorumlanmıştır.

Bilgisayar biliminin bir alt dalı olarak düşünülen ve tanımlanan yapay zeka çalışmaları paralel olarak iki koldan ilerlemektedir. Bunlardan biri daha çok sektördeki problemlere cevap veren ve daha proje süresine bağlı olarak sistematik bir hal alması değişkenlik gösteren uygulama alanıdır. Bu alan daha çok ortaya çıkan problemi çözmeye yöneliktir ve hızlı aksiyon alınarak ilerletilmektedir. Bir diğeri ise biraz daha araştırmaya ve mümkün olduğu kadar modellere dökülerek açıklamaya yönelik olan teorik kısımdır. Bu kısımda yenilikler ortaya konulmak istenmektedir ve bundan dolayı problem çözüm süreci içerisinde belki de yeni problemler ile daha geniş bir araştırma süreciyle çözümler sunulmaktadır. Bu iki alan paralelde ilerletilirken aslında oldukça önemli olan bir kesişimleri de vardır. Örnek vermek gerekirse bilgisayarlardaki görü üzerine ortaya çıkarılan ve dokümanite edilmiş bir makale çok kısa süre içerisinde sektördeki problemlere çözüm olarak kullanılmaya başlayabilir. Aynı şekilde eşlenik olarak işin doğasında barındırdığı şekilde sektördeki bir

problem daha akademik bazda çözüme kavuşturulabilecek bir yapıda olduğu için akademik manada hemen araştırılmaya başlanabilir. Yapay zeka bilgisayar biliminin ve yazılımın sektördeki ihtiyaçlara verdiği cevap hızı ve ortaya çıkardığı karlılık sayesinde çok hızlı bir gelişim ile ilerlemektedir. Bu gelişim sürecinde kendi için 4 tip ayrıma gidilmiştir. Bunlar Reaktif Makineler, Limitli Hafıza, Zihin Teorisi ve Öz Farkındalıktır.

3.7.1 Reaktif Makineler

Oldukça basit bir konsept üzerine kurulu olan bu tür herhangi bir şekilde yeni hafıza oluşumuna veya geçmişe yönelik deneyim analizine ihtiyaç duymamaktadır. Bu tür doğrudan gördükleri ve karşılaştığı durumlara cevap vermek üzerine bir işletim gerçekleştirmektedir. En bilindik örneklerinden biri IBM tarafından geliştirilen satranç oynayabilen “Derin Mavi” makine teknolojisidir.

3.7.2 Limitli Hafıza

Bu teknoloji önceden tanıtılmış bazı cisimleri tanıyan ve bunlar üzerinden görev olarak verilmiş bazı algoritmaları ilgili değişkenleri girdi olarak kullanarak görevinin tamamlayan sohbet robotları ve otonom araçlarda oldukça fazla görmekteyiz.

3.7.3 Zihin Teorisi

Bu teknolojiyi daha çok gelecek üzerine kurgulanmış filmlerde görmekteyiz. Yapay zeka insan zihnini ve davranışlarını anlamaya ve buna göre kendi aksiyonlarına, davranışlarına yön vermeye odaklanmıştır.

3.7.4 Öz Farkındalık

Öz farkındalık daha çok bilinç varlığı üzerine gelişmektedir. Kendi duygularını tanımlayabilen ve diğerlerini tahmin edebilen yapılar için geçerlidir. (Saygılı, 2020 :14)

3.8 Araştırmada Kullanılan Veriler ve Araçlar

Sosyal medya üzerinde son kullanıcının ihtiyacını karşılamak amacıyla münferit alanlarda içerikler üretilmektedir. Bu şekilde son kullanıcının algısı odaklı bir şekilde bu araçlar üzerinde olmaktadır. Gerek bireyler gerek kurumlar tarafından kullanılan, bireylerin fikir paylaşımlarında bulunarak iletişim sağladığı diğer bir deyişle içerik ürettiği, kurumların hem pazarlama iletişimlerin de bulunduğu hem de esas çekirdek ürünü olan içerikler üretebildiği mecralar artık sosyal medya araçlarıdır.

Küresel olarak en çok kullanılan sosyal medya araçlarından biri olan Twitter uygulama aşamasında kullanıldı. Twitter geliştirici olarak nitelendirdiği hesap üzerinden araştırmacılara veri sağlayabilmektedir. Son kullanıcıların ürettiği bu içerikler üzerinden fikirler desteklenmektedir. Sağlanan veri baz alınarak açık kaynak kodlu olarak son dönemde en çok tercih edilen yazılım dilleri sayesinde verilerin analizi gerçekleştirilerek ihtiyaç bazında durum tespitleri, araştırmacı deneyimi doğrultusunda yorumlamalar gerçekleştirilmektedir.

Bu dillere örnek olarak Python, R, Julia verilebilir. Araştırmanın uygulama kısmında yazılım dili olarak Python kullanıldı. Araştırmada çeşitlilik olması ve farklı bölgelerdeki durum tespitini sağlamak amacıyla Türkiye, Çin, Amerika Birleşik Devletleri ve İngiltere ülkeleri baz alındı. Bu ülkelerdeki insanlar üzerinden Twitter sosyal medya mecrası ile yapılan paylaşımların veya diğer bir deyişle üretilen metin içeriklerinin içerisinde frekans baz alınmış ve görselleştirilmiştir.

Böylece üretilen içerikler, manidar olarak bu süreç içerisinde frekans yoğunluğu bazında ülkelerin durumlara yaklaşım tarzını, aldıkları aksiyonları, tepkileri, kültürlerini yansıtmaktadır. İlk Covid-19 kaynaklı ölümün gerçekleştiği Türkiye’de üzerine konuşulan yegane başlık Covid-19 iken takip eden ve kişiler nezdinde enformasyon bakımından daha olgunlaşan Nisan ayı araştırmanın uygulanmasında baz alınan tarih olmuştur. Bu tarihten itibaren yukarıdaki çeşitli lokasyonlarda üretilen içerikler çalışmanın ana verisi için kullanılan filtrelerden birkaçıdır.

Ana veriyi oluşturan diğer filtre ise sağlık kelimesidir. Her sosyal medya aracının içerisinde olduğu gibi, uygulamada kullanılan aracın ekosisteminde bazı fonksiyonlar bulunmaktadır. Sağlık konusunda atılan veriler bizim ana veri popülasyonunu oluşturacağı için #sağlık veya #health gibi etiketleri kullanarak son kullanıcının bu

bazda ürettiği içeriklere ulaşıp ana veri oluşturuldu. Özetle hedeflediğimiz lokasyonlardaki içeriklere, veriye ulaşmak amacıyla kullandığımız sosyal medya aracının özelliği sayesinde sağlık filtreleriyle durum tespiti yapmak istediğimiz tarih başlangıç olacak şekilde çekim sağlanmıştır.

Veri olarak her bir ülke için içerik üreticinin kurum veya birey üretim tarihi diğer bir deyişle başlangıç tarihi nisan olacak şekilde ülke bazında paylaşım ana verimizi oluşturmaktadır. Belli bir tarih aralığı için durum tespiti yapılmak istenmesidir. Böylece tarih aralığından sapmadan diğer lokasyon, üretilen içeriğe temel oluşturan kelime gibi filtreler ile oldukça yüksek bir popülasyonda veri seti oluşturuldu.

Daha sonrasında veri üzerindeki gürültüğü azaltmak için daha anlamlı bir sonucun çıkması ve uygulama kısmında verimli olması için genelde veri bilimi ve veri madenciliği projelerinde kullanılan veri temizliği gerçekleştirildi. Durum tespiti için kullanılacak olan ham veriler gürültüden arındırılıp üzerinde fonksiyon uygulanabilecek hale getirilip sonrasında tabular yapıya büründürüldü. Hedef doğrultusunda frekansla korelasyonu olacak şekilde görselleştirme araçlarıyla veri yorumlanabilir hale geldi. Veri gürültüsü ve veri temizleme aşamaları veri madenciliği, veri bilimi projelerini yüzlük değer verebileceğimiz bir süreç olarak nitelersek oldukça büyük bir kısmını ortalama olarak 60 – 70 diyebileceğimiz bir parçasını almaktadır. Verinin üzerine model geliştirilebilir hale gelmesi için bu yapıyı kullanmak oldukça elzemdir. Bu işlem sonrasında doğruluk oranlarında, durum tespitlerinde ise anlamlı sonuçlar üreten yapıya ulaşmak çok daha mümkündür.

DÖRDÜNCÜ BÖLÜM

UYGULAMA

4.1. Veriseti Frekans Analizi, Görselleştirme ve Kalitatif Bazda Yorumlanması

4.1.1 Çin

Şekil 4.1: Çin Görselleştirme



Ülke bazlı durum tespiti analizlerimize öncelikle salgının mazhar olduğu ülke olan Çin ile başlayalım. Çin’de başlangıç olarak nisan ayını baz aldığımız bu tarih aralığında sağlık konusunda en çok kullanılan kelimeler disease, human, health, buddha ve artan vaka sayılarını temsilen her gün yönünde yatay bir zaman çizelgesi gibi tarihi vurgulayan every kelimesi kullanılmıştır. Salgın ve insan üzerine yoğunluk gösteren kelimelere ek olarak buddha kelimesinin kullanılması, kelime anlamı bilinçlenmiş ve uyanmış olarak kullanılabilecek fakat daha çok dini bir boyutu

simgeleyen yönüyle Çin'deki sosyal medya aracındaki son kullanıcıların bu tarz potansiyel belirsizlik durumlarındaki algısını ortaya koymaktadır.

4.1.2 İngiltere

Şekil 4.2: İngiltere Görselleştirmesi



Durum tespiti yaptığımız diğer bir ülke İngiltere ilk görülen vaka ile birlikte sıkı önlemler alan İngiltere’de sosyal medyadaki son kullanıcılar sağlık bazında ürettikleri içeriklerde wellbeing, health systems gibi doğrudan salgın ile alakalı olan kelimeler kullanılmıştır. Kişiler içerisinde buldukları olumsuzlukları deneyimlerken psikolojik olarak kendilerini rahatlatmak için belli bir frekansa sahip meditasyon müziklerini angel frequency ve solfeggio angel kelimeleri ile aramışlardır. Bilindiği gibi hızla artan vakalar ve potansiyel mutasyonlar nedeniyle birçok ülke sağlık sistemlerinin yeterliliği konusundaki çeşitli risklerle baş başa kaldılar. Durumun bu şekilde ilgili ülke vatandaşlarına yansımaları ise sosyal medya üzerindeki içerik üretim hacmine, erişimdeki kolaylığın ve yayılan kitle tabakasının genişliği gibi bazı dönemin isteklerinin neden olduğu şekilde, kişilerin bunu paylaşması enformasyona neden olmasını sağlamaktadır. Mevcut kelimelerini ilgili dönemde en çok kullanılan kelimeler arasında bulunması ise durum üzerine yapılan potansiyel tartışmaların çeşitli

sosyoloji terimleriyle harmanlanmasına ve sađlık konularıyla kesişmesine neden olmuştur.

4.1.3 Türkiye

Şekil 4.3: Türkiye Görselleştirmesi



Türkiye ise mart ayıyla beraber oldukça kritik bir döneme girdiğimiz durumda son kullanıcılar sosyal medya platformları üzerinde ürettiği içeriklerde en çok Dokuz Eylül Üniversitesi, Araştırma, Uygulama, Kontrollü, Düzenleme, Kapsamında ve Sosyal gibi sosyoloji, bilim ve sađlık alanlarında doğrudan ilişkili kelimeler kullanmışlardır. Ülkede aratılan kelimeler bazında değerlendirmeler yapıldığında etkisini dünyada göstermeye başlayan ilgili virüse karşı önlemlerin alınmaya başlandığı öngörülmektedir.

4.1.4 Amerika Birleşik Devletleri

ABD lokasyonunda ise New Normal, Early Pack, bölümlendirmeyi temsilen Part kelimeleri sosyal medyadaki son kullanıcılar tarafından en çok sıklıkla kullanılan kelimeler olarak görülmektedir.

Şekil 4.4: Amerika Birleşik Devletleri Görselleştirmesi

Lovin Early pack
new normal
part
vicegandako Early

Örnekteki diğer ülkelere göre başlarda yönetimde algısı nedeniyle daha yavaş aksiyon alınan ülkede halk tarafından oluşturulan içerikler, onların düşüncelerine baktığımızda aslında gayet farkındalık söz konusu görünüyor. Alınan aksiyonlar ve aşı üretimi ile birlikte, ki aşı konusunda en fazla satış ortaklığı Almanya ile birlikte Amerika Birleşik Devletleri yaptı, aşılama oranlarındaki yoğunluğu en fazla olan ülkelerden biri diyebiliriz.

4.2 Duygu Analizi ve Ülke Karşılaştırması

Duygu analizinde Python dili kullanılmıştır. NLTK kütüphanesi içerisinde kelime bazında hassas analiz içeren VADER modeli duygu analizinin bazını oluşturmaktadır. Türkiye, Çin, İngiltere, ABD gibi farklı ülkeler bazında Covid 19 süreci içerisinde yapılan paylaşımları baz alan veri seti üzerinden geliştirilen model ile yapılan duygu analizi sonucunda farklı bakış açıları ve tepkiler ortaya çıkmıştır. Bu tarz analizlerde ekseriyetle ülkelerinde ortak noktası olarak nötr tepki oranının yoğunluğunu pivot edilerek yorumlanabilmektedir. Çin devletinin lokasyonunda yapılan paylaşımlarda durum 0.116 pozitif, 0.059 negatif, 0.825 nötr olarak rakamlarla ifade edilmektedir. İngiltere devletinin lokasyonunda yapılan paylaşımlar üzerinden elde edilen verilerin duygu durumları ise 0.157 pozitif, 0.05 negatif, 0.157 nötr olmak olarak ifade edilmektedir. Türkiye devletinin lokasyonunda yapılan paylaşımlar üzerinden elde edilen verilerin duygu durumları ise 0.004 pozitif, 0.003 negatif, 0.993 nötr olarak

ortaya çıkmıştır. ABD lokasyonunda ise bu rakamlar 0.128 pozitif, 0.062 negatif, 0.81 nötr olmak üzere nihai analiz sonucunu göstermektedir.

Çin

Out[10]: {'neg': 0.059, 'neu': 0.825, 'pos': 0.116}

İngiltere

Out[12]: {'neg': 0.05, 'neu': 0.792, 'pos': 0.157}

Türkiye

Out[14]: {'neg': 0.003, 'neu': 0.993, 'pos': 0.004}

ABD

Out[16]: {'neg': 0.062, 'neu': 0.81, 'pos': 0.128}

Çıktılar ülkelerin durumlara farkı duygular yoğunluklarıyla tepki verdiklerini göstermektedir. Çin pozitif duygu yoğunluğu İngiltere'den düşükken, Türkiye ve ABD analizi sonucundaki pozitif duygu yoğunluğu rakamlarından yüksektir. Bu durum Çin devletinin potansiyel risk durumlarına karşı verdiği pozitif duygu yoğunluğunun görece ortalama üzerinde olduğunu göstermektedir. Negatif duygu yoğunluğunu karşılaştıracak olursak bu sefer Çin devleti pozitif duygu yoğunluğuna göre daha farklı bir kombinasyon göstermektedir. Bu sefer ABD negatif duygu yoğunluğu oranına göre daha düşük olduğunu, İngiltere ve Türkiye ülkelerine göre daha yüksek olduğunu görmekteyiz. Nötr duygu yoğunluğu ise Türkiye'den daha düşük, İngiltere ve ABD nötr duygu durumu oranlarından daha yüksektir. İngiltere pozitif duygu yoğunluğu Çin, Türkiye ve ABD ülkelerinin pozitif duygu oranlarına göre daha yüksek olduğu saptanmıştır. Negatif duygu durumu oranına baktığımızda Türkiye'nin oranına göre daha yüksek olduğunu görürken aynı zamanda ABD ve Çin ülkelerinin de negatif duygu oranlarından daha düşük olduğu saptanmıştır. İngiltere nötr duygu durumu analizinde ise Çin, Türkiye ve ABD ülkelerine göre daha düşük oranda bir skora sahip olduğu tespit edilmiştir. Türkiye pozitif duygu analizinde İngiltere, Çin ve ABD ülkelerinin pozitif duygu analizi oranlarına göre daha az bir oran olarak analiz sonucunda saptanmıştır. Negatif duygu analizinde ise pozitif duygu analizine benzer

şekilde İngiltere, Çin ve ABD ülkelerinin negatif duygu analizine göre daha düşük bir oran ile karşımıza gelmektedir. Pozitif ve negatif oranlarının yakınlığından da anlayacağımız şekilde Türkiye nötr duygu analizi ile ön plana çıkmış ve İngiltere, Çin ve ABD ülkelerinin nötr duygu analizi oranlarında göre yüksek bir oran olarak tespit edilmiştir. ABD pozitif duygu analizi oranı olarak İngiltere'den düşük Çin ve Türkiye ülkelerinin oranlarına göre daha yüksek bir pozitif duygu oranı ile ortaya çıkmıştır. negatif duygu oranı analizinde ise Türkiye, Çin ve İngiltere'ye nazaran daha yüksek bir negatif duygu oranına sahip olduğu tespit edilmiştir. Böylece potansiyel risk durumlarına karşı ABD'nin daha yüksek negatif duygu durumun olduğu yorumlanabilir. Potansiyel risk durumlarına karşı pozitif duygu durumu oranı daha yüksek olan ise İngiltere ülkesidir. Potansiyel risk durumlarına karşı nötr duygu durumu oranı daha yüksek olan ise Türkiye'dir. ABD nötr duygu durumu analizi ise İngiltere'den yüksek iken, Türkiye ve Çin'den geridedir.

SONUÇ

Sağlık sektörü konusunda münferit haber kaynakları tarafında risk içerisinde olan ülkeler olarak duyduklarımızı sosyal medya üzerindeki ilgili ülkelerdeki kişiler tarafından üretilen, yapılan fikir paylaşımlarında aslında sadece İngiltere konu üzerine eğilmiş görünmektedir. Bu durum ile beraber münferit haber kaynaklarının son kullanıcı tarafından yaşanan deneyimle bazen kesişmediğini son zamanlarda sosyal medya araçlarının belki de temel haber kaynağı olarak kullanılabilecek tek yer olmasının tesadüfi olmadığı savını da ortaya koyabiliriz. İlgili paylaşımlara olumlu olumsuz doğrudan kritiklerin gelmesi bu durumu daha verimli bir hale de sokmaktadır.

Süreç içerisinde deneyimlediğimiz gibi toplulukların bilgi paylaşımı dönemin sorunlarına çözüm bulunmasını, fikirleri daha kaliteli bir hale getirilmesini, potansiyel bilgi kirliliğini ve bilgi gürültüsünü engelleyerek küresel manada bir iletişim sağlamaktadır. Enformasyon oluşturan bu akışlar ise ülkelerin durumlara karşı gösterdiği tutumları, kültürlerini, yaşadıkları olumlu olumsuz durumları karşılaştırmalı olarak ortaya koyabilmektedir.

Gelişmiş ülkelerdeki yaşanan sağlık sorunlarının, sosyal medya üzerindeki paylaşımlardan sadece Avrupa'daki kısımda bir temel bulduğu, diğer tarafta ilgili dönemdeki algının aslında yoğun bir algı ile mazhar olmadığını görebilmekteyiz. Çin'de konuya yaklaşım çok daha felsefi ve dini olarak görünürken Amerika Birleşik Devletleri üzerinde daha ekonomik ve hızlı geçiş aşamaları üzerinde durulmaktadır. İngiltere daha çok sistem sorun çözümü, refah gibi kelimeler ile sosyolojik bazda konu üzerinde durmaktadır. Türkiye'de ise yaşanan deneyimler sonucunda kurumlar ve alınan tedbirler üzerine paylaşılan düşünceler temel oluşturmaktadır. Bu da bizlere sağlığın olumlu/olumsuz durumlardaki insan hayatında olan önemli değişikliklerin günlük hayatta kullanılan kelimelere kadar olan etkisini göstermektedir.

Salgının yaşandığı günlerde farklı ülkelerin paylaşımlarını barındıran veri seti üzerine yapılan duygu analizine baktığımızda, Türkiye, Çin, ABD, İngiltere ülkeleri arasında pozitif duygu yoğunluğu oranı en yüksek İngiltere olarak görünmektedir. Negatif duygu yoğunluğu oranı bakımından ABD oransal olarak ön plana çıkmaktadır. Nötr duygu yoğunluğu oranına baktığımızda Türkiye diğer ülkelere nazaran daha yüksek bir orana sahip olduğu araştırmada tespit edilmiştir. Çin oransal bazda diğer ülkelere nazaran bir çıkış ortaya koymamıştır fakat kendi içerisinde değerlendirdiğimizde Çin nötr duygu yoğunluğu, pozitif duygu yoğunluğu ve negatif duygu yoğunluğu şeklinde sıralayabileceğimiz oransal değerlere sahiptir.

Bu durumda negatif duygu yoğunluğuna sahip olan Amerika Birleşik Devletleri, yeni süreç için hızlı geçiş planları ve ekonomik paketler ile soruna karşılık vermiş olarak görünmektedir. Pozitif duygu yoğunluğuna sahip İngiltere soruna çözüm, refah ve sanatsal bir bakış açısıyla karşılık verirken alternatiflere nazaran nötr duygu yoğunluğuna sahip Türkiye deneyimlediği süreçler ve kurumlar üzerinden içerikler aktararak soruna karşılık vermiştir.

Çin duygu yoğunluğu analizi bakımından diğer ülkelere nazaran bir oran ile ön plana çıkmamaktadır fakat kendi içerisinde duygu yoğunluğu sıralaması nötr başta olacak şekilde pozitif ve negatif olarak sıralanabilir. Kendi içerisindeki bu analiz sonucu nötr duygu yoğunluğuna sahip Çin soruna karşı felsefi ve dini bir bakış açısıyla karşılık vermiş olduğu görünmektedir.

Ülkelerin bireysel farklı analiz çeşitlerine ek olarak kendi aralarında farklı analiz teknikleri kullanılarak geliştirilen model sonuçlarına, kartezyen çarpıma benzetebileceğimiz, çapraz yorum değerlendirmesi yaptığımızda İngiltere ve Çin refah, stabil kelimeleri ile etiketlenebilecek bir tutum sergilerken, Amerika Birleşik Devletleri'nde panik ve plan kelimeleri ile etiketlenebilecek bir tutum analiz edilmiştir. Türkiye ise stabil, hızlı aksiyon ve kritik kelimeleri ile etiketlenebilecek düzeyde tutum sergilemiş olarak görünmektedir.

KAYNAKÇA

Talih Akkaya, D. , “Sosyal Medya Reklamcılığında Tüketici Algılarının Tutum,
Akyol, A. & Davranış Ve Satın Alma Niyeti Üzerine Etkisi”, Marmara
Gölbaşı Şimşek, Üniversitesi İktisadi ve İdari Bilimler Dergisi, 2017
G.

Abdallah Alalwan, “Investigating The Impact Of Social Media Advertising
A. Features on Customer Purchase Intention”, International
Journal of Information Management, 2018

Jones, C., Ryan, S. “Understanding Digital Marketing”, Kogan Pages Inc, 2012

Hazar, M. “Sosyal Medya Bağımlılığı – Bir Alan Çalışması”, İletişim
Kuram ve Araştırma Dergisi, 2011

Mayfield, A. “What is Social Media”, London, 2008

Gül Yüzüak, A. “Investigation of User Attitudes and Perceptions Toward
Brands On Social Media in Relation to Purchase Intention
Case Study: Analysis Of Koton’s Social Media Strategies”,
İstanbul Bilgi Üniversitesi, 2018

- Eren Erdogmus, I., "Social media marketing: Exploring the user typology in Turkey", International Journal of Technology Marketing, 2013
- Çiçek, M.
- Karlıklı, U. "Chatbot Nedir?", 2017, (Çevrimiçi)
<https://lechatbot.com/chatbot-nedir-53f2638db5d7>
- Karlıklı, U. "Bir Chatbotun İyi Çalıştığını Nasıl Anlarız?", 2019, (Çevrimiçi)
<https://lechatbot.com/bir-chat-botun-iyi-calistigini-nasil-anlariz-161f35c89a7a>
- Hearst, M. "What is Text Mining?", 2003, (Çevrimiçi)
<http://people.ischool.berkeley.edu/~hearst/text-mining.html>
- Waegel, D. "The Development of Text-Mining Tools and Algorithms", Ursinus College, 2006
- Marr, B. "What's The Difference Between Structured, Semi-Structured And Unstructured Data?", 2019, (Çevrimiçi)
<https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data>

- Kumar, L. “Text Mining: Concepts, Process And Applications”, Information Extraction, 2013
- Witten, I., Bray, Z., Mahoul, M. “Text Mining: A New Frontier For Lossless Compression”, University of Waikato, 2000
- Scholtes, J. “Text-Mining: The Next Step in Search Technology”, DESI-III Workshop Barcelona ,2009
- Mohri, M., Rostamizadeh, A., Talwalkar, A. “Foundations of Machine Learning, Second Edition (Adaptive Computation And Machine Learning Series)”, MIT Press ,2018
- Gençalp, E. “Atm Demand Prediction with Machine Learning Methods / Makine Öğrenmesi Yöntemleri İle Atmlerde Talep Tahmini”, Galatasaray Üniversitesi ,2020
- Maglogiannis, I. “Emerging Artificial Intelligence Applications in Computer Engineering”,IOS Press ,2007
- Mshir, S., Kaya, M. “Signature Recognition Using Machine Learning”, IEEE ,2020

Makridakis, S. “The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms”, Futures, 2017

Haton, J. P. “A Brief Introduction To Artificial Intelligence”, IFAC Proceedings Volumes, 2006

Turing, A. “Computing Machinery and Intelligence”, Mind, 1950

Boden M. “Mind as Machine - A History of Cognitive Science”, Oxford University Press, 2006

MacGuire B., Huang, T., Yang, G., Smith, C. “The History of Artificial Intelligence”, University of Washington, 2006

Saygılı, H. “Public Policies on Artificial Intelligence: Comparison of the USA and China’s AI Public Policies”, Ankara Yıldırım Beyazıt Üniversitesi, 2020

Kaplan, A., Haenlein, M. “Users of the World, Unite! The Challenges and Opportunities of Social Media”, Business Horizons, 2010

Smith, D. “The History of R”, 2020, (Çevrimiçi)
<https://blog.revolutionanalytics.com>

Deligianni, F., “Computing Research Repository. Artificial Intelligence and
Ravi, D., & Yang, Robotics”, Computing Research Repository, 2018
G.

Doğan, İ. “The Application Of Artificial Intelligence In New Age Of
Marketing: An Analysis On AI Mobile Banking Apps”,
İstanbul Bilgi Üniversitesi, 2018

Jacob, V., Moore, “Artificial Intelligence and the Management Science
Practitioner: Rational Choice and Artificial Intelligence”,
J., Whinston, A. Interfaces, 1988

Akın, A., “Toplumsal Cinsiyet Ayrımcılığı ve Sağlık”, Toplum
Hekimliği Bülteni, 2007

Sabuncu, N., “Sağlık Bakım Sistemi, Sağlık Ekibi ve Hemşirenin Rollerini,
Babadağ, K., Hemşirelik Esasları”, Anadolu Üniversitesi, 1994
Taşocak, G. ve
Atabek, T.

Neyzi, O. “Erken Yaşam ile İleri Yaşlardaki Sağlık Arasındaki İlişki”,
Sürekli Tıp Eğitim Dergisi, 2000

Güler, Ç. ve Çobanoğlu, Z., “Kentleşme Ve Çevre Sağlığı”, Çevre Sağlığı Temel
Kaynaklar Dizisi No: 26, 1994

Vogel, L. “İş Sağlığı: Topluluk Politikası İçin Sekiz Öncelikli Eylem
Alanı”, 2004, (Çevrimiçi)
hesa.etui-rehs.org/uk/publications/files/Turc.pdf.

Yıldırım, İ., “Algılanan Sosyal Destek Ölçeğinin Geliştirilmesi
Güvenirliliği ve Geçerliliği”, Hacettepe Üniversitesi Eğitim
Fakültesi Dergisi, 1997

Batlaş, A., Batlaş, Z. “Stres ve Başa Çıkma Yolları”, Remzi Kitabevi, 2004

Dinçer, Ö. ve Fidan, Y., “İşletme Yönetimine Giriş”, Beta, 1999

Ayata, D. “Applying Machine Learning And Natural Language
Processing
Techniques To Twitter Sentiment Classification For Turkish
And English”, İstanbul Teknik Üniversitesi, 2000

Tuğrul, B. “Açık Kaynak İstihbaratı: Twitter Analizi”, Ankara Üniversitesi, 2020

Banavand, R. “Political Marketing And The U.S. Presidential Campaign Strategies: A Functional And Rhetorical Analysis Of Donald Trump’s And Hillary Clinton’s Discourse On Twitter”,Dokuz Eylül Üniversitesi ,2020

Kousar Aslam “Comparative Analysis of Active Learning Strategies in Twitter Domain”, Sabancı Üniversitesi, 2015

Gebreyesus, Y. “Sentiment Analysis in Social Media: A Comparative Study”, Atılım Üniversitesi, 2018

“What is Python?”, <https://pythoninstitute.org/>

“What is R?”, <https://mran.microsoft.com/>

“Julia (programlama dili)”, <https://tr.wikipedia.org/>

“Cholera”, <https://www.who.int/>

“Healthcare in China”, <https://tr.wikipedia.org/>

“Top Computer Languages”, <https://statisticstimes.com/>

- Yüksel, N., Tan, F. “Metin Madenciliği Teknikleri İle Sosyal Ağlarda Bilgi Keşfi”, Mühendislik Bilimleri ve Tasarım Dergisi, 2018
- Güvercin, M., Ferhatosmanoğlu, H., Gedik, B. “Çağrı Merkezi Metin Madenciliği Yaklaşımı”, İhsan Dogramacı Bilkent Üniversitesi, 2017
- Artsın, M. “Bir Metin Madenciliği Uygulaması: Vosviewer”, Eskişehir Teknik Üniversitesi Bilim Ve Teknoloji Dergisi, 2020
- Atan, S. “Metin Madenciliği: İmkanlar, Yöntemler Ve Kısıtlar”, Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 2020
- Evrenseker. S. “Sosyal Ağlarda Akan Veri Madenciliği”, Orta Doğu Amerikan Üniversitesi, 2014
- Zontul, M., Aydın, G. “NoSQL Veri Tabanları Üzerinde Bir Metin Madenciliği Uygulaması”, Altınbaş Üniversitesi Mühendislik Sistemleri Ve Mimarlık Dergisi, 2017

- Küçük, M. “Metin Madenciliği Tabanlı Bildirim Takip Sistemi”, PAMUKKALE ÜNİVERSİTESİ, 2019
- Yıldız, O. “Metin Madenciliğinde Anahtar Kelime Seçimi Bir Üniversite Örneği”, Yönetim Bilişim Sistemleri Dergisi, 2016
- Beşkirli, A.,
Gülbandılar, E.,
Dağ, İ.
Kocatekin, T. “Metin Madenciliği Yöntemleri ile Twitter Verilerinden Bilgi Keşfi”, Eskişehir Osmangazi Üniversitesi, 2021
- “Türkçe Radyoloji Raporlarında Metin Madenciliği”, Bahçeşehir Üniversitesi, 2013
- Kaşıkçı, T.,
Gökçen, H. “Metin Madenciliği ile E-Ticaret Sitelerinin Belirlenmesi”, Bilişim Teknolojileri Dergisi, 2014
- Çalış, K.,
Gazdağı, O.,
Yıldız, O.
Kılınç, D.,
Borandağ, E.,
Yücalar, F.,
Tunalı, V.,
Şimşek, M.,
Özçift, A.
Melek, C. “Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri ile Otomatik Tespiti”, Bilişim Teknolojileri Dergisi, 2013
- “KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi”, Marmara Fen Bilimleri Dergisi, 2016
- “Metin Madenciliği Teknikleri İle Şirketlerin Vizyon İfadelerinin Analizi”, Dokuz Eylül Üniversitesi, 2012
- Yıldız, B.,
Ağdeniz, Ş. “Muhasebede Analiz Yöntemi Olarak Metin Madenciliği”, Muhasebe Bilim Dünyası Dergisi, 2018

Kobayashi, V., “Text Mining in Organizational Research”, Organizational
Mol, S., Berkers, Research Methods, 2018
H., Kismiho’k, G.,
Hartog, D.

Lee, J., Yoon, W., “Biobert: A Pre-Trained Biomedical Language
Kim S., Sunkyu, Representation Model For Biomedical Text Mining”,
K., Kang, J. Bioinformatics, 2019

Bach, M., Krsti’c, “Text Mining for Big Data Analysis in Financial Sector”,
Z., Seljan, S., Sustainability, 2019
Turulja, L.

Kontostathis, A., “Text Mining and CyberCrime”, Amanda Leatherman
Edwards, L. Ursinus College, 2010

Spasic, I., “Text Mining and Ontologies in Biomedicine: Making Sense
Ananiadou, S., Of Raw Text”, Briefings In Bioinformatics, 2005
McNaught, J.,
Kumar, A.

Vijaymeena, M., “A Survey On Similarity Measures In Text Mining”, Nandha
Kavitha, K., Engineering College, 2016

Driel, M., “A Text-Mining Analysis Of The Human Phenome”,
Bruggeman, J., European Journal of Human Genetics, 2006
Vriend, G.,