

T.C.
KAYSERİ ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
SAĞLIK YÖNETİMİ ANABİLİM DALI

HASTA PROFİLLERİNİ ANALİZ ETMEK İÇİN
BİRLİKTELİK KURALLARI MADENCİLİĞİ:
KAYSERİ'DEKİ BİR HASTANE ÖRNEĞİ

Hazırlayan

Salih Murat GÜRBÜZ

Danışman

Doç. Dr. Yılmaz DELİCE

Dr. Öğr. Üyesi Salih HİMMETOĞLU

Yüksek Lisans Tezi

Ekim 2023, KAYSERİ

T.C.
KAYSERİ ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
SAĞLIK YÖNETİMİ ANABİLİM DALI

HASTA PROFİLLERİNİ ANALİZ ETMEK İÇİN
BİRLİKTELİK KURALLARI MADENCİLİĞİ:
KAYSERİ'DEKİ BİR HASTANE ÖRNEĞİ

Hazırlayan

Salih Murat GÜRBÜZ

Danışman

Doç. Dr. Yılmaz DELİCE

Dr. Öğr. Üyesi Salih HİMMETOĞLU

Ekim 2023, KAYSERİ

BİLİMSEL ETİĞE UYGUNLUK

Bu tezin kendi çalışmam olduğunu, planlanmasından yazımına kadar hiçbir aşamasında etik dışı davranışımın olmadığını, tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları kaynaklar listesine aldığımı, tez çalışması ve yazımı sırasında patent ve telif haklarımı ihlal edici bir davranışımın olmadığını beyan ederim.

Salih Murat GÜRBÜZ

İmza:

T.C.
Kayseri Üniversitesi
Lisansüstü Eğitim Enstitüsü Müdürlüğü

Anabilim Dalı : Sağlık Yönetimi
Program Adı : Sağlık Yönetimi
Tez Başlığı : Hasta Profillerini Analiz Etmek için Birliktelik Kuralları Madenciliği:
Kayseri'deki Bir Hastane Örneği

Yukarıda bilgileri verilen tez çalışmasının a) Giriş, b) Ana bölümler ve c) Sonuç kısımlarından oluşan (Kapak, Önsöz, Özet, İçindekiler ve Kaynakça hariç) toplam 79 sayfalık kısmına ilişkin 13/12/2023 tarihinde **Turnitin** intihal programından aşağıda belirtilen filtreleme uygulanarak alınmış olan özgünlük raporuna göre tezin benzerlik oranı: **% 15'** dir.

Uygulanan filtrelemeler:

- 1- Giriş dahil
- 2- Ana Bölümler dahil
- 3- Sonuç dahil
- 4- Alıntılar dahil
- 5- Kapak hariç
- 6- Önsöz ve Teşekkür hariç
- 7- İçindekiler hariç
- 8- Kaynakça hariç
- 9- Özet hariç
- 10- Yedi (7) kelimedenden daha az örtüşme içeren metin kısımları hariç

Kayseri Üniversitesi Lisansüstü Eğitim Enstitüsü Tez Önerisi Hazırlama, Tez Yazma ve Teslim Yönergesini inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini, aksinin tespit edileceği muhtemel durumlarda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini bilgilerinize arz ederim. 14/12/2023

Danışman
Doç. Dr. Yılmaz DELİCE
İmza

Öğrenci
Salih Murat GÜRBÜZ
İmza

YÖNERGEYE UYGUNLUK SAYFASI

Hasta Profillerini Analiz Etmek için Birliktelik Kuralları Madenciliği: Kayseri'deki Bir Hastane Örneği adlı Yüksek Lisans tezi, Kayseri Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi'ne uygun olarak hazırlanmıştır.

Hazırlayan

Salih Murat GÜRBÜZ

İmza

Danışman

Doç Dr. Yılmaz DELİCE

İmza

2. Danışman

Dr. Öğr. Üyesi Salih HİMMETOĞLU

İmza

Sağlık Yönetimi ABD Başkanı

Prof. Dr. Özgür DEMİRTAŞ

İmza

KABUL VE ONAY TUTANAĐI

Doç. Dr. Yılmaz DELİCE danışmanlığında **Salih Murat GÜRBÜZ** tarafından hazırlanan “..Hasta Profillerini Analiz Etmek için BirlikteK Kuralları Madenciligi: Kayseri'deki Bir Hastane Örneđi” adlı bu çalıřma, jürimiz tarafından Kayseri Üniversitesi Lisansüstü Eğitim Enstitüsü **Sađlık Yönetimi** Anabilim Dalında **Yüksek Lisans** tezi olarak kabul edilmiştir.

...29../.12../2023....
(Tez Savunma Sınav Tarihi Yazılacak)

JÜRİ:

Danışman

Üye

Üye

ONAY:

Bu tezin kabulü Enstitü Yönetim Kurulunun/...../..... tarih ve sayılı kararı ile onaylanmıştır.

...../...../.....

Doç. Dr. Ahmet TERZİ
Enstitü Müdürü

ÖNSÖZ / TEŞEKKÜR

Bu tez çalışmamda katkı ve yardımlarından dolayı hocalarım Doç. Dr. Yılmaz DELİCE ve Dr. Öğr. Üyesi Salih HİMMETOĞLU'na, mesai arkadaşım ve değerli katkıları ile her daim yanımda olan Dr. Öğretim Üyesi Ayhan RENKLİER, Dr. Öğretim Üyesi Ömer Faruk SARITAŞ'a ve Sağlık Yönetimi Ana Bilim Dalındaki çok değerli hocalarıma, destekleriyle hep yanımda olan kıymetli eşim, kızlarım ve aileme teşekkür ederim. Ayrıca tezimde kullanılan verilerin sağlanmasında katkıda bulunan değerli Erciyes Üniversitesi Sağlık Uygulama ve Araştırma Merkezi Müdürlüğü yöneticilerine teşekkür ederim.

Salih Murat GÜRBÜZ, Kayseri, 2023

HASTA PROFİLLERİNİ ANALİZ ETMEK İÇİN BİRLİKTELİK KURALLARI MADENCİLİĞİ: KAYSERİ'DEKİ BİR HASTANE ÖRNEĞİ

Salih Murat GÜRBÜZ

Kayseri Üniversitesi Lisansüstü Eğitim Enstitüsü

Yüksek Lisans Tezi, Eylül 2023

Danışman: Doç. Dr. Yılmaz DELİCE, Dr. Öğr. Üyesi Salih HİMMETOĞLU

KISA ÖZET

Bilişim teknolojilerinde yaşanan hızlı gelişmeler günümüz sağlık işletmelerinde de yazılımsal çalışmaların yürütülerek yenilikçi karar destek sistemlerinin geliştirilmesine imkân sağlamaktadır. Bu çalışmaların arasında sağlık işlemelerine gelen hastaların profillerinin belirlenmesine yönelik uygulamalar da yer almaktadır. Hasta profillerinin çıkarılması, karar vericilerin yönetsel açıdan daha doğru kararlar vermelerine yardımcı olacak enstürümanlardan biridir. Bu çalışmada bir sağlık işletmesinden alınan hasta verilerine veri madenciliği yöntemlerinden biri olan birliktelik kuralı yaklaşımı uygulanmıştır. Ele alınan veriler için Apriori Algoritması kullanılarak veriler arasındaki olası gizli örüntüleri tespit edebilen etkin kurallar elde edilmesi amaçlanmıştır. Çalışma kapsamında Erciyes Üniversitesi Tıp Fakültesi Hastaneleri bünyesinde faaliyet gösteren Çocuk Hastanesinin 2021-2022 yıllarına ait hasta kayıt bilgileri analiz edilmiştir. Bu amaçla öncelikle mevcut veri yığını üzerinde veri ön işleme prosedürleri uygulanarak anlamlı örüntülerin keşfedilmesini engelleyebilecek hatalı / eksik verilerin düzeltilmesi sağlanmıştır. Ön işleme aşaması ile gürültüden arındırılmış olan hedef veri kümesi, WEKA veri madenciliği yazılımı üzerinden Apriori Algoritmasına tabi tutularak 10 ayrı örüntü deseni elde edilmiştir. Elde edilen sonuçlar incelendiğinde önerilen yaklaşımın hastane yöneticilerine karar verme süreçlerinde farklı bakış açıları ve yön gösterici fikirler sunabileceği görülmüştür.

Anahtar Kelimeler: Sağlık Yönetimi, Veri Madenciliği, Apriori, Hasta Profili, Birliktelik Kuralı

ASSOCIATION RULES MINING TO ANALYZE PATIENT PROFILES: AN EXAMPLE OF A HOSPITAL IN KAYSERİ

Salih Murat GÜRBÜZ

Kayseri University, Institute of Graduate Education

Program Thesis, September 2023

Supervisor: Doç Dr. Yılmaz DELİCE, Dr. Öğr Üyesi Salih HİMMETOĞLU

ABSTRACT

Rapid developments in information technologies enable the development of innovative decision support systems by carrying out software studies in today's healthcare enterprises. Among these studies, there are also applications to determine the profiles of patients coming to healthcare facilities. Creating patient profiles is one of the instruments that will help decision-makers make more accurate administrative decisions. In this study, the association rule approach, one of the data mining methods, was applied to patient data received from a healthcare institution. It is aimed to obtain effective rules that can detect possible hidden patterns among the data by using the Apriori Algorithm for the data considered. Within the scope of the study, patient registration information of the Children's Hospital operating within Erciyes University Faculty of Medicine Hospitals for the years 2021-2022 was analyzed. For this purpose, first of all, data pre-processing procedures were applied to the existing data stack to correct erroneous/missing data that could prevent the discovery of meaningful patterns. The target data set, which was cleared of noise with the pre-processing stage, was subjected to the Apriori Algorithm via WEKA data mining software and 10 different pattern patterns were obtained. When the results obtained were examined, it was seen that the proposed approach could offer different perspectives and guiding ideas to hospital managers in their decision-making processes.

Keywords: Healthcare Management, Data Mining, Apriori, Patient Profile, Association Rule

İÇİNDEKİLER

HASTA PROFİLLERİNİ ANALİZ ETMEK İÇİN BİRLİKTELİK KURALLARI MADENCİLİĞİ: KAYSERİ'DEKİ BİR HASTANE ÖRNEĞİ

BİLİMSEL ETİĞE UYGUNLUK	i
TEZ ÖZGÜNLÜK SAYFASI.....	ii
YÖNERGEYE UYGUNLUK SAYFASI	iii
KABUL VE ONAY SAYFASI.....	iv
ÖNSÖZ / TEŞEKKÜR.....	v
KISA ÖZET.....	vi
ABSTRACT	vii
İÇİNDEKİLER.....	viii
KISALTMALAR ve SİMGELER.....	xi
TABLolar LİSTESİ	xii
ŞEKİLLER LİSTESİ.....	xiii
1.GİRİŞ.....	1
2.VERİ MADENCİLİĞİ	4
2.1 Veri Madenciliği Tanımı	4
2.2. Veri Madenciliği Tarihçesi	9
2.3. Veri Madenciliği Uygulama Alanları	11
2.4. Sağlık Alanında Veri Madenciliği Uygulamaları	14
2.5 Veri Madenciliğinde Kullanılan Programlar	17
2.5.1 IBM SPSS Statistics.....	18
2.5.2 IBM SPSS Modeler.....	19
2.5.3 Intelligent Miner	20
2.5.4. Analysis Manager Analysis Manager	21
2.5.5 Keel	21

2.5.6 Knime.....	21
2.5.7 Orange.....	22
2.5.8 R.....	22
2.5.9 RapidMiner	22
2.5.10 WEKA.....	23
2.6 Veri Madenciliği Yöntemleri.....	24
2.6.1. Veri Ön İşleme	26
2.6.2. Tahmin Edici Veri Madenciliği Yöntemleri.....	28
2.6.2.1. Sınıflandırma.....	28
2.6.2.1.1 Karar Ağaçları	30
2.6.2.1.2. Genetik Algoritmalar.....	31
2.6.2.1.3. Naïve-bayes	32
2.6.2.1.4. Yapay Sinir Ağları.....	32
2.6.2.1.5. K-en En Yakın Komşu	33
2.6.2.1.6. Regresyon.....	33
2.6.2. Tanımlayıcı Veri Madenciliği Yöntemleri.....	34
2.6.2.1. Kümeleme	34
2.6.2.2. Birliktelik Kuralı	35
2.6.2.3. Carma Algoritması	37
2.6.2.4. GRI Algoritması.....	37
2.6.2.5. Eclat Algoritması	38
2.6.2.6. FP-Growth Algoritması.....	38
2.6.2.7. Apriori Algoritması.....	38
3. ÇOCUK HASTANE VERİLERİNİN VERİ MADENCİLİĞİ TEKNİKLERİ İLE ANALİZİ.....	43
4. DEĞERLENDİRME VE SONUÇ	74

KAYNAKÇA	77
TURNITIN RAPORU	88
ÖZ GEÇMİŞ.....	89
ETİK KURUL ONAYI	88



KISALTMALAR ve SİMGELER

AI	Yapay Zekâ
CRM	Customer Relations Management
E.Ü.	Erciyes Üniversitesi
ESK	Elektronik Sağlık Kaydı
Gb	Gigabyte
KDS	Karar Destek Sistemi
s.	Sayfa
Vb.	Ve benzeri
Vd.	Ve diğerleri
Vs.	Vesaire
WEKA	Waikato Environment for Knowledge Analysis

TABLolar LİSTESİ

Tablo 1.	Veri İşleme Tekniklerinin Gelişimi ve Veri Madenciliği Yolcuğu.....	10
Tablo 2.	Veri Madenciliğinde Kullanılan Uygulamalar.....	18
Tablo 3.	Veri Kalitesini Etkileyen Boyutlar.....	27
Tablo 4.	Veri Ambarı Setleri Kategori Tablosu	43
Tablo 5.	Çocuk Hastanesi Tüm Veriler kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüler	46
Tablo 6.	Çocuk Hastanesi Tüm Veriler Acil Departmanı Hariç Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüler	50
Tablo 7.	Çocuk Hastanesi Tüm Veriler Bölge Kayseri Bilgisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüler	53
Tablo 8.	Çocuk Hastanesi Kayseri ili Acil Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüler	55
Tablo 9.	Çocuk Hastanesi Kayseri ili Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüler	57
Tablo 10.	Çocuk Hastanesi Kayseri ili Hariç Acil Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüleri.....	60
Tablo 11.	Çocuk Hastanesi Sadece İç Anadolu Bölgesi Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüleri.....	64
Tablo 12.	Çocuk Hastanesi Sadece İç Anadolu Bölgesi Acil Bölümü Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüleri	67
Tablo 13.	Çocuk Hastanesi İç Anadolu Bölgesi Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüleri.....	70
Tablo 14.	Çocuk Hastanesi İç Anadolu Bölgesi Hariç ve Acil Bölümü Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüleri	72

ŞEKİLLER LİSTESİ

Şekil 1. Veri Enformasyon Bilgi	5
Şekil 2. Veri Ambarı Şeması	7
Şekil 3. Veri Madenciliği Süreci	9
Şekil 4. IBM SPSS Statics MLP Çok Katmanlı Algılayıcı Sinir Ağı Model Ekranı	19
Şekil 5. IBM SPSS Modeller En Üst Düzeyde Akış Tuval Ekranı	20
Şekil 6. WEKA Program Ekranı	24
Şekil 7. Veri Madenciliği Yöntemleri Şeması.....	26
Şekil 8. Veri Ön İşleme Yöntemleri Şeması.....	28
Şekil 8. Karar Ağacı Modeli Şeması	30
Şekil 9. Genetik Algoritma Genel Akış Şeması	31
Şekil 10. Çok katmanlı Yapay Sinir Ağları Şeması	33
Şekil 11. Kümeleme Şeması.....	35
Şekil 12. Dört nesne içeren bir kümenin farklı alt nesne oluşumu.....	37
Şekil 13. Apirio Algoritması Akış Diyagramı.....	40
Şekil 15. Örnek Uygulama	41
Şekil 16. Çocuk Hastanesi Tüm Veriler kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü	45
Şekil 17. Çocuk Hastanesi Tüm Veriler Acil Departmanı Hariç Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü.....	49
Şekil 18. Çocuk Hastanesi Tüm Veriler Bölge Kayseri Bilgisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü.....	52
Şekil 19. Çocuk Hastanesi Kayseri ili Acil Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü.....	54
Şekil 20. Çocuk Hastanesi Kayseri ili Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü.....	56
Şekil 21. Çocuk Hastanesi Kayseri ili Hariç Acil Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü.....	59

- Şekil 22. Çocuk Hastanesi Sadece İç Anadolu Bölgesi Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü..... 63
- Şekil 23. Çocuk Hastanesi Sadece İç Anadolu Bölgesi Acil Bölümü Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü.. 66
- Şekil 24. Çocuk Hastanesi İç Anadolu Bölgesi Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü..... 69
- Şekil 25. Çocuk Hastanesi İç Anadolu Bölgesi Hariç ve Acil Bölümü Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü.. 73



1.GİRİŞ

Günümüzde birçok sektörde olduğu gibi sağlık alanında da veri madenciliği yöntemlerinin etkin bir şekilde kullanılması sağlık yönetimi ve sağlık çalışanları açısından büyük faydalar sağlamaktadır. Bu yöntemler genellikle karar vericilerin doğru kararlar vermelerine yardımcı olacak çeşitli yönetsel araçlar olarak karşımıza çıkmaktadır. Bu yönetsel araçları kullanarak büyük verilerden faydalı örüntüler ortaya çıkaran işletmeler, rakiplerine göre farklılık oluşturmakta ve bu durum işletme karlılığı ve verimini olumlu yönde etkilemektedir. Boyutları her geçen gün biraz daha artan sağlık verisinin uzman bir veri madenciliği tekniği ile işlenerek verimlilik ve kalite avantajı sağlayacak bir güce dönüştürülmesi, sağlık işletmelerine yapılan büyük bilişim yatırımlarının bir sonucu olarak karar vericilerin daha etkin kararlar verebilmesini sağlamaktadır. Sağlık işletmelerinde veri madenciliği çalışmalarına gerekli önem verildiğinde rekabet ve hizmet kalitesi açısından rakiplere karşı üstünlük elde edileceği açıktır.

Sağlık işletmelerinde doğru şekilde planlanmış bir veri madenciliği tekniğinin kullanımı, doğru şekilde kayıt altında tutulup yorumlanan verilerin derinlerinde bulunan kıymetli, daha önce hiç farkına varılmamış örüntüleri ortaya çıkaracak ve bu örüntülerden faydalı davranış modelleri oluşturmayı sağlayacaktır. Burada sorulması gereken soru doğru veri madenciliği çalışmasının nasıl yürütüleceği meselesidir. Doğru veri madenciliği uygulamasının ilk adımı, yapılacak çalışmanın hangi amaç doğrultusunda gerçekleştirileceğinin ve hangi sorun ile ilgilenileceğinin belirlenmesi olmalıdır. Doğru yürütülemeyen veri madenciliği, madencinin sorunu anlamaya çalışırken ele aldığı veriden sonuç elde edemediği ve sorunun katlanarak büyüdüğü karşılıksız bir çabadan öteye gidemeyecektir. Amacın ve sorunun belirlenmesi iki ana unsur olarak ortaya konduktan sonra büyük verinin bir dizi işleme tabi tutulması ile veri madenciliğine başlanmış olmaktadır. Bu aşamada öncelikle veri madenciliği yapılacak verinin toplanması ve uyumlaştırılması gerekmektedir. Toplanan veriden yola çıkılarak yürütülecek olan veri madenciliği çalışmaları için yeterli miktarda verinin barındırılıp

barındırılmadığı hususuna oldukça dikkat edilmelidir. Eksik ya da yetersiz veri ile başlanan çalışmalarda başarı olasılığı azalacak ve verimsizlikler oluşabilecektir. Bu adımın tamamlanması sonrasında veri birleştirme ve temizleme adımları ile toplanan veriler düzenlenip amaca uygun hale getirilerek belirlenen sorunun cevabına yönelik tek bir veri tabanında toplanmış veri ambarı haline getirilmektedir.

Veri ambarı oluşturulduktan sonra amaca ve sorunun çözümüne uygun bir veri madenciliği modelinin belirlenmesi ve bu modele uygun veri formatlarının hazırlanması aşamaları devreye girmektedir. Büyük veri içerisinde örüntü keşfetmek amacıyla uygun veri formatının oluşturulması aşamasında dikkat edilmesi gereken en önemli husus, veri madenciliğinde en çok hataya yol açtığı düşünülen hatalı ya da eksik verilerin dikkatli bir şekilde ele alınması gerekliliğidir. Veri ön işleme sürecinde eksik verilerin doğru şekilde tamamlanması bir metot olarak karşımıza çıkmaktadır. İlginç ve keşfedilmemiş örüntüleri elde etmeye yönelik çalışmalarda güvenilir olmayan hatalı verileri temizlemek ve eksik verileri tamamlamak yapılacak çalışmanın başarılı olmasında önemli bir yer tutmaktadır. Ayrıca, kullanılacak olan veri tabanının yeterli büyüklükte olması avantaj iken verinin rastgeleliğini bozmadan örneklemeler yapılabilmesi ve seçilen veri setinin tüm popülasyonu temsil edip etmediğinin kontrolü yapılan çalışmanın güvenilirlik düzeyini artırmaktadır.

Önerilen tez çalışması kapsamında ele alınan veriden etkin kurallar elde edebilmek amacıyla tanımlayıcı veri madenciliği yöntemlerinden birisi olan *birliktelik kuralı* yöntemi tercih edilmiştir. Çalışma kapsamında Erciyes Üniversitesi Tıp Fakültesi Hastaneleri bünyesinde faaliyet gösteren Çocuk Hastanesi'nin 2021-2022 yıllarına ait hasta verilerinin yer aldığı bir veri kümesi kullanılmıştır. Büyük boyutlu veri içerisinde bulunan keşfedilmemiş kısımlara ulaşarak anlamlı örüntüler oluşturabilmek amacıyla elde edilen veride bulunması muhtemel eksik değerler uygun yöntemlerle tamamlanmış ve daha gerçekçi analiz sonuçları elde edilebilmesi amacıyla hassas bir dizi gürültü azaltma prosedürü uygulanmıştır. Ön işleme süreci sonunda elde edilen uygun veri WEKA açık kaynak kodlu veri madenciliği yazılımında bulunan Apriori algoritmasına tabi tutularak 10 farklı örüntü deseni elde edilmiştir. Bu desenler ile ilgili detaylı açıklamalar tezin değerlendirme ve sonuç kısmında detaylı bir biçimde ele alınmıştır.

Tezin ikinci bölümünde literatür çalışmasına yer verilmiştir. Bu bölümde konuyla ilgili yerli ve yabancı literatür taranarak veri madenciliği konusu güncel bilgiler ışığında sunulmaktadır. Veri madenciliğinin tanımı, tarihçesi, uygulama alanları ile sağlıkta veri madenciliği uygulamaları konuları ayrıntılı ve olabildiğince örneğe dayalı olarak anlatılmıştır. Veri madenciliğinde kullanılan programların en son versiyonları ve üreticilerin kendi resmî web sitelerinden elde edilen güncel bilgiler sunulmuş olup literatüre katkı sağlaması amacı ile programların ekran görüntüleri ve yazılımı üreten firmaların sunduğu güncel bilgilere de yer verilmiştir. Veri madenciliği yöntemleri konusunda; veri ön işleme konusu, tahmin edici yöntemlerden sınıflandırma, yapay sinir ağları ve regresyon konularına yer verilirken tanımlayıcı yöntemler konusunda ise kümeleme ve birliktelik kuralına değinilmiştir. Son olarak hastane verilerinin veri madenciliği teknikleri ile analizi ile ilgili literatürde yer alan çalışmalara yer verilmiştir.

Tezin üçüncü bölümünde, WEKA Veri Madenciliği yazılımı kullanılarak Erciyes Üniversitesi Tıp Fakültesi Hastaneleri'nin 2021 ve 2022 yıllarına ait 1.110.322 satır hasta verisinden veri ön işleme çalışması yapılarak Çocuk Hastanesi verileri ayrıştırılmış gerekli olan veri birleştirme ve temizleme işlemleri yapılarak 27.000 satırlık hasta veri seti oluşturulmuştur. Veri ön işleme aşamasından geçirilen ham veriler temizlenerek planlanan 10 veri seti üzerinde WEKA Apriori algoritmasıyla birliktelik kuralı yöntemi uygulanarak veri madenciliği sonuçları elde edilmiştir. Her bir veri setinden elde edilen örüntülerden çıkan keşfedilmemiş bilgilerin değerlendirilmesi ilk 3 örüntü üzerinden yapılmış olup sonrasında genel örüntülerden elde edilen bilgiler ışığında ortaya çıkan analizler yapılmıştır.

Dördüncü bölümde veri madenciliği yöntemi ile elde edilen örüntüler ışığında keşfedilen bilgilerin sağlık işletmelerine sağlayacağı katkılar istatistiki bilgi olarak ortaya konulmuş ve işletmelerdeki karar vericilerin doğru adımlar atması amacı ile elde edilen bilgileri kullanmaları konusunda yol gösterici yorumlar, öneriler ve değerlendirmeler yapılmıştır.

2.VERİ MADENCİLİĞİ

2.1 Veri Madenciliği Tanımı

Şirketlerin dijital sistemler kullanılarak geliştirilebilecek pek çok noktası bulunmakta ve her bölümün iyileştirilebilmesi uzun yıllar alabilmektedir. Doğru ve etkin yönetim için bir şirketteki en ufak veri parçacığı bile dijital formata aktarılmalı ve kolayca erişilebilir olmalıdır. Bu veriler mevcut işle ilgili her türlü dosya, eposta, kayıt ve ilişkili bilgileri barındırabilmektedir. Bu bakış açısıyla şirket içerisindeki her işlem dijitalleştirilmeli ve diğer işlemlerle bütünleştirilmelidir. İş ortakları ve müşteriler açısından bakıldığında da tüm işletme süreçlerine ait işlemlerin dijital ortama taşınması ve ilgili tüm verilerin işlenerek erişilebilir olmasının sağlanması büyük önem arz etmektedir (Gates, 1999, s.378). *‘Dijital sinir sistemiyle düşünce hızında çalışmak’* kitabında Gates, 1999’lu yıllarda başarılı bir işletme yönetimi ve müşteri memnuniyeti gereklilikleri için dijital sinir sistemi diye adlandırdığı verilerden bahsetmiş ve aralarında kurulabilecek bağların başarının anahtarı olabileceğinden ve o zamanki kısıtlı verilerle kurulacak ilişkilerin işletmelere sağlayacağı faydalara değinmiştir.

Hal böyle iken veri madenciliği tanımına girmeden önce verinin ne olduğunun bilinmesi gerekmektedir. Veri ve bilgi kavramları aynı şeyi ifade ettiği düşünülen ama aslında farklı şeyleri ifade eden kavramlardır. En basit ifadeyle veri, çeşitli analiz yolları ile işlendiğinde bilgiye dönüşen bir kavramdır. Bir diğer deyişle veri bilginin ham işlenmemiş halidir. Bütün bu kapsamda verinin tanımını yapmamız gerekirse: analiz neticesinde bilgiye ulaşmayı sağlayan tek başına anlamlı çıkarım yapılamayacak parçalardır. Verilerin toplanma biçimi gözlem ve kayıt işlemleri ile gerçekleşmekte ve

toplandıktan sonra bir kısım işlemler uygulanarak (veri ön işleme ve analiz) çeşitli çıkarımlarda bulunulabilmekte, raporlanabilmekte ya da görselleştirilebilmektedir (Yalçın, 2019, s. 9). Aşağıda verilen veri, enformasyon, bilgi şekline bakıldığında bilginin temelini veri ve enformasyonun oluşturduğu görülmektedir. Bu kavramlardan enformasyon ve bilgi çokça birbirine karıştırılan kavramlar arasında yer almaktadır. Enformasyon denilen kavram verinin rasyonel biçimde akıl süzgecinden geçirilerek yorumlanması ile ortaya çıkmaktadır. Bilgi ise karar verme süreçlerinde planlama ve karşılaştırmaya dayanak oluşturacak eylem ve uygulamaların temelini oluşturmakta ve bu temele göre analiz değerlendirme ve tahminler yapılabilmektedir (Çapar, 2005, s. 176).

Şekil 1. Veri, Enformasyon ve Bilgi



Veri madenciliğini daha iyi anlayabilmek için veri tabanı kavramının iyi şekilde anlaşılması gerekmektedir. Veri tabanı belirli özelliklerde örgütlenmiş verilerin bir arada bulunduğu yapıya verilen isimdir. Ele alınan veri miktarı ile işlem yapan kullanıcı sayısı büyüklüklerine bakılmaksızın tüm veri operasyonlarının güvenilir şekilde ve makul sürelerde işletilmesi gerekmektedir. Veritabanına aynı anda birden çok kullanıcının erişebilmesi, farklı uygulamaların veri sorgusu yapabilmesi, veri üzerinde yapılan operasyonların tutarlı ve veriye etkisinin bir bütünlük içinde yürütülmesi de veri tabanı yönetim sistemi dediğimiz yapıların vazgeçilmez özelliklerindedir. Veri tabanı yönetim sistemleri, veri tabanını oluşturan verilerin tüm özelliklerinin korunmasını sağlayacak şekilde tasarlanmışlardır. Herhangi bir verinin sahip olduğu veri türü özelliği ile o verinin değeri, fiziksel konumu, uzunluğu, indeks yapısı, log kaydı, yetkilendirme

özellikleri ya da erişim kısıtları, parametreleri, birleştirme (join) yapıları gibi birçok özelliği bulunmaktadır. Bu özellikler sayesinde veri tabanlarından veri madenciliği teknikleriyle veri içerisinde saklı, anlamlı bilgilerin gün yüzüne çıkarılmasını sağlanmaktadır (Yorulmaz, 2018, s. 10).

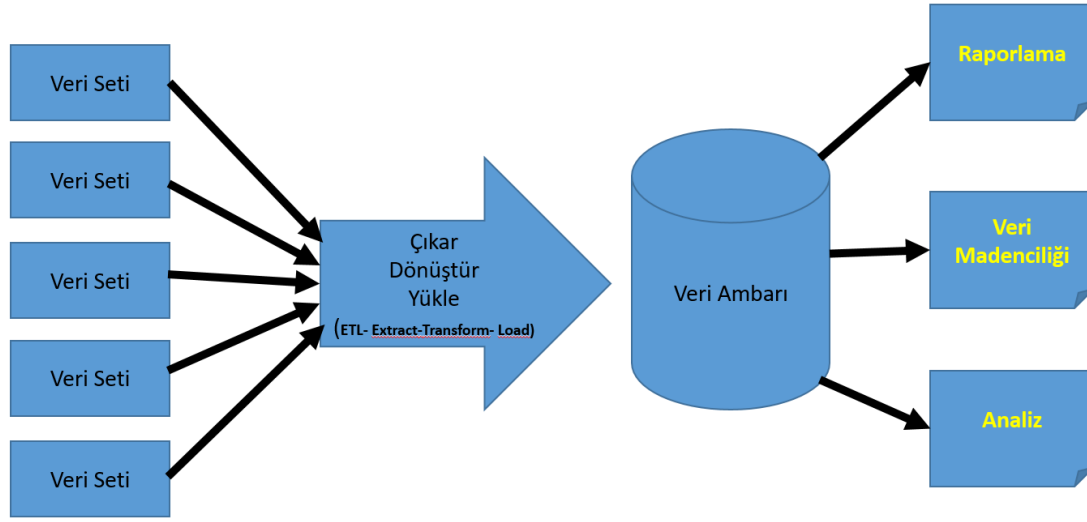
Bir başka ifadeyle veri tabanı, veriyi kataloglamayı, sınıflandırmayı, veri ile ilgili ilişkilendirmeler oluşturmayı hedefleyen, veriye hızlı erişilebilmesini ve depolanmasını sağlayan fiziksel sistemlerin adıdır. Veri tabanı verilerinin kayıt ortamında en az hafızayı kaplayacak şekilde saklanmasını amaçlarken aynı zamanda verimliliğin ve tutarlılığın sağlanması için veri tabanlarını oluşturan ve yöneten veri tabanı yönetim sistemi yazılımlarına ihtiyaç duyulmaktadır (Yalçın, 2019, s. 9).

Veri madenciliğinin gerçekleştirilebilmesi için veri, veri tabanı ve veri tabanının veri ambarına dönüştürülmesi için yapılması gereken çeşitli ön işlemler gerekmektedir. İşte veri madenciliğine uygun hale getirilmiş zaman değeri barındıran ve çeşitli sınıflandırmaların gerçekleştirildiği veri tabanlarına veri ambarı denilmektedir (Silahtaroglu, 2016, s. 15).

Inmon ise veri ambarını kitabında şöyle tanımlamaktadır: “*Karar alma sürecinde olan yöneticilerin karar alma sürecini desteklemek için, belli bir konuya odaklanarak, zaman bağımlı, entegre, ayrıntılı kurumsal verileri kapsayan kalıcı veri bütününe veri ambarı denir*” (Inmon, 2002, s. 33). Bu bağlamda veri ambarının oluşturulabilmesi için veri yüklemesi, veri entegrasyonu, veri temizleme, veri dönüşümü gibi adımların bulunduğu sürecin işletilmesi gerekmektedir (Han ve Kamber, 2006, s. 39).

Aşağıdaki şekilde bir veri ambarı oluşturulması aşamaları anlatılmaktadır. Öncelikli olarak oluşturulan veri setleri veri ön işleme adımlarından geçirilerek gerekli işlemler tamamlanmaktadır. Bu işlemlerden sonra değiştirme, çıkarma, temizleme ve aktarma işlemleri gerçekleştirilmektedir. Sonrasında ön işleme sürecinden geçirilen veriler veri ambarına dönüştürülerek oluşturulan veri ambarından veri madenciliği teknikleri kullanılarak gerekli olan raporlama ve analiz işlemleri gerçekleştirilmektedir.

Şekil 2. Veri Ambarı Şeması



Veri tabanları kamu kuruluşları ve ticari işletmelerin ana faaliyetlerini devam ettirebilmeleri için elzem olan verileri depolamakta kullanılmaktadır. Günlük yaşantımızın bir parçası olan alışverişlerimizde kredi kartı kullanımının artması, bankacılık işlem verileri, telefon görüşme detayları olmak üzere birçok farklı alanda oluşturulan verilerin yanında uzay gözlemleri, sağlık verileri ve bilimsel araştırma temelli depolanan çok çeşitli özelliklere sahip veriler, amaçları farklı veritabanlarında yönetilmekte ve gelişmiş dijital saklama ortamlarında saklanabilmektedir (Aydın, 2007, s. 3).

Veri madenciliği son zamanlarda bilgi endüstrisinde büyük önem verilen konular arasında yer almaktadır. Bunun nedeni, büyük miktarda verinin yaygın olarak bulunması ve bu tür verileri faydalı bilgiye dönüştürmeye yönelik acil ihtiyaçlardan kaynaklanmaktadır. Veri madenciliği, büyük miktardaki verilerden anlamlı bilgileri çıkarmak için istatistiksel ve matematiksel yöntemlerin kullanıldığı bir disiplindir. Veri madenciliği, karmaşık veri kümelerini analiz ederek desenleri, ilişkileri ve trendleri keşfetmeyi amaçlamaktadır. Veri madenciliği, çeşitli veri kaynaklarından bilgi toplar, veri ön işleme adımlarıyla veriyi hazırlar, istatistiksel ve veri madenciliği algoritmalarını uygular ve sonuçları yorumlar. Bu süreçte, sınıflandırma, kümeleme, regresyon, ilişki analizi, zaman serisi analizi gibi teknikler kullanılmaktadır.

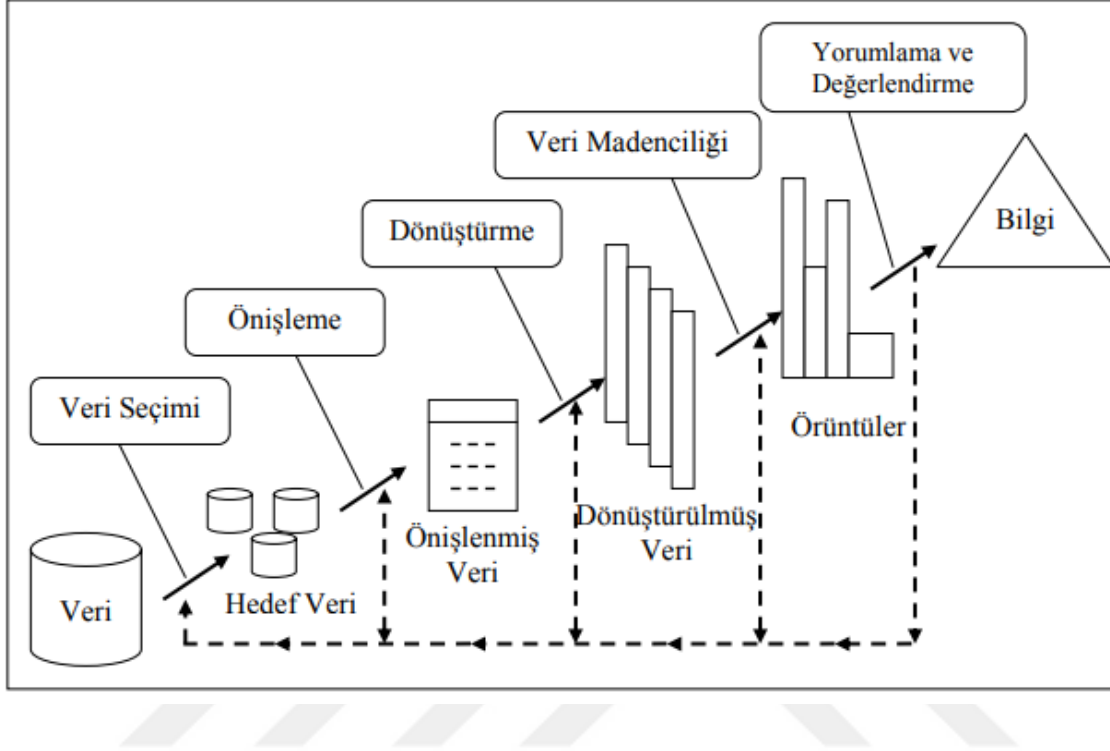
Veri madenciliği günümüzde pazarlama, finans, sağlık, sosyal medya analitiği, müşteri ilişkileri yönetimi gibi büyük verilerin söz konusu olduğu birçok alanda kullanılarak eğilimleri tahmin etme, müşteri davranışlarını analiz etme ve dolandırıcılık tespiti gibi faydalı araçlar sunmaktadır.

Günümüzde başlı başına bir konu olan veri madenciliği ile ilgili ilk teknik uygulamalar yaklaşık 30 yıl öncesine dayanırken kökenleri, 1950'lerde yapay zekânın (AI) erken gelişim dönemlerine kadar gidebilmektedir. Bu dönemde, örüntü tanıma ve kural tabanlı muhakemedeki gelişmeler, veri madenciliğinin dayanacağı temel yapı taşlarını sağlamaktadır. İlişkisel veri tabanının ortaya çıkışı ile ticari kuruluşlar daha büyük hacimlerde verileri depolamaya başlamışlardır. Büyük verilerin birikmesi ve bunları analiz edilip daha sonra anlamlı bilgilere dönüştürme ihtiyacının artması sonucunda veri madenciliği kavramı ortaya çıkmıştır.

Büyük miktardaki verilerden anlamlı bilgileri çıkarmak için istatistiksel ve matematiksel yöntemlerin kullanıldığı önemli bir disiplin olarak veri madenciliği yerini almaktadır. Veri madenciliği için çok sayıda tanım bulunmaktadır. Örneğin veri madenciliği “yeni bilgiyi keşfetmek” olarak tanımlanırken (Becerra, vd., 2004), “bilinmeyi bilmek, anlamak” olarak da tanımlanmaktadır (Awad ve Ghaziri, 2004). Ayrıca, “büyük veri tabanlarından veri nitelikleri, eğilimler veya örüntüler gibi gizli bilgileri, verileri farklı açılardan analiz ederek ve yararlı bilgilere özetleyerek çıkarma işlemidir” ifadesi bir başka tanım olarak karşımıza çıkmaktadır (Buruncuk, 2006). Örüntüleri ve ilişkileri bulmak, verileri sınıflandırmak ve yapılandırılmış verilerin büyük veri kümelerindeki sonuçlarını tahmin ederek sonuçları analiz etmek, veri madenciliği kavramının genel bir tanımı olarak ifade edilebilmektedir. Bilgisayar bilimi, makine öğrenimi ve istatistiğin ortak paydalarda bulunduğu bir alan olarak görülebilmektedir. Bu kategoriye uyan farklı teknikler arasında ilişki öğrenme, veri sınıflandırma ve kümeleme ile regresyon yer almaktadır. Aşağıdaki şekilde verinin anlamlı ve ilginç veriye dönüşümü aşamaları ve veri madenciliği süreçleri gösterilmektedir. Bu şekil incelendiğinde ham veri alınarak öncelikle veri seçimi yapılır ve veri ön işleme süreçlerinden geçirilerek karar verilen veri madenciliği amacına uygun olarak dönüştürme işlemleri gerçekleştirilir. Dönüştürülmüş veri üzerinde yapılan veri madenciliği işlemi örüntülerin ortaya çıkarılmasına ve bu örüntüler üzerinde

yapılan yorumlama ve değerlendirme işlemleri ile birlikte ilginç ve anlamlı bilgiye ulaşmamızı sağlamaktadır.

Şekil 3. Veri Madenciliği Süreci Kaynak: (Fayyad vd., 1996, s. 37)



2.2. Veri Madenciliği Tarihçesi

Gazi Güder'in Bilgi İşlem Terimleri Sözlüğü'nde veri, insan ya da makine tarafından üretilen açıklama, iletişim, durum, koşul vb. açıklamak için kullanılan sayı, harf ya da sembollerin oluşturduğu işlenmemiş ham bilgiye verilen ad olarak tanımlanmaktadır (Güder, 1986, s. 109). Tanım 1986 yılında yapılmış olsa da günümüzde de geçerliliğini korumaktadır. Bunun yanında o yıllardan günümüze geldiğimizde en büyük değişim ise üretilen verilerin çeşitliliği ve boyutlarındaki devasa artıştır. Hızlı gelişen bilişim teknolojileri çok büyük boyuttaki verilerin depolanmasına imkân sağlamıştır. İnternet, sosyal medya, IOT (nesnelerin interneti) cihazları ve IP (internet protokolü) ile konuşan her cihaz artan dijitalleşme çerçevesinde çok farklı ve büyük boyutta verinin kaynağını oluşturmaktadır.

İlk bilgisayarlar lambalı teknolojiye sahip sadece sayım yapmak amacı ile 1950'lerde kullanılmaya başlanmıştır. 1960'lı yıllara geldiğimizde teknoloji daha da gelişmiş ve

veri tabanı ile verilerin depolanması konuşulur olmuştur. 1960'ların sonu 1970'lerin başında bilim insanları transistörden entegre denilen yapılara geçerken basit öğrenme modeline sahip bilgisayarları geliştirmeyi başardılar. 1980'li yıllara gelindiğinde ise şirketler, farklı verilerden oluşan (rakip verileri ya da müşteri verileri) veri tabanları oluşturmaya başladılar. 1980'lerden sonra veri tabanı yönetim sistemleri daha çok bilimsel çalışmalarda kullanılmaya başladı ve yaygınlaştı. 1990'lı yılların başlarında veri ve veriden nasıl anlamlı bilgi üretileceği konusundaki ilk çalışmalar üretilerek veri tabanları üzerinde öncü çalışmalar başlamıştır. 1992 yılına geldiğimizde veri madenciliği için önemli bir atılım gerçekleşerek bu çalışmalara ivme kazandıracak ilk yazılım üretilmiş ve kullanılmaya başlanmıştır. 2000'li yıllar veri madenciliği tekniklerinin farklılaştığı ve sürekli gelişim gösterdiği yıllar olmuştur. Çalışmalar sonucu ulaşılan neticelerin oldukça tatmin edici olması farklı disiplinlerden de bu alana yönelik ilginin artmasını sağlamış ve bu kapsamdaki çalışmalar hızlanmıştır (Savaş vd., 2013, s. 5).

İnsanlar geçmişte olduğu gibi bugün de verileri yorumlayıp anlamlı bilgiler çıkarmaya çalışmakta ve bunun için çeşitli metotlar ile donanımlar oluşturmaktadırlar. Bu donanımın en önemlisi teknolojik gelişmeler sonucu bulunan bilgisayar olup bu teknoloji sayesinde bilginin depolanması ve taşınması daha sistematik bir hale getirilmiştir. Zamanla her alanda önemli ölçüde veri toplanmaya ve bu verilerin bilgiye dönüşümü konusunda çalışmalar yapılmaya başlamıştır (Farboudi, 2009, s. 55).

Aşağıdaki Tablo1'de veri işleme tekniklerinin gelişimi ve veri madenciliğinin tarihsel yolculuğu özetlenmiştir.

Tablo 1. Veri İşleme Tekniklerinin Gelişimi ve Veri Madenciliğinin Tarihsel Yolcuğu

Gelişim Sıralaması	İlgili Teknolojiler	Ürün Tedarikçileri	Karakteristikler
1960'lar Veri Toplama	Bilgisayarlar Kasetler Diskler	IBM CDC	Statik veri ulaşımı
1970'ler İlişkisel Veri Modeli	OLTP (Online Transaction Processing)	IBM CDC	İlişkisel Veri Modeli
1980'ler Veriye Ulaşım	İlişkisel veri tabanları (RDMS) SQL ODBC	IBM, Sybase Informix, Oracle Microsoft	Veri kayıt anında dinamik ulaşımın sağlanması
1990'lar Veri Depolama ve	OLAP(Online Analytical Process)	Cornshare, Arbor, Cognos	Her seviyedeki veriye dinamik ulaşımın

Karar Destek	Çok boyutlu veri tabanları Veri ambarları	Pilot, Microstrategy	sağlanması
2000'ler ve Günümüz Veri Madenciliği	Gelişmiş algoritmalar Çok işlemcili bilgisayarlar Büyük hacimli veri tabanları	IBM ,Pilot Lockheed, SGI	Bilgiye önceden ulaşım Coğrafi engellerin kalkması

Kaynak: (Çoban, 2006, s. 35)

2.3. Veri Madenciliği Uygulama Alanları

Veri madenciliği uygulama alanları hızla değişen dijital dünya gibi değişmekte ve bu değişimlerin oluşturduğu avantajlarla birlikte her geçen gün yeni fırsatlar doğurmaktadır. Bize sunduğu imkânların yanında kişisel verilerimizin hangi izinlerle nasıl işlendiği ve nerelerde kullanıldığı konusu da dikkat edilmesi gereken bir başka husus olarak karşımıza çıkmaktadır. Farklı uygulama alanlarında kullanılmaya başlanan veri tabanları günümüzde her tür iş kolunda kendisine yer bulmakta ve her tür işletmenin vazgeçilmez unsuru haline gelmektedir. Kategorik anlamda bakıldığında veri madenciliğinin kullanım alanları şu şekildedir:

Pazarlama: Veri tabanları, müşterilerin demografik satın alma özellikleri arasındaki bağlantıların belirlenmesi, kampanya ürünlerinin seçimi, satın alma eğitimlerinin tespiti, mevcut müşterileri kaybetmeden yeni müşteriler kazanmaya yönelik yapılacak analizler, pazarlama kampanyası analizi gibi çalışmalarda çokça kullanılmaktadır. Bunun yanında müşteri ilişkileri yönetimi (CRM) kavramı ile başlayan müşteri odaklı davranışların incelendiği (satış tahmini, müşteri sadakati vb.) çalışmalar en yaygın veri madenciliği uygulama alanları olarak karşımıza çıkmaktadır.

Banka ve sigortacılık: Hem kişisel hem de kurumsal anlamda finansal araçlar arasında yer alan bankacılık ve sigortacılık faaliyetlerinin verimli şekilde yapılabilmesi amacı ile bu kategorilerde yapılan iş ve işlemler arasındaki ilişkilerin incelenmesi bu ilişkiler arasında kurulan bağların tespiti açısından veri tabanı teknolojileri kullanılmaktadır. Bankacılık işlemleri, kredi kartı dolandırıcılığı, kredi taleplerinde meydana gelen usulsüzlüklerin tespiti vb. gibi konularda yapılan analizler ve bu analizler sonucu elde edilen bilgiler hem sektörel açıdan önem arz etmekte hem de yapılabilecek olan olası dolandırıcılık faaliyetlerini engellemek amacı ile önemli bir enstürüman olarak kullanılmaktadır.

Biyoloji, genetik ve kimya: Bitki türleri, çeşitlilik analizi, gen haritasının çıkarılması gibi konuların tespit edilmesi, analizi ve sınıflandırılmasına yönelik yapılan çalışmaların yanında kanserli hücrelerin tespit edilmesi, virüs keşifleri, sınıflandırılması gibi karışık parametrelerin analizi ve değerlendirilmesinde kullanılmaktadır. Moleküllerin kategorizasyonu ve çeşitlendirilmesi gibi konularda yeni keşiflerin önünü açmaktadır.

Yüzey analizi ve coğrafi bilgi sistemleri: Dünya üzerindeki yerlerin coğrafi ayrımlarına göre sınıflandırılması, iklim ve yağış haritalarının oluşturulması, şehirlerde yerleşke yerlerinin tespiti, şehir suç oranları ve ekonomik göstergeleri ile demografik yapı, kentlerde posta hizmetleri, ATM hizmetleri, otobüs durakları vb. gibi belediyeçilik hizmetlerin yerinde ve doğru şekilde konumlandırılmasında kullanılmaktadır.

Bilişim ve mühendislik: Tespit edilen görüntülerin analizi ile suçlu tanıma, yol bilgisi, yüz tanıma ve doğal afet tahmini gibi işlemlerde kullanılmaktadır. Bunun yanında internet işlemlerinde yapılan dolandırıcılığın tespitinde ve bilgisayar ağlarına yetkisiz erişim tespitinden yüz şeklinden kimlik tespitine kadar çeşitli yapay zekâ uygulamalarında kullanılmaktadır.

Uzay bilimleri ve teknoloji: Gezegenlerde yer alan yüzey oluşumlarının ve gezegenlerin uzaydaki konumu, uzayda yapılacak yeni keşifler, gök cisimlerinin bulunduğu konumun sınıflandırılmasında ve astronomi alanında yapılacak tahmin ve hesaplamalar gibi farklı alanlarda kullanılmaktadır.

Sosyal bilimler ve davranış bilimleri: Başta siyasi anlamdaki anket ve yoklamalar, kişilerin genel eğilimleri, seçim davranış öngörülleri oluşturma ve davranış modelleri, toplumsal davranış ve alışkanlıklar gibi psikolojik, sosyolojik, istatistik bilimlerini içeren çeşitli analizleri gerçekleştirmek amacıyla kullanılmaktadır. Bu analizler sonucu sosyal bilimcilerin insan ve davranışları anlamalarına yardımcı olacak çıktılar elde edilebilmektedir.

Perakendecilik: Bir işletmenin kuracağı satış noktasının en uygun yer ve maliyete sahip olması amacı ile yapılacak veri analizleri, müşterilerin alış-veriş alışkanlıklarının analiz edilerek bir ürün ya da hizmetin daha çok nasıl satılabileceği, tedarik ve mağaza yerleşim yerlerinin seçimi optimizasyonu gibi birçok alanda kullanılmaktadır. Bu alanda

elde edilen veriler ve arařtırmalar büyük tedarik zincirleri ve mağazalarca satın alınmakta ve aktif olarak iřletme karlılıđını artırmak için kullanılmaktadır.

Borsa: Borsada bulunan hisse senedi fiyat tahminleri, para piyasaları analizi, hangi finansal aracın tercih edilmesi gerektiđi ve bu finansal araçlarla yapılacak olan kısa, orta ve uzun vadeli alım-satım stratejilerinin belirlenmesi gibi bir iřletme için büyük önem arz eden finansal yatırım argümanlarının analizinde kullanılmaktadır.

Endüstri: Bir ürünün üretilme aşamasından başlayarak kalite kontrol analizlerinde, ürünün dođru şekilde paketlenmesi, en hızlı ve ekonomik lojistik hizmetinin alınması ve ürün hakkındaki geri dönüşlerin analizleri gibi birçok alanda kullanılmaktadır.

Telekomünikasyon: Telekomünikasyon sürecinde verilen hizmetlerin devamlılıđının sağlanması, kalitesi ve sürdürülebilir olması gibi konularda elde edilen verilerin, iyileřtirme çalışmalarında, hatların yoğunluk tahmini analizlerinde kullanılması ve telefon dolandırıcılıđının tespit edilmesi amacı ile yapılan arama ve aranan numaralar ile karakteristiklerin çıkarılması ile analiz edilmesinde kullanılmaktadır.

İstihbarat: Devletlerin iç ve dış güvenlikleri için birçok metot bulunmaktadır. Bunlardan birisi de sinyal iz sürme, şüpheli telefon görüşmeleri gibi gelişmiş istihbarat teknikleridir. Ülkelerin ulusal güvenlik teşkilatları bu bağlamda birçok telefon görüşmesini kayıt altına almakta ve bu biriktirilen veriler kendi analistleri tarafından veri madenciliđi teknikleri kullanılarak analiz edilmekte ve istihbari bilgiye dönüřtürölmektedir.

İnternet ve web madenciliđi: İnternet ve web üzerindeki veriler gün geçtikçe logaritmik olarak artmakta ve bilgilerin kategorizasyonu karmaşıklaşmaktadır. İnternet üzerinde yer alan bilgi dökümanları ve resimlerden başka, hareketli görüntüler (streaming) ve sayısal veriler de web verileri olarak görölmektedir. İnternetin üzerinde yer alan verilerin erişim hızının artırılmasına yönelik algortima ve veri madenciliđi çalışmalarından yola çıkan web madenciliđi konusu üzerinde çalışılmaktadır (Akman, 2010, s. 10).

Sosyal medya: Sosyal medya ve bu medyalarda paylaşılan içerikler ile ilgili çok büyük bir veri trafiđi söz konusudur. Bu büyük verinin analizi ve dođru kullanımı ile kişilerin

gönüllü olarak paylaştıkları veriler üzerinden anlamlı bilgiye ulaşmak ve istenilen bu anlamlı bilginin birçok pazarlama faaliyetinde kullanımını günümüzde geleneksel pazarlama araçlarının (televizyon, radyo, gazete, dergi, broşür vb.) oldukça önüne geçmiştir (Kirby, 2006, s. 34). Günlük yaşantımızın sonucu olan ve sosyal mühendisliğin en fazla uygulandığı sosyal medya veri madenciliğinden elde edilen izlenme oranları, yaş, eğitim ve ilgi alanları gibi verileri analiz ederek yukarıda bahsedilen alanlarda perakendecilik, banka, istihbarat, pazarlama ve sosyal davranışlar açısından insanları yönlendirmektedirler.

2.4. Sağlık Alanında Veri Madenciliği Uygulamaları

Her alanda yaygın şekilde kullanımı olan veri madenciliği şüphesiz sağlık alanında da oldukça önemli bir yere sahiptir. Sağlık alanının tanı teşhis ve tedavideki kalite ve performansının artırılmasına yönelik yapılan veri madenciliği çalışmalarının önemi yadsınamaz durumdadır. Bunun yanında veri madenciliği uygulamaları ile hedeflenen bir önemli nokta da sağlık yönetimi açısından işletmenin karlılığının artırılması, operasyonel maliyetlerin azaltılması ve personel performansının artırılarak daha memnun müşteri portföyünün sağlanarak tedavi sürecinin bu kapsamda tamamlanabilmesidir. Müşteri memnuniyeti, tercih edilebilirlik ve sürdürülebilir kalite kavramlarının sağlanabilmesi için gerekli olan analizler sağlık kurumları için çok kıymetli ve büyük öneme sahip bilgilerdir.

Hal böyle iken sağlık alanında veri madenciliği uygulamalarına göz atacak olursak bu çalışmaların bilişim teknolojileri kullanılmadan önceki zamanlarda da yapılmaya çalışıldığını görmekteyiz. Sağlık alanında veri madenciliğinin temelini oluşturan ilk çalışma 1961 yılında Julian Samora ve arkadaşları tarafından gerçekleştirilen istatistiksel bir çalışmadır. Bu çalışmada davranışsal temel baz alınıp hastaların tedavisinde doktor-hasta ilişkisine dikkat çekilmiş ve teşhis-tedavi sürecinin verimi için 125 kişiden oluşan hastalarda yapılmış olan istatistiksel çalışma ile geçmişteki tıbbi ve demografik veriler ışığında bu tür çalışmaların yapılması gerektiği savunulmuştur (Samora, 1961, s. 83). Bu çalışmalar zaman içerisinde devam etmiş ve 1963 yılında Baldwin sağlık sektörü için bilginin önemine vurgu yaparak düşük maliyetli tedavi

planlaması için kullanılacak olan tıbbi verilerin ulařımındaki zorluęa dikkat çekmiřtir (Baldwin, 1963, s. 1237).

Bu çalıřmalar devam etmiř ve 90'lı yıllara geldiđimizde saęlık alanında veri madencilięi teknikleri hem yönetimsel anlamda hem de tanı-teřhis-tedavi baęlamında kullanılmaya bařlanmıřtır. 1988 yılında medikal acil karar destek sistemlerinde iliřkisel veri tabanı kullanımı fikri uluslararası konferanslarda tartıřılmaya bařlanmıřtır (Naeymi, 1988, s. 203). Aynı yıllarda uluslararası yapay zekâ konferansında medikal destek sistemlerinden anlamlı verinin elde edilmesi konuları görüřülmeye bařlanmıřtır (McLeish, 1988, s. 778). Daha sonra veri madencilięini kullanarak apriori yöntemi ile nicel birliktelik kurallarında hiyerarři oluřturma süreçleri bařlamıřtır (Agrawal, 1996, S.307). Birliktelik kuralları analizinde yapılan arařtırmalarda dünyada ve ölkemizde en çok kullanılan algoritmanın apriori yöntemi olduęu görölmektedir (Eker, 2016, s. 7).

Modern dünyadaki en önemli arařtırma alanlarından birisi olan saęlık alanında biliřim teknolojileri konusunda son otuz yılda büyük ilerlemeler kat edilmiřtir. Bilgi teknolojilerinde ve tanı araçlarındaki geliřim saęlık kuruluřlarının kalite ve performans deęerlerinin olumlu yönde artırmasına katkı saęlamıřtır (Kudyba, 2004, s. 146). Saęlık alanında kullanılan bilgi çeřitlilięinin artması ve bilgi depolama alanlarının geliřmesi sonucunda bu bilgilerin yorumlanması ve anlamlı hale gelmesi daha da kolaylařmıřtır. Bu sayede saęlık alanında kullanılan veri tabanları ile anlamlı veriler elde edilmekte teřhis, tedavi gibi operasyonel faaliyetlerin yanında, hastane yönetimi, müřteri memnuniyeti ve kalite kontrol gibi birçok alanda da iyileřtirme faaliyetleri yürütölmektedir (Esen, 2009, s. 29). Ayrıca saęlık alanında veri madencilięi yönetiminin kullanılması ile analize dayalı belirli bir alana odaklanmış ekip çalıřmaları yaygın hale gelmiř bu sayede yapılan arařtırmalarda ilgilenilen odak noktasına göre daha verimli ve detaylı problem çözümlerine ulařılmıřtır. Son arařtırmalar, saęlıkta veri madencilięi uygulamalarının belirli konulara odaklandıęını göstermektedir. Bunlardan en dikkat çekenler hastane kaynaklarının etkin ve verimli kullanılması, hastane deęerlendirmesi, tedavi tekniklerinin geliřtirilmesi, tıbbi süreçlerin kalitesinin artırılması, etkin maliyet ve yüksek verimlilik politikaları vb. konular olarak sıralanabilmektedir (Obenshain, 2004, s. 692).

Sağlıkta kullanılan veri madenciliğinin ülke sağlık politikaları kapsamında tedavi tekniklerini geliştirmeye ve bu tekniklerin kalitesini artırmaya yönelik önemli katkıları ve girişimleri bulunmaktadır. Veri madenciliği sayesinde tıp uzmanları ve hasta/hasta ailesi tedavi tekniklerini karşılaştırarak daha uygun maliyetli ve doğru tedavi yöntemine karar verebilmektedir. Hasta bakım planları, hastaların sosyo-kültürel ve fizyolojik durumları değerlendirilerek yapılmalıdır. Bakımın kalitesi, hastaların tedavi ve iyileşme sürecini doğrudan etkilemektedir. İyi bir sağlık hizmeti verebilmek için hastaların ihtiyaçlarının doğru şekilde tespiti, tercihleri ve tedavi şekilleri değerlendirilmelidir. O zaman ülke ve kurum bazlı etkin maliyetli ve yüksek verimliliğe sahip sağlık politikaları oluşturularak uygulanabilir olmaktadır (Tomar ve Agarwal, 2013, s. 1135)

Ülkemizde veri madenciliğinin sağlık alanında hayati öneme sahip olduğu bilinmekte olup Sağlık Bakanlığı'na bağlı Veri Yönetimi Daire Başkanlığı bünyesinde veri madenciliği konusunda çalışmalar yürütülmektedir¹. Bu kapsamda Sağlık Bakanlığı tarafından yapılan çalışmalar sonucunda 2023 yılında sağlık hizmetleri konusunda “Sağlıkta Dönüşüm Programı” ortaya konulmuştur. Ülkemiz için çok büyük öneme sahip sağlık verilerinin toplanması, saklanması ve analiz edilmesi sürecinde toplanan verilerin ulusal veya uluslararası standartlarda olmadığı tespit edilmiş ve bu standardın olmayışından kaynaklı oluşan karmaşanın engellenmesi amacı ile çalışma başlatılmıştır. Bu program çerçevesinde “Karar Süreçlerinde Bilgiye Etkili Erişim: SBS” hedefi ile yola çıkan Sağlık Bakanlığı, verilerinin toplandığı Elektronik Sağlık Kaydı (ESK) veritabanı, Ulusal Sağlık Veri Sözlüğü başta olmak üzere veri setleri, sağlık kodlama konusunda sektör çalışanlarına yol göstermek amacı ile Karar Destek Sistemi (KDS) gibi çalışmalar yaparak bu programı desteklemiştir². Sağlık Bakanlığı bünyesinde toplanan verilerin veri madenciliği ile anlamlı bilgiye dönüşümü, ülkemizin sağlık sisteminde atılacak adımlar ve belirlenecek yol haritası ile sosyal devletin sağlık hizmetini sunma noktasındaki sorumluluğuna şüphesiz katkıda bulunmaktadır.

¹ İnternet Erişim: <https://www.saglik.gov.tr/TR,11415/saglikta-donusum-programi.html> [Erişim tarihi: 18.07.2023]

² İnternet Erişim: <https://sbsgm.saglik.gov.tr/TR,32360/veri-denetimi-birimi.html> [Erişim tarihi: 18.07.2023]

Bunun yanında sađlık hizmetlerinde kullanılan veri madenciliđinin önünde çeşitli engeller de bulunmaktadır. Hastadan alınan verilerin farklı biçimlerde tutulması, veri setlerini oluşturmadaki en önemli engellerden biridir. Sađlık sektöründe kullanılan yazılım ve donanımların farklı standartlardaki veriyi tutması veri tabanlarının entegrasyonunu zorlaştıran bir diđer faktördür. Literatüre baktığımızda ise yüksek hacim, farklı veri tabanları ve güncelleme sorunları, çok veya az sayıda deđişkenin bulunması, eksik ya da tutarsız veri, veri tabanlarının entegrasyonundaki güçlükler ve anlaşılamayan sonuçların miktarı (Hosseinkhah vd., 2009, s. 189) olarak belirtilmiştir. Bir başka engel ise medikal alanında kaliteli veriye ulaşmadaki güçlüktür. Sađlık verileri doğası geređi karmaşıktır (Bellazzi ve Zupan, 2008, s.85). Terminoloji standardizasyonunun sağlanamaması, kelime bilgi eksiklikleri ve veri tabanlarındaki veri formatlarındaki hatalar veri işlemlerini zorlaştırmakta (Esfandiari vd., 2014, s.84) ve anlamlı bilgiye ulaşılmasını engellemektedir. Sađlık hizmetleri alanındaki bir başka engel de hasta hakları, hastanın gizlilik politikaları, kişisel verileri koruma kanunu gibi hususlardır. Bu engel hem tıbbi hem de akademik araştırmaların önünde önemli bir sorun olarak durmakta ve çalışmayı yapacak akademisyen için dikkat edilmesi gereken bir husus olmaktadır. Bilişim sistemlerinin oluşturulması maliyeti yüksek kalifiye personel çalıştırılması gereken bir iştir. Bu bileşenlerin iyi uygulandıđı sistemler, sistemin etkinliğini ve başarısını etkileyecektir. Veri madenciliđi sonrası elde edilen analizin başarısı ya da başarısızlıđı, sistem kadar sistemi kullanan kişilerin teknik ve bilgi becerileri yani kalifiye olmaları ile doğrudan bađlantılıdır. Hastanelerde kullanılan hastane bilgi sistemleri, bilişim personeli ve son kullanıcılar tarafından iyi anlaşılmalı ve benimsenen sistemlerden oluşturulmalıdır. Ancak bu sayede başarı elde edilebilmektedir. Uygulamaların hayata geçirilmesi konusunda yürütülecek faaliyetler uzman kişiler tarafından gerçekleştirilmelidir. Ayrıca sađlık sistemlerindeki dijitalleşme her geçiş sađlık personeline iyi anlatılmalı ve sistemin bir parçası olduđu vurgusunun yapılması sistemin çalışabilir halde kalmasının önemli unsurlarındandır (Niaksu ve Kurasova, 2010, s.58).

2.5 Veri Madenciliđinde Kullanılan Programlar

Günümüzde yaptığımız işin özelliklerine göre doğru ve hızlı çözümler bulma arayışımızın vazgeçilmez yardımcıları olan bilgisayar programları, veri madenciliđi

çalışmalarında da imdadımıza yetişmektedir. Birbirinden farklı teknolojiler ve algoritmalar kullanılarak kodlanmış birçok veri madenciliği uygulaması geliştirilmiş ve kullanım için bizlere sunulmuştur.

Veri madenciliği için uygun yazılım; en çok özelliğe sahip olan yazılım değil, kullanıcı dostu, yapılacak çalışmanın özelliklerine en uygun ve yapılan çalışmada maksimum ölçüde doğruluk sağlayacak özellikleri barındıran programlardır (Dolgun, 2006: 50).

Bu kapsamda her geçen gün değişen ihtiyaç ve yöntemlerin uygulanmasını kolaylaştırmak adına yepyeni yazılımlar ortaya çıkmaktadır. Bu yazılımların bir kısmı açık kaynaklı kodlu olduğu gibi bir kısmı ise lisanslı ürünler olarak veri madenciliği uzmanlarının hizmetine sunulmaktadır.

Tablo 2. Veri Madenciliğinde Kullanılan Uygulamalar

Ticari	Açık Kaynak Kodlu
SPSS Modeler (Clementine)	Orange,
Excel	RapidMiner
SPSS	WEKA
SAS	R
Angoss	Keel
KXEN	Knime
MS SQL Server	Tanagra
MATLAB	Scriptella ETL
Oracle Modülleri	jHepWork

Kaynak: (Kaya ve Özel, 2014, s. 2)

Bu yazılımlar çok çeşitli olmakla birlikte en çok kullanılan yazılımlar aşağıda genel olarak anlatılmıştır:

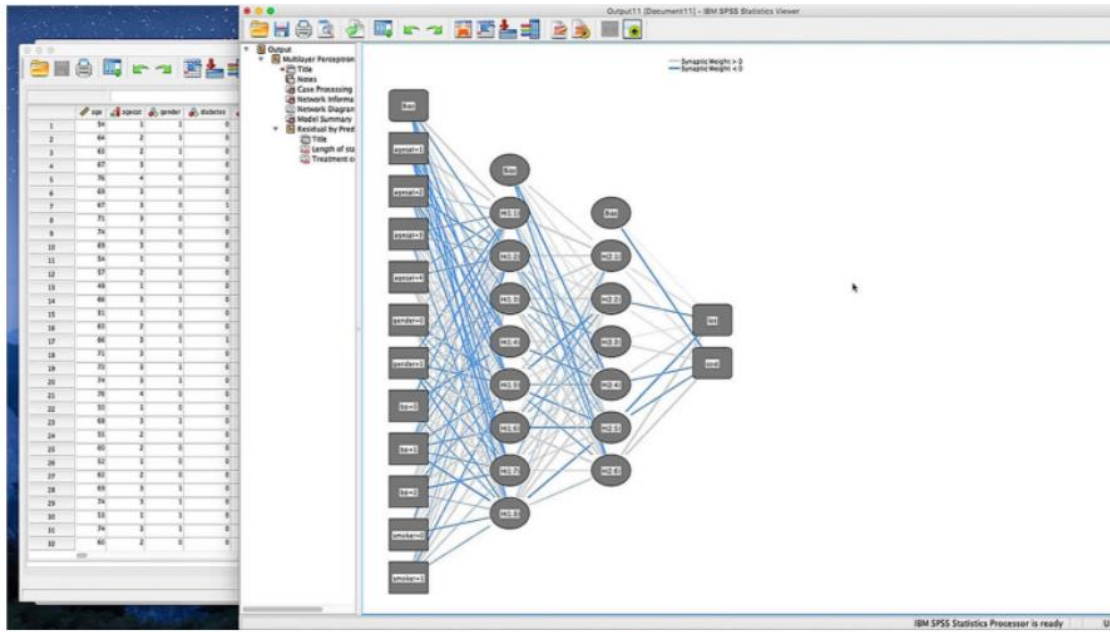
2.5.1 IBM SPSS Statistics

SPSS bilgisayar yazılımının ilk sürümü 1968 yılında Statistical Package for the Social Sciences kısaltması kullanılarak piyasaya çıkarılan istatistik analizlerin yapıldığı bir programdır. 2009 Mart ayından sonra firma Amerika Birleşik Devletleri şirketi olan

IBM'e satılmış geçiş döneminde 2009 - 2010 yılları arasında PASW Statistics adıyla anılıp Ağustos 2010 yılından itibaren IBM SPSS Statistics olarak ismi değiştirilmiştir³.

SPSS programın güncel versiyonu 29.0'dır. Bu program özellikle anket uygulamalarında kullanılmakta ve eğitim şirketleri başta olmak üzere pek çok kamu kuruluşunda çokça tercih edilen bir program olma özelliğine sahiptir. Bu yazılım sosyal ve sağlık bilimleri disiplinlerinde çokça kullanılmaktadır⁴.

Şekil 4. IBM SPSS Statics MLP Çok Katmanlı Algılayıcı Sinir Ağı Model Ekranı



Kaynak: (IBM SPSS Statics MLP Çok Katmanlı Algılayıcı Sinir Ağı Model Ekranı)⁵

2.5.2 IBM SPSS Modeller

IBM firması tarafından üretilen IBM SPSS Modeller programı, iş uzmanlığı temelli, tahmine dayalı modelleri hızla geliştirmemizi sağlayan ve karar alma işlem kalitesini geliştirmek için bunları iş operasyonlarına yerleştirmemizi sağlayan veri madenciliği araçları kümesi olarak tanımlanmaktadır. CRISP-DM modeli çevresinde tasarlanan IBM

³ İnternet Erişim: <https://tr.wikipedia.org/wiki/SPSS> [Erişim tarihi: 20.07.2023]

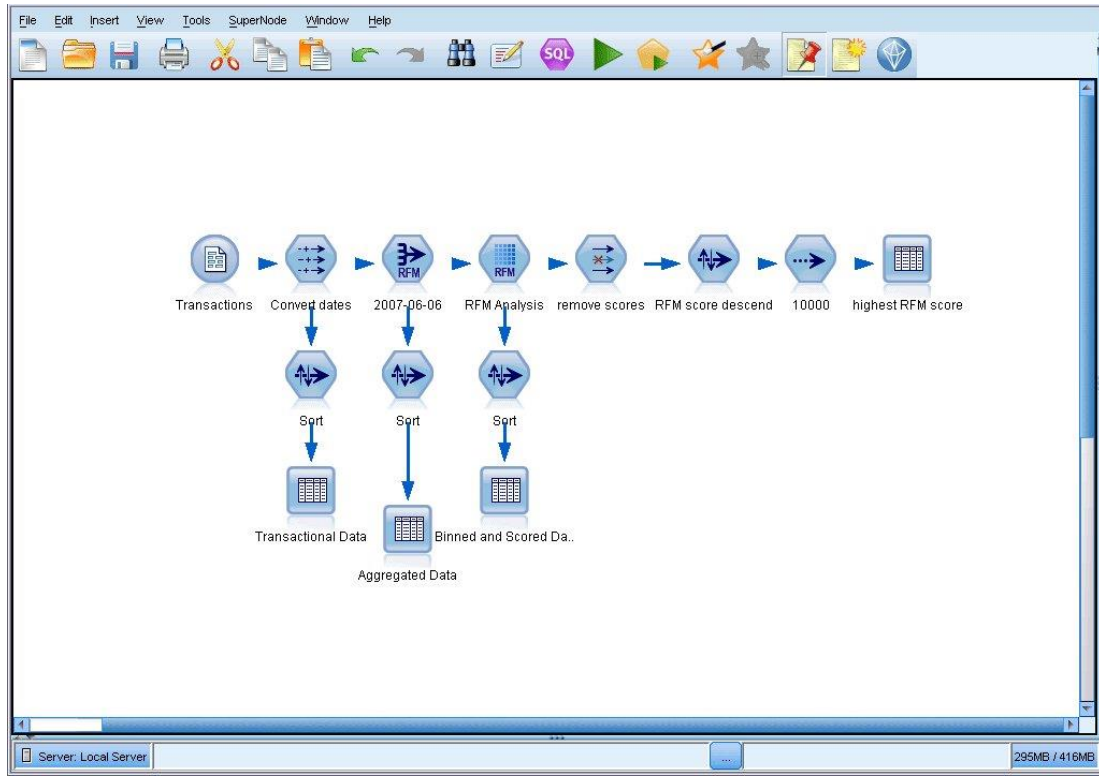
⁴ İnternet Erişim: https://community.ibm.com/community/user/ai-datascience/blogs/nitin-mathur1/2019/11/14/spss-statistics-vs-modeler?_gl=1*qfmqw5*_ga*MTMxNzgxODcwLjE2ODk4MDgyNDI.*_ga_FYECCCS21D*MTY4OTgwODI0Mi4xLjEuMTY4OTgwOTkwMi4wLjAuMA.. [Erişim tarihi: 20 Temmuz 2023]

⁵ İnternet Erişim: <https://www.ibm.com/products/spss-statistics> [Erişim tarihi: 20.07.2023]

SPSS Modeler, verilerin daha iyi iş sonuçları üretmesi amacı ile tüm veri madenciliği sürecini desteklemektedir⁶.

IBM SPSS Modeler, makine öğrenimi, yapay zekâ ve istatistik temelli çeşitli modelleme yöntemleri sunmaktadır. Modelleme kataloğunda bulunan yöntemler, verilerinizden yeni bilgiler türetmenize ve tahmine dayalı modeller geliştirmenize olanak sağlamaktadır. Belirli güçlü yönleri bulunması nedeniyle her bir yöntem, belirli sorun tipleri için daha uygundur.

Şekil 5. IBM SPSS Modeler En Üst Düzeyde Akış Tuval Ekranı



Kaynak: (IBM SPSS Modeler En Üst Düzeyde Akış Tuval Ekranı⁷)

2.5.3 Intelligent Miner

IBM şirketi tarafından üretilen Intelligent Miner isimli yazılım gelişmiş analitik ve modelleme tekniklerini kullanan ve veri madenciliği algoritmaları ile çalışan bir

⁶ İnternet Erişim: <https://www.ibm.com/docs/tr/spss-modeler/18.4.0?topic=guide-about-spss-modeler> [Erişim tarihi: 20.07.2023]

⁷ İnternet Erişim: <https://www.ibm.com/docs/tr/spss-modeler/18.4.0?topic=glance-customizing-spss-modeler-window> [Erişim tarihi: 20.07.2023]

yazılımdır. IBM firması tarafından geliştirilen DB2 veri tabanı IBM sistemleri olan işletmelerde ölçeklenebilirlik ve yüksek performans sağlamaktadır. Bu program ile istatistik becerileri gerektiren zaman serisi tahminleri otomatik olarak gerçekleştirilebilmektedir. Ambar, stok yönetimi gibi işletmelerde hayati öneme sahip süreçlerin etkin şekilde yönetimini sağlamak için zaman serisi tahminleri kullanılması programın güçlü yanlarını temsil etmektedir⁸.

2.5.4. Analysis Manager

Microsoft şirketinin geliştirdiği, karar ağacı ve kümeleme analizi algoritmalarını barındıran yazılımdır. Kullanıcı dostu ve görsel yönden güçlü olmasına rağmen her tür veri madenciliği algoritmasını desteklememesi eksik yönünü oluşturmaktadır (Can, 2017, s. 26). Günümüzde bu yazılım Microsoft Azure platformu ile entegre edilerek Azure Analysis Services ürünü olarak bulut hizmetlerinde veri madenciliği imkanını sunmaktadır⁹.

2.5.5 Keel¹⁰

Granada Üniversitesi tarafından İspanya Ulusal Bilim Projeleri Kurumu destek projesi ile geliştirilmiş bir programdır. Java alt yapısı kullanılan program, kümeleme başta olmak üzere sınıflandırma alanlarında kullanılan klasik veri madenciliği algoritmalarında yetersiz, veri görselleştirmede zayıf olan bir yazılımdır. Bu yazılım özellikle kural tabanlı kümelelerde bulanık sınıflandırıcılar ile yapay zekâ tabanlı sınıflandırma algoritmalarını kapsamaktadır (Bilgin, 2009, s. 808).

2.5.6 Knime¹¹

Görsel veri madenciliği araştırmacıları tarafından Konstanz Üniversitesi'nde geliştirilen bir programdır. Knime yazılım geliştirme kiti ile araştırmacılara kendi modüllerini oluşturma imkânı sağlamaktadır. Kurulum yapılmadan çalışabilen *.txt, *.arff, *.table

⁸ İnternet Erişim: <https://www.ibm.com/docs/en/db2/11.1?topic=api-data-mining-intelligent-miner> [Erişim tarihi: 20.07.2023]

⁹ İnternet Erişim: <https://azure.microsoft.com/tr-tr/products/analysis-services> [Erişim tarihi: 20.07.2023]

¹⁰ İnternet Erişim: <http://www.keel.es/> [Erişim tarihi: 21.07.2023]

¹¹ İnternet Erişim: <http://www.knime.com/> [Erişim tarihi: 21.07.2023]

formatı üzerinden veri alabilen zengin görselleştirme yeteneğine sahip olan bir yazılımdır (Bilgin, 2009, s. 810).

2.5.7 Orange¹²

C++ yazılım geliştirme dili ile yazılan Orange yazılımı Bilgisayar ve Enformatik Bilimleri Mühendisliği akademisyenleri tarafından Slovenya Ljubljana Üniversitesi'nde geliştirilmiştir. Bu program yapay zekâ araştırmaları ekibinin uzun uğraşları sonucu ortaya çıkarılmış ara yüz ve grafik ortamında Qt3 kütüphanesi kullanılarak yazılmıştır. Python'da geliştirilen yazılıma Orange adı verilmiştir (Bilgin, 2009, s.812).

2.5.8 R¹³

Auckland Üniversitesi araştırmacılarından olan istatistik bilimciler Ross Ihaka, Robert Gentleman'ın ortak çalışması neticesinde Unix makinalarda yaygın olarak kullanılan istatistiksel hesaplama, grafik ve veri analizleri üzerine geliştirilen bir yazılımdır (Tekerek, 2011, s. 167).

2.5.9 RapidMiner¹⁴

RapidMiner, 2001 yılında Dortmund Üniversitesi'ndeki bir veri bilimi projesinden doğmuştur. Yapay zekâyı yeniden keşfetme misyonu ile sürekli yenilik ve gelişmeyi hedefleyen Java dili kullanılarak geliştirilen bir yazılımdır¹⁵. Açık kaynak kod yazılımları arasında yer alan RapidMiner, 22 adet farklı dosya formatındaki veriyi işleyebilmektedir. Oracle, MS SQL Server, PostgreSQL, MySQL, JDBC, Sybase, Access, IBM DB2, Ingres gibi yaygın şekilde kullanılan veri tabanlarını destekleyen bir yazılımdır (Bilgin, 2009, s. 810). Elektronik tablolama yazılımları arasında en sık kullanılan Excel dosyalarından veri alabilmektedir. Yaygın birçok işletim sistemlerinde hatasız çalışan kullanıcı dostu bir yazılımdır. Açık kaynaklı kodlu yazılımlar arasında WEKA gibi kullanım kolaylığı ve geniş veri tabanı desteği ile dünya genelinde bilinen

¹² İnternet Erişim: <https://orangedatamining.com/> [Erişim tarihi: 21.07.2023]

¹³ İnternet Erişim: <https://www.r-project.org/> [Erişim tarihi: 21.07.2023]

¹⁴ İnternet Erişim: <https://rapidminer.com/> [Erişim tarihi: 21.07.2023]

¹⁵ İnternet Erişim: <https://rapidminer.com/about-us/> [Erişim tarihi: 21.07.2023]

ve veri madenciliği konusunda çalışmalar yapan RapidMiner¹⁶ script yazma imkânı ile de araştırmacılara esneklik sağlamaktadır.

RapidMiner veri madenciliği konusunda veri bilimi öğrenimini yaygınlaştırmak için RapidMiner Akademisi adı altında ortakları aracılığı ile canlı eğitimler verilmekte bu eğitimlerde örnek uygulamalar ve yöntemeler üzerinden bilgiler aktarılmaktadır. Bu eğitimler sonunda RapidMiner sertifikasyonları verilmekte ve bu sertifikasyonlar araştırmacıların veri madenciliği konusundaki etkinliklerini artırmaktadır¹⁷.

2.5.10 WEKA¹⁸

Programı Waikato Environment for Knowledge Analysis (WEKA) kelimesinin kısaltılması ile Yeni Zelanda da yaşayan ve uçamayan meraklı bir kuşun adı verilmiştir. Açık kaynak kodu esas alınarak Java teknolojilerinden faydalanılarak yazılan WEKA yazılımı Waikato Üniversitesi'nde çalışmalarını sürdüren araştırmacılar tarafından geliştirilmiştir. Yazılım veri madenciliği konusunda çeşitli araçları içermektedir. Bu araçlar veri hazırlama aşamasından başlayarak sınıflandırma, regresyon, vb. gibi konuları içermekte olup, ön işlenmiş veriler '.arff' uzantılı bir dosyadan okunarak veri madenciliği analizleri gerçekleştirilmektedir (Witten, 2005).

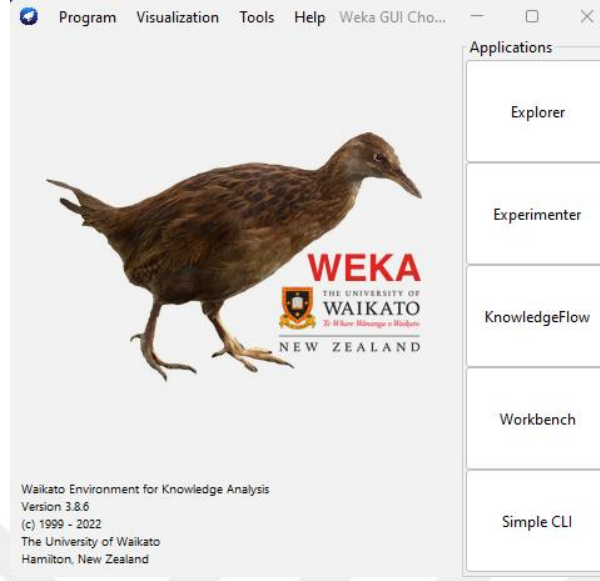
WEKA yazılımı, makine öğrenmesi, istatistik gibi alanların ilgi duyduğu çok çeşitli algortima ve kütüphaneleri barındırmaktadır.

¹⁶ İnternet Erişim: <https://rapidminer.com/partners/> [Erişim tarihi: 21.07.2023]

¹⁷ İnternet Erişim: <https://academy.rapidminer.com/> [Erişim tarihi: 21.07.2023]

¹⁸ İnternet Erişim: <https://www.cs.waikato.ac.nz/ml/weka/index.html> [Erişim tarihi: 22.07.2023]

Şekil 6. WEKA Program Ekranı



2.6 Veri Madenciliği Yöntemleri

IBM şirketinin 2014 yılı araştırmasına göre dünya üzerindeki verinin tahmini olarak %90'ı sadece son 2 yıl içerisinde üretilmiştir. Günümüzde 2,5 ekzabayt boyutunda verinin günlük üretimi söz konusudur. 2003 yılına kadar üretilen toplam veri miktarı 5 ekzabayt boyutundayken, 2014 yılında aynı miktardaki veri 2 gün içerisinde üretilmektedir (Sağiroğlu ve Sinanç, 2013, s. 42). Bu durum bir veri seline maruz kaldığımızı ve bu selin kontrol ve ıslah edilmemesi durumunda veri çöplüğüne dönüşen bir dijital dünyaya neden olacağı düşünülmektedir. Üretilen verinin doğru analiz edilememesi durumunda anlamsız bir biçimde dijital evrende başıboş gezen göktaşı misali var olacağı gerçeği düşünüldüğünde veri madenciliğinin önemi daha da iyi anlaşılmaktadır. 2023 yılına geldiğimizde ise dünya üzerinde oluşturulan % 90 verinin 2018-2023 yılları arasında oluşturulduğu ve gün başına 2,5 milyar GB boyutunda veri üretildiği bilinmektedir. Forbes'da 2022'de yayınlanan bir habere göre; 2023 yılı Büyük Veri Analitiği (Big Data Analytics) piyasası 103 milyar dolar civarında bir hacme sahip olacak¹⁹ ve bu büyük veriyi anlamlandıran veri madenciliğinde isabetli hamleler yapan firmalar pazar paylarını büyüterek dünya genelinde marka değerlerini arttıracaklardır.

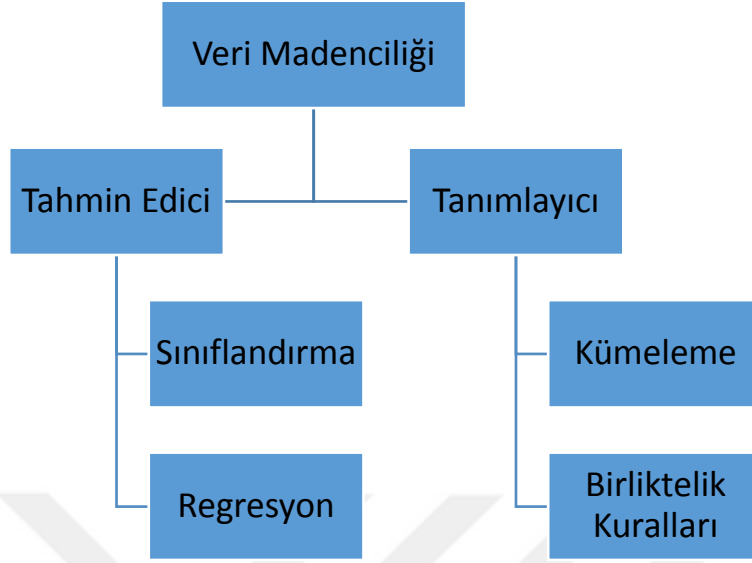
¹⁹ İnternet erişim: <https://www.mysoft.com.tr/buyuk-veri-big-data> [Erişim tarihi:21.07.2023]

Veri madenciliği sadece bir veri kümesine çeşitli metotlarla uygulanan istatistik, korelasyon ya da sıralama raporu olarak anlaşılmalıdır. Veri madenciliği geleneksel yöntemler kullanıldığında anlamlandırılmayan dolayısıyla değerlendirilemeyen verilerden yeni eğilim ve örüntülerin çıkarılması olarak anlamak daha doğru olmaktadır. Veri madenciliği tek bir yöntem ve kurala bağlı belirli bir kurallar manzumesi de değildir. Veri madenciliği uzmanının sezgilerine, işlem tekrar etme konusundaki sabrına ve esnek düşünme yeteneğine göre anlamlı sonuçlar verebilen bir tekniktir. Bu nedenle bazı durumlarda veri tabanı yönetiminin doğru seçilerek özenle çalışma yapılması durumunda örüntüye hızlıca ulaşılabilmektedir (Westphal ve Blaxton, 1998, s. 36).

Veri madenciliği kavramsal olarak büyük bir veri madeninden kazı denilen süreci yürüterek anlamlı verileri ayıklama ve ortaya çıkarma tekniği olarak düşünülebilmektedir. Veri madenciliği görüldüğü üzere iyi anlaşılması gereken doğru yöntemlerin uygulanması ile anlamlı bilgiye ulaşımın sağlanacağı farklı yöntemleri ve sezgisel çalışmayı gerektiren bir disiplinler bütünüdür. Dolayısıyla her veri madenciliğinin mihenk noktası veri setini temsil edecek bir model oluşturmaktan geçmektedir. Veri setinin doğru şekilde modellenememesi doğru analizlerin de elde edilememesine neden olacaktır. Modellere uygulanacak birçok yöntem bulunmakla beraber veri madenciliği konusu yapay zekâ ve istatistik temelli çok çeşitli algoritmalar barındırmaktadır (Durdu, 2012, s. 20).

Veri madenciliği Şekil 7’de de görüldüğü üzere tahmin edici (predictive) ve tanımlayıcı (descriptive) diye adlandırılan 2 ana başlıkta değerlendirmektedir (Zhong ve Zhou, 2003). Veri madenciliği modelleri kümeleme, regresyon, sınıflandırma ve birliktelik kuralları olmak üzere dört başlıkta incelenmektedir (Akpınar, 2000). Sınıflandırma ve Regresyon modelleri tahmin edici, Kümeleme ve Birliktelik Kuralları modelleri ise tanımlayıcı modellerdir (Akpınar, 2000).

Şekil 7. Veri Madenciliği Yöntemleri Şeması



Tahmin edici diye adlandırılan modellerde, sonuçları hakkında daha önceden bilgi sahibi olunan veriler kullanılarak çeşitli modeller geliştirilmektedir. Daha sonra geliştirilen modelden faydalanılarak neticesi bilinmeyen veriler için neticelerin tahmin edilmesi hedeflenmektedir. Örneğin, geçmişte satış verileri bilinen bir ürünün geçmişteki sonuçları değerlendirilerek gelecekte gerçekleşmesi muhtemel satışları tahmin edilebilmektedir. Tanımlayıcı modellerde ise eldeki veriler ile karar vermeye öncülük edecek veri düzenlerinin tanımlanması esas alınmaktadır.

2.6.1 Veri Ön İşleme

Kullanılacak verinin analiz edilmeden önce çeşitli süreçler aracılığıyla uygunlaştırılmasına veri ön işleme denilmekte olup bu işlem yapılacak analizin daha doğru ve güvenilir olması ve sonuçların kalitesi için oldukça büyük önem arz etmektedir. Doğru uygulanan veri ön işleme aşaması ile yürütülen veri temizliği, sonuçların doğruluğuna olumlu yönde etki etmektedir (Tüzüntürk, 2010, s. 66). Veri ön işleme ile makine öğrenmesi için kullanılacak ham verinin uygunlaştırma süreci başlatılarak hatalı, uygun olmayan ya da eksik verilerin temizlenmesi işlemi gerçekleştirilmektedir. Bu aşamada veri setinde bulunan eksik veriler (missing value) uygun değerler ile doldurulmakta (fitting) ya da bir bir sınıfa ya da niteliğe ait eksik veri çok fazla ise bu durumda veriler silinerek (cleaning) veri ön işleme işlemi tamamlanmış olmaktadır. Bu süreçler için açık kaynak kodlu ya da lisanslı birçok program kullanılmaktadır. WEKA, RapidMiner Studio,

Orange, ELKI, Jhep Work, KNIME yazılımları sıklıkla bu amaçla kullanılan yazılımlar arasındadır (Tekerek, 2011, s. 163). Modelleme konusundaki çalışmalarda genellikle verilerin ham halinin tahmin modeline uygulanması aşamasında önce veri ön işleme aşamasının uygulandığı bu işlemten sonra ise veri madenciliği programları vasıtasıyla ilgili örüntülerin oluşturulduğu görülmektedir (Qui, vd., 2018, 1342). Yapılan çalışmalarda veri ön işlemenin önemi yadsınamaz ve bilgi keşfi sürecinin %50 ile %80' ini etkileyebileceği belirtilmiş olup doğru veri işlemedeki önemine değinilmiştir (Akpınar, 2000, s. 14).

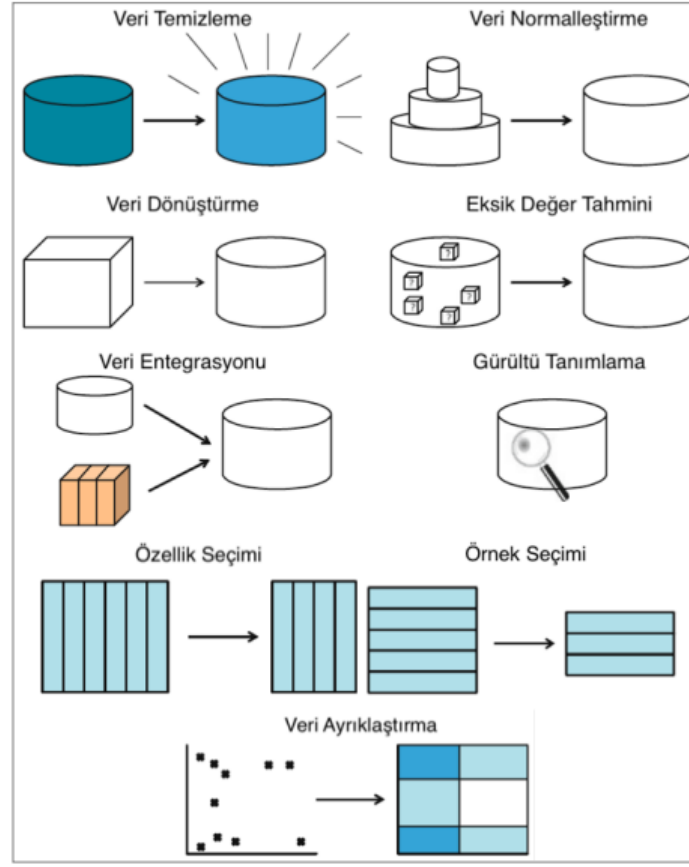
Tablo 3: Veri Kalitesini Etkileyen Boyutlar

Boyutlar	Tanımlamalar
Anlaşılabilirlik	Verinin rahat bir şekilde anlaşılabilmesi
Değer katabilirlik,	Veri kullanıldığında fayda sağlayabilmesi, katkı düzeyi
Doğruluk	Verinin doğru ve güvenilir olması
Erişilebilirlik	Verinin kolaylıkla elde edilebilmesi, erişilebilmesi
Güncellik	Verinin azami miktarda güncel olması
Güvenlik	Veriye harici erişimin uygun şekilde kısıtlanmış olması ve veri güvenliği
İnanılrlık	Verinin doğruluğunun genel kabul görmesi
Nesnellik	Verinin tarafsız olması
Özlu Temsil	Verinin toplu ve düzgün gösterilmesi
Saygınlık	Veri içeriğinin ve kaynağının saygın olması
Tamlık	Verinin tam, eksiksiz olması
Tutarlılık	Verinin tek formatta hazırlanmış olması
Uygunluk	Verinin ilgili işe uygun olması
Veri miktarının uygunluğu	Verinin yeterli miktarda olması
Yorumlanabilirlik	Verinin uygun dilde olması ve tanımların açık olması

Kaynak: (Pipino, vd, 2002)

Veri ön işleme aşamasında veri temizleme, normalleştirme, dönüştürme, eksik değer tahmini, veri entegrasyonu, gürültü tanımlama, veri indirgeme, ayrıklaştırma, özellik seçimi, örnek seçimi gibi adımlar bulunmakta olup bunlar ön işleme sürecinin etkin şekilde yürütülüp doğru analiz ve örüntüler elde edebilmek için oldukça önem arz etmektedir. Aşağıda yer alan Şekil 8'de veri ön işleme yöntemleri görsel olarak ifade edilmektedir (García, vd., 2015).

Şekil 8. Veri Ön İşleme Yöntemleri Şeması



Kaynak: (Garcia, vd., 2015)

2.6.2 Tahmin Edici Veri Madenciliği Yöntemleri

Tahmin edici modeller adından da anlaşılacağı üzere sonuçları bilinen verilerden sonuçları bilinmeyen verilerin belirlenmesine yönelik modeller geliştirilmesini ifade etmektedir. Bu modeller sayesinde örüntüler, bilinmeyen eğilimler ve bağlantılar bulunmaya çalışılmaktadır (Han, 2012, 328).

2.6.2.1 Sınıflandırma

Tahmin edici yöntemlerden sınıflandırma işlemi, nesnenin niteliklerinden faydalanarak, o nesnenin daha önce tanımlanmış sınıflardan hangilerine sahip olduğu sonucuna ulaşılması gereken hallerde kullanılmaktadır (Zhang, 2000). Sınıflandırma işleminin nitelikleri ve hangi sınıfa ait olduğu bilinen veri elemanları baz alınarak nitelikleri bilinen fakat ne tür bir sınıfa ait olduğu belli olmayan veri elemanlarının sınıf değerleri

tahmin yolu ile bulunmaya çalışılmaktadır. Bu işlem için modelin geliştirilme sürecinde veri kümesindeki eğitim verisi olarak belirlenip ayrılmış ve bu sınıf değerleri ile özellikleri bilinen veri elemanlarının eğitilmesiyle süreç başlatılmaktadır.

Sınıflandırma ve regresyon yöntemleri, verileri sınıflandırarak önemi saptanan ya da sonuç verileri tahmin edilen modellerin kurulduğu veri analizi türüdür (Han & Kamber, 2000). Sınıflandırma ile kategorik veriler tahmin edilirken, regresyon sürekli verilerin tahmin edilmesi durumunda uygulanmaktadır (Han & Kamber, 2000). Sınıflandırma ve regresyon modelleri incelendiğinde yapay sinir ağları, genetik algoritmalar, karar ağaçları, naïve-bayes, K-en yakın komşu vb. diye adlandırılan çeşitli yöntemler uygulanmaktadır.

Sınıflandırma iki adımdan oluşmaktadır. Öncelikle belirlenen veri sınıflarına göre model geliştirilmekte ve bu geliştirilen model sınıflandırma için kullanılmaktadır. Bu veri kümeleri önceden belirlenip tanımlanmış belirli bir sınıf içeriğine göre etiketlenmektedir. Bu anlamda ikiye ayrılan veri kümelerinde *eğitim verisi* ve *test verisi* kümeleri belirlenerek çalışmaya başlanmaktadır. Eğitim veri kümesi temel alınarak test veri kümesindeki verilerin hangi sınıfa ait olduğu bilgisine ulaşılmaktadır. Sınıflandırma yöntemleri kullanılırken aşağıda belirtilen karşılaştırma kriterleri önerilmektedir [Han, 2001]:

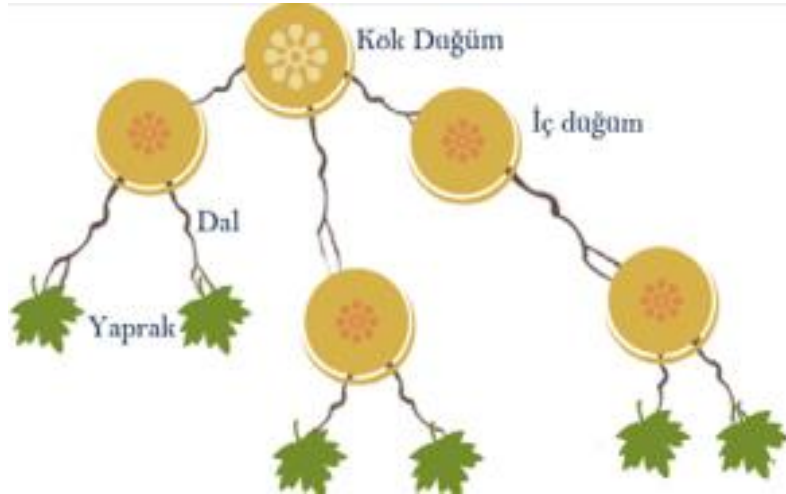
- Tahmin doğruluğu (Predictive Accuracy): Verinin yeni olması veya önceden bilinmeyen verilerin sınıf etiketlemesi, modelin doğru tahmin edilebilme kabiliyedir.
- Hız (speed): Model geliştirilirken oluşan hesaplama maliyetine verilen addır.
- Sağlamlık (Robustness): Kayıp veya gürültülü veri olması durumunda dahi modelin tahmini doğru yapma kabiliyetidir.
- Ölçeklendirilebilirlik (Scalability): Model hacmi büyük verilerden oluşması durumunda, etkin çalışma kabiliyetine sahip olmaktadır.
- Yorumlanabilirlik (Interpretability): Modelin sağladığı anlaşılma seviyesine verilen addır.

2.6.2.1.1 Karar Ağaçları

Sınıflandırma modelleri arasında en çok tercih edilenlerden birisi karar ağaçlarıdır. Veri tabanına rahat entegre edilebilmesi, ucuzluğu, kolay yorumlanabilirliği ve yüksek güvenilirliği gibi avantajlarının olması tercih edilme sebebinin açıklanmaktadır (Gürsoy, 2009). Temelde karar ağaçları veriyi farklı kümeler ayırmaktadır. Karar ağaçlarında dalların her birinin bir olasılığı bulunmaktadır. Bütün olasılıklar dallardan köke doğru 20'ye yakın hesaplanabilir ve uygun olan kural oluşturur ve bu kural denetlenebilir yapıdadır (Silahtaroglu, 2008, s.105).

Karar ağacı kavramını; "Olası tüm eylemlerin yönlerini, eylemlerin yönlerine etkisi olabilecek tüm olası faktörleri ve tüm bu faktörlere dayanan her bir olası sonucu, verilere bağlı olarak değerlendiren, çizgi, kare, daire gibi geometrik semboller kullanımı yoluyla karar vericiye sorunu anlamada kolaylık sağlayan düzenleme" olarak da tanımlamak mümkündür. Bu tanımlamaya göre karar ağaçlarının çeşitli eylem seçeneklerini, farklı olası etkenlerin ve eylemlerin sonuçlarını içerdiğini söyleyebiliriz. (Gürsoy, 2009)

Şekil 8. Karar Ağacı Modeli Şeması



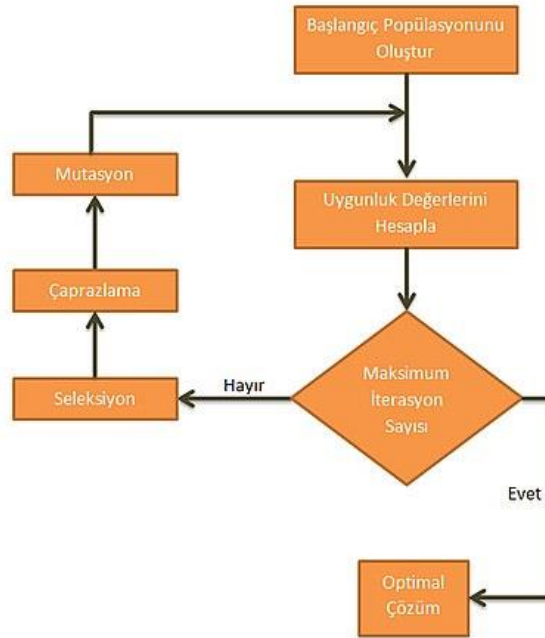
Kaynak: (Kavzaoglu, vd., 2012, s. 36)

2.6.1.1.2 Genetik Algoritmalar

Genetik Algoritmalar, günümüz algoritma yöntemleri arasında yapay zekânın sürekli gelişen bir türü olan evrimsel hesaplama tekniğinin alt başlıkları arasında yer almaktadır. “Evrimsel kuramını ortaya koyan Darwin “doğada en sağlam ve en iyinin yaşaması“ prensibinden yola çıkarak oluşturduğu kuramda bir veri kümesinden özel bir veriye ulaşmak için oluşturulan metasezgisel arama felsefesini ortaya koymuştur (Gürsoy, 2009).

Genetik algoritma, genetik mühendisliği ile biyoloji alanında yapılan çalışmalar neticesinde ortaya konan teknik ile canlıların yapılarında var olan birtakım özelliklerin sanal ortamda taklit edilmesi ile geliştirilmiş bir tekniktir. Geleneksel yöntemlerle problemin çözümünün zor olduğu durumlarda tercih edilmektedir. Otomatik programlama, öğrenme kabiliyeti olan makineler, planlama, üretim hattı tasarımı, ekonomi, çevre bilimi gibi alanlarda da uygulanan genetik algoritma veri madenciliği uygulamalarında başarılı bir yöntem olarak kullanılmaktadır (Esen,2009).

Şekil 9. Genetik Algoritma Genel Akış Şeması



Kaynak: (Wikipedia)²⁰

²⁰ İnternet erişim: https://tr.wikipedia.org/wiki/Genetik_algoritma [Erişim tarihi:22.07.2023]

2.6.1.1.3 Naïve-bayes

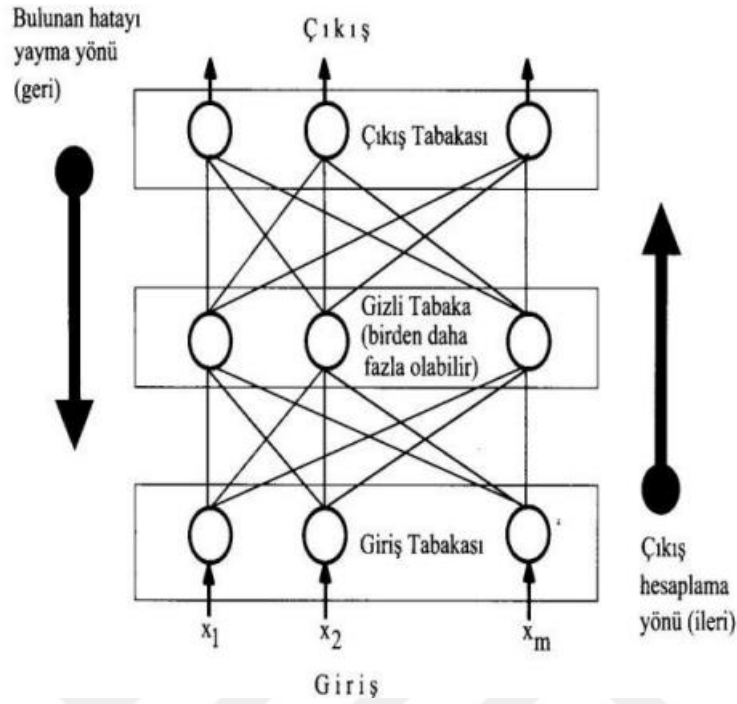
İstatistikteki Bayes teoremine dayanarak oluşturulmuş bir yöntemdir. Hali hazırda kullanılan belirlenmiş sınıflara ait verilerin gerçekleşme ihtimali üzerine öngörüler oluşturulmaktadır. Bu teoreme herhangi bir durumun belirsizlik taşıması durumunda modeli oluşturularak, ilgili durumla evrensel doğrular ve gerçekçi gözlemler incelenerek belli sonuçlar oluşturulmasına imkân sağlamaktadır. Özellikle belirsizlik olan durumlarda karar verme aşamasında çokça kullanılmaktadır (Altıntaş,2010). Aynı zamanda Naive Bayes istatistiksel yöntemler kullanılarak sınıflandırma yapan uygulama aşamasında niteliklerin birbirinden etkilenmemesine dayalı bir algoritmadır. Belirlenen nitelikler birbirinden bağımsız olmaması durumunda hesaplamalar doğru yapılamamaktadır. Bu yüzden verilerin hepsi aynı öncelikte olmalı, biri diğerinden daha önemli veya önemsiz olmamalıdır. Bu yöntemde frekans ağırlıklandırma ile bit ağırlıklandırma yöntemleri kullanılmaktadır (Poyraz, 2012, s.40).

2.6.1.1.4 Yapay Sinir Ağları

Yapay sinir ağları, sırayla bir şeyi öğrenme, hatırlama ve öğrenilen şeyi genelleme yaparak çevresinde topladığı yeni verilerden veri üretme gibi temellere dayanan, insan insan beynindeki öğrenme işlemlerinin matematiksel şekilde modellenmesi sonucunda geliştirilmiştir (Kabalcı, 2014, s. 45). Yapay Sinir Ağları teknikleri tahmin yapmak, veriyi sınıflandırmak, veriyi ilişkilendirmek, veriyi filtrelemek ve veriyi yorumlamak için uygulanmaktadır (Ağyar, 2015, s. 22).

Biyolojik sinir ağlarımızda hücrelerimizin bir diğer hücreyi tetiklemesi, sinirsel aktarımı devam ettirmesi için bir eşik değer kadar uyarılması gerekmektedir. Yapay sinir hücresinin uyarımı da aynı model örnek alınarak gerçekleştirilmektedir. Şekil 10'da da görülen bir yapay sinir ağı modeli, n adet katman (layer) ve her katmanda değişik sayılarda olabilen hesaplama elemanlarından meydana gelmektedir. Yapay sinir ağı modelinde bulunan hesaplama elemanları ise, yapay sinir hücresi (artificial neuron), düğüm (node), birim (unit) ya da işlem elemanı (processing element) olarak adlandırılabilir (Akpınar, 2014).

Şekil 10. Çok katmanlı Yapay Sinir Ağları Şeması



Kaynak: (Sağiroğlu, vd., 2003)

2.6.1.1.5 K-en En Yakın Komşu

Modelde yeni bir olayın gerçekleşmesi durumunda, tüm veriyi inceleyen algoritma bu duruma en çok benzeyen olayların bir altkümelerini oluşturduğu ve bu oluşturduğu bu alt kümeleri çıktığı tahmin etmek için kullandığı benzerlik ile öğrenme prensibine dayanan bir yöntemdir (Aydoğan, 2008, s. 21). K en yakın komşu yöntemi, verileri sınıflandırmak için kullanılan yöntemler arasında yer almakta sınıflandırma yapılırken, öğrenme kümesindeki verilere benzeyen en yakın k adet verinin ortalaması alınarak eşik değere göre sınıflandırma yapılmaktadır (Shah ve Kusiak, 2004, s. 187).

2.6.1.1.6 Regresyon

Tahmin edici modellerden birisi de regresyondur. Bu yöntemle sonucu bilinenden yola çıkılarak sonucu bilinmeyen değerlerin çıkarımı yapılabilmektedir. Regresyon yönteminde değerlerde süreklilik aranmakta ve bu süreklilik üzerine tahminleme işlemi yapılmaktadır. Modelde girdi ile çıktı arasında bağlantı olup olmadığı belirlenerek tahminlerde bulunmaktadır. Regresyon analizine baktığımızda girdiler “bağımsız

değişken” ve “bağımlı değişken” olarak iki tanım ortaya çıkmaktadır. Bağımlı değişken sonucu ifade etmektedir. Girdi değişkeninin sayısı birden çok olabilir ve önemli olan bu girdilerin çözümlemede katkı sağlayıp sağlamadığı ve katkısı az olan girdilerin modelden çıkarılmasıdır (Silahtaroglu, 2013, s. 304).

Regresyon yöntemi çoğunlukla değişkenler arasında bulunan belirli ilişkileri incelemeyi hedeflemektedir (Parvez, 2015, s. 120). Kullanılan bağımlı değişken sürekli değişken yapıda olduğunda doğrusal regresyon modeli uygulanmaktadır ve yer alacak bağımsız değişkenler hem kesikli hem de sürekli değişken yapısında olabilmektedir. Bağımlı değişkenin ikili ya da kategorik değişken olması durumunda doğrusal regresyon uygulanamaz ve bu durumda diğer regresyon modelleri uygulanabilmektedir. Uygulanacak modeller arasında en çok tercih edilen model ise lojistik regresyon modelidir. Bu model ile doğrusal regresyon modeli arasındaki en büyük fark, bağımlı değişkenin ikili veya kategorik yapıda olmasıdır (Günaydın, 2022, s. 145).

2.6.2 Tanımlayıcı Veri Madenciliği Yöntemleri

Tanımlayıcı model ile yapılan veri madenciliği uygulamalarında verilerin arasındaki gizli bilgiler tespit edilmeye çalışılmaktadır. Tanımlayıcı modellerin tahmin modellerinden temel farkı mevcut durum analizinin daha doğru yapılmasını sağlamayı hedeflemesidir (Arslan, 2020, s. 33).

2.6.2.1 Kümeleme

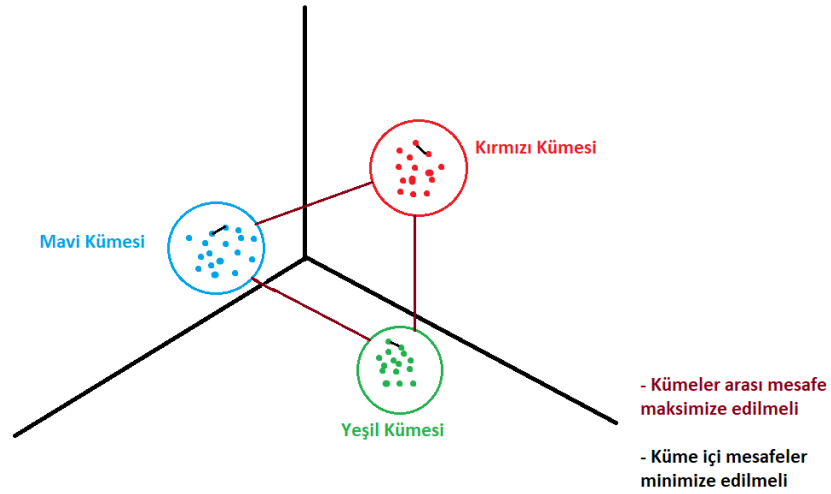
Veri tabanında bulunan verilerin sahip oldukları benzer özelliklere göre anlamlı ve kullanışlı gruplara ayrılmasını sağlayan modelin adıdır. Günümüzde veri akış hızının çok yoğun olmasından dolayı işletmeler verileri doğru anlamlandırabilmek için gruplama ihtiyacı duymaktadır. Bu nedenle veriler kümeleme yöntemleri ile gruplara ayrılmaktadır. Verilerin küme sayısı, ne tür kümelerden oluşacağı bilgisi verilerin birbirleri ile olan benzerlik durumlarına göre saptanmaktadır (Han, 2012, s. 445).

Kümeleme yönteminde tahmin yönteminde olduğu gibi hedef seçilmemekte, var olan verilerden homojen gruplar elde etmeye çalışılmaktadır. Bu açıdan değerlendirildiğinde kümelemenin denetimsiz öğrenmeye dayalı bir yöntem olduğu görülmektedir (Han,

2012, s. 446). Kümeleme yönteminde grupların oluşturulması aşamasında uzaklık belirleme tekniklerinden yararlanılmaktadır.

Kümeler, verilerin birbirleriyle olan benzerlik ve uzaklık özelliklerine bağlı olarak oluştuğundan, küme içi benzerliğin en yüksek olması ve kümeler arası benzerliğin ise en düşük seviyede tutulması hedeflenmektedir (Han, 2012, s. 448).

Şekil 11. Kümeleme Şeması



Kümeleme yöntemleri verinin tipi, amacı ve oluşturulma şekline göre ikiye ayrılmaktadır: Bunlar hiyerarşik ve hiyerarşik olmayan yöntemlerdir. Kümeleme yöntemleri genel olarak şu şekilde sıralanmaktadır (Bolat, 2019, s. 23):

- Hiyerarşik yöntemler
- Bölme yöntemleri
- Yoğunluk tabanlı yöntemler
- Izgara tabanlı yöntemler
- Model tabanlı yöntemler

2.6.2.2 Birliktelik Kuralı

Birliktelik kuralı yöntemi, veri kümelerindeki kayıtlar arasındaki bağlantıları çözmeye çalışan bir veri madenciliği tekniğidir. 1993 yılında ilk defa kullanılan bu yöntem, veri nitelikleri arasındaki kombinasyonları çıkararak niteliklerin bütün değerlerinin bütün kombinasyonlarda denenerek örüntülerin bulunmaya çalışıldığı bir tekniktir (Agrawal,

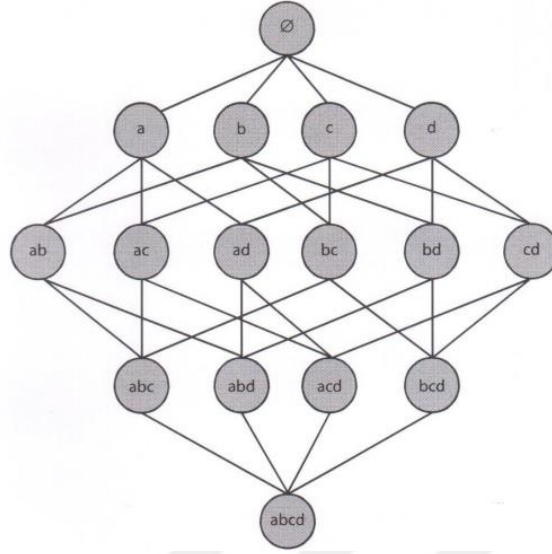
vd., 1993, s. 310). Birliktelik kuralını anlayabilmek için bir pazar sepeti düşünülebilir. Bu sepet hedef markete giden müşterilerin satın aldıkları ürünler arasındaki birliktelik ilişkisini ortaya koymak ve müşterilerin satın alma alışkanlıklarını belirlemeye çalışmak amacıyla kullanılmaktadır. Örneğin X ürününü alan müşterilerin çoğunlukla Y ve Z ürünlerini de satın aldıklarının tespit edilmesi durumunda bireylerin gelecekteki satın alma alışkanlıkları konusunda yönlendirme yapılabileceği görülmektedir. Örneğin hamile kadınların çocuk kıyafetleri, sağlıklı beslenme, egzersiz ve organik gıda ürünlerinin bulunduğu reyonlara ilgi duymalarının bu yöntemle ortaya çıkarılması sağlanabilmektedir.

Gelişen ve değişen dijital dünya şirketlerin bilişim altyapılarında topladıkları ve depoladıkları verilerin oldukça büyümesine yol açmıştır. Bu verilerin anlamlı hale getirilmesi amacı ile aralarındaki bağı ve birlikteliği ortaya çıkarmak bir başka deyişle birliktelik ilişkilerini keşfetmek, şirket yöneticilerinin karar alma süreçlerinde doğru ve verimli hamleler yapmalarını sağlamaktadır. Bu yöntem, birlikte hareket eden verilerin davranış biçimlerini anlamak ve çıkan bu birliktelik sonuçlarını değerlendirerek gelecekte bu tarz işlemlerin çıkarımını yapmak olarak da tanımlanabilmektedir (Sözen, vd., 2017, s. 680).

Ardışık analiz ise birbiri ile bağlantısı olan fakat birbirini izleyen dönemlerde gerçekleşen bağlantıların anlamlandırılmasında kullanılmaktadır (Akpınar, 2000). Bu yöntem bazen farklı bir veri madenciliği operasyonu olarak da değerlendirilebilmektedir.

Sağlık alanında sağlık yönetimi kapsamında verilen hizmetlerin niteliklerinde bir arada olma şartlarını belirlemek amacıyla birliktelik kuralı yönteminin kullanıldığı çeşitli uygulamalar görülmektedir (Bekki, 2016, s. 110). Önerilen çalışma kapsamında sağlık yönetimi alanında çok yeni olan veri madenciliği uygulamaları arasında yer alan birliktelik kuralı yöntemiyle hasta verileri üzerinde birliktelik analizi çalışması yapılarak veri ambarındaki verilerin belirli seviyede anlamlı hale getirilmesi planlanmıştır.

Şekil 12. Dört nesne içeren bir kümenin farklı alt nesne oluşumu



Kaynak: (Bekki, 2016, s. 120)

2.6.2.3 Carma Algoritması

Büyük veri kümesi hesaplamasını çevrimiçi hale getirmek için, Carma (Sürekli Birlikte Kuralı Madenciliği Algoritması) adlı bir algoritma kullanılmaktadır. Bu algoritma, tüm büyük veri kümelerini üretmek için yaptığı işlem adımlarını iki kez yürütmektedir. İlk taramada, algoritma sürekli olarak tüm potansiyel büyük veri kümelerinden kafes oluşturmakta ve her set için deterministik alt ve üst sınır belirlemektedir. Bu işlemlerden sonra ortaya çıkan ilişkilendirme kuralları, her kuralın destek ve güven sınırlılıkları ile kullanıcıya verilmektedir. Destek ve güven eşikleri kullanıcı tarafından ayarlanabilmektedir. Üretilen kurallar ve sınırlardan memnun olduğunda araştırmacı kural silsilesindeki madenciliği erkenden sonlandırabilmektedir. Bunun yanında Carma'nın ilk tarama algoritması kullanılarak bir işlem akışı sürekli olarak işleyebilir ve yeniden taramaya gerek duyulmadan sonuçtaki ilişkilendirme kurallarını çevrimiçi oluşturabilmektedir (Hidber, 1999, s. 145).

2.6.2.4 GRI Algoritması

GRI Smyth ve Goodman tarafından bir kural veya hipotezin bilgi içeriğini ölçmek için 1992 yılında geliştirilmiş bir algoritmadır (Kokoç, vd., 2016, s. 95). GRI Algoritmasının niceliksel bir ölçüsü ile sınırları bulunmaktadır ve elde edilen kurallar 2 sınıfa ayrılmaktadır. İlginç ve ilginç olmayan olarak sınıflandırması yapılan kurallardan ilginç

olmayan kurallar ortadan kaldırılmaktadır. Bu eleme neticesinde yüksek kaliteli ve düşük nicelikli kurallara ulaşılmaktadır (Özceyhan, vd., 2012, s.265).

2.6.2.5 Eclat Algortiması

Sık öge kümeleri madenciliği konusunda birçok algoritma olsa da bu konuda yapılan çalışmalarda belirsiz veritabanları üzerinden sık öge kümelerinin madenciliğindeki bazı yeni bulgulara dayanan yeni algoritmalar önerildi. Bu algoritmalar temel olarak, Apriori, FP-growth ve Eclat olmak üzere üç farklı algoritmanın varyantları ya da uzantıları olarak sınıflandırılabilir (Yu ve Wang, 2014, s. 2116).

Bu bağlamda öge kümeleme madenciliği konusunda Eclat algoritması, denklik sınıfı veri kümeleme ve aşağıdan yukarıya kafes geçişlerinin özetlenmesidir. Bu algoritma sayesinde tüm farklı atom çiftlerinin tid listelerinin kesişmesi ve ortaya çıkan tid listesinin kategorik değişkenimizdeki farklı etiket sayısının kontrol edilmesiyle sık öğeler oluşturulmaktadır. Geçerli seviyedeki öge kümeleri sık olduğunda ise tekrarlı bir prosedür çağrılmaktadır. Bu işlem, tüm sık öge kümelerinin numaralandırılması yapılarına kadar tekrar edilmektedir (Agrawal, vd., 1994, s. 490).

2.6.2.6 FP-Growth Algoritması

Sık kullanılan öge veri tabanlarında yaygın bir şekilde kullanılan diğer algoritma FP-Growth algoritmasıdır. Bu algoritma ile hedeflenen şey yararlı öge kümesi ilişkilendirme bilgilerini tutmak ve bir sık model ağacına sıkıştırmaktır. Bu sıkıştırılan veri tabanı bir dizi koşullu veritabanına bölünür ve her bir başka sık örüntü parçasıyla ilişkilendirilerek ayrı ayrı incelenir. FP-Growth ağacının iyileştirmeleri, koşullu birçok FP-Ağacının oluşturulması aşamasındaki yükün azaltılmasını amaçlar çünkü modeller uzadıkça ya da destek seviyesi düştükçe FP-Growth algoritmasının performansı da azalacaktır (Han, vd., 2000, s. 6).

2.6.2.7 Apriori Algoritması

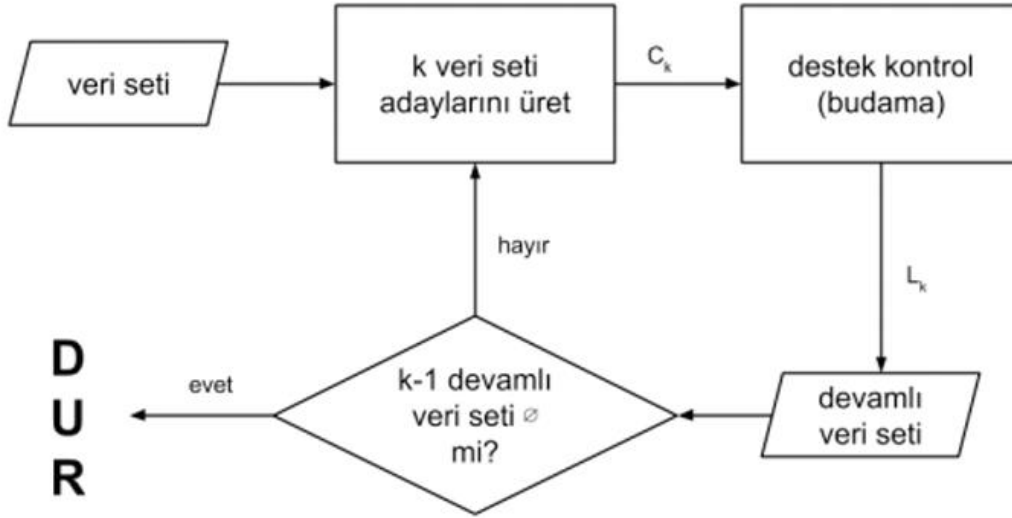
Apriori algoritması, ilk önce küçük kümelere bakma ve tüm alt kümeleri sık olamayacak büyük kümeleri tanıma yolunda en büyük aday kümeleri eleyerek çalışmaktadır. Temel Apriori algoritmasındaki çeşitli iyileştirmeler, veri tabanı

üzerindeki tarama sayısını azaltmaya ve aday k-öge kümelerinin boyutunu azaltmaya odaklanmaktadır (Zaki, 2000, s. 376).

R. Agrawal ve arkadaşları tarafından 1993 yılında önerilen Apriori algoritması, tek boyutlu, tek katmanlı ve Boolean birliktelik kuralları madenciliği için uygun bir algoritmadır (Agrawal, 1993, s. 216). Klasik algoritmalar arasında yer alan Apriori algoritması veri madenciliğinde sıkça kullanılmaktadır. Birliktelik kuralı uygulanan verilerde en popüler metodlardan biri olan bu algortima ilişki kurallarını tespit etmek için kullanılmaktadır. Amacı geniş nesne kümelerinin ilişki analiziyle tespit edilerek ortaya çıkarılmasıdır. Apriori algoritması daha çok müşteri bilgileri, network üzerindeki loglar, ayrıntılı bilgilerin olduğu veri tabanları üzerinde çalışmak üzere tasarlanmıştır. Örnek olarak, bir müşterinin alışveriş alışkanlıkları arasındaki ilişkiyi gösterebilmektedir. Algoritmanın temel ana fikri, özyinelemeli olarak katman katman alt kümeler üzerinde arama yönteminin kullanılmasıdır.

Geniş nesne kümelerini keşfetmek amacıyla ilk olarak her bir nesnenin destek oranı üzerinden matematiksel bir işlem yapılarak hesaplanmakta ve daha önce belirlenen destek oranı ile mukayese edilmektedir. Destek oranı hesaplanan kümeye aday nesne kümesi denilmektedir. Bunun yanında destek oranını aşan nesne kümeye de geniş nesne kümesi adı verilmektedir. Bu geniş nesne kümelerini ortaya çıkarmak ve destek seviyesinin altında kalan nesne kümelerini bir sonraki adımda değerlendirmeyerek en geniş nesne kümesini buluncaya kadar tüm veriler içerisinde tarama yapmak, Apriori algoritmasının çalışma prensibidir.

Şekil 13. Apriori Algoritması Akış Diyagramı



Apriori algoritmasının temel adımları aşağıda sıralanmıştır.

- Veri seti üzerinde minimum destek eşikini belirle (Öğe seti içerisinde sık olması gereken minimum sıklık)
- Veri seti üzerinde sıklık öğelerini tek tek belirle (Veri setinde her bir öğenin oluşumunu saydır).
- Veri setinden iki dizi boyutunda aday öğe kümelerini oluştur (İlişkili olabilecek veri öğe çiftlerini oluştur.)
- Veri setinde sık olmayan öğe kümelerini çıkar (Destek eşik seviyelerini karşılamayan öğe kümelerini çıkar)
- Daha büyük boyutlarda öğe kümeleri oluştur (3, 4 ve benzeri boyutlardaki sık öğe kümelerini birleştir).
- Budama işlemini tekrarlayın (Destek eşik seviyelerini karşılamayan öğe kümelerini elemeye devam et)
- Daha fazla sık öğe kümesi oluşturulamayana kadar yinele.
- Aralarındaki ilişkiyi ifade eden birliktelik kurallarını oluştur (Bu kuralların gücünü ve önemini değerlendirmek için ölçütleri hesapla)

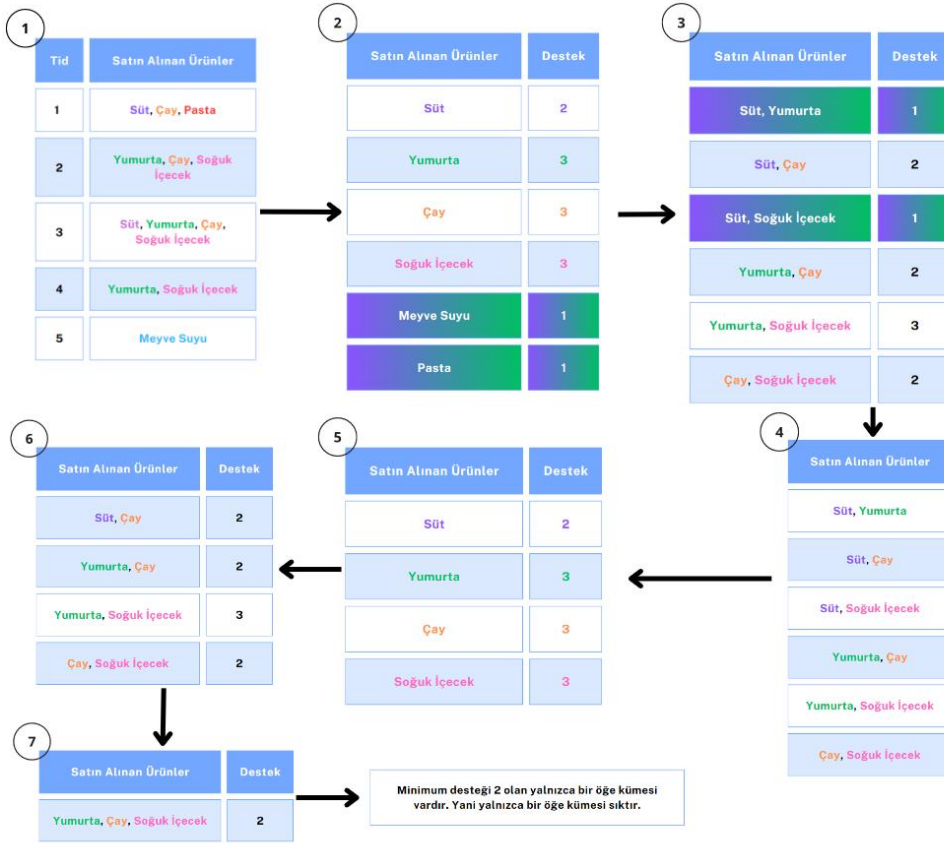
Apriori algoritmasının R. Agrawal ve arkadaşları tarafından oluşturulan pseudo kodu aşağıda verilmiştir (Agrawal, 1993, s. 210).

Şekil 14. Pseudo Kodu

```

Apriori( $T, \epsilon$ )
 $L_1 \leftarrow \{\text{large 1 - itemsets}\}$ 
 $k \leftarrow 2$ 
while  $L_{k-1} \neq \text{emptyset}$ 
     $C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \in \bigcup L_{k-1} \wedge b \notin a\}$ 
    for transactions  $t \in T$ 
         $C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}$ 
        for candidates  $c \in C_t$ 
             $\text{count}[c] \leftarrow \text{count}[c] + 1$ 
     $L_k \leftarrow \{c \mid c \in C_k \wedge \text{count}[c] \geq \epsilon\}$ 
     $k \leftarrow k + 1$ 
return  $\bigcup_k L_k$ 
    
```

Şekil 15. Örnek Uygulama



Adım 1: Veritabanındaki veriler

Adım 2: Tüm öğelerin destek/frekansını hesaplayın

Adım 3: Minimum desteęi 2'den az olan öğeleri atın

Adım 4: İki öğeyi birleştirin

Adım 5: Tüm öğelerin destek/frekansını hesaplayın

Adım 6: Minimum desteęi 2'den az olan öğeleri atın, combi Üç öğeyi birleştirin ve desteklerini hesaplayın.

Adım 7: Minimum desteęi 2'den az olan öğeleri atın

Sonuç: Yalnızca bir öğe kümesi sıktır (Yumurta, Çay, Soęuk İçecek) çünkü bu öğe kümesi minimum 2 desteęe sahiptir

Kaynak: (url: <https://t4tutorials.com/apriori-algorithm-in-data-mining-with-examples/>)

3. ÇOCUK HASTANE VERİLERİNİN VERİ MADENCİLİĞİ TEKNİKLERİ İLE ANALİZİ

Bu tez çalışmasında Erciyes Üniversitesi Çocuk Hastanesi'ne gelen hastaların profillerini analiz etmek için birliktelik kuralı yaklaşımı kullanılmıştır. Birliktelik kuralı olarak literatürde yaygın bir şekilde kullanılan Apriori Algoritması tercih edilmiştir. Weka 3.9.0 programı kullanılarak uygulanan Apriori Algoritmasında en düşük güven düzeyi (Confidence level) %60 seviyesi belirlenerek örüntüler analiz edilmiştir.

Sonuçları elde etme aşamasından önce sağlıklı ve tutarlı sonuçlar elde edebilmek amacıyla veri tipleri ve veri yapıları irdelenmiştir. Erciyes Üniversitesi Hastanelerinin 2022 yılı içerisinde bir milyondan fazla hasta kaydı bulunmaktadır. Önerilen çalışmada daha anlamlı veriler elde etmek amacıyla veri daraltma işlemi uygulanmıştır. Bu kapsamda hasta verileri Çocuk Hastanesi örneğine indirgenerek 228.052 hasta kaydından 3.050 adet eksik veri temizlenerek 225.002 hasta kaydı üzerinde işlem yapılmıştır. Büyük bir veri hacmine sahip olan veri tabanı içerisinde örüntülerin engellenmemesi adına bir takım veri ön işleme adımları gerçekleştirilmiştir. Bu adımlardan biri veri temizleme adımıdır. Mevcut veri setinde bazı eksik verilerin mevcut olduğu fark edilmiştir. Bu eksik verilerin türetilmesi yerine ilgili veri satırının silinmesi kararlaştırılmıştır. Bunun en büyük sebebi mevcut veri setindeki satır sayısının çok yüksek olmasıdır. Bu verilerin silinmesi veri setinde büyük etkiler oluşturmamakla birlikte eksik verilerin tamamlanarak veri setine eklenmesinin var olan örüntülerin kaybına neden olabileceği ihtimalidir. Böylece mevcut veri setinin güvenilirliği artırılmıştır. Bu tez çalışmasında Tablo 5'te verilen cinsiyet, yaş zaman dilimi, bölge, ay ve bölüm bilgisinin bulunduğu 6 adet farklı veri tipi dikkate alınarak birliktelik kuralı yaklaşımı uygulanmıştır.

Tablo 4. Veri Ambarı Setleri Kategori Tablosu

Cinsiyet	Yaş	Zaman Dilimi	Bölge	Ay	Bölüm
Kadın	0-3	Öğleden	Pilot (Kayseri)	Ocak	Acil
Erkek	4-10	önce	İç Anadolu	Şubat	Alerji
	11-14	Öğleden	Güneydoğu	Mart	Beslenme
		Sonra	Anadolu,	Nisan	Metabolizma
		Mesai Dışı	Doğu Anadolu,	Mayıs	Cerrahi
			Ege,	Haziran	Endokrin
			Marmara,	Temmuz	Enfeksiyon
			Akdeniz,	Ağustos	Gastroenteroloji
			Karadeniz	Eylül	Göğüs
				Ekim	Hastalıkları
				Kasım	Hematoloji
				Aralık	Onkoloji
					İmmünoloji
					Kardiyoloji
					Kemik İliği
					Nefroloji
					Nöroloji
					Pediyatri (Genel)
					Psikiyatri
					Radyoloji
					Romatoloji
					Üroloji
					Yeni Doğan

Yukarıda tanımlanan veri özelliklerine göre 10 farklı alt veri seti belirlenmiştir. Bu veri setleri şu şekildedir:

1. Çocuk Hastanesi Tüm Veriler
2. Çocuk Hastanesi Tüm Veriler Acil Departmanı Hariç Verisi
3. Çocuk Hastanesi Tüm Veriler Bölge: Kayseri Verisi
4. Çocuk Hastanesi Kayseri İli Acil Hariç Verisi
5. Çocuk Hastanesi Kayseri İli Hariç Verisi
6. Çocuk Hastanesi Kayseri İli Hariç Acil Hariç Verisi
7. Çocuk Hastanesi Sadece İç Anadolu Bölgesi Verisi
8. Çocuk Hastanesi Sadece İç Anadolu Bölgesi Acil Bölümü Hariç Verisi
9. Çocuk Hastanesi İç Anadolu Bölgesi Hariç Verisi
10. Çocuk Hastanesi İç Anadolu Bölgesi Hariç ve Acil Bölümü Hariç Verisi

Her bir veri seti için yapılan birliktelik kuralı çıkarımı çalışmalarında elde edilen örüntüler ve yorumları aşağıda verilmiştir.

Şekil 16. Çocuk Hastanesi Tüm Veriler kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori-N 100-T 0-C 0.6-D 0.05-U 1.0-M 0.1-S-1.0-c-1

Start Stop

Associator output

Result list (right-click...)

13:58:21 - Apriori

Minimum support: 0.1 (22500 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 30
Size of set of large itemsets L(3): 18

Best rules found:

1. yas_kategorisi=0-3_yas_arasi bolum_kategorisi=ACIL DEPARTMANI 26896 ==> bolge_kategorisi=PILOT 25050 <conf:(0.93)> lift:(1.27) lev:(0.02) [5399] conv:(3.92)
2. saat_kategorisi=mesai_disi bolum_kategorisi=ACIL DEPARTMANI 43063 ==> bolge_kategorisi=PILOT 36254 <conf:(0.84)> lift:(1.15) lev:(0.02) [4792] conv:(1.7)
3. CINSIYET=Hadin bolum_kategorisi=ACIL DEPARTMANI 34802 ==> bolge_kategorisi=PILOT 29180 <conf:(0.84)> lift:(1.15) lev:(0.02) [3753] conv:(1.67)
4. bolum_kategorisi=ACIL DEPARTMANI 74131 ==> bolge_kategorisi=PILOT 62043 <conf:(0.84)> lift:(1.15) lev:(0.04) [7883] conv:(1.65)
5. CINSIYET=Erkek bolum_kategorisi=ACIL DEPARTMANI 39329 ==> bolge_kategorisi=PILOT 32863 <conf:(0.84)> lift:(1.14) lev:(0.02) [4129] conv:(1.64)
6. CINSIYET=Erkek yas_kategorisi=0-3_yas_arasi 38050 ==> bolge_kategorisi=PILOT 31003 <conf:(0.81)> lift:(1.12) lev:(0.01) [3203] conv:(1.45)
7. yas_kategorisi=0-3_yas_arasi 66399 ==> bolge_kategorisi=PILOT 54031 <conf:(0.81)> lift:(1.11) lev:(0.02) [5520] conv:(1.45)
8. CINSIYET=Hadin yas_kategorisi=0-3_yas_arasi 28349 ==> bolge_kategorisi=PILOT 23028 <conf:(0.81)> lift:(1.11) lev:(0.01) [2316] conv:(1.44)
9. CINSIYET=Erkek saat_kategorisi=mesai_disi 35070 ==> bolge_kategorisi=PILOT 27160 <conf:(0.77)> lift:(1.06) lev:(0.01) [1537] conv:(1.19)
10. saat_kategorisi=mesai_disi 66263 ==> bolge_kategorisi=PILOT 51281 <conf:(0.77)> lift:(1.06) lev:(0.01) [2869] conv:(1.19)
11. CINSIYET=Hadin saat_kategorisi=mesai_disi 31193 ==> bolge_kategorisi=PILOT 24121 <conf:(0.77)> lift:(1.06) lev:(0.01) [1331] conv:(1.19)
12. yas_kategorisi=4-10_yas_arasi bolum_kategorisi=ACIL DEPARTMANI 32889 ==> bolge_kategorisi=PILOT 25233 <conf:(0.77)> lift:(1.05) lev:(0.01) [1204] conv:(1.16)
13. saat_kategorisi=ogleden_sonra 55287 ==> bolge_kategorisi=PILOT 41554 <conf:(0.75)> lift:(1.03) lev:(0.01) [1161] conv:(1.08)
14. CINSIYET=Erkek 119793 ==> bolge_kategorisi=PILOT 87822 <conf:(0.73)> lift:(1) lev:(0) [301] conv:(1.01)
15. CINSIYET=Hadin 105209 ==> bolge_kategorisi=PILOT 76564 <conf:(0.73)> lift:(1) lev:(-0) [-301] conv:(0.99)
16. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 33088 ==> bolge_kategorisi=PILOT 23442 <conf:(0.71)> lift:(0.97) lev:(-0) [-732] conv:(0.92)
17. yas_kategorisi=11-18_yas_arasi 69723 ==> bolge_kategorisi=PILOT 49388 <conf:(0.71)> lift:(0.97) lev:(-0.01) [-1551] conv:(0.92)
18. CINSIYET=Hadin yas_kategorisi=11-18_yas_arasi 36635 ==> bolge_kategorisi=PILOT 25946 <conf:(0.71)> lift:(0.97) lev:(-0) [-819] conv:(0.92)
19. saat_kategorisi=mesai_disi bolge_kategorisi=PILOT 51281 ==> bolum_kategorisi=ACIL DEPARTMANI 36254 <conf:(0.71)> lift:(2.15) lev:(0.09) [19358] conv:(2.29)
20. CINSIYET=Erkek saat_kategorisi=ogleden_once 54937 ==> bolge_kategorisi=PILOT 38233 <conf:(0.7)> lift:(0.95) lev:(-0.01) [-1903] conv:(0.89)
21. saat_kategorisi=ogleden_once 103452 ==> bolge_kategorisi=PILOT 71551 <conf:(0.69)> lift:(0.95) lev:(-0.02) [-4030] conv:(0.87)
22. CINSIYET=Hadin saat_kategorisi=ogleden_once 48515 ==> bolge_kategorisi=PILOT 33318 <conf:(0.69)> lift:(0.94) lev:(-0.01) [-2126] conv:(0.86)
23. CINSIYET=Erkek yas_kategorisi=4-10_yas_arasi 48655 ==> bolge_kategorisi=PILOT 33377 <conf:(0.69)> lift:(0.94) lev:(-0.01) [-2170] conv:(0.86)
24. yas_kategorisi=4-10_yas_arasi 88880 ==> bolge_kategorisi=PILOT 60967 <conf:(0.69)> lift:(0.94) lev:(-0.02) [-3968] conv:(0.86)
25. CINSIYET=Hadin yas_kategorisi=4-10_yas_arasi 40225 ==> bolge_kategorisi=PILOT 27590 <conf:(0.69)> lift:(0.94) lev:(-0.01) [-1798] conv:(0.86)
26. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 34361 ==> bolge_kategorisi=PILOT 23197 <conf:(0.68)> lift:(0.92) lev:(-0.01) [-1907] conv:(0.83)
27. CINSIYET=Erkek saat_kategorisi=mesai_disi 35070 ==> bolum_kategorisi=ACIL DEPARTMANI 22923 <conf:(0.65)> lift:(1.98) lev:(0.05) [11368] conv:(1.94)
28. yas_kategorisi=4-10_yas_arasi saat_kategorisi=ogleden_once 39821 ==> bolge_kategorisi=PILOT 25898 <conf:(0.65)> lift:(0.89) lev:(-0.01) [-3195] conv:(0.77)
29. saat_kategorisi=mesai_disi 66263 ==> bolum_kategorisi=ACIL DEPARTMANI 43063 <conf:(0.65)> lift:(1.97) lev:(0.09) [21231] conv:(1.92)

Status

OK Log

Tablo 5. Çocuk Hastanesi Tüm Veriler kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüler

1. yas_kategorisi=0-3_yas_arasi bolum_kategorisi=ACIL DEPARTMANI 26896 ==> bolge_kategorisi=PILOT 25050 <conf:(0.93)> lift:(1.27) lev:(0.02) [5399] conv:(3.92)
2. saat_kategorisi=mesai_disi bolum_kategorisi=ACIL DEPARTMANI 43063 ==> bolge_kategorisi=PILOT 36254 <conf:(0.84)> lift:(1.15) lev:(0.02) [4792] conv:(1.7)
3. CINSIYET=Kadin bolum_kategorisi=ACIL DEPARTMANI 34802 ==> bolge_kategorisi=PILOT 29180 <conf:(0.84)> lift:(1.15) lev:(0.02) [3753] conv:(1.67)
4. bolum_kategorisi=ACIL DEPARTMANI 74131 ==> bolge_kategorisi=PILOT 62043 <conf:(0.84)> lift:(1.15) lev:(0.04) [7883] conv:(1.65)
5. CINSIYET=Erkek bolum_kategorisi=ACIL DEPARTMANI 39329 ==> bolge_kategorisi=PILOT 32863 <conf:(0.84)> lift:(1.14) lev:(0.02) [4129] conv:(1.64)
6. CINSIYET=Erkek yas_kategorisi=0-3_yas_arasi 38050 ==> bolge_kategorisi=PILOT 31003 <conf:(0.81)> lift:(1.12) lev:(0.01) [3203] conv:(1.45)
7. yas_kategorisi=0-3_yas_arasi 66399 ==> bolge_kategorisi=PILOT 54031 <conf:(0.81)> lift:(1.11) lev:(0.02) [5520] conv:(1.45)
8. CINSIYET=Kadin yas_kategorisi=0-3_yas_arasi 28349 ==> bolge_kategorisi=PILOT 23028 <conf:(0.81)> lift:(1.11) lev:(0.01) [2316] conv:(1.44)
9. CINSIYET=Erkek saat_kategorisi=mesai_disi 35070 ==> bolge_kategorisi=PILOT 27160 <conf:(0.77)> lift:(1.06) lev:(0.01) [1537] conv:(1.19)
10. saat_kategorisi=mesai_disi 66263 ==> bolge_kategorisi=PILOT 51281 <conf:(0.77)> lift:(1.06) lev:(0.01) [2869] conv:(1.19)
11. CINSIYET=Kadin saat_kategorisi=mesai_disi 31193 ==> bolge_kategorisi=PILOT 24121 <conf:(0.77)> lift:(1.06) lev:(0.01) [1331] conv:(1.19)
12. yas_kategorisi=4-10_yas_arasi bolum_kategorisi=ACIL DEPARTMANI 32889 ==> bolge_kategorisi=PILOT 25233 <conf:(0.77)> lift:(1.05) lev:(0.01) [1204] conv:(1.16)
13. saat_kategorisi=ogleden_sonra 55287 ==> bolge_kategorisi=PILOT 41554 <conf:(0.75)> lift:(1.03) lev:(0.01) [1161] conv:(1.08)

14. CINSIYET=Erkek 119793 ==> bolge_kategorisi=PILOT 87822 <conf:(0.73)>
lift:(1) lev:(0) [301] conv:(1.01)
15. CINSIYET=Kadin 105209 ==> bolge_kategorisi=PILOT 76564 <conf:(0.73)>
lift:(1) lev:(-0) [-301] conv:(0.99)
16. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 33088 ==>
bolge_kategorisi=PILOT 23442 <conf:(0.71)> lift:(0.97) lev:(-0) [-732]
conv:(0.92)
17. yas_kategorisi=11-18_yas_arasi 69723 ==> bolge_kategorisi=PILOT 49388
<conf:(0.71)> lift:(0.97) lev:(-0.01) [-1551] conv:(0.92)
18. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 36635 ==>
bolge_kategorisi=PILOT 25946 <conf:(0.71)> lift:(0.97) lev:(-0) [-819]
conv:(0.92)
19. saat_kategorisi=mesai_disi bolge_kategorisi=PILOT 51281 ==>
bolum_kategorisi=ACIL DEPARTMANI 36254 <conf:(0.71)> lift:(2.15)
lev:(0.09) [19358] conv:(2.29)
20. CINSIYET=Erkek saat_kategorisi=ogleden_once 54937 ==>
bolge_kategorisi=PILOT 38233 <conf:(0.7)> lift:(0.95) lev:(-0.01) [-1903]
conv:(0.89)
21. saat_kategorisi=ogleden_once 103452 ==> bolge_kategorisi=PILOT 71551
<conf:(0.69)> lift:(0.95) lev:(-0.02) [-4030] conv:(0.87)
22. CINSIYET=Kadin saat_kategorisi=ogleden_once 48515 ==>
bolge_kategorisi=PILOT 33318 <conf:(0.69)> lift:(0.94) lev:(-0.01) [-2126]
conv:(0.86)
23. CINSIYET=Erkek yas_kategorisi=4-10 _yas_arasi 48655 ==>
bolge_kategorisi=PILOT 33377 <conf:(0.69)> lift:(0.94) lev:(-0.01) [-2170]
conv:(0.86)
24. yas_kategorisi=4-10 _yas_arasi 88880 ==> bolge_kategorisi=PILOT 60967
<conf:(0.69)> lift:(0.94) lev:(-0.02) [-3968] conv:(0.86)
25. CINSIYET=Kadin yas_kategorisi=4-10 _yas_arasi 40225 ==>
bolge_kategorisi=PILOT 27590 <conf:(0.69)> lift:(0.94) lev:(-0.01) [-1798]
conv:(0.86)
26. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 34361 ==>
bolge_kategorisi=PILOT 23197 <conf:(0.68)> lift:(0.92) lev:(-0.01) [-1907]
conv:(0.83)
27. CINSIYET=Erkek saat_kategorisi=mesai_disi 35070 ==>
bolum_kategorisi=ACIL DEPARTMANI 22923 <conf:(0.65)> lift:(1.98)
lev:(0.05) [11368] conv:(1.94)

28. yas_kategorisi=4-10 _yas_arasi saat_kategorisi=ogleden_once 39821 ==>
bolge_kategorisi=PILOT 25898 <conf:(0.65)> lift:(0.89) lev:(-0.01) [-3195]
conv:(0.77)

29. saat_kategorisi=mesai_disi 66263 ==> bolum_kategorisi=ACIL DEPARTMANI
43063 <conf:(0.65)> lift:(1.97) lev:(0.09) [21231] conv:(1.92)

Tablo 5’te görüldüğü gibi Çocuk Hastanesi ‘Tüm Veriler’ kategorisine ait 29 örüntü tespit edilmiştir. İlk örüntüye göre yaşı 0-3 yaş aralığında olan, bölüm kategorisi Acil Departmanı olan 26896 hastanın, 25050 kişininin Pilot (Kayseri) ilinden geldiği %93 güven düzeyi ile anlaşılmaktadır.

İkinci örüntüye göre mesai dışı acil departmanına gelen 43.063 hastanın, 36.254 adedi % 84 güven düzeyi ile Pilot (Kayseri) ilinden geldiği anlaşılmaktadır.

Üçüncü örüntüye göre cinsiyeti kadın olan acil departmanına gelen 34.802 hastanın, 29.180 hastası % 84 güven düzeyi ile yine Pilot (Kayseri) ilinden geldiği anlaşılmaktadır.

Geri kalan 26 örüntü için yapılan veri madenciliği sonuçları incelendiğinde Pilot (Kayseri) ili ve Acil Departmanının toplam veri içerisindeki ağırlığının oldukça fazla olduğu için diğer kayıtlar üzerine baskın olduğu anlaşılmaktadır.

Bu durum farklı örüntülerin engellenmesine neden olabileceği için pilot bölgeleri içeren veriler çıkartılarak yeniden bir birliktelik kuralı çıkarımı gerçekleştirilmiştir.

Şekil 17. Çocuk Hastanesi Tüm Veriler Acil Departmanı Hariç Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Associator

Choose Apriori -N 100 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start Stop

Result list (right-click...)

- 14:05:58 - Apriori
- 14:06:19 - Apriori

Associator output

Apriori
=====

Minimum support: 0.1 (15087 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 14
Size of set of large itemsets L(2): 23
Size of set of large itemsets L(3): 13

Best rules found:

1. CINSIYET=Erkek yas_kategorisi=0-3_yas_arasi 23346 ==> bolge_kategorisi=PILOT 17285 <conf:(0.74)> lift:(1.09) lev:(0.01) [1448] conv:(1.24)
2. yas_kategorisi=0-3_yas_arasi saat_kategorisi=ogleden_once 24702 ==> bolge_kategorisi=PILOT 18270 <conf:(0.74)> lift:(1.09) lev:(0.01) [1513] conv:(1.24)
3. yas_kategorisi=0-3_yas_arasi 39503 ==> bolge_kategorisi=PILOT 28981 <conf:(0.73)> lift:(1.08) lev:(0.01) [2184] conv:(1.21)
4. saat_kategorisi=ogleden_sonra 37786 ==> bolge_kategorisi=PILOT 26944 <conf:(0.71)> lift:(1.05) lev:(0.01) [1311] conv:(1.12)
5. CINSIYET=Erkek 80464 ==> bolge_kategorisi=PILOT 54959 <conf:(0.68)> lift:(1.01) lev:(0) [376] conv:(1.01)
6. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 26053 ==> bolge_kategorisi=PILOT 17709 <conf:(0.68)> lift:(1) lev:(0) [36] conv:(1)
7. yas_kategorisi=11-18_yas_arasi 55377 ==> bolge_kategorisi=PILOT 37628 <conf:(0.68)> lift:(1) lev:(0) [63] conv:(1)
8. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 29324 ==> bolge_kategorisi=PILOT 19919 <conf:(0.68)> lift:(1) lev:(0) [27] conv:(1)
9. CINSIYET=Erkek saat_kategorisi=ogleden_once 47782 ==> bolge_kategorisi=PILOT 32342 <conf:(0.68)> lift:(1) lev:(-0) [-70] conv:(1)
10. CINSIYET=Kadin 70407 ==> bolge_kategorisi=PILOT 47384 <conf:(0.67)> lift:(0.99) lev:(-0) [-376] conv:(0.98)
11. saat_kategorisi=ogleden_once 89885 ==> bolge_kategorisi=PILOT 60372 <conf:(0.67)> lift:(0.99) lev:(-0) [-601] conv:(0.98)
12. CINSIYET=Kadin saat_kategorisi=ogleden_once 42103 ==> bolge_kategorisi=PILOT 28030 <conf:(0.67)> lift:(0.98) lev:(-0) [-530] conv:(0.96)
13. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 31762 ==> bolge_kategorisi=PILOT 21078 <conf:(0.66)> lift:(0.98) lev:(-0) [-467] conv:(0.96)
14. CINSIYET=Erkek yas_kategorisi=4-10_yas_arasi 31065 ==> bolge_kategorisi=PILOT 19965 <conf:(0.64)> lift:(0.95) lev:(-0.01) [-1107] conv:(0.9)
15. yas_kategorisi=4-10_yas_arasi 55991 ==> bolge_kategorisi=PILOT 35734 <conf:(0.64)> lift:(0.94) lev:(-0.01) [-2247] conv:(0.89)
16. CINSIYET=Kadin yas_kategorisi=4-10_yas_arasi 24926 ==> bolge_kategorisi=PILOT 15769 <conf:(0.63)> lift:(0.93) lev:(-0.01) [-1139] conv:(0.88)
17. yas_kategorisi=0-3_yas_arasi bolge_kategorisi=PILOT 28981 ==> saat_kategorisi=ogleden_once 18270 <conf:(0.63)> lift:(1.06) lev:(0.01) [1003] conv:(1.09)
18. yas_kategorisi=4-10_yas_arasi saat_kategorisi=ogleden_once 33421 ==> bolge_kategorisi=PILOT 21024 <conf:(0.63)> lift:(0.93) lev:(-0.01) [-1647] conv:(0.87)
19. yas_kategorisi=0-3_yas_arasi 39503 ==> saat_kategorisi=ogleden_once 24702 <conf:(0.63)> lift:(1.05) lev:(0.01) [1167] conv:(1.08)
20. bolge_kategorisi=IC ANADOLU BOLGESI 32844 ==> saat_kategorisi=ogleden_once 20267 <conf:(0.62)> lift:(1.04) lev:(0) [699] conv:(1.06)
21. CINSIYET=Kadin yas_kategorisi=4-10_yas_arasi 24926 ==> saat_kategorisi=ogleden_once 15091 <conf:(0.61)> lift:(1.02) lev:(0) [240] conv:(1.02)

Status

OK Log

Tablo 6. Çocuk Hastanesi Tüm Veriler Acil Departmanı Hariç Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüler

1. CINSIYET=Erkek yas_kategorisi=0-3_yas_arasi 23346 ==>
bolge_kategorisi=PILOT 17285 <conf:(0.74)> lift:(1.09) lev:(0.01) [1448]
conv:(1.24)
2. yas_kategorisi=0-3_yas_arasi saat_kategorisi=ogleden_once 24702 ==>
bolge_kategorisi=PILOT 18270 <conf:(0.74)> lift:(1.09) lev:(0.01) [1513]
conv:(1.24)
3. yas_kategorisi=0-3_yas_arasi 39503 ==> bolge_kategorisi=PILOT 28981
<conf:(0.73)> lift:(1.08) lev:(0.01) [2184] conv:(1.21)
4. saat_kategorisi=ogleden_sonra 37786 ==> bolge_kategorisi=PILOT 26944
<conf:(0.71)> lift:(1.05) lev:(0.01) [1311] conv:(1.12)
5. CINSIYET=Erkek 80464 ==> bolge_kategorisi=PILOT 54959 <conf:(0.68)>
lift:(1.01) lev:(0) [376] conv:(1.01)
6. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 26053 ==>
bolge_kategorisi=PILOT 17709 <conf:(0.68)> lift:(1) lev:(0) [36] conv:(1)
7. yas_kategorisi=11-18_yas_arasi 55377 ==> bolge_kategorisi=PILOT 37628
<conf:(0.68)> lift:(1) lev:(0) [63] conv:(1)
8. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 29324 ==>
bolge_kategorisi=PILOT 19919 <conf:(0.68)> lift:(1) lev:(0) [27] conv:(1)
9. CINSIYET=Erkek saat_kategorisi=ogleden_once 47782 ==>
bolge_kategorisi=PILOT 32342 <conf:(0.68)> lift:(1) lev:(-0) [-70] conv:(1)
10. CINSIYET=Kadin 70407 ==> bolge_kategorisi=PILOT 47384 <conf:(0.67)>
lift:(0.99) lev:(-0) [-376] conv:(0.98)
11. saat_kategorisi=ogleden_once 89885 ==> bolge_kategorisi=PILOT 60372
<conf:(0.67)> lift:(0.99) lev:(-0) [-601] conv:(0.98)
12. CINSIYET=Kadin saat_kategorisi=ogleden_once 42103 ==>
bolge_kategorisi=PILOT 28030 <conf:(0.67)> lift:(0.98) lev:(-0) [-530]
conv:(0.96)
13. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 31762 ==>
bolge_kategorisi=PILOT 21078 <conf:(0.66)> lift:(0.98) lev:(-0) [-467]
conv:(0.96)
14. CINSIYET=Erkek yas_kategorisi=4-10 _yas_arasi 31065 ==>
bolge_kategorisi=PILOT 19965 <conf:(0.64)> lift:(0.95) lev:(-0.01) [-1107]
conv:(0.9)

15. yas_kategorisi=4-10 _yas_arasi 55991 ==> bolge_kategorisi=PILOT 35734
<conf:(0.64)> lift:(0.94) lev:(-0.01) [-2247] conv:(0.89)
16. CINSIYET=Kadin yas_kategorisi=4-10 _yas_arasi 24926 ==>
bolge_kategorisi=PILOT 15769 <conf:(0.63)> lift:(0.93) lev:(-0.01) [-1139]
conv:(0.88)
17. yas_kategorisi=0-3_yas_arasi bolge_kategorisi=PILOT 28981 ==>
saat_kategorisi=ogleden_once 18270 <conf:(0.63)> lift:(1.06) lev:(0.01) [1003]
conv:(1.09)
18. yas_kategorisi=4-10 _yas_arasi saat_kategorisi=ogleden_once 33421 ==>
bolge_kategorisi=PILOT 21024 <conf:(0.63)> lift:(0.93) lev:(-0.01) [-1647]
conv:(0.87)
19. yas_kategorisi=0-3_yas_arasi 39503 ==> saat_kategorisi=ogleden_once 24702
<conf:(0.63)> lift:(1.05) lev:(0.01) [1167] conv:(1.08)
20. bolge_kategorisi=IC ANADOLU BOLGESI 32844 ==>
saat_kategorisi=ogleden_once 20267 <conf:(0.62)> lift:(1.04) lev:(0) [699]
conv:(1.06)
21. CINSIYET=Kadin yas_kategorisi=4-10 _yas_arasi 24926 ==>
saat_kategorisi=ogleden_once 15091 <conf:(0.61)> lift:(1.02) lev:(0) [240]
conv:(1.02)

Tablo 6’da görüldüğü gibi Çocuk Hastanesi Acil Bölümü dışındaki diğer bölüm kategorisine ait 21 örüntü tespit edilmiştir. İlk örüntüye göre yaşı 0-3 yaş aralığında cinsiyeti erkek olan 23.346 hastanın, 17.285 hastasının Pilot (Kayseri) ilinden geldiği %74 güven düzeyi ile anlaşılmaktadır.

İkinci örüntüye göre yaşı 0-3 yaş aralığında öğleden önce gelen 24.702 hastanın, 18.270 hastası % 74 güven düzeyi ile Pilot (Kayseri) ilinden geldiği anlaşılmaktadır.

Üçüncü örüntüye göre yaşı 0-3 yaş aralığında gelen 39.503 hastanın, 28.981 hastası % 73 güven düzeyi ile yine Pilot (Kayseri) ilinden geldiği anlaşılmaktadır.

Geri kalan 18 örüntü için yapılan veri madenciliği incelendiğinde Pilot (Kayseri) ilinin toplam veri içerisinde baskın olduğu anlaşılmaktadır.

Şekil 18. Çocuk Hastanesi Tüm Veriler Bölge Kayseri Bilgisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Choose **Apriori -N 100 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1**

Start | Stop

Associator output

Result list (right-click...)

14:14:54 - Apriori

```
=== Run information ===
Scheme:      weka.associations.Apriori -N 100 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    2)164386_veri_sadece_kayseri_weka
Instances:   164386
Attributes:  5
             ay_kategorisi
             CINSIYET
             yas_kategorisi
             saat_kategorisi
             bolum_kategorisi
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (16439 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10
Size of set of large itemsets L(2): 22
Size of set of large itemsets L(3): 2

Best rules found:
1. CINSIYET=Erkek saat_kategorisi=mesai_disi 27160 ==> bolum_kategorisi=ACIL DEPARTMANI 19270 <conf:(0.71)> lift:(1.88) lev:(0.05) [9019] conv:(2.14)
2. saat_kategorisi=mesai_disi 51281 ==> bolum_kategorisi=ACIL DEPARTMANI 36254 <conf:(0.71)> lift:(1.87) lev:(0.1) [16899] conv:(2.12)
3. CINSIYET=Kadin saat_kategorisi=mesai_disi 24121 ==> bolum_kategorisi=ACIL DEPARTMANI 16984 <conf:(0.7)> lift:(1.87) lev:(0.05) [7880] conv:(2.1)
```

Status: OK | Log | x0

Tablo 7. Çocuk Hastanesi Tüm Veriler Bölge Kayseri Bilgisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüler

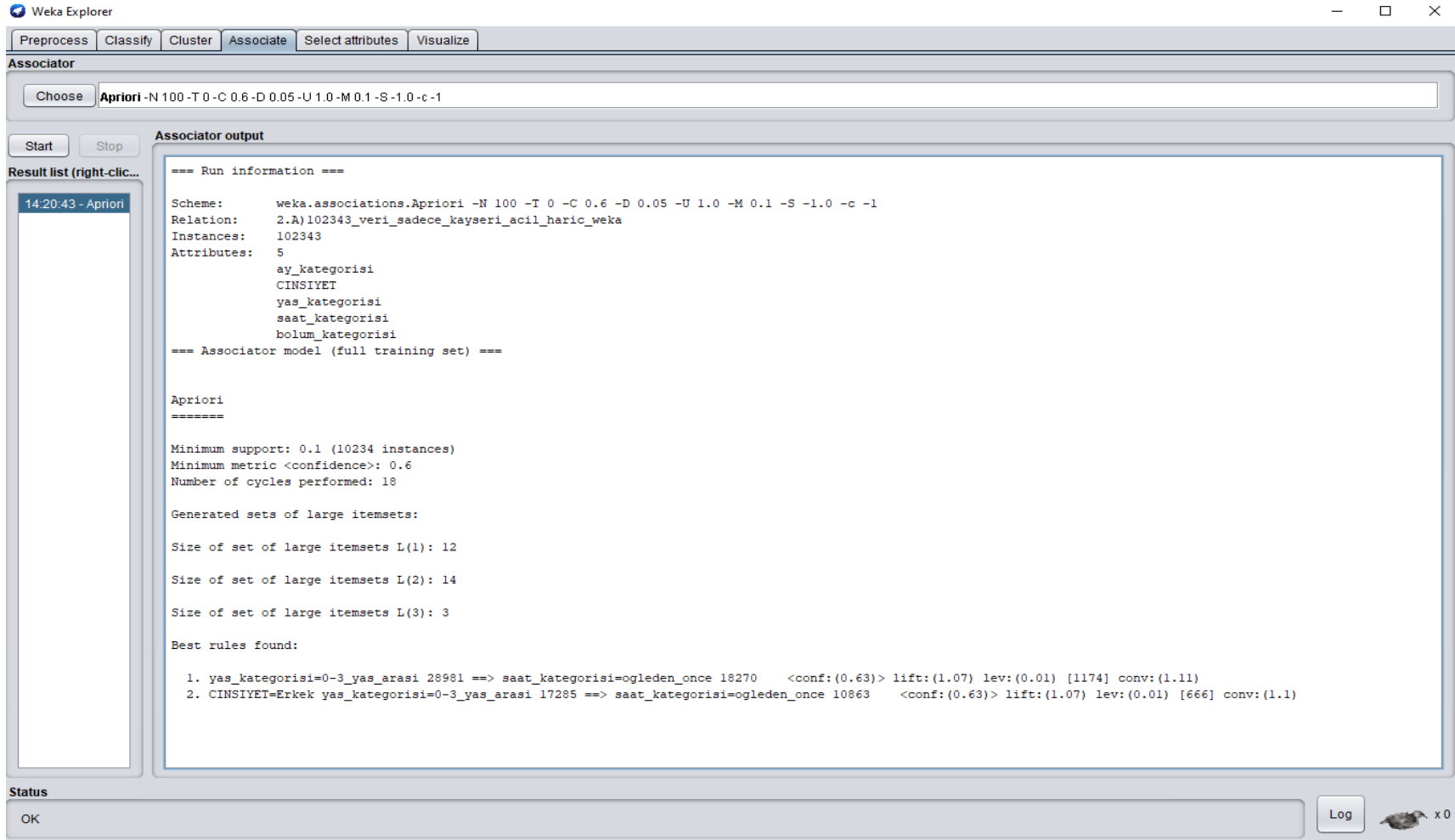
1. CINSİYET=Erkek saat_kategorisi=mesai_disi 27160 ==> bolum_kategorisi=ACIL DEPARTMANI 19270 <conf:(0.71)> lift:(1.88) lev:(0.05) [9019] conv:(2.14)
2. saat_kategorisi=mesai_disi 51281 ==> bolum_kategorisi=ACIL DEPARTMANI 36254 <conf:(0.71)> lift:(1.87) lev:(0.1) [16899] conv:(2.12)
3. CINSİYET=Kadin saat_kategorisi=mesai_disi 24121 ==> bolum_kategorisi=ACIL DEPARTMANI 16984 <conf:(0.7)> lift:(1.87) lev:(0.05) [7880] conv:(2.1)

Tablo 7’de görüldüğü gibi Çocuk Hastanesi Kayseri kategorisi kullanılarak 3 örüntü tespit edilmiştir. İlk örüntüye göre cinsiyeti erkek olan mesai dışı gelen 27.160 hastanın, 19.270 hastası acil departmanına geldiği %71 güven düzeyi ile anlaşılmaktadır.

İkinci örüntüye göre mesai dışı gelen 51.281 hastanın, 36254 hastası acil bölümüne geldiği %71 güven düzeyi ile anlaşılmaktadır.

Üçüncü örüntüye göre cinsiyeti kadın olan mesai dışı gelen 24.121 hastanın, 16.984 hastanın acil bölümüne %70 güven düzeyi ile geldiği anlaşılmaktadır.

Şekil 19. Çocuk Hastanesi Kayseri ili Acil Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü



Tablo 8. Çocuk Hastanesi Kayseri ili Acil Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüler

1. yas_kategorisi=0-3_yas_arasi 28981 ==> saat_kategorisi=ogleden_once 18270
<conf:(0.63)> lift:(1.07) lev:(0.01) [1174] conv:(1.11)
2. CINSIYET=Erkek yas_kategorisi=0-3_yas_arasi 17285 ==>
saat_kategorisi=ogleden_once 10863 <conf:(0.63)> lift:(1.07) lev:(0.01) [666]
conv:(1.1)

Tablo 8’de görüldüğü gibi Çocuk Hastanesi Kayseri ili acil hariç kategorisine ait 2 örüntü tespit edilmiştir. İlk örüntüye göre 0-3 yaş aralığındaki 28.981 hastanın, 18.270 hastası öğleden önce geldiği %70 güven düzeyi ile anlaşılmaktadır.

İkinci örüntüye göre cinsiyeti erkek olan 0-3 yaş aralığındaki 17.285 hastanın, 10.863 hastası öğleden önce geldiği %63 güven düzeyi ile anlaşılmaktadır.

Şekil 20. Çocuk Hastanesi Kayseri ili Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -N 100 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start Stop

Associator output

Result list (right-click...)

15:51:52 - Apriori

```
bolge_kategorisi
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (6062 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Size of set of large itemsets L(2): 31

Size of set of large itemsets L(3): 10

Best rules found:

1. yas_kategorisi=0-3_yas_arasi 12368 ==> bolge_kategorisi=IC ANADOLU BOLGESI 9053 <conf:(0.73)> lift:(1.17) lev:(0.02) [1345] conv:(1.41)
2. CINSIYET=Kadin saat_kategorisi=ogleden_once 15197 ==> bolge_kategorisi=IC ANADOLU BOLGESI 10287 <conf:(0.68)> lift:(1.09) lev:(0.01) [815] conv:(1.17)
3. saat_kategorisi=ogleden_once 31901 ==> bolge_kategorisi=IC ANADOLU BOLGESI 21338 <conf:(0.67)> lift:(1.07) lev:(0.02) [1456] conv:(1.14)
4. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 11164 ==> bolge_kategorisi=IC ANADOLU BOLGESI 7435 <conf:(0.67)> lift:(1.07) lev:(0.01) [477] conv:(1.13)
5. CINSIYET=Erkek saat_kategorisi=ogleden_once 16704 ==> bolge_kategorisi=IC ANADOLU BOLGESI 11051 <conf:(0.66)> lift:(1.06) lev:(0.01) [640] conv:(1.11)
6. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 9646 ==> bolge_kategorisi=IC ANADOLU BOLGESI 6120 <conf:(0.63)> lift:(1.02) lev:(0) [108] conv:(1.03)
7. CINSIYET=Kadin 28645 ==> bolge_kategorisi=IC ANADOLU BOLGESI 18145 <conf:(0.63)> lift:(1.02) lev:(0) [292] conv:(1.03)
8. bolum_kategorisi=ACIL DEPARTMANI 12088 ==> yas_kategorisi=4-10_yas_arasi 7656 <conf:(0.63)> lift:(1.38) lev:(0.03) [2089] conv:(1.47)
9. yas_kategorisi=11-18_yas_arasi 20335 ==> bolge_kategorisi=IC ANADOLU BOLGESI 12848 <conf:(0.63)> lift:(1.01) lev:(0) [174] conv:(1.02)
10. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 10689 ==> bolge_kategorisi=IC ANADOLU BOLGESI 6728 <conf:(0.63)> lift:(1.01) lev:(0) [66] conv:(1.02)
11. CINSIYET=Erkek 31971 ==> bolge_kategorisi=IC ANADOLU BOLGESI 19632 <conf:(0.61)> lift:(0.99) lev:(-0) [-292] conv:(0.98)
12. yas_kategorisi=4-10_yas_arasi saat_kategorisi=ogleden_once 13923 ==> bolge_kategorisi=IC ANADOLU BOLGESI 8527 <conf:(0.61)> lift:(0.98) lev:(-0) [-150] conv:(0.97)
```

Status

OK Log x0

Tablo 9. Çocuk Hastanesi Kayseri ili Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüler

1. yas_kategorisi=0-3_yas_arasi 12368 ==> bolge_kategorisi=IC ANADOLU BOLGESI 9053 <conf:(0.73)> lift:(1.17) lev:(0.02) [1345] conv:(1.41)
2. CINSIYET=Kadin saat_kategorisi=ogleden_once 15197 ==> bolge_kategorisi=IC ANADOLU BOLGESI 10287 <conf:(0.68)> lift:(1.09) lev:(0.01) [815] conv:(1.17)
3. saat_kategorisi=ogleden_once 31901 ==> bolge_kategorisi=IC ANADOLU BOLGESI 21338 <conf:(0.67)> lift:(1.07) lev:(0.02) [1456] conv:(1.14)
4. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 11164 ==> bolge_kategorisi=IC ANADOLU BOLGESI 7435 <conf:(0.67)> lift:(1.07) lev:(0.01) [477] conv:(1.13)
5. CINSIYET=Erkek saat_kategorisi=ogleden_once 16704 ==> bolge_kategorisi=IC ANADOLU BOLGESI 11051 <conf:(0.66)> lift:(1.06) lev:(0.01) [640] conv:(1.11)
6. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 9646 ==> bolge_kategorisi=IC ANADOLU BOLGESI 6120 <conf:(0.63)> lift:(1.02) lev:(0) [108] conv:(1.03)
7. CINSIYET=Kadin 28645 ==> bolge_kategorisi=IC ANADOLU BOLGESI 18145 <conf:(0.63)> lift:(1.02) lev:(0) [292] conv:(1.03)
8. bolum_kategorisi=ACIL DEPARTMANI 12088 ==> yas_kategorisi=4-10_yas_arasi 7656 <conf:(0.63)> lift:(1.38) lev:(0.03) [2089] conv:(1.47)
9. yas_kategorisi=11-18_yas_arasi 20335 ==> bolge_kategorisi=IC ANADOLU BOLGESI 12848 <conf:(0.63)> lift:(1.01) lev:(0) [174] conv:(1.02)
10. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 10689 ==> bolge_kategorisi=IC ANADOLU BOLGESI 6728 <conf:(0.63)> lift:(1.01) lev:(0) [66] conv:(1.02)
11. CINSIYET=Erkek 31971 ==> bolge_kategorisi=IC ANADOLU BOLGESI 19632 <conf:(0.61)> lift:(0.99) lev:(-0) [-292] conv:(0.98)
12. yas_kategorisi=4-10_yas_arasi saat_kategorisi=ogleden_once 13923 ==> bolge_kategorisi=IC ANADOLU BOLGESI 8527 <conf:(0.61)> lift:(0.98) lev:(-0) [-150] conv:(0.97)

Tablo 9’da görüldüğü gibi Çocuk Hastanesi Kayseri ili hariç kategorisine ait 12 örüntü tespit edilmiştir. İlk örüntüye göre 0-3 yaş aralığındaki 12.368 hastanın, 9.053 hastası İç Anadolu Bölgesinden geldiği %73 güven düzeyi ile anlaşılmaktadır.

İkinci örüntüye göre cinsiyeti kadın ve öğleden önce gelen 15.197 hastanın, 10.287 hastası İç Anadolu Bölgesinden geldiği %68 güven düzeyi ile anlaşılmaktadır.

Üçüncü örüntüye göre öğleden önce 31.901 hastanın, 21.338 hastası İç Anadolu Bölgesinden geldiği %68 güven düzeyi ile anlaşılmaktadır.

Geri kalan 9 örüntü için yapılan veri madenciliği incelendiğinde Pilot (Kayseri) ili dışındaki veriler incelendiğinde İç Anadolu Bölgesinin baskın şekilde yer aldığı anlaşılmaktadır.

Şekil 21. Çocuk Hastanesi Kayseri ili Hariç Acil Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örintülerin Ekran Görüntüsü

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -N 100 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start Stop

Associator output

Result list (right-click...)

15:56:33 - Apriori

Size of set of large itemsets L(2): 24

Size of set of large itemsets L(3): 13

Best rules found:

1. yas_kategorisi=0-3_yas_arasi saat_kategorisi=ogleden_once 6432 ==> bolge_kategorisi=IC ANADOLU BOLGESI 5197 <conf:(0.81)> lift:(1.19) lev:(0.02) [843] conv:(1.68)
2. yas_kategorisi=0-3_yas_arasi 10522 ==> bolge_kategorisi=IC ANADOLU BOLGESI 8342 <conf:(0.79)> lift:(1.17) lev:(0.03) [1220] conv:(1.56)
3. bolum_kategorisi=NOROLOJI DEPARTMANI 6835 ==> bolge_kategorisi=IC ANADOLU BOLGESI 4859 <conf:(0.71)> lift:(1.05) lev:(0) [233] conv:(1.12)
4. saat_kategorisi=mesai_disi 8173 ==> bolge_kategorisi=IC ANADOLU BOLGESI 5718 <conf:(0.7)> lift:(1.03) lev:(0) [186] conv:(1.08)
5. CINSIYET=Kadin saat_kategorisi=ogleden_once 14073 ==> bolge_kategorisi=IC ANADOLU BOLGESI 9764 <conf:(0.69)> lift:(1.03) lev:(0) [239] conv:(1.06)
6. saat_kategorisi=ogleden_once 29513 ==> bolge_kategorisi=IC ANADOLU BOLGESI 20267 <conf:(0.69)> lift:(1.01) lev:(0.01) [292] conv:(1.03)
7. CINSIYET=Kadin 23023 ==> bolge_kategorisi=IC ANADOLU BOLGESI 15794 <conf:(0.69)> lift:(1.01) lev:(0) [211] conv:(1.03)
8. CINSIYET=Erkek saat_kategorisi=ogleden_once 15440 ==> bolge_kategorisi=IC ANADOLU BOLGESI 10503 <conf:(0.68)> lift:(1.01) lev:(0) [53] conv:(1.01)
9. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 10684 ==> bolge_kategorisi=IC ANADOLU BOLGESI 7206 <conf:(0.67)> lift:(1) lev:(-0) [-24] conv:(0.99)
10. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 8344 ==> bolge_kategorisi=IC ANADOLU BOLGESI 5627 <conf:(0.67)> lift:(1) lev:(-0) [-20] conv:(0.99)
11. CINSIYET=Erkek 25505 ==> bolge_kategorisi=IC ANADOLU BOLGESI 17050 <conf:(0.67)> lift:(0.99) lev:(-0) [-211] conv:(0.97)
12. yas_kategorisi=11-18_yas_arasi 17749 ==> bolge_kategorisi=IC ANADOLU BOLGESI 11834 <conf:(0.67)> lift:(0.99) lev:(-0) [-178] conv:(0.97)
13. CINSIYET=Kadin yas_kategorisi=4-10_yas_arasi 9157 ==> bolge_kategorisi=IC ANADOLU BOLGESI 6060 <conf:(0.66)> lift:(0.98) lev:(-0) [-137] conv:(0.96)
14. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 9405 ==> bolge_kategorisi=IC ANADOLU BOLGESI 6207 <conf:(0.66)> lift:(0.98) lev:(-0) [-158] conv:(0.95)
15. yas_kategorisi=4-10_yas_arasi saat_kategorisi=ogleden_once 12397 ==> bolge_kategorisi=IC ANADOLU BOLGESI 7864 <conf:(0.63)> lift:(0.94) lev:(-0.01) [-526] conv:(0.88)
16. saat_kategorisi=ogleden_sonra 10842 ==> bolge_kategorisi=IC ANADOLU BOLGESI 6859 <conf:(0.63)> lift:(0.93) lev:(-0.01) [-478] conv:(0.88)
17. yas_kategorisi=4-10_yas_arasi 20257 ==> bolge_kategorisi=IC ANADOLU BOLGESI 12668 <conf:(0.63)> lift:(0.92) lev:(-0.02) [-1042] conv:(0.86)
18. yas_kategorisi=0-3_yas_arasi bolge_kategorisi=IC ANADOLU BOLGESI 8342 ==> saat_kategorisi=ogleden_once 5197 <conf:(0.62)> lift:(1.02) lev:(0) [123] conv:(1.04)
19. CINSIYET=Kadin yas_kategorisi=4-10_yas_arasi 9157 ==> saat_kategorisi=ogleden_once 5700 <conf:(0.62)> lift:(1.02) lev:(0) [131] conv:(1.04)
20. yas_kategorisi=4-10_yas_arasi bolge_kategorisi=IC ANADOLU BOLGESI 12668 ==> saat_kategorisi=ogleden_once 7864 <conf:(0.62)> lift:(1.02) lev:(0) [159] conv:(1.03)
21. CINSIYET=Kadin bolge_kategorisi=IC ANADOLU BOLGESI 15794 ==> saat_kategorisi=ogleden_once 9764 <conf:(0.62)> lift:(1.02) lev:(0) [158] conv:(1.03)
22. bolge_kategorisi=IC ANADOLU BOLGESI 32844 ==> saat_kategorisi=ogleden_once 20267 <conf:(0.62)> lift:(1.01) lev:(0.01) [292] conv:(1.02)
23. CINSIYET=Erkek bolge_kategorisi=IC ANADOLU BOLGESI 17050 ==> saat_kategorisi=ogleden_once 10503 <conf:(0.62)> lift:(1.01) lev:(0) [133] conv:(1.02)
24. yas_kategorisi=4-10_yas_arasi 20257 ==> saat_kategorisi=ogleden_once 12397 <conf:(0.61)> lift:(1.01) lev:(0) [77] conv:(1.01)
25. yas_kategorisi=0-3_yas_arasi 10522 ==> saat_kategorisi=ogleden_once 6432 <conf:(0.61)> lift:(1.01) lev:(0) [32] conv:(1.01)
26. CINSIYET=Kadin 23023 ==> saat_kategorisi=ogleden_once 14073 <conf:(0.61)> lift:(1.01) lev:(0) [71] conv:(1.01)
27. yas_kategorisi=11-18_yas_arasi bolge_kategorisi=IC ANADOLU BOLGESI 11834 ==> saat_kategorisi=ogleden_once 7206 <conf:(0.61)> lift:(1) lev:(0) [8] conv:(1)
28. CINSIYET=Erkek 25505 ==> saat_kategorisi=ogleden_once 15440 <conf:(0.61)> lift:(1) lev:(-0) [-71] conv:(0.99)
29. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 9405 ==> saat_kategorisi=ogleden_once 5676 <conf:(0.6)> lift:(0.99) lev:(-0) [-43] conv:(0.99)
30. CINSIYET=Erkek yas_kategorisi=4-10_yas_arasi 11100 ==> saat_kategorisi=ogleden_once 6697 <conf:(0.6)> lift:(0.99) lev:(-0) [-53] conv:(0.99)
31. yas_kategorisi=11-18_yas_arasi 17749 ==> saat_kategorisi=ogleden_once 10684 <conf:(0.6)> lift:(0.99) lev:(-0) [-110] conv:(0.98)
32. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 8344 ==> saat_kategorisi=ogleden_once 5008 <conf:(0.6)> lift:(0.99) lev:(-0) [-66] conv:(0.98)

Status

OK Log x0

Tablo 10. Çocuk Hastanesi Kayseri ili Hariç Acil Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örintüleri

1. yas_kategorisi=0-3_yas_arasi saat_kategorisi=ogleden_once 6432 ==> bolge_kategorisi=IC ANADOLU BOLGESI 5197 <conf:(0.81)> lift:(1.19) lev:(0.02) [843] conv:(1.68)
2. yas_kategorisi=0-3_yas_arasi 10522 ==> bolge_kategorisi=IC ANADOLU BOLGESI 8342 <conf:(0.79)> lift:(1.17) lev:(0.03) [1220] conv:(1.56)
3. bolum_kategorisi=NOROLOJI DEPARTMANI 6835 ==> bolge_kategorisi=IC ANADOLU BOLGESI 4859 <conf:(0.71)> lift:(1.05) lev:(0) [233] conv:(1.12)
4. saat_kategorisi=mesai_disi 8173 ==> bolge_kategorisi=IC ANADOLU BOLGESI 5718 <conf:(0.7)> lift:(1.03) lev:(0) [186] conv:(1.08)
5. CINSIYET=Kadin saat_kategorisi=ogleden_once 14073 ==> bolge_kategorisi=IC ANADOLU BOLGESI 9764 <conf:(0.69)> lift:(1.03) lev:(0) [239] conv:(1.06)
6. saat_kategorisi=ogleden_once 29513 ==> bolge_kategorisi=IC ANADOLU BOLGESI 20267 <conf:(0.69)> lift:(1.01) lev:(0.01) [292] conv:(1.03)
7. CINSIYET=Kadin 23023 ==> bolge_kategorisi=IC ANADOLU BOLGESI 15794 <conf:(0.69)> lift:(1.01) lev:(0) [211] conv:(1.03)
8. CINSIYET=Erkek saat_kategorisi=ogleden_once 15440 ==> bolge_kategorisi=IC ANADOLU BOLGESI 10503 <conf:(0.68)> lift:(1.01) lev:(0) [53] conv:(1.01)
9. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 10684 ==> bolge_kategorisi=IC ANADOLU BOLGESI 7206 <conf:(0.67)> lift:(1) lev:(-0) [-24] conv:(0.99)
10. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 8344 ==> bolge_kategorisi=IC ANADOLU BOLGESI 5627 <conf:(0.67)> lift:(1) lev:(-0) [-20] conv:(0.99)
11. CINSIYET=Erkek 25505 ==> bolge_kategorisi=IC ANADOLU BOLGESI 17050 <conf:(0.67)> lift:(0.99) lev:(-0) [-211] conv:(0.97)
12. yas_kategorisi=11-18_yas_arasi 17749 ==> bolge_kategorisi=IC ANADOLU BOLGESI 11834 <conf:(0.67)> lift:(0.99) lev:(-0) [-178] conv:(0.97)
13. CINSIYET=Kadin yas_kategorisi=4-10 _yas_arasi 9157 ==> bolge_kategorisi=IC ANADOLU BOLGESI 6060 <conf:(0.66)> lift:(0.98) lev:(-0) [-137] conv:(0.96)
14. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 9405 ==> bolge_kategorisi=IC ANADOLU BOLGESI 6207 <conf:(0.66)> lift:(0.98) lev:(-0) [-158] conv:(0.95)

15. yas_kategorisi=4-10 _yas_arasi saat_kategorisi=ogleden_once 12397 ==> bolge_kategorisi=IC ANADOLU BOLGESI 7864 <conf:(0.63)> lift:(0.94) lev:(-0.01) [-526] conv:(0.88)
16. saat_kategorisi=ogleden_sonra 10842 ==> bolge_kategorisi=IC ANADOLU BOLGESI 6859 <conf:(0.63)> lift:(0.93) lev:(-0.01) [-478] conv:(0.88)
17. yas_kategorisi=4-10 _yas_arasi 20257 ==> bolge_kategorisi=IC ANADOLU BOLGESI 12668 <conf:(0.63)> lift:(0.92) lev:(-0.02) [-1042] conv:(0.86)
18. yas_kategorisi=0-3_yas_arasi bolge_kategorisi=IC ANADOLU BOLGESI 8342 ==> saat_kategorisi=ogleden_once 5197 <conf:(0.62)> lift:(1.02) lev:(0) [123] conv:(1.04)
19. CINSIYET=Kadin yas_kategorisi=4-10 _yas_arasi 9157 ==> saat_kategorisi=ogleden_once 5700 <conf:(0.62)> lift:(1.02) lev:(0) [131] conv:(1.04)
20. yas_kategorisi=4-10 _yas_arasi bolge_kategorisi=IC ANADOLU BOLGESI 12668 ==> saat_kategorisi=ogleden_once 7864 <conf:(0.62)> lift:(1.02) lev:(0) [159] conv:(1.03)
21. CINSIYET=Kadin bolge_kategorisi=IC ANADOLU BOLGESI 15794 ==> saat_kategorisi=ogleden_once 9764 <conf:(0.62)> lift:(1.02) lev:(0) [158] conv:(1.03)
22. bolge_kategorisi=IC ANADOLU BOLGESI 32844 ==> saat_kategorisi=ogleden_once 20267 <conf:(0.62)> lift:(1.01) lev:(0.01) [292] conv:(1.02)
23. CINSIYET=Erkek bolge_kategorisi=IC ANADOLU BOLGESI 17050 ==> saat_kategorisi=ogleden_once 10503 <conf:(0.62)> lift:(1.01) lev:(0) [133] conv:(1.02)
24. yas_kategorisi=4-10 _yas_arasi 20257 ==> saat_kategorisi=ogleden_once 12397 <conf:(0.61)> lift:(1.01) lev:(0) [77] conv:(1.01)
25. yas_kategorisi=0-3_yas_arasi 10522 ==> saat_kategorisi=ogleden_once 6432 <conf:(0.61)> lift:(1.01) lev:(0) [32] conv:(1.01)
26. CINSIYET=Kadin 23023 ==> saat_kategorisi=ogleden_once 14073 <conf:(0.61)> lift:(1.01) lev:(0) [71] conv:(1.01)
27. yas_kategorisi=11-18_yas_arasi bolge_kategorisi=IC ANADOLU BOLGESI 11834 ==> saat_kategorisi=ogleden_once 7206 <conf:(0.61)> lift:(1) lev:(0) [8] conv:(1)
28. CINSIYET=Erkek 25505 ==> saat_kategorisi=ogleden_once 15440 <conf:(0.61)> lift:(1) lev:(-0) [-71] conv:(0.99)

29. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 9405 ==>
saat_kategorisi=ogleden_once 5676 <conf:(0.6)> lift:(0.99) lev:(-0) [-43]
conv:(0.99)
30. CINSIYET=Erkek yas_kategorisi=4-10 _yas_arasi 11100 ==>
saat_kategorisi=ogleden_once 6697 <conf:(0.6)> lift:(0.99) lev:(-0) [-53]
conv:(0.99)
31. yas_kategorisi=11-18_yas_arasi 17749 ==> saat_kategorisi=ogleden_once 10684
<conf:(0.6)> lift:(0.99) lev:(-0) [-110] conv:(0.98)
32. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 8344 ==>
saat_kategorisi=ogleden_once 5008 <conf:(0.6)> lift:(0.99) lev:(-0) [-66] conv

Tablo 10'da görüldüğü gibi Çocuk Hastanesi Kayseri ili hariç Acil Departmanı Hariç kategorisine ait 32 örüntü tespit edilmiştir. İlk örüntüye göre 0-3 yaş aralığında öğleden önce gelen 6.432 hastanın, 5.197 hastası İç Anadolu Bölgesinden geldiği %81 güven düzeyi ile anlaşılmaktadır.

İkinci örüntüye göre 0-3 yaş aralığındaki 10.522 hastanın, 8.342 hastası İç Anadolu Bölgesinden geldiği %79 güven düzeyi ile anlaşılmaktadır.

Üçüncü örüntüye göre Nöroloji Bölümüne gelen 6.835 hastanın, 4.859 hastası İç Anadolu Bölgesinden geldiği %71 güven düzeyi ile anlaşılmaktadır.

Geri kalan 29 örüntü için yapılan veri madenciliği incelendiğinde Pilot (Kayseri) ili ve Acil Bölümü dışındaki veriler incelendiğinde İç Anadolu Bölgesinin baskın şekilde yer aldığı anlaşılmaktadır.

Şekil 22. Çocuk Hastanesi Sadece İç Anadolu Bölgesi Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -N 100 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start Stop

Result list (right-click...)

- 16:05:15 - Apriori
- 16:05:39 - Apriori
- 16:05:57 - Apriori

Associator output

```
ay_kategorisi
CINSIYET
yas_kategorisi
saat_kategorisi
bolum_kategorisi

=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (3778 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11
Size of set of large itemsets L(2): 16
Size of set of large itemsets L(3): 3

Best rules found:

1. yas_kategorisi=0-3_yas_arasi 9053 ==> saat_kategorisi=ogleden_once 5376 <conf:(0.59)> lift:(1.05) lev:(0.01) [262] conv:(1.07)
2. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 6728 ==> saat_kategorisi=ogleden_once 3915 <conf:(0.58)> lift:(1.03) lev:(0) [114] conv:(1.04)
3. yas_kategorisi=11-18_yas_arasi 12848 ==> saat_kategorisi=ogleden_once 7435 <conf:(0.58)> lift:(1.02) lev:(0) [177] conv:(1.03)
4. yas_kategorisi=0-3_yas_arasi 9053 ==> CINSIYET=Erkek 5205 <conf:(0.57)> lift:(1.11) lev:(0.01) [500] conv:(1.13)
5. CINSIYET=Kadin 18145 ==> saat_kategorisi=ogleden_once 10287 <conf:(0.57)> lift:(1) lev:(0) [37] conv:(1)
6. CINSIYET=Erkek 19632 ==> saat_kategorisi=ogleden_once 11051 <conf:(0.56)> lift:(1) lev:(-0) [-37] conv:(1)
7. CINSIYET=Kadin yas_kategorisi=4-10_yas_arasi 7569 ==> saat_kategorisi=ogleden_once 4119 <conf:(0.54)> lift:(0.96) lev:(-0) [-156] conv:(0.95)
8. yas_kategorisi=4-10_yas_arasi 15876 ==> saat_kategorisi=ogleden_once 8527 <conf:(0.54)> lift:(0.95) lev:(-0.01) [-440] conv:(0.94)
9. CINSIYET=Erkek yas_kategorisi=4-10_yas_arasi 8307 ==> saat_kategorisi=ogleden_once 4408 <conf:(0.53)> lift:(0.94) lev:(-0.01) [-284] conv:(0.93)
10. saat_kategorisi=ogleden_sonra 8069 ==> CINSIYET=Erkek 4280 <conf:(0.53)> lift:(1.02) lev:(0) [86] conv:(1.02)
11. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 7435 ==> CINSIYET=Kadin 3915 <conf:(0.53)> lift:(1.1) lev:(0.01) [343] conv:(1.1)
12. yas_kategorisi=11-18_yas_arasi 12848 ==> CINSIYET=Kadin 6728 <conf:(0.52)> lift:(1.09) lev:(0.01) [556] conv:(1.09)
13. yas_kategorisi=4-10_yas_arasi 15876 ==> CINSIYET=Erkek 8307 <conf:(0.52)> lift:(1.01) lev:(0) [56] conv:(1.01)
14. saat_kategorisi=ogleden_once 21338 ==> CINSIYET=Erkek 11051 <conf:(0.52)> lift:(1) lev:(-0) [-37] conv:(1)
15. yas_kategorisi=4-10_yas_arasi saat_kategorisi=ogleden_once 8527 ==> CINSIYET=Erkek 4408 <conf:(0.52)> lift:(0.99) lev:(-0) [-23] conv:(0.99)
16. saat_kategorisi=mesai_disi 8370 ==> CINSIYET=Erkek 4301 <conf:(0.51)> lift:(0.99) lev:(-0) [-48] conv:(0.99)
```

Status

OK Log x0

Tablo 11. Çocuk Hastanesi Sadece İç Anadolu Bölgesi Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüleri

1. yas_kategorisi=0-3_yas_arasi 9053 ==> saat_kategorisi=ogleden_once 5376
<conf:(0.59)> lift:(1.05) lev:(0.01) [262] conv:(1.07)
2. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 6728 ==>
saat_kategorisi=ogleden_once 3915 <conf:(0.58)> lift:(1.03) lev:(0) [114]
conv:(1.04)
3. yas_kategorisi=11-18_yas_arasi 12848 ==> saat_kategorisi=ogleden_once 7435
<conf:(0.58)> lift:(1.02) lev:(0) [177] conv:(1.03)
4. yas_kategorisi=0-3_yas_arasi 9053 ==> CINSIYET=Erkek 5205 <conf:(0.57)>
lift:(1.11) lev:(0.01) [500] conv:(1.13)
5. CINSIYET=Kadin 18145 ==> saat_kategorisi=ogleden_once 10287
<conf:(0.57)> lift:(1) lev:(0) [37] conv:(1)
6. CINSIYET=Erkek 19632 ==> saat_kategorisi=ogleden_once 11051
<conf:(0.56)> lift:(1) lev:(-0) [-37] conv:(1)
7. CINSIYET=Kadin yas_kategorisi=4-10 _yas_arasi 7569 ==>
saat_kategorisi=ogleden_once 4119 <conf:(0.54)> lift:(0.96) lev:(-0) [-156]
conv:(0.95)
8. yas_kategorisi=4-10 _yas_arasi 15876 ==> saat_kategorisi=ogleden_once 8527
<conf:(0.54)> lift:(0.95) lev:(-0.01) [-440] conv:(0.94)
9. CINSIYET=Erkek yas_kategorisi=4-10 _yas_arasi 8307 ==>
saat_kategorisi=ogleden_once 4408 <conf:(0.53)> lift:(0.94) lev:(-0.01) [-284]
conv:(0.93)
10. saat_kategorisi=ogleden_sonra 8069 ==> CINSIYET=Erkek 4280
<conf:(0.53)> lift:(1.02) lev:(0) [86] conv:(1.02)
11. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 7435 ==>
CINSIYET=Kadin 3915 <conf:(0.53)> lift:(1.1) lev:(0.01) [343] conv:(1.1)
12. yas_kategorisi=11-18_yas_arasi 12848 ==> CINSIYET=Kadin 6728
<conf:(0.52)> lift:(1.09) lev:(0.01) [556] conv:(1.09)
13. yas_kategorisi=4-10 _yas_arasi 15876 ==> CINSIYET=Erkek 8307
<conf:(0.52)> lift:(1.01) lev:(0) [56] conv:(1.01)
14. saat_kategorisi=ogleden_once 21338 ==> CINSIYET=Erkek 11051
<conf:(0.52)> lift:(1) lev:(-0) [-37] conv:(1)
15. yas_kategorisi=4-10 _yas_arasi saat_kategorisi=ogleden_once 8527 ==>
CINSIYET=Erkek 4408 <conf:(0.52)> lift:(0.99) lev:(-0) [-23] conv:(0.99)

16. saat_kategorisi=mesai_disi 8370 ==> CINSIYET=Erkek 4301 <conf:(0.51)>
lift:(0.99) lev:(-0) [-48] conv:(0.99)

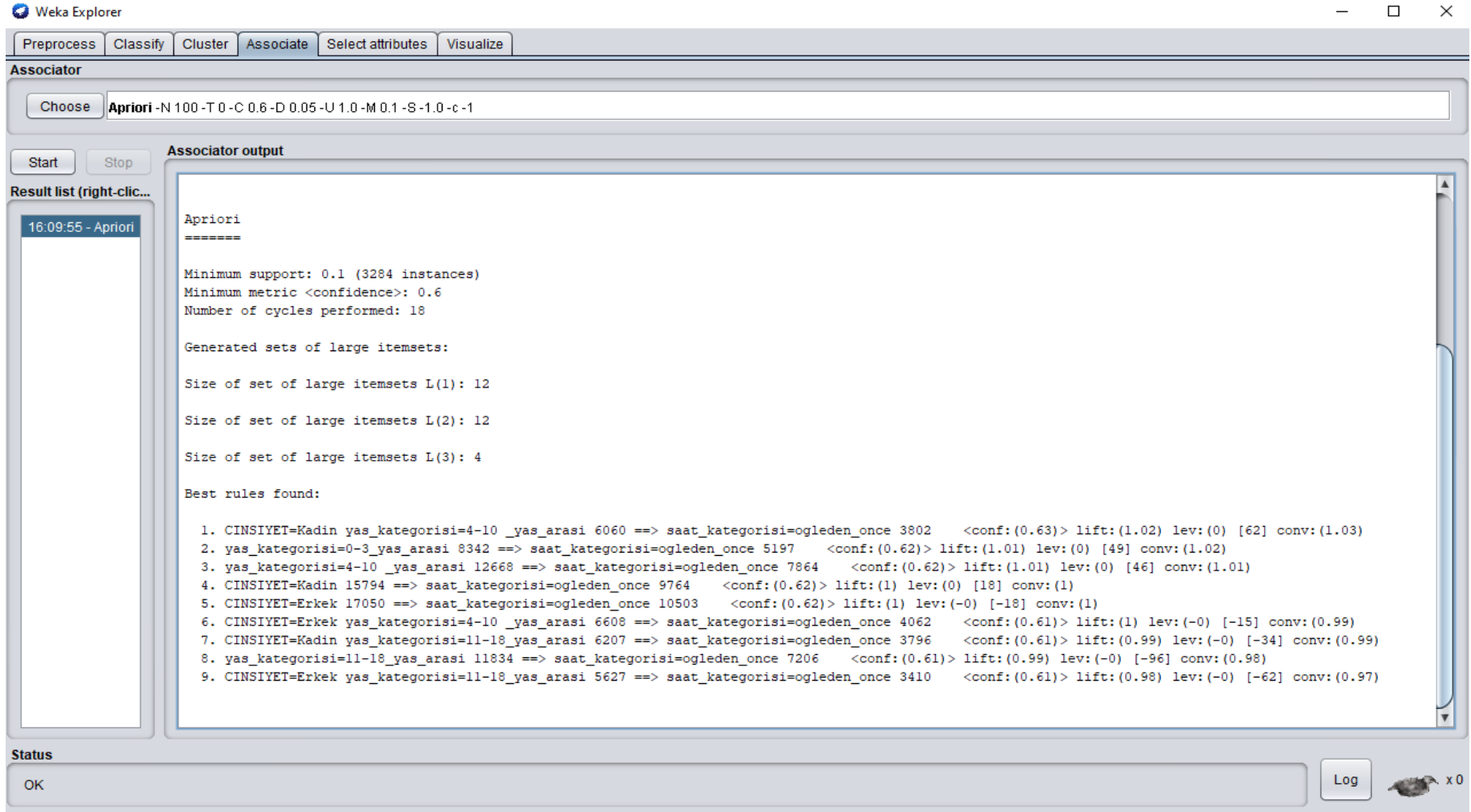
Tablo 11’de görüldüğü gibi Çocuk Hastanesi İç Anadolu Bölgesi kategorisine ait 16 örüntü tespit edilmiştir. İlk örüntüye göre 0-3 yaş aralığında 9.053 hastanın, 5.376 ‘sının öğleden önce hastaneye geldiği %59 güven düzeyi ile anlaşılmaktadır.

İkinci örüntüye göre cinsiyeti kadın ve 11-18 yaş aralığındaki 6.728 hastanın, 3.915 hastası öğleden önce hastaneye geldiği %58 güven düzeyi ile anlaşılmaktadır.

Üçüncü örüntüye göre 11-18 yaş aralığındaki 12.848 hastanın, 7.435 hastası öğleden önce hastaneye geldiği %58 güven düzeyi ile anlaşılmaktadır.

Geri kalan 13 örüntü için yapılan veri madenciliği İç Anadolu Bölgesinden 11-18 yaş aralığında hasta sayısının daha çok olduğu ve cinsiyet ağırlığının erkek olduğu bilgisine ulaşılmıştır.

Şekil 23. Çocuk Hastanesi Sadece İç Anadolu Bölgesi Acil Bölümü Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -N 100 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start Stop

Associator output

Result list (right-click...)

16:09:55 - Apriori

```
Apriori
=====

Minimum support: 0.1 (3284 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 12
Size of set of large itemsets L(3): 4

Best rules found:

1. CINSIYET=Kadin yas_kategorisi=4-10 _yas_arasi 6060 ==> saat_kategorisi=ogleden_once 3802 <conf:(0.63)> lift:(1.02) lev:(0) [62] conv:(1.03)
2. yas_kategorisi=0-3 _yas_arasi 8342 ==> saat_kategorisi=ogleden_once 5197 <conf:(0.62)> lift:(1.01) lev:(0) [49] conv:(1.02)
3. yas_kategorisi=4-10 _yas_arasi 12668 ==> saat_kategorisi=ogleden_once 7864 <conf:(0.62)> lift:(1.01) lev:(0) [46] conv:(1.01)
4. CINSIYET=Kadin 15794 ==> saat_kategorisi=ogleden_once 9764 <conf:(0.62)> lift:(1) lev:(0) [18] conv:(1)
5. CINSIYET=Erkek 17050 ==> saat_kategorisi=ogleden_once 10503 <conf:(0.62)> lift:(1) lev:(-0) [-18] conv:(1)
6. CINSIYET=Erkek yas_kategorisi=4-10 _yas_arasi 6608 ==> saat_kategorisi=ogleden_once 4062 <conf:(0.61)> lift:(1) lev:(-0) [-15] conv:(0.99)
7. CINSIYET=Kadin yas_kategorisi=11-18 _yas_arasi 6207 ==> saat_kategorisi=ogleden_once 3796 <conf:(0.61)> lift:(0.99) lev:(-0) [-34] conv:(0.99)
8. yas_kategorisi=11-18 _yas_arasi 11834 ==> saat_kategorisi=ogleden_once 7206 <conf:(0.61)> lift:(0.99) lev:(-0) [-96] conv:(0.98)
9. CINSIYET=Erkek yas_kategorisi=11-18 _yas_arasi 5627 ==> saat_kategorisi=ogleden_once 3410 <conf:(0.61)> lift:(0.98) lev:(-0) [-62] conv:(0.97)
```

Status

OK Log x0

Tablo 12. Çocuk Hastanesi Sadece İç Anadolu Bölgesi Acil Bölümü Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüleri

1. CINSİYET=Kadin yas_kategorisi=4-10 _yas_arasi 6060 ==> saat_kategorisi=ogleden_once 3802 <conf:(0.63)> lift:(1.02) lev:(0) [62] conv:(1.03)
2. yas_kategorisi=0-3_yas_arasi 8342 ==> saat_kategorisi=ogleden_once 5197 <conf:(0.62)> lift:(1.01) lev:(0) [49] conv:(1.02)
3. yas_kategorisi=4-10 _yas_arasi 12668 ==> saat_kategorisi=ogleden_once 7864 <conf:(0.62)> lift:(1.01) lev:(0) [46] conv:(1.01)
4. CINSİYET=Kadin 15794 ==> saat_kategorisi=ogleden_once 9764 <conf:(0.62)> lift:(1) lev:(0) [18] conv:(1)
5. CINSİYET=Erkek 17050 ==> saat_kategorisi=ogleden_once 10503 <conf:(0.62)> lift:(1) lev:(-0) [-18] conv:(1)
6. CINSİYET=Erkek yas_kategorisi=4-10 _yas_arasi 6608 ==> saat_kategorisi=ogleden_once 4062 <conf:(0.61)> lift:(1) lev:(-0) [-15] conv:(0.99)
7. CINSİYET=Kadin yas_kategorisi=11-18_yas_arasi 6207 ==> saat_kategorisi=ogleden_once 3796 <conf:(0.61)> lift:(0.99) lev:(-0) [-34] conv:(0.99)
8. yas_kategorisi=11-18_yas_arasi 11834 ==> saat_kategorisi=ogleden_once 7206 <conf:(0.61)> lift:(0.99) lev:(-0) [-96] conv:(0.98)
9. CINSİYET=Erkek yas_kategorisi=11-18_yas_arasi 5627 ==> saat_kategorisi=ogleden_once 3410 <conf:(0.61)> lift:(0.98) lev:(-0) [-62] conv:(0.97)

Tablo 12’de görüldüğü gibi Çocuk Hastanesi İç Anadolu Bölgesi Acil Hariç kategorisine ait 9 örüntü tespit edilmiştir. İlk örüntüye göre cinsiyeti kadın olan 4-10 yaş aralığında 6.060 hastanın, 3.802 ‘sının öğleden önce hastaneye geldiği %63 güven düzeyi ile anlaşılmaktadır.

İkinci örüntüye göre 0-3 yaş aralığındaki 8.342 hastanın, 5.197 ‘sının öğleden önce hastaneye geldiği %62 güven düzeyi ile anlaşılmaktadır.

Üçüncü örüntüye göre 4-10 yaş aralığındaki 12.668 hastanın, 7.864 hastası öğleden önce hastaneye geldiği %62 güven düzeyi ile anlaşılmaktadır.

Geri kalan 6 örüntü için yapılan veri madenciliđi hastaların genelinin öğleden önce hastaneye gelip muayene olduđu bilgisine ulařılmıştır.



Şekil 24. Çocuk Hastanesi İç Anadolu Bölgesi Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -N 100 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start Stop

Associator output

Result list (right-click...)

18:05:19 - Apriori

Minimum metric <confidence>: 0.6
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 26
Size of set of large itemsets L(3): 14

Best rules found:

1. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 3961 ==> bolge_kategorisi=AKDENIZ BOLGESI 2837 <conf:(0.72)> lift:(1.16) lev:(0.02) [386] conv:(1.34)
2. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 3729 ==> bolge_kategorisi=AKDENIZ BOLGESI 2625 <conf:(0.7)> lift:(1.14) lev:(0.01) [317] conv:(1.29)
3. yas_kategorisi=11-18_yas_arasi 7487 ==> bolge_kategorisi=AKDENIZ BOLGESI 5203 <conf:(0.69)> lift:(1.12) lev:(0.02) [570] conv:(1.25)
4. yas_kategorisi=4-10_yas_arasi saat_kategorisi=mesai_disi 3701 ==> bolum_kategorisi=ACIL DEPARTMANI 2555 <conf:(0.69)> lift:(2.2) lev:(0.06) [1395] conv:(2.22)
5. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 3526 ==> bolge_kategorisi=AKDENIZ BOLGESI 2366 <conf:(0.67)> lift:(1.08) lev:(0.01) [184] conv:(1.16)
6. bolum_kategorisi=ACIL DEPARTMANI bolge_kategorisi=AKDENIZ BOLGESI 4147 ==> yas_kategorisi=4-10_yas_arasi 2780 <conf:(0.67)> lift:(1.27) lev:(0.03) [594] conv:(1.43)
7. CINSIYET=Kadin yas_kategorisi=4-10_yas_arasi 5066 ==> bolge_kategorisi=AKDENIZ BOLGESI 3348 <conf:(0.66)> lift:(1.07) lev:(0.01) [213] conv:(1.12)
8. CINSIYET=Kadin saat_kategorisi=ogleden_once 4910 ==> bolge_kategorisi=AKDENIZ BOLGESI 3196 <conf:(0.65)> lift:(1.05) lev:(0.01) [157] conv:(1.09)
9. yas_kategorisi=4-10_yas_arasi saat_kategorisi=ogleden_once 5396 ==> bolge_kategorisi=AKDENIZ BOLGESI 3510 <conf:(0.65)> lift:(1.05) lev:(0.01) [171] conv:(1.09)
10. yas_kategorisi=4-10_yas_arasi 12037 ==> bolge_kategorisi=AKDENIZ BOLGESI 7767 <conf:(0.65)> lift:(1.04) lev:(0.01) [318] conv:(1.07)
11. yas_kategorisi=4-10_yas_arasi saat_kategorisi=mesai_disi 3701 ==> bolge_kategorisi=AKDENIZ BOLGESI 2385 <conf:(0.64)> lift:(1.04) lev:(0) [94] conv:(1.07)
12. CINSIYET=Erkek bolum_kategorisi=ACIL DEPARTMANI 3884 ==> yas_kategorisi=4-10_yas_arasi 2479 <conf:(0.64)> lift:(1.21) lev:(0.02) [431] conv:(1.31)
13. saat_kategorisi=ogleden_once 10563 ==> bolge_kategorisi=AKDENIZ BOLGESI 6725 <conf:(0.64)> lift:(1.03) lev:(0.01) [188] conv:(1.05)
14. CINSIYET=Kadin 10500 ==> bolge_kategorisi=AKDENIZ BOLGESI 6675 <conf:(0.64)> lift:(1.03) lev:(0.01) [177] conv:(1.05)
15. CINSIYET=Erkek yas_kategorisi=4-10_yas_arasi 6971 ==> bolge_kategorisi=AKDENIZ BOLGESI 4419 <conf:(0.63)> lift:(1.02) lev:(0) [105] conv:(1.04)
16. saat_kategorisi=mesai_disi 6612 ==> bolum_kategorisi=ACIL DEPARTMANI 4157 <conf:(0.63)> lift:(2.01) lev:(0.09) [2085] conv:(1.85)
17. yas_kategorisi=4-10_yas_arasi bolum_kategorisi=ACIL DEPARTMANI 4448 ==> bolge_kategorisi=AKDENIZ BOLGESI 2780 <conf:(0.63)> lift:(1.01) lev:(0) [27] conv:(1.02)
18. CINSIYET=Erkek saat_kategorisi=ogleden_once 5653 ==> bolge_kategorisi=AKDENIZ BOLGESI 3529 <conf:(0.62)> lift:(1.01) lev:(0) [31] conv:(1.01)
19. bolum_kategorisi=ACIL DEPARTMANI 7155 ==> yas_kategorisi=4-10_yas_arasi 4448 <conf:(0.62)> lift:(1.18) lev:(0.03) [677] conv:(1.25)
20. saat_kategorisi=mesai_disi bolum_kategorisi=ACIL DEPARTMANI 4157 ==> yas_kategorisi=4-10_yas_arasi 2555 <conf:(0.61)> lift:(1.17) lev:(0.02) [364] conv:(1.23)
21. saat_kategorisi=mesai_disi bolge_kategorisi=AKDENIZ BOLGESI 3983 ==> bolum_kategorisi=ACIL DEPARTMANI 2428 <conf:(0.61)> lift:(1.95) lev:(0.05) [1180] conv:(1.76)
22. saat_kategorisi=ogleden_sonra 5664 ==> bolge_kategorisi=AKDENIZ BOLGESI 3424 <conf:(0.6)> lift:(0.98) lev:(-0) [-80] conv:(0.96)
23. CINSIYET=Erkek 12339 ==> bolge_kategorisi=AKDENIZ BOLGESI 7457 <conf:(0.6)> lift:(0.98) lev:(-0.01) [-177] conv:(0.96)
24. saat_kategorisi=mesai_disi 6612 ==> bolge_kategorisi=AKDENIZ BOLGESI 3983 <conf:(0.6)> lift:(0.97) lev:(-0) [-108] conv:(0.96)

Status

OK Log x0

Tablo 13. Çocuk Hastanesi İç Anadolu Bölgesi Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüleri

1. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 3961 ==>
bolge_kategorisi=AKDENIZ BOLGESI 2837 <conf:(0.72)> lift:(1.16)
lev:(0.02) [386] conv:(1.34)
2. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 3729 ==>
bolge_kategorisi=AKDENIZ BOLGESI 2625 <conf:(0.7)> lift:(1.14) lev:(0.01)
[317] conv:(1.29)
3. yas_kategorisi=11-18_yas_arasi 7487 ==> bolge_kategorisi=AKDENIZ
BOLGESI 5203 <conf:(0.69)> lift:(1.12) lev:(0.02) [570] conv:(1.25)
4. yas_kategorisi=4-10 _yas_arasi saat_kategorisi=mesai_disi 3701 ==>
bolum_kategorisi=ACIL DEPARTMANI 2555 <conf:(0.69)> lift:(2.2)
lev:(0.06) [1395] conv:(2.22)
5. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 3526 ==>
bolge_kategorisi=AKDENIZ BOLGESI 2366 <conf:(0.67)> lift:(1.08)
lev:(0.01) [184] conv:(1.16)
6. bolum_kategorisi=ACIL DEPARTMANI bolge_kategorisi=AKDENIZ
BOLGESI 4147 ==> yas_kategorisi=4-10 _yas_arasi 2780 <conf:(0.67)>
lift:(1.27) lev:(0.03) [594] conv:(1.43)
7. CINSIYET=Kadin yas_kategorisi=4-10 _yas_arasi 5066 ==>
bolge_kategorisi=AKDENIZ BOLGESI 3348 <conf:(0.66)> lift:(1.07)
lev:(0.01) [213] conv:(1.12)
8. CINSIYET=Kadin saat_kategorisi=ogleden_once 4910 ==>
bolge_kategorisi=AKDENIZ BOLGESI 3196 <conf:(0.65)> lift:(1.05)
lev:(0.01) [157] conv:(1.09)
9. yas_kategorisi=4-10 _yas_arasi saat_kategorisi=ogleden_once 5396 ==>
bolge_kategorisi=AKDENIZ BOLGESI 3510 <conf:(0.65)> lift:(1.05)
lev:(0.01) [171] conv:(1.09)
10. yas_kategorisi=4-10 _yas_arasi 12037 ==> bolge_kategorisi=AKDENIZ
BOLGESI 7767 <conf:(0.65)> lift:(1.04) lev:(0.01) [318] conv:(1.07)
11. yas_kategorisi=4-10 _yas_arasi saat_kategorisi=mesai_disi 3701 ==>
bolge_kategorisi=AKDENIZ BOLGESI 2385 <conf:(0.64)> lift:(1.04) lev:(0)
[94] conv:(1.07)
12. CINSIYET=Erkek bolum_kategorisi=ACIL DEPARTMANI 3884 ==>
yas_kategorisi=4-10 _yas_arasi 2479 <conf:(0.64)> lift:(1.21) lev:(0.02) [431]
conv:(1.31)

13. saat_kategorisi=ogleden_once 10563 ==> bolge_kategorisi=AKDENIZ BOLGESI 6725 <conf:(0.64)> lift:(1.03) lev:(0.01) [188] conv:(1.05)
14. CINSIYET=Kadin 10500 ==> bolge_kategorisi=AKDENIZ BOLGESI 6675 <conf:(0.64)> lift:(1.03) lev:(0.01) [177] conv:(1.05)
15. CINSIYET=Erkek yas_kategorisi=4-10 _yas_arasi 6971 ==> bolge_kategorisi=AKDENIZ BOLGESI 4419 <conf:(0.63)> lift:(1.02) lev:(0) [105] conv:(1.04)
16. saat_kategorisi=mesai_disi 6612 ==> bolum_kategorisi=ACIL DEPARTMANI 4157 <conf:(0.63)> lift:(2.01) lev:(0.09) [2085] conv:(1.85)
17. yas_kategorisi=4-10 _yas_arasi bolum_kategorisi=ACIL DEPARTMANI 4448 ==> bolge_kategorisi=AKDENIZ BOLGESI 2780 <conf:(0.63)> lift:(1.01) lev:(0) [27] conv:(1.02)
18. CINSIYET=Erkek saat_kategorisi=ogleden_once 5653 ==> bolge_kategorisi=AKDENIZ BOLGESI 3529 <conf:(0.62)> lift:(1.01) lev:(0) [31] conv:(1.01)
19. bolum_kategorisi=ACIL DEPARTMANI 7155 ==> yas_kategorisi=4-10 _yas_arasi 4448 <conf:(0.62)> lift:(1.18) lev:(0.03) [677] conv:(1.25)
20. saat_kategorisi=mesai_disi bolum_kategorisi=ACIL DEPARTMANI 4157 ==> yas_kategorisi=4-10 _yas_arasi 2555 <conf:(0.61)> lift:(1.17) lev:(0.02) [364] conv:(1.23)
21. saat_kategorisi=mesai_disi bolge_kategorisi=AKDENIZ BOLGESI 3983 ==> bolum_kategorisi=ACIL DEPARTMANI 2428 <conf:(0.61)> lift:(1.95) lev:(0.05) [1180] conv:(1.76)
22. saat_kategorisi=ogleden_sonra 5664 ==> bolge_kategorisi=AKDENIZ BOLGESI 3424 <conf:(0.6)> lift:(0.98) lev:(-0) [-80] conv:(0.96)
23. CINSIYET=Erkek 12339 ==> bolge_kategorisi=AKDENIZ BOLGESI 7457 <conf:(0.6)> lift:(0.98) lev:(-0.01) [-177] conv:(0.96)
24. saat_kategorisi=mesai_disi 6612 ==> bolge_kategorisi=AKDENIZ BOLGESI 3983 <conf:(0.6)> lift:(0.97) lev:(-0) [-108] conv:(0.96)

Tablo 13’de görüldüğü gibi Çocuk Hastanesi İç Anadolu Bölgesi Hariç kategorisine ait 24 örüntü tespit edilmiştir. İlk örüntüye göre cinsiyeti kadın olan 11-18 yaş aralığında 3.961 hastanın, 2.837 hastasının Akdeniz Bölgesinden geldiği %72 güven düzeyi ile anlaşılmaktadır.

İkinci örüntüye göre 11-18 yaş aralığında öğleden önce gelen 3.729 hastanın, 2.625 hastasının Akdeniz Bölgesinden geldiği %70 güven düzeyi ile anlaşılmaktadır.

Üçüncü örüntüye göre 11-18 yaş aralığındaki 7.487 hastanın, 5203 hastası Akdeniz Bölgesinden geldiği %69 güven düzeyi ile anlaşılmaktadır.

Geri kalan 21 örüntü için yapılan veri madenciliği İç Anadolu bölgesi hariç tutularak yapılan değerlendirmede hastaların genelinin Akdeniz Bölgesinden muayeneye gelen hastaların oluşturduğu anlaşılmaktadır.



Şekil 25. Çocuk Hastanesi İç Anadolu Bölgesi Hariç ve Acil Bölümü Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntülerin Ekran Görüntüsü

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -N 100 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start Stop

Associator output

Result list (right-click...)

18:09:14 - Apriori

Minimum support: 0.1 (1568 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 15
Size of set of large itemsets L(2): 17
Size of set of large itemsets L(3): 12
Size of set of large itemsets L(4): 1

Best rules found:

1. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 3198 ==> bolge_kategorisi=AKDENIZ BOLGESI 2279 <conf:(0.71)> lift:(1.12) lev:(0.02) [243] conv:(1.26)
2. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 3478 ==> bolge_kategorisi=AKDENIZ BOLGESI 2444 <conf:(0.7)> lift:(1.1) lev:(0.01) [229] conv:(1.22)
3. yas_kategorisi=11-18_yas_arasi 5915 ==> bolge_kategorisi=AKDENIZ BOLGESI 4094 <conf:(0.69)> lift:(1.09) lev:(0.02) [328] conv:(1.18)
4. CINSIYET=Kadin yas_kategorisi=4-10_yas_arasi 3097 ==> bolge_kategorisi=AKDENIZ BOLGESI 2070 <conf:(0.67)> lift:(1.05) lev:(0.01) [98] conv:(1.09)
5. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 2717 ==> bolge_kategorisi=AKDENIZ BOLGESI 1815 <conf:(0.67)> lift:(1.05) lev:(0.01) [85] conv:(1.09)
6. CINSIYET=Kadin saat_kategorisi=ogleden_once 4309 ==> bolge_kategorisi=AKDENIZ BOLGESI 2849 <conf:(0.66)> lift:(1.04) lev:(0.01) [105] conv:(1.07)
7. yas_kategorisi=4-10_yas_arasi 7589 ==> bolge_kategorisi=AKDENIZ BOLGESI 4987 <conf:(0.66)> lift:(1.03) lev:(0.01) [155] conv:(1.06)
8. yas_kategorisi=4-10_yas_arasi saat_kategorisi=ogleden_once 4533 ==> bolge_kategorisi=AKDENIZ BOLGESI 2976 <conf:(0.66)> lift:(1.03) lev:(0.01) [90] conv:(1.06)
9. CINSIYET=Kadin 7229 ==> bolge_kategorisi=AKDENIZ BOLGESI 4720 <conf:(0.65)> lift:(1.03) lev:(0.01) [117] conv:(1.05)
10. CINSIYET=Erkek yas_kategorisi=4-10_yas_arasi 4492 ==> bolge_kategorisi=AKDENIZ BOLGESI 2917 <conf:(0.65)> lift:(1.02) lev:(0) [57] conv:(1.04)
11. CINSIYET=Erkek yas_kategorisi=4-10_yas_arasi saat_kategorisi=ogleden_once 2635 ==> bolge_kategorisi=AKDENIZ BOLGESI 1703 <conf:(0.65)> lift:(1.02) lev:(0) [25] conv:(1.03)
12. saat_kategorisi=ogleden_once 9246 ==> bolge_kategorisi=AKDENIZ BOLGESI 5960 <conf:(0.64)> lift:(1.01) lev:(0) [73] conv:(1.02)
13. CINSIYET=Erkek saat_kategorisi=ogleden_once 4937 ==> bolge_kategorisi=AKDENIZ BOLGESI 3111 <conf:(0.63)> lift:(0.99) lev:(-0) [-32] conv:(0.98)
14. CINSIYET=Erkek 8455 ==> bolge_kategorisi=AKDENIZ BOLGESI 5265 <conf:(0.62)> lift:(0.98) lev:(-0.01) [-117] conv:(0.96)
15. saat_kategorisi=ogleden_sonra 3983 ==> bolge_kategorisi=AKDENIZ BOLGESI 2470 <conf:(0.62)> lift:(0.97) lev:(-0) [-65] conv:(0.96)
16. CINSIYET=Kadin yas_kategorisi=4-10_yas_arasi 3097 ==> saat_kategorisi=ogleden_once 1898 <conf:(0.61)> lift:(1.04) lev:(0) [72] conv:(1.06)
17. CINSIYET=Kadin bolge_kategorisi=AKDENIZ BOLGESI 4720 ==> saat_kategorisi=ogleden_once 2849 <conf:(0.6)> lift:(1.02) lev:(0) [66] conv:(1.03)

Status

OK Log x 0

Tablo 14. Çocuk Hastanesi İç Anadolu Bölgesi Hariç ve Acil Bölümü Hariç Verisi Kullanılarak Apriori Algoritması ile Wekada Üretilen Örüntüleri

1. CINSIYET=Kadin yas_kategorisi=11-18_yas_arasi 3198 ==>
bolge_kategorisi=AKDENIZ BOLGESI 2279 <conf:(0.71)> lift:(1.12)
lev:(0.02) [243] conv:(1.26)
2. yas_kategorisi=11-18_yas_arasi saat_kategorisi=ogleden_once 3478 ==>
bolge_kategorisi=AKDENIZ BOLGESI 2444 <conf:(0.7)> lift:(1.1) lev:(0.01)
[229] conv:(1.22)
3. yas_kategorisi=11-18_yas_arasi 5915 ==> bolge_kategorisi=AKDENIZ
BOLGESI 4094 <conf:(0.69)> lift:(1.09) lev:(0.02) [328] conv:(1.18)
4. CINSIYET=Kadin yas_kategorisi=4-10 _yas_arasi 3097 ==>
bolge_kategorisi=AKDENIZ BOLGESI 2070 <conf:(0.67)> lift:(1.05)
lev:(0.01) [98] conv:(1.09)
5. CINSIYET=Erkek yas_kategorisi=11-18_yas_arasi 2717 ==>
bolge_kategorisi=AKDENIZ BOLGESI 1815 <conf:(0.67)> lift:(1.05)
lev:(0.01) [85] conv:(1.09)
6. CINSIYET=Kadin saat_kategorisi=ogleden_once 4309 ==>
bolge_kategorisi=AKDENIZ BOLGESI 2849 <conf:(0.66)> lift:(1.04)
lev:(0.01) [105] conv:(1.07)
7. yas_kategorisi=4-10 _yas_arasi 7589 ==> bolge_kategorisi=AKDENIZ
BOLGESI 4987 <conf:(0.66)> lift:(1.03) lev:(0.01) [155] conv:(1.06)
8. yas_kategorisi=4-10 _yas_arasi saat_kategorisi=ogleden_once 4533 ==>
bolge_kategorisi=AKDENIZ BOLGESI 2976 <conf:(0.66)> lift:(1.03)
lev:(0.01) [90] conv:(1.06)
9. CINSIYET=Kadin 7229 ==> bolge_kategorisi=AKDENIZ BOLGESI 4720
<conf:(0.65)> lift:(1.03) lev:(0.01) [117] conv:(1.05)
10. CINSIYET=Erkek yas_kategorisi=4-10 _yas_arasi 4492 ==>
bolge_kategorisi=AKDENIZ BOLGESI 2917 <conf:(0.65)> lift:(1.02) lev:(0)
[57] conv:(1.04)
11. CINSIYET=Erkek yas_kategorisi=4-10 _yas_arasi saat_kategorisi=ogleden_once
2635 ==> bolge_kategorisi=AKDENIZ BOLGESI 1703 <conf:(0.65)>
lift:(1.02) lev:(0) [25] conv:(1.03)
12. saat_kategorisi=ogleden_once 9246 ==> bolge_kategorisi=AKDENIZ BOLGESI
5960 <conf:(0.64)> lift:(1.01) lev:(0) [73] conv:(1.02)

13. CINSIYET=Erkek saat_kategorisi=ogleden_once 4937 ==> bolge_kategorisi=AKDENIZ BOLGESI 3111 <conf:(0.63)> lift:(0.99) lev:(-0) [-32] conv:(0.98)
14. CINSIYET=Erkek 8455 ==> bolge_kategorisi=AKDENIZ BOLGESI 5265 <conf:(0.62)> lift:(0.98) lev:(-0.01) [-117] conv:(0.96)
15. saat_kategorisi=ogleden_sonra 3983 ==> bolge_kategorisi=AKDENIZ BOLGESI 2470 <conf:(0.62)> lift:(0.97) lev:(-0) [-65] conv:(0.96)
16. CINSIYET=Kadin yas_kategorisi=4-10 _yas_arasi 3097 ==> saat_kategorisi=ogleden_once 1898 <conf:(0.61)> lift:(1.04) lev:(0) [72] conv:(1.06)
17. CINSIYET=Kadin bolge_kategorisi=AKDENIZ BOLGESI 4720 ==> saat_kategorisi=ogleden_once 2849 <conf:(0.6)> lift:(1.02) lev:(0) [66] conv:(1.03)

Tablo 14'de görüldüğü gibi Çocuk Hastanesi İç Anadolu Bölgesi ve Acil Bölümü Hariç kategorisine ait 17 örüntü tespit edilmiştir. İlk örüntüye göre cinsiyeti kadın olan 11-18 yaş aralığında 3.198 hastanın, 2.279 hastasının Akdeniz Bölgesinden geldiği %71 güven düzeyi ile anlaşılmaktadır.

İkinci örüntüye göre 11-18 yaş aralığında öğleden önce gelen 3.478 hastanın, 2.444 hastasının Akdeniz Bölgesinden geldiği %70 güven düzeyi ile anlaşılmaktadır.

Üçüncü örüntüye göre 11-18 yaş aralığındaki 5.915 hastanın, 4094 hastası Akdeniz Bölgesinden geldiği %69 güven düzeyi ile anlaşılmaktadır.

Geri kalan 14 örüntü için yapılan veri madenciliği İç Anadolu bölgesi, Acil Bölümü hariç tutularak yapılan değerlendirmede hastaların genelinin Akdeniz Bölgesinden muayeneye gelen hastaların oluşturduğu anlaşılmaktadır.

4. DEĞERLENDİRME VE SONUÇ

Makine öğrenme yöntemleri hem hasta verilerinin içinde barındırdıkları yeni modelleri ve ilginç bilgileri keşfedip anlamlandırabilmeyi hem de elde edilen neticeleri ve/veya risk değerlendirmelerini sınıflandırma ve tahmine yönelik çalışmalar için kullanabilmeyi sağlamaktadır. Hasta verileri kimi zaman sağlık personeline tıbbi bakım için yol göstermesinin yanında kimi zaman da sağlık hizmetlerinin daha iyi gerçekleştirilmesi ve hasta memnuniyetinin sağlanması amacı ile hasta profilinin ortaya konması şeklinde de gerçekleşebilmektedir.

Bu çalışma ile karar alma süreçlerinde hastane kullanımı konusunda yapılan örüntüler incelenerek hastane yöneticilerine hastaların cinsiyeti, yaş aralıkları, geliş zamanları, geldikleri bölgeler, geldikleri aylar ve yararlandıkları bölümler konularında ışık tutmak amacı ile birliktelik kuralları kullanılarak veri madenciliği çalışması yürütülmüştür. Birliktelik kuralları yöntemi uygulanırken günümüzün algoritmaları içinde yer alan Apriori algoritması ile yapılan veri madenciliğinin en kullanışlı ve elde edilen neticeler bağlamında en anlamlı ve tutarlı sonuçları verdiği söylenebilmektedir (Parraga-Alava, vd., 2018, s. 24). Özellikle sağlık hizmetleri alanında büyük veri miktarlarının olduğu düşünülürse bu algoritmanın güvenle kullanılan bir algoritma olarak çalışmalarda kullanılmak üzere uygulanabileceği görülmektedir. Bu çalışma kapsamında, Kayseri’de bulunan Erciyes Üniversitesi hastaneleri verileri kullanılarak, hasta profil ilişkileri birliktelik kurallarının var olup olmadığı anlaşılmasına çalışılmıştır. Apriori algoritmasının sağlık hizmetlerinde kullanılması, sonuçların sezgisel ve anlaşılabilir olarak sunulduğu bir veri madenciliği yaklaşımını ve veri ön işleme ile karışıklığı oluşturan birçok kuralın sadeleştirilmesini sağlamıştır (Mahmoodi, vd., 2016, s.14).

Çalışmaya başlandığında 2022 yılına ait bir milyondan fazla ham hasta kaydı ve bu hastalara ait 9 veri seti bulunmaktaydı. Bu veri setleri; gün, hasta adı (K.V.K.K gereği

maskelenmiş olarak Adın ilk harfi ve Soyadın ilk harfi şeklinde), cinsiyet, yaş, saat, il, ilçe, mahalle ve poliklinik veri setlerinden oluşmaktaydı. Hasta profillerinin çıkarılması ve hasta profil analizlerinin güvenli şekilde yapılabilmesi amacı ile ham verinin işlenmesi ve bir dizi veri ön işleminden geçirilmesi gerekmiştir. 9 veri seti içinde Bölüm Kategorisi 20 farklı bölüm değeri alabilmekte olduğu için anlamlı veriye ulaşmak için bölüm kategorisi Çocuk Hastanesi ile sınırlandırılmıştır. Gereksiz olan veriler temizlenerek; adı soyadı, ilçe, mahalle bilgileri silinmiştir.

Büyük bir veri hacmine sahip olan veri tabanı içerisinde örüntülerin engellenmemesi adına bir takım veri ön işleme adımları gerçekleştirilmiştir. Bu adımlardan biri veri temizleme adımıdır. Mevcut veri setinde eksik verilerin bulunduğu fark edilmiştir. Bu eksik verilerin türetilmesi yerine ilgili veri satırının silinmesi kararlaştırılmıştır. Bunun en büyük sebebi mevcut veri setindeki satır sayısının çok fazla olmasıdır. Bu verilerin silinmesi veri setinde büyük etkiler oluşturmamakla birlikte eksik verilerin tamamlanarak veri setine eklenmesi örüntülerin kaybına da neden olabilmektedir. Böylece mevcut veri setinin güvenilirliği artırılmıştır. Eksik bilgi içeren satırlar (yaş bilgisi, şehir bilgisi vb.) silinmiştir. Veri setleri içinde yer alan veri tiplerinin azaltılmasına yönelik olarak veri dönüştürme işlemleri gerçekleştirilmiştir. Bu bağlamda yer alan cinsiyet bilgisi 2 kategoriye, yaş ve saat bilgileri 3 ana kategoriye indirgenmiş, 81 il bilgisi Kayseri (Pilot) olmak üzere 8 kategoride incelenmiş, Çocuk Hastanesinde yer alan 20 poliklinik veri dönüştürme ile veri ön işleme süreçlerine tabi tutulmuştur.

Veri ön işleme adımları esnasında gürültü oluşturabilecek uyumsuz veriler güvenilirliğin ve anlamlı veri bütünlüğünün sağlanması için silinmiştir. Örneğin Çocuk Hastanesi'nde 18 yaşından büyük hasta olmaması gerekirken hatalı kayıtların varlığı tespit edilmiş ve bu veriler değerlendirmeye alınmamıştır.

Sonuçları elde etme aşamasından önce sağlıklı ve tutarlı sonuçlar elde edebilmek için verilerin tipleri ve verinin yapısı irdelenmiştir. Erciyes Üniversitesi Hastaneleri'nin 2022 yılı içerisindeki bir milyondan fazla hasta kaydı anlamlı veriler elde etmek için bahsedilen veri ön işlem süreçleri ile daraltılarak; Çocuk Hastanesi örneğine indirgenmiş 228.052 hasta kaydından 3.050 eksik veri temizlenerek 225.002 hasta kaydı üzerinde işlem yapılmıştır.

Birliktelik Kuralı yaklaşımında literatürde yaygın kullanım alanı olan Apriori Algoritmasının uygulanmasına karar verilmiş ve bu işlemler Weka 3.9.0 programı kullanılarak yürütülmüştür. Uygulanan Apriori Algoritmasında en düşük güven düzeyi (Confidence level) %60 olarak örüntüler analiz edilmiştir. Weka yazılımında analiz yapılabilmesi için elde edilen veri ambarında bulunan veri setleri tek satır haline getirilerek Weka yazılımının işleyeceği *.Csv formatına dönüştürülmüştür. Weka yazılımında açılan *.csv Wekanın kendi formatı olan *.arff formatına çevrilerek daha önce veri ön işlemeden geçen veriler veri madenciliği analizleri yapılarak örüntülerin oluşturulmasını sağlamıştır. Elde edilen örüntüler bu çalışmada 10 ana kategoride değerlendirilmiştir. 10 kategori içinde 165 örüntü kuralı oluşturulmuştur. Veri madenciliği ile oluşturulan bu kategoriler incelendiğinde; veri setlerinde yer alan Kayseri (Pilot) verisinin ve Çocuk Hastanesi Acil Bölümünün verilerinin diğer verilere göre çok fazla olduğu tespit edilmiştir. Bu baskınlık çıkacak örüntüleri anlamlandırılmasını engelleyebileceği düşünülmüştür. Bu bağlamda veri madenciliğinde ilginç neticeler elde etmek için baskın olan verilerin adım adım çıkarılarak doğru ve anlamlı kombinasyonlara ulaşılması için oluşturulan veri madenciliği örüntüleri incelenmiştir.

İnceleme neticesinde Erciyes Üniversitesi Çocuk Hastanesi'nin hasta profili çıkarımı yapılmaktadır. Hastane Kayseri ve İç Anadolu Bölgesi hariç tutulduğunda en çok Akdeniz Bölgesinden gelen hastalara hizmet vermekte olduğu görülmüştür. Bu bağlamda Çocuk Hastanesi Akdeniz Bölgesi'nden gelen hastaların daha verimli bir muayene süreci yaşamaları amacı ile terminalden özellikle sabah saatlerinde geldikleri bilgisi düşünülerek servis hizmeti sunabileceği bilgisine ulaşılmıştır. Ayrıca Büyükşehir Belediyesi ile yapılabilecek görüşmeler neticesinde bu hastaların ulaşımı için ring seferleri düzenlenebilecektir.

KAYNAKÇA

- Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data* (pp. 207-216).
- Agrawal, R. S., & Srikant, R. (1994, September). R. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB* (pp. 487-499).
- Agrawal, R. v.d., "Fast discovery of association rules. Advances in Knowledge Discovery and Data Mining", AAAI/MIT Press, Cambridge, s.307-328
- Ağyar, Z. (2015). Yapay sinir ağlarının kullanım alanları ve bir uygulama. *Mühendis ve Makine*, 56(662), 22-23
- Ahmad, P., Qamar, S., & Rizvi, S. Q. A. (2015). Techniques of data mining in healthcare: a review. *International Journal of Computer Applications*, 120(15).
- Akman, M. (2010). Veri Madenciliğine Genel Bakış ve Random Forests Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama. Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara.
- Akman, M. (2010). *Veri madenciliğine genel bakış ve random forests yönteminin incelenmesi: sağlık alanında bir uygulama* (Master's thesis, Sağlık Bilimleri Enstitüsü). Ankara.
- Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, İstanbul Üniversitesi, İşletme Fakültesi Dergisi, , 29(1), 1–22, İstanbul.
- Akpınar, H. (2014). Data veri madenciliği - veri analizi. İstanbul: Papatya Yayıncılık.
- Altıntaş, Y. (2010). Veri madenciliğinin tıpta kullanımı ve bir uygulama: hemodiyaliz hastaları için risk seviyelerine göre risk faktörlerinin etkileşimlerinin incelenmesi. *Ulusal Tez Merkezi*, 269710, 1-3.
- Arslan, O. M., (2020) Veri Madenciliği İle Talep Tahmini Ve Bir İşletmede Uygulama, Gazi Üniversitesi, Bilişim Enstitüsü, Yüksek Lisans Tezi, Ankara.
- Awad, E.M. and H.M.GAHAZIRI (2004), Knowledge Management, Pearson, Prentice Hall.
- Aydın, S. (2007). Veri Madenciliği ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama. Eskişehir Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Eskişehir.
- Aydoğan, E. (2008). Veri madenciliğinde sınıflandırma problemleri için evrimsel algoritma tabanlı yeni bir yaklaşım: Rough-Mep algoritması. *Gazi Üniversitesi Fen Bilimleri Enstitüsü: Yayımlanmamış doktora tezi*. Ankara.
- Baldwin, R. Impediments to Acquisition and Use of Medical Knowledge, *Science*, Volume 141, Issue 3587, 1963, s.1237-1238
- Becerra-Fernandez, I et. al. (2004), Knowledge Management, Pearson, Prentice Hall.
- Bekki, A. (2016). İlişki kuralları. F. Er (Editör), Veri madenciliği (s. 103-125). Eskişehir: Anadolu Üniversitesi Açıköğretim Yayınları

- Bilgin, T.T., “Veri Akışı Diyagramları Tabanlı Veri Madenciliği Araçları ve Yazılım Geliştirme Ortamları”, Akademik Bilişim’09 - XI. Akademik Bilişim Konferansı Bildirileri, Şanlıurfa, 807-814 (2009)
- Bolat, H., (2019) Kayseri İl Emniyet Müdürlüğü Trafik Kaza Verilerinden Veri Madenciliği Yaklaşımları, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kayseri
- Buruncuk, G. (2006). *Data mining for customer segmentation and profiling: a case study for a fast moving consumer goods (FMCG) company* (Master's thesis, Sosyal Bilimler Enstitüsü).
- Can, O. (2017). *Türkiye sağlık araştırmasının veri madenciliği teknikleri ile incelenmesi* (Master's thesis, Sosyal Bilimler Enstitüsü).
- Çapar, S. G., & Szefer, P. (2005). Determination and speciation of trace elements in foods (pp. 111-158). CRC Press, Boca Raton, FL.
- Çoban, A. (2006). İmalat Sanayinde Veri Madenciliği Destekli Tedarikçi Seçimi Uygulaması. Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, Sakarya.
- Doğrul, G., Akay, D., & Kurt, M. (2015). Analysis Of Traffic Accidents By Rules Of Association. syf, 265, 284.
- Durdu, M. (2012). *Application of data mining in customer relationship management market basket analysis in a retailer store* (Doctoral dissertation, DEÜ Fen Bilimleri Enstitüsü). İzmir.
- Eker, 2016, S.7 <https://ab.org.tr/ab16/bildiri/46.pdf> .
- Esen, Fevzi, M. (2009). Veritabanlarından Bilgi Keşfi Veri Madenciliği ve Bir Sağlık Uygulaması İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İstanbul
- Esen, M.F., Veritabanlarında Bilgi Keşfi: Veri Madenciliği ve Bir Sağlık Uygulaması, Yüksek Lisans Tezi, İstanbul Üni., Sosyal Bil. Ens., 2009
- Esfandiari, N., Babavalian, M. R., Moghadam, A. M. E., & Tabar, V. K. (2014). Knowledge discovery in medicine: Current issue and future trend. *Expert Systems with Applications*, 41(9), 4434-4463. Bellazzi, Riccardo, and Blaz Zupan. "Predictive data mining in clinical medicine: current issues and guidelines." *International journal of medical informatics* 77.2 (2008): 81-97.
- Farboudi , S., Tıp Bilişiminde İstatistiksel Veri Madenciliği, Yüksek Lisans Tezi, Hacettepe Üni., Fen Bil. Ens., 2009
- Farboudi, S. (2009). Tıp bilişiminde istatistiksel veri madenciliği. *Yayımlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara.*
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-37.
- Gates, B. “Dijital Sinir Sistemiyle Düşünce Hızında Çalışmak”, İstanbul, Doğan Kitapçılık A.Ş, 1999
- Güder, Gazi, S.K., “ Bilgi İşlem Terimleri Sözlüğü”, İstanbul, Birsen Yayınları, 1986

- Günaydın, Ecem Büşra. (2022) Sağlık Verilerinde Veri Madenciliği Teknikleri İle Sağkalımı Etkileyen Faktörlerin Seçimi Performanslarının Değerlendirilmesi, EGE ÜNİVERSİTESİ, SAĞLIK BİLİMLERİ ENSTİTÜSÜ, Yüksek Lisans Tezi, İzmir.
- Gürsoy Şimşek, U., T.,” Veri Madenciliği ve Bilgi Keşfi”, Pegem Akademi, 1.Baskı, Ankara, 2009
- Güven, A., Bozkurt, Ö., Kalıpsız O. (2007). Veri madenciliğinin geleceği. IX. Akademik Bilişim Konferansı Bildirileri, Dumlupınar Üniversitesi, , s. 150- 158, Kütahya.
- Han, J. And Kamber, M., “Data mining concepts and techniques”, Morgan Kaufmann Publishers, Tokyo, 30-33 (2001).
- Han, J., & Kamber, M. (2000) Data Mining: Concepts and Techniques, Simon Fraser University
- Han, J., Kamber, M., & Pei, J. (2011). Data mining concepts and techniques third edition. The Morgan Kaufmann Series in Data Management Systems, 5(4), 83-124.
- Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. ACM sigmod record, 29(2), 1-12.
- Han, J., ve Kamber, M., Data Mining Concepts and Techniques, 2. Baskı, San Francisco:
- Hidber, C. (1999). Online association rule mining. ACM Sigmod Record, 28(2), 145-156.
- Hosseinkhah, F., H. Ashktorab, and R. Veen. "MM Owrang O (2009),“Challenges in Data Mining on Medical Databases”." 502-511.
- Kabalıcı, E. (2014). Yapay Sinir Ağları. Ders Notları <https://ekblc.files.wordpress.com/2013/09/ysa.Pdf>
- Kaya, Özel (2014) (<https://ab.org.tr/ab14/bildiri/79.pdf> Erişim 20.07.2023)
- Kirby, J., & Marsden, P. (Eds.). (2006). Connected marketing: the viral, buzz and word of mouth revolution. Elsevier
- Kokoç, M. E. L. D. A., Ersöz, S., Aktepe, A., & Türker, A. K. (2016). Improvement of facility layout by using data mining algorithms and an application. International Journal of Intelligent Systems and Applications in Engineering, 4(Special Issue-1), 92-100.
- Kudyba, S. Managing Data Mining: Advice from Experts, CyberTech Publishing, 2004, s.146- 163
- M.J.Zaki, (2000), “IEEE Transactions on Knowledge and Data Engineering”, vol.3, pp.372–390, 2000. doi:10.1109/69.846291
- Mahmoodi, S. A., Mirzaie, K., & Mahmoudi, S. M. (2016). A new algorithm to extract hidden rules of gastric cancer data based on ontology. *SpringerPlus*, 5(1), 1-21.
- Mcleish,M. “Exploring Knowledge Acquisition tools for a veterinary medical expert system”, Proceedings of the First International Conference on Industrial and

Engineering Applications of Artificial Intelligence and Expert Systems, Volume 2, ACM Publishers, 1988, s.778-788

- Niaksu, O., Kurasova, O. (2010). Data Mining Applications in Healthcare : Research vs Practices. *International Journal of Bio-Science & Bio- Technology*, pp. 58-70.
- Obenshain K. M.,(2004) “Application of Data Mining Techniques To Healthcare”, *Data Infect Control Hosp Epidemiol*, 25, 690–695, DOI: 10.1086/502460.
- Özseyhan, C., Badur, B., & Darcan, O. N. (2012). An association rule-based recommendation engine for an online dating site. *Communications of the IBIMA*, 1.
- Parraga-Alava, J., Dorn, M. and Inostroza-Ponta, M. (2018). A multi-objective gene clustering algorithm guided by apriori biological knowledge with intensification and diversification strategies. *BioData Min.*, 11, 16-38.
- Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211-218.
- Poyraz, O., 2012. Tıp’da Veri Madenciliği Uygulamaları: Meme Kanseri Veri Seti Analizi, Trakya Üniversitesi, Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, Edirne.
- Sagiroglu, S., Sinanc, D. (2013, May). Big data: A review. In *2013 international conference on collaboration technologies and systems (CTS)* (pp. 42-47). IEEE.
- Sağiroğlu, Ş.; Beşdok E. & Erler, M. (2003). Mühendislikte Yapay Zeka Uygulamaları I: Yapay Sinir Ağları. Kayseri: Ufuk Kitap Kıratsiye-Yayıncılık.
- Samora, J. Saunders, L. Larson, F “Medical Vocabulary Knowledge Among Hospital Patients”, *Journal of Health and Human Behavior*, Volume 2, No 2, 1961, s.83-92
- Savaş, S., Topaloğlu, N. ve Yılmaz, M. (2012). “ Veri Madenciliği ve Türkiye’deki Uygulama Örnekleri”. *İstanbul Üniversitesi Fen Bilimleri Dergisi*, 21: 1-23.
- Shah, S. C., & Kusiak, A. (2004). Data mining and genetic algorithm based gene/SNP selection. *Artificial intelligence in medicine*, 31(3), 183-196
- Silahtaroglu, G., 2008. Veri Madenciliği. Papatya Yayınları, İstanbul.
- Silahtaroglu, G., 2013. Veri Madenciliği Kavram ve Algoritmaları, 2. Basım, Papatya Yayıncılık Eğitim, İstanbul, 304 s
- Silahtaroglu, G., Veri Madenciliği Kavram ve Algoritmaları, 3. Baskı, İstanbul, Papatya Yayıncılık Eğitim, 2016. İstanbul.
- Sözen, E., Bardak, T., Peker, H., & Bardak, S. (2017). Apriori Algoritması Kullanılarak Mobilya Seçimde Etkili Olan Faktörlerin Analizi. *İleri Teknoloji Bilimleri Dergisi*, 6(3), 679-684.
- Tekerek, A. (2011). Veri Madenciliği Süreçleri Ve Açık Kaynak Kodlu Veri Madenciliği Araçları. Akademik Bilişim 11 - XIII. Akademik Bilişim Konferansı Bildirileri 2 - 4 Şubat 2011 İnönü Üniversitesi, 161-169, Malatya.
- Tomar, N.S. and Agarwal, R.M. (2013) Influence of Treatment of *Jatropha curcas* L. Leachates and Potassium on Growth and Phytochemical Constituents of Wheat (*Triticum aestivum* L.). *American Journal of Plant Sciences*, 4, 1134-1150. <https://doi.org/10.4236/ajps.2013.45140>

- Tüzüntürk, S, (2010), “Veri Madenciliği ve İstatistik,” Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,29(1), 65-90, Bursa.
- Qiu Y, Li J, Huang X, Shi H. A Feasible Data-Driven Mining System to Optimize Wastewater Treatment Process Design and Operation. *Water* 2018, 10, 1342; doi:10.3390/w10101342
- Westphal, C., Blaxton, T. (1998). *Data mining solutions: methods and tools for solving real-world problems*. John Wiley & Sons, Inc..
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2005). *Data Mining: Practical machine learning tools and techniques*. San Francisco. *Morgan Kaufmann*.
- Yalçın, L. (2019). Sağlık Sektöründe Veri Madenciliği, Milli Savunma Üniversitesi Hezârfen Havacılık Ve Uzay Teknolojileri Enstitüsü, Yüksek Lisans Tezi, Ankara.
- Yorulmaz, A. (2018) Veri Tabanı Sistemlerinde Yüksek Performans Ve Erişilebilirlik İle Güvenliğin Oracle Veri Tabanı Sistemi Özelinde İncelenmesi. Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Sakarya.
- Yu, X., & Wang, H. (2014). Improvement of Eclat Algorithm Based on Support in Frequent Itemset Mining. *J. Comput.*, 9(9), 2116-2123. “Scalable algorithms for association mining
- Zhang, G. P. (2000). Neural networks for classification: A survey. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 30(4), 451–462. <https://doi.org/10.1109/5326.897072>
- Zhong, N., & Zhou, L. (Eds.). (2003). *Methodologies for Knowledge Discovery and Data Mining: Third Pacific-Asia Conference, PAKDD'99, Beijing, China, April 26-28, 1999, Proceedings*. Springer.

Ek-1 Turnitin Raporu

HASTA PROFİLLERİNİ ANALİZ ETMEK İÇİN BİRLİKTELİK KURALLARI MADENCİLİĞİ: KAYSERİ'DEKİ BİR HASTANE ÖRNEĞİ

ORJİNALLİK RAPORU

% 15 BENZERLİK ENDEKSİ	% 14 İNTERNET KAYNAKLARI	% 9 YAYINLAR	% 9 ÖĞRENCİ ÖDEVLERİ
----------------------------------	------------------------------------	------------------------	--------------------------------

BİRİNCİL KAYNAKLAR

1	hdl.handle.net İnternet Kaynağı	% 2
2	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Öğrenci Ödevi	% 2
3	acikbilim.yok.gov.tr İnternet Kaynağı	% 1
4	repositorio.utfpr.edu.br İnternet Kaynağı	% 1
5	repository.library.teimes.gr İnternet Kaynağı	% 1
6	Alireza Hamidieh, Maliheh Ghanbari. "Developing a Decision Model for Enhancing the Safety of CNG Stations Using Data Mining and Safety Indicators Classification", Research Square Platform LLC, 2023 Yayın	% 1
7	pastebin.com	

	İnternet Kaynađı	% 1
8	Elisandra Duarte Pinheiro, Celso Bilynkievycz dos Santos, Caroline Lievore, Luiz Alberto Pilatti. "UTILIZAÇÃO DE REDES SOCIAIS POR DOCENTES DOS PROGRAMAS DE PÓS GRADUAÇÃO DA ÁREA DE ENGENHARIAS III", Interfaces Científicas - Educação, 2017 Yayın	% 1
9	Submitted to Infile Öđrenci Ödevi	% 1
10	t4tutorials.com İnternet Kaynađı	% 1
11	Submitted to Istanbul Aydin University Öđrenci Ödevi	<% 1
12	Submitted to Middlesex University Öđrenci Ödevi	<% 1
13	Leonor Adriana Cárdenas-Robledo, Alejandro Peña-Ayala. "Ubiquitous learning: A systematic review", Telematics and Informatics, 2018 Yayın	<% 1
14	9lib.net İnternet Kaynađı	<% 1
15	s3-ap-southeast-1.amazonaws.com İnternet Kaynađı	<% 1

16	Submitted to University of South Florida Öğrenci Ödevi	<% 1
17	openaccess.hacettepe.edu.tr:8080 İnternet Kaynağı	<% 1
18	Submitted to Aksaray Aniversitesi Öğrenci Ödevi	<% 1
19	Submitted to University of Computer Studies Öğrenci Ödevi	<% 1
20	toad.halileksi.net İnternet Kaynağı	<% 1
21	Submitted to UI, Springfield Öğrenci Ödevi	<% 1
22	abis-files.kayseri.edu.tr İnternet Kaynağı	<% 1
23	Submitted to Bournemouth University Öğrenci Ödevi	<% 1
24	Submitted to Carnegie Mellon University Öğrenci Ödevi	<% 1
25	www.tuicool.com İnternet Kaynağı	<% 1
26	acikerisim.erbakan.edu.tr İnternet Kaynağı	<% 1
27	eei.fei.tuke.sk İnternet Kaynağı	<% 1

28	geka.gov.tr İnternet Kaynađı	<% 1
29	Submitted to Bahcesehir University Öđrenci Ödevi	<% 1
30	Submitted to Istanbul Medipol Āniversitesi Öđrenci Ödevi	<% 1
31	libweb.kpfu.ru İnternet Kaynađı	<% 1
32	Submitted to Anadolu University Öđrenci Ödevi	<% 1
33	Submitted to Hacettepe University Öđrenci Ödevi	<% 1
34	Submitted to Istanbul Medeniyet Āniversitesi Öđrenci Ödevi	<% 1
35	Submitted to Omer Halisdemir University Öđrenci Ödevi	<% 1
36	balikoltada.tr.gg İnternet Kaynađı	<% 1
37	Submitted to Erciyes Āniversitesi Öđrenci Ödevi	<% 1
38	Submitted to New York Institute of Technology Öđrenci Ödevi	<% 1
39	Submitted to University of Southampton	

Öğrenci Ödevi

<% 1

40

Submitted to Postgraduate Institute of
Medicine

Öğrenci Ödevi

<% 1

41

Submitted to Utah Education Network

Öğrenci Ödevi

<% 1

42

[silo.tips](#)

İnternet Kaynağı

<% 1

43

Submitted to Flinders University

Öğrenci Ödevi

<% 1

44

Michael Jordan Bianchi. "Recomendação de
práticas de gestão de projetos baseadas em
algoritmos e evidência científica:
contribuições para proposição de modelos
híbridos e >i/i<", Universidade de Sao Paulo,
Agencia USP de Gestao da Informacao
Academica (AGUIA), 2022

Yayın

<% 1

45

Submitted to Universitaet Hamburg

Öğrenci Ödevi

<% 1

46

[www.slideserve.com](#)

İnternet Kaynağı

<% 1

47

[abis-files.erciyes.edu.tr](#)

İnternet Kaynağı

<% 1

48

igs2018.mu.edu.tr
İnternet Kaynağı

<% 1

49

www.acarindex.com
İnternet Kaynağı

<% 1

50

ÖZSEVEN, Turgut and ERSÖZ, Taner. "E-
TİCARET VERİLERİNİN MÜŞTERİ PROFİLİ
AÇISINDAN DEĞERLENDİRİLMESİ", Bülent
Ecevit Üniversitesi, 2016.
Yayın

<% 1

Alıntıları çıkart

Kapat

Eşleşmeleri çıkar

Kapat

Bibliyografyayı Çıkart üzerinde

ETİK KURUL ONAYI

Bu çalışma için Etik Kurul onayı Kayseri Üniversitesi Etik Kurulu'nun 11/09/2023 tarihli ve 64/2023 numaralı kararı ile alınmıştır



ÖZ GEÇMİŞ

1. İLETİŞİM BİLGİLERİ

Adı Soyadı : Salih Murat GÜRBÜZ

Uyruğu : T.C.

2. ÖĞRENİM DURUMU

Derece	Alan	Üniversite	Yıl
Lisans	İşletme Fakültesi	Anadolu Üniversitesi	2006
Yüksek	Sanat Tarihi	Erciyes Üniversitesi	2014

3. YABANCI DİL : Orta

4. İŞ DENEYİMLERİ

1995-1996

Kayseri Valiliği İl Özel İdare Müdürlüğü

Bilgi İşlem Sorumlusu

1996-1998

Yozgat İl Jandarma Komutanlığı

Bilgi Toplama İşleme Merkezi Sorumlusu

1998-2020

Erciyes Üniversitesi Bilgi İşlem Daire Başkanlığı

Network Sorumlusu

2020-

Kayseri Üniversitesi Bilgi İşlem Daire Başkanlığı

Şube Müdürü