

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**GEYVE DEVLET HASTANESİNDE YATAN COVID-19 HASTA
VERİLERİNİN VERİ MADENCİLİĞİ YÖNTEMİYLE
İNCELENMESİ**

YÜKSEK LİSANS TEZİ

Seda UÇAR

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Bilim Dalı

ŞUBAT 2023

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**GEYVE DEVLET HASTANESİNDE YATAN COVID-19 HASTA
VERİLERİNİN VERİ MADENCİLİĞİ YÖNTEMİYLE
İNCELENMESİ**

YÜKSEK LİSANS TEZİ

Seda UÇAR

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Bilim Dalı

Tez Danışmanı: Doç. Dr. Nilüfer YURTAY

ŞUBAT 2023

Seda UÇAR tarafından hazırlanan “Veri Madenciliği Yöntemleriyle Covid-19 Yatan Hasta Verilerinin İncelenmesi” adlı tez çalışması 07.02.2023 tarihinde aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Bilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı : **Doç. Dr. Nilüfer YURTAY (Danışman)**
Sakarya Üniversitesi

Jüri Üyesi : **Dr. Öğr. Üyesi Serap ÇAKAR**
Sakarya Üniversitesi

Jüri Üyesi : **Dr. Öğr. Üyesi Ziyet PAMUK**
Zonguldak Bülent Ecevit Üniversitesi

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “GEYVE DEVLET HASTANESİNDE YATAN COVID-19 HASTA VERİLERİNİN VERİ MADENCİLİĞİ YÖNTEMİYLE İNCELENMESİ” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’ de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığımı, etik kurul onay belgesi aldığımı, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim. (07/02/2023).

(imza)

Seda UÇAR

TEŐEKKÖR

Yüksek lisans eğitiminde bende değerli bilgi ve deneyimlerini esirgemeyen, her konuda fikir danışmaktan çekinmediğim, tezimin hazırlanmasında bana sunduğu katkılarından dolayı değerli danışman hocam Doç. Dr. Nilüfer Yurtay'a teşekkürlerimi sunarım.

Seda Uçar

İÇİNDEKİLER

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	viii
İÇİNDEKİLER	ix
KISALTMALAR	xii
TABLO LİSTESİ	xiii
ŞEKİL LİSTESİ	xv
ÖZET	xvii
SUMMARY	xix
1. GİRİŞ	1
1.1. Tezin Kapsamı	2
1.2. Tezin Amacı	3
1.3. Literatür Araştırması	3
1.4. Hipotez	5
2. VERİ MADENCİLİĞİ	7
2.1. Veri Madenciliği Veri Kaynakları	7
2.1.1. Veri ambarları	7
2.1.2. Veri kümeleri	8
2.1.3. Veritabanı sistemleri	8
2.2. Veri Madenciliği Uygulama Alanları	9
2.3. Veri Madenciliği Yöntemleri	10
2.3.1. Sınıflandırma	10
2.3.2. Kümeleme	10
2.3.3. Birliktelik analizi	11
2.4. Veri Madenciliği Süreci	12
2.4.1. İş anlama	12
2.4.2. Veriyi anlama	13
2.4.3. Veri hazırlık süreci	13
2.4.3.1. Verilerin temizlenmesi	13
2.4.3.2. Verilerin tekrar yapılandırılması	14
2.4.4. Modelleme	16
2.4.5. Değerlendirme	16
2.4.6. Uygulama	16
3. BİRLİKTELİK ANALİZİ	17
3.1. Destek ve Güven Değeri	17
3.2. Lift Değeri	19
3.3. Birliktelik Analizi Algoritmaları	20
3.3.1. Apriori algoritması	20
4. UYGULAMA	25
4.1. Veriyi Anlama ve Hazırlık Süreci	25
4.2. Analiz Sonuçları	28
4.3. Apriori Algoritmasının Uygulanması	34

4.3.1. Birliktelik Kuralları	34
5. SONUÇ VE ÖNERİLER.....	45
KAYNAKLAR.....	49
ÖZGEÇMİŞ.....	51

KISALTMALAR

CRISP-DM : Cross-Industry Standart Process for Data Mining

Covid-19 : Yeni Koronavirüs Hastalığı

PCR : Polimeraz Zincir Reaksiyonu

TABLO LİSTESİ

Sayfa

Tablo 3.1. Hasta şikâyet tablosu.....	20
Tablo 3.2. Destek değerleri.	21
Tablo 3.3. Eşik destek değerine eşit ve büyük olan şikâyetler.....	21
Tablo 3.4. İkili şikâyet grupları ve destek değerleri.....	22
Tablo 3.5. Eşik destek değerine eşit ve büyük olan ikili şikâyet grupları.....	22
Tablo 3.6. Eşik destek değerine eşit ve büyük olan üçlü şikâyet grupları.....	22
Tablo 3.7. Oluşturulan birliktelik kuralları.....	23
Tablo 4.1. Aşı durumları.	25
Tablo 4.2. Yaş kategori tablosu.....	25
Tablo 4.3. Yatış süresi kategori tablosu.	26
Tablo 4.4. Tanı tablosu.....	26

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1. Veritabanı ve kullanıcı arasındaki süreç.....	8
Şekil 2.2. CRISP-DM süreci.....	12
Şekil 4.1. Cinsiyet dağılım grafiği.....	28
Şekil 4.2. Yatış süresi dağılım grafiği.....	29
Şekil 4.3. Yaş dağılım grafiği.....	29
Şekil 4.4. Aşı tercihi dağılım grafiği.....	30
Şekil 4.5. Hastaneden çıktıktan sonraki aşı tercihleri.....	30
Şekil 4.6. Aşı olmayanların hastanede yatış süresi.....	31
Şekil 4.7. Aşısız olanların hastanede yatış süresi.....	31
Şekil 4.8. Aşı türlerine göre ortalama hastanede yatış süresi.....	32
Şekil 4.9. Hastalanmadan önce aşısız olmayan hastaların yatış süresi ve aşı tercihi grafiği.....	33
Şekil 4.10. Hastanede yatış süresi ve cinsiyet dağılım grafiği.....	33
Şekil 4.11. Hastanede yatış süresi ve tanı dağılım grafiği.....	34
Şekil 4.12. Aylara göre dağılım grafiği.....	34

GEYVE DEVLET HASTANESİNDE YATAN COVID-19 HASTA VERİLERİNİN VERİ MADENCİLİĞİ YÖNTEMİYLE İNCELENMESİ

ÖZET

Koronavirüs tüm dünyada olduğu gibi ülkemizde de hızlı bir şekilde yayılarak insanların sağlığını olumsuz yönde etkilemiştir. Ülkemizde hastalığa yakalanan insanlar arasında, hastalığı hafif atlatanlar olduğu gibi hastanede yatmak zorunda kalan kişiler de olmuştur. Hastalığa yakalanıp hastaneye yatış yapan kişilere ait verilerin tutulması Covid-19 servisleri için önemli bir parametre haline gelmiştir.

Bu çalışmada; 2 Nisan 2020 ile 30 Eylül 2021 tarihleri arasında Geyve Devlet Hastanesi Covid-19 servisinde yatan hastaların ortak özelliklerini belirlemek amacıyla, hastalara ait veriler incelenmiştir. Elde edilen veri kümesine birliktelik analizi yöntemlerinden biri olan Apriori algoritması uygulanmıştır ve Covid-19 hastalığına yakalanan kişilerin birliktelik kuralları çıkarılmıştır.

Bunun için hastaların aşı sayısı, aşı türü, yaşı, cinsiyeti, hastanede kalma süresi ve hastaneden çıkış türü parametreleri veri madenciliği yazılımlarından biri olan RapidMiner uygulaması kullanılarak incelenmiştir. Bu çalışma sonucunda; aşı sayısı, aşı türü, yaş, cinsiyet, hastanede kalma süresi ve hastaneden çıkış türü parametrelerinin birbirleriyle ilişkileri hakkında çıkarımlarda bulunulmuştur.

Hasta profillerinin incelenip hastaların hastanede ne kadar kalacağını öngörmek yatan hasta servislerindeki planlama açısından önemli olabilmektedir. Bu çalışmadan Sakarya Geyve Devlet Hastanesi Covid-19 yatan hasta servisinden alınan 556 hastanın verileri birliktelik kural analizi yöntemlerinden olan apriori algoritması ile incelenmiş ve birliktelik kuralları çıkarılmıştır.

Analiz sonucunda hastaneye yatan 556 hastadan 506'sı hastaneden sağlıklı bir şekilde taburcu edilmiştir. Hastalığa yakalandığında aşısı olmayan 398 hastadan 198'i hastaneden taburcu olduktan sonra iki Biontech aşısı olmayı tercih etmiştir.

Hastaneye yatan 556 hastanın 336'sı, 0-7 gün hastanede yatmıştır. Sadece 1 kişi 30 günden fazla hastane yatmıştır. Bu kişinin cinsiyeti kadındır, 60-74 yaş aralığındadır ve hastaneden çıktıktan sonra iki Sinovac ve bir Biontech aşısı olmayı tercih etmiştir.

Hastanede yatış yapan hastalar 164 kişi ile en çok 45-74 yaş aralığında kümelenmiştir. Bu 164 kişiden 126'sının hastalanmadan önce aşı bilgisi yoktur.

Çalışma sonucunda elde edilen bilgilerin hastane yönetimine ve doktorlara fayda sağlayacağı ve hastaların aldığı hizmete artı yönde katkı yapacağı düşünülmektedir.

EXAMINATION OF INPATIENT COVID-19 PATIENT DATA IN GEYVE STATE HOSPITAL BY DATA MINING METHOD

SUMMARY

Covid-19 is a respiratory illness that first emerged in the city of Wuhan, China in late December 2019. Upon investigation, the virus was officially identified on January 13, 2020. The virus quickly spread from China to the rest of the world.

Covid-19 is transmitted through droplets among people. While some people are at higher risk for Covid-19, about 80% of people who contract the virus have mild symptoms. About 20% of patients may experience more serious symptoms and require hospitalization.

In our country, there have been people who have contracted the disease and either had mild symptoms or needed to be hospitalized. The data of people who were hospitalized after contracting the disease has become an important parameter for Covid-19 services.

With the increasing use of computers in the healthcare field, all hospital processes are now kept digitally. The amount of data kept digitally is increasing every day. This data can be made meaningful through data mining studies.

The data kept in hospitals can be in various forms, such as diagnoses, surgeries, lengths of hospital stays, and results of imaging procedures. Making this data meaningful can improve the quality of work in hospitals and allow patients to receive quality services.

More and more data is being produced in the healthcare sector every day. This data is very extensive and complex. Analyzing and using this data correctly can improve hospital performance.

The use of data mining techniques in the healthcare field can provide more effective treatments, group patients with similar diseases, and be useful in predicting length of hospital stays.

In this study, the data of patients who were hospitalized in the Covid-19 unit at Geyve State Hospital between April 2, 2020 and September 30, 2021 was used as the data set. The first hospitalization due to Covid-19 at Geyve State Hospital occurred on April 2, 2020 during the pandemic period. The hospital provided the data as an Excel file. The data for each patient was not fully filled out on each row of the file, and missing data was obtained from the hospital system and processed individually. Records of the same patient were detected in the data set and removed. The records of patients who were hospitalized multiple times were merged. After all these processes, columns that will not be used in our study were removed, and we were left with 8 attributes: vaccine type, gender, age, type of vaccine received before hospitalization (if applicable), type of vaccine received after discharge (if applicable), length of hospital stay, type of discharge, and diagnosis.

The vaccine information for the patients was processed individually from the hospital system on October 1, 2021. The vaccine information for the patients on that date was entered into the file. The vaccination process in Turkey began with the Sinovac CoronaVac vaccine on January 14, 2021 by the Ministry of Health. In this study, the CoronaVac vaccine will be referred to as Sinovac. Biontech vaccine began to be used in Turkey from April 12, 2021. On June 30, 2021, a call was made for those over 50 years old and healthcare workers to receive the third dose of vaccine. On August 16, 2021, some countries called for two Biontech vaccines for those coming to their countries, and as a result, those who received two Sinovac and one Biontech vaccine were given the right to receive a fourth dose.

The age data was calculated from the date of birth data of the patient and created as a separate column. Then, the age data was categorized and organized.

The length of hospital stay data was calculated by taking the difference between the admission date and discharge date of the patient, and then categorizing and organizing the lengths of stay, creating a separate column for the range of days the patient was hospitalized.

The diagnosis column was entered in two separate ways: only Covid-19 diagnosis and Covid-19 diagnosis with additional disease record. The type of discharge data was entered in three different forms: discharge from the hospital, transfer to another service, and referral.

The gender column was entered as female and male.

Columns were added that show the vaccine preferences before and after illness based on the date of vaccination and the date of hospitalization. If the patient had no vaccination, it was recorded as not vaccinated. If the patient had a vaccination, it was recorded as not having a vaccination before or after the illness.

After the data was cleaned and processed, analyses were conducted on the data.

The distribution of patients according to gender is 271 male and 285 female.

The distribution of hospitalization periods of patients is as follows. 336 people have stayed in the hospital for 0-7 days.

The age distribution of patients hospitalized in the hospital is as follows: 164 people were hospitalized between the ages of 45 and 59. The most accumulation was in this age range. Then, with 154 people, the most hospitalizations were in the age range of 60-74.

The vaccine preferences of hospitalized patients were as follows: 215 people most preferred the combination of two Biontech vaccines. Then the most preferred vaccine combination was 76 people with two Sinovac vaccines.

Of the 556 hospitalized patients, 398 of them preferred to be vaccinated after being discharged from the hospital. There is no vaccine registration information for these 398 people before being hospitalized. Of the 398 people who have never been vaccinated, 213 of them preferred to receive two Biontech vaccines after overcoming the illness.

The distribution of hospitalization periods for the 496 patients who had no vaccine registration before falling ill is as follows: The most accumulation, 303 people, were hospitalized in the range of 0-7 days.

The distribution of hospitalization periods for the 60 patients who were vaccinated before falling ill is as follows: 33 of the 60 people were hospitalized in the range of 0-7 days, while 27 were hospitalized in the range of 8-15 days. 19 of the 60 people who were vaccinated and hospitalized had two Sinovac vaccines when they were caught with the illness.

The average hospitalization periods by vaccine type are as follows: The average hospitalization period for 215 people who preferred two Biontech vaccines is 7.54 days. The average hospitalization period for 98 people who do not have a vaccine is 6.49 days.

Before being hospitalized, 398 people who had no vaccine were hospitalized, and their choice of vaccination was as follows. 121 people who were hospitalized in the 0-7 day range preferred to have two Biontech vaccines afterwards. 83 people who were hospitalized in the 8-15 day range preferred to have two Biontech vaccines afterwards.

The distribution of hospitalization duration and diagnosis was as follows. Out of 556 people hospitalized, 336 people had a hospitalization duration of 0-7 days. Of these, 127 people were hospitalized only with a diagnosis of Covid-19. 209 people were hospitalized with Covid-19 and an additional illness.

The distribution of hospitalized patients by month was as follows. The busiest hospitalization periods in 2020 were in April and November. The busiest hospitalization period in 2021 was in April. June 2021 was the month with the least hospitalization. No one was hospitalized in this month.

After the analysis, the apriori algorithm was applied to the data that was made suitable for the study. The Rapidminer application was used when applying the apriori algorithm, and association rules were obtained. The rules with the highest lift value are given among the obtained rules.

It is thought that the information obtained as a result of the study will be beneficial to the hospital management and doctors, and will contribute positively to the service received by the patients.

1. GİRİŞ

Covid-19 hastalığı ilk olarak Çin'in Wuhan kentinde 2019 yılının aralık ayının sonlarında solunum yolu hastalığı belirtileri şeklinde ortaya çıkmıştır. Yapılan araştırmalar sonucunda virüs 13 Ocak 2020'de resmi olarak tanımlanmıştır [1]. Çin'de ortaya çıkan bu virüs daha sonra tüm dünyaya hızlı bir şekilde yayılmıştır.

Covid-19 hastalığı insanlar arasında damlacık yoluyla bulaşmaktadır. Covid-19 hastalığında bazı insanlar daha fazla risk altında olmakla birlikte, hastaların %80'i hastalığı hafif geçirmektedir. Hastaların %20'lik bir kısmı ise daha ciddi semptomlar gösterip hastanede tedavi olmak zorunda kalabilmektedir [1].

Covid-19 hastalığını belirti göstermeden geçirenler olsada, en sık karşılaşılan belirtiler ateş, öksürük ve nefes darlığı olmaktadır. Daha şiddetli geçirenlerde ise zatürre, ağır solunum yetmezliği, böbrek yetmezliği ve ölüm gelişebilmektedir [1].

Hastalıktan en çok etkilenen kişiler genellikle;

- 60 yaş ve üzeri yaş grubunda olan kişiler
- Ciddi kronik tıbbi rahatsızlıkları olan insanlar,
- Kalp hastalığı,
- Hipertansiyon,
- Diyabet,
- Kronik solunum yolu hastalığı,
- Kanser,
- Sağlık Çalışanları

olarak listelenmektedir [1].

Covid-19 tanısı PCR (Polymerase Chain Reaction) denilen testler ile konulmaktadır.

Covid-19 aşılarının tümü, bağışıklık sistemimize Covid-19'a neden olan virüsü güvenli bir şekilde tanıtmayı ve yok etmeyi öğretecek şekilde tasarlanmıştır.

Çalışmada ismi geçecek aşuların türleri kısaca aşağıda açıklanmaktadır [2].

- Baęışıklık oluřturmak için kullanılan etkin olmayan virüs ieren ařılar (İnaktif ařılar, örneęin Conovac-Sinovac ařısı) .
- Tek başına güvenli bir baęışıklık oluřturan protein üretmek için virüsün genetik kodunu barındıran RNA'yı ieren ařılar (m-RNA ařıları, Pfizer-Biontech ařısı)

Türkiye'de Covid-19 sürecinde önemli olan tarihler ařaęıda kronolojik sıra ile verilmiřtir [3].

- Dünya Saęlık Örgütü tarafından 30 Ocak 2020'de halk saęlığı acil durumu olarak bildirilmiřtir.
- 11 Mart 2020'de küresel salgın yani pandemi olarak ilan edilmiřtir.
- 11 Mart 2020'de Türkiye'de ilk Covid-19 vakası bildirilmiřtir.
- 13 Ocak 2021'de ilk Covid-19 ařılması Saęlık Bakanlıęının onayı ile Coronavac ařısı ile bařlamıřtır.
- 12 Nisan 2021 tarihiyle beraber Pfizer-Biontech ařısının kullanımına bařlanmıřtır.
- 30 Haziran 2021 tarihinden itibaren 50 yař ve üzerindekiilerin ve saęlık alıřanlarının 3. doz ařı olması yönünde karar verilmiřtir.
- 16 Aęustos 2021'de yurtdıřına ıkıřlarda bazı ölkelerin 2 doz Biontech ařısı istemesi üzerine 2 doz Sinovac ařısı ve 1 doz Biontech ařısı olanlara 4. doz ařı olma hakkı tanınmıřtır.
- 3 Kasım 2021'de 2 doz Biontech ařısı yaptıranlar için 6 ay sonra 3. doz ařı yapılabileceęi açıklanmıřtır.

Bu alıřmada Covid-19 PCR tanı testi pozitif olup, hastanede yatarak tedavi gören hastaların bilgileri incelenecektir. Yatıř yapan hastaların cinsiyet, yař, ařı bilgisi, hastanede yatıř süresi ve tanı nitelikleri kullanılarak birliktelik analizi alıřması yapılacaktır.

1.1. Tezin Kapsamı

Geyve Devlet Hastanesi Covid-19 servisinde 2 Nisan 2020 ile 30 Eylül 2021 tarihleri arasında yatan hastalara ait elde ettięimiz ham veriler bir veri madencilięi metodolojisi olan CRISP-DM (Cross-Industry Standard Process for Data Mining) kullanılarak veri analizi yapılmıřtır.

Yapılan analizler sonucunda elde edilen veri 556 adet satırdan oluşmaktadır. Çalışmada kullanılan nitelikler ise yaş, cinsiyet, aşı durumu, hastanede kalma süresi, hastaneden çıkış türü, tanı, hastalanmadan önce aşırı durumu ve hastalandıktan sonra aşı durumu olmak üzere 8 adettir. Elde edilen verilere yatan hastaların ortak özellikleri belirlemek amacıyla birliktelik analizi algoritmalarından olan Apriori algoritması uygulanmıştır ve birliktelik kuralları çıkarılmıştır.

1.2. Tezin Amacı

Ülkemizde Covid-19 vaka sayılarının yüksek olduğu dönemlerde hastaneye yatış oranları artabilmektedir. Hasta profillerinin incelenip hastaların hastanede ne kadar kalacağını öngörmek yatan hasta servislerindeki planlama açısından önemli olabilmektedir. Bu çalışmada Sakarya Geyve Devlet Hastanesi Covid-19 yatan hasta servisinden alınan 556 hastanın verileri birliktelik kural analizi yöntemlerinden olan apriori algoritması ile incelenmiş ve birliktelik kuralları çıkarılmıştır.

Çalışma sonucunda elde edilen bilgilerin hastane yönetimine ve doktorlara fayda sağlayacağı ve hastaların aldığı hizmete artı yönde katkı yapacağı düşünülmektedir.

1.3. Literatür Araştırması

Günümüzde sağlık alanında bilgisayar kullanımının artması ile hastanelerdeki her işlem artık dijital ortamda tutulmaktadır. Dijital ortamda tutulan veriler her geçen gün artmaktadır. Bu verilerin kullanılması ve anlamlı hale gelmesi veri madenciliği çalışmaları ile mümkün olmaktadır.

Hastanelerde tutulan veriler, yapılan teşhisler, yapılan operasyonlar, hastanede yattıkları süreler, uygulanan görüntüleme işlemlerinin sonuçları gibi çeşitli şekilde olabilmektedir. Bu verilerin anlamlı hale gelmesi hem hastanelerdeki çalışmaların kalitesinin artmasına hem de hastaların nitelikli hizmetler almasına olanak sağlamaktadır.

Sağlık sektöründe her geçen gün daha fazla veri üretilmektedir. Üretilen bu veriler oldukça geniş kapsamlı ve karmaşık olmaktadır. Bu verileri analiz etmek ve doğru kullanımını sağlamak hastanelerin performansını artırabilmektedir [4].

Sağlık alanında veri madenciliği tekniklerinin kullanılması ile, daha etkin tedaviler sunulabilmekte, benzer hastalığa sahip hastaların gruplandırılmasında, hastanede kalış sürelerini tahmin etmede yararlı olabilmektedir [5].

Covid-19 hastalığı hem ülkemiz hem dünya için yeni bir hastalık olduğundan dolayı bu alanda yapılan çalışmalar hastalık ile alakalı yol gösterici olmaktadır.

Literatürde Covid-19 ve veri madenciliği alanında çok çeşitli çalışmalar yapılmıştır.

Silahtaroglu ve arkadaşının yaptığı çalışmada, covid-19 test sonucu pozitif olan 1313 hastanın şikayetleri ve gösterdiği semptomlar üzerine çalışmaya yapılmış. Ayrıca bu veriler üzerinde Fuzzy C-Means ve PCA veri indirgeme algoritması kullanılarak bir kümeleme modeli sunulmuştur. Kümeleme sonucunda, öksürük, boğaz ağrısı ve halsizliğin en yaygın semptomlar olduğu gözlemlenmiştir [6].

Çelik yaptığı çalışmada, Covid-19 semptomları üzerine apriori algoritması uygulamıştır. Çalışma sonucunda A01 (Typhoid and paratyphoid fevers), A02 (Other salmonella infections) ve A04 (Diğer bakteriyel bağırsak enfeksiyonları) kodları ile belirtilen semptomların bir arada olduğu hastaların Covid-19' a kesin olarak yakalandığını söylemiştir [7].

Pinter ve arkadaşlarının yaptığı çalışmada Macaristandan gelen verileri kullanarak Covid-19'u tahmin etmek için hibrit bir makine öğrenimi yaklaşımı önermişlerdir.[8]

Kalezhi ve arkadaşları, Covid-19 salgınının Zambiya' da nasıl yayıldığını daha iyi ortaya çıkarmak için bir çalışma yapmışlardır. Bu çalışmadan WEKA uygulamasının içindeki J48 karar ağacı, çok katmanlı algılayıcı ve naive bayes tekniklerini kullanmışlardır ve iyi sonuçlar elde etmişlerdir [9].

Shanin ve arkadaşlarının yaptığı çalışmada Spark kullanarak apriori ve fp-growth algoritmalarının performansını ve aralarındaki farkları değerlendirmiş ve sonucunda FP-Growth algoritmasının bellek ihtiyacı, veritabanına başvuru ve zaman konusunda daha üstün olduğunu ortaya koymuştur [10].

Bozkurt ve arkadaşı, kişilerin tam kan sayımı sonuçları üzerinden destek vektör makineleri, rastgele orman, gaussian ve naive bayes sınıflandırma algoritmaları kullanarak Covid-19 pozitif olan hastaları tahmin etmeye çalışmışlardır. Çalışmanın sonucunda destek vektör makineleri %92 başarı, rastgele orman algoritması %90 başarı, naive bayes algoritması ise %86 başarı elde etmiştir [11].

Ahouz ve arkadaşı, least-square boosting algoritmasını kullanarak iki haftalık bir süre için Covid-19 insidansını tahmin etmeyi amaçlamışlardır. Çalışma sonucunda ise %98,45 oranıyla yüksek doğruluğa ulaşmışlardır [12].

Tandan ve arkadaşları yaptığı çalışmada ortalama yaşı 52 olan 1560 hasta verisine bir makine öğrenme tekniği olan birliktelik kural madenciliği tekniği uygulamıştır. Çalışma sonucunda hastalarda en sık bildirilen semptomların ateş, öksürük, zatürre ve boğaz ağrısı olduğu sadece %1'inde septik şok, solunum sıkıntısı sendromu ve solunum yetmezliği gibi ciddi semptomlar olduğu ortaya çıkmıştır. Ortaya çıkan kurallar yaş ve cinsiyete bağlı olarak değişmektedir. Kronik rahatsızlığı olan ve Covid-19'dan ölen hastalarda daha ciddi semptom birliktelikleri olduğu görülmüştür [13].

Gök'ün hazırlamış olduğu yüksek lisans tez çalışmasında, Brezilya Einstein Hastanesi'nden alınan veri seti hastaların kan ve laboratuvar sonuçlarını içermektedir. Bu veriler makine öğrenmesi algoritmaları kullanılarak hastaların Covid-19 hastalık seviyesi tahmin edilmiştir. Çalışmanın sonucunda, Rassal Orman algoritması, 0.979 doğruluk değeri ile hastaların Covid-19 hastalık seviyesini başarılı bir şekilde tahmin etmiştir [14].

Oğuz yapmış olduğu çalışmada Siirt Eğitim ve Araştırma Hastanesinden aldığı radyografi verilerini kullanmıştır. Bu verilerden elde edilen 1345 adet BT görüntüsü üzerine çeşitli derin öğrenme modelleri uygulamıştır. Çalışma sonucundaki verilere göre en iyi performans sonucu ResNet-50 ve DVM yöntemleri ile elde edilmiştir. Performans sonucu yüksek olan metod Covid-19 hastalığı için gereksiz yapılacak testleri önlemede yardımcı olarak başvurulabilir [15].

1.4. Hipotez

Çalışmanın ana amacı Covid-19 pandemisinde 2 Nisan 2020 ile 30 Eylül 2021 tarihleri arasında Geyve Devlet Hastanesi Covid-19 hasta servisinde yatan hastaların ortak özelliklerini çıkarmaktır.

2. VERİ MADENCİLİĞİ

Veriler üzerinde analizler yapmak amacıyla çeşitli istatistiksel ve matematiksel yöntemler kullanılabilir. Ancak tutulan veri miktarı arttıkça bu veriler üzerinde analizler yapmak oldukça karmaşık hale gelmektedir. Bu veriler üzerinde doğru bir şekilde çözümlenmeler ve analizler yapmak için veri ambarı ve veri madenciliği kavramları ortaya atılmıştır [16].

Günümüz teknolojileri büyük miktarda verileri toplamak için gerekli ortamı sağlamaktadır. Fakat bu veriler işlenmedikçe toplanan verilerin bir anlamı olmamaktadır. Veri madenciliği sayesinde veriler arasındaki anlamlı olan ilişkiyi ortaya çıkarmak, veriler üzerinden çıkarımlar yapmak mümkün olmaktadır [15].

Veri madenciliği büyük ölçekli veriler arasından anlamlı olan bilgiye ulaşma yöntemi olarak tanımlanmaktadır. Günümüzde bilgisayarların hayatımıza dahil olmasıyla her yapılan işlem dijital ortamda kayıt altına alınmaktadır. Hastanelerde, mağazalarda, bankalarda, devlet dairelerinde yaptığımız her işlem veritabanlarında tutulmaktadır.

Tutulan bu verilerin hepsi veri madenciliği için birer kaynaktır [17]. Veri madenciliği, üretilen tüm bu verileri belirli yöntemler kullanarak, verilerin içindeki görülemeyen, gizli kalan bilgileri ortaya çıkarma süreci olarak değerlendirilmektedir [16].

2.1. Veri Madenciliği Veri Kaynakları

2.1.1. Veri ambarları

Veritabanlarında tutulan veriler doğrudan veri madenciliği uygulamalarında kullanılamaz. Bu verilerin doğru veri madenciliği süreçlerinden geçirilip kullanılması gerekir. Verilerin belirli bir zamana ait, yapılacak çalışmaya uygun şekilde konu odaklı, birleştirilmiş ve sabitlenmiş veritabanlarına veri ambarları nedir [17].

Veri ambarlarında aynı amaç ile ilgili olan veriler birbirine bağlanmıştır.

Veri ambarlarında veriler bir veya birden fazla veritabanından alınıp bir araya getirebilir. Bu veriler gerekli veri temizleme, dönüştürme gibi veri madenciliği süreçlerinden geçirilerek bir araya gelmiş verilerdir.

Veri ambarlarında bulunan veriler belirli bir zamana ait verilerdir. Örneğin son 1 yıllık veya 5 yıllık döneme ait veriler bir araya getirilmiş ve zaman içindeki değişimlerini gözlemlemeye uygun yapıda bulunmaktadır.

Veri ambarlarında bulunan veriler silinemez ve ambara yeni veri eklenemez. Veri ambarları veri giriş çıkışına kapalıdır [16].

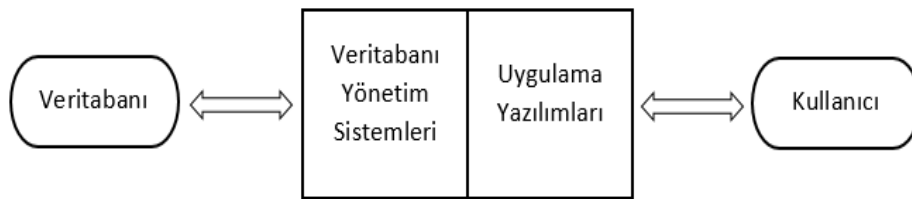
2.1.2. Veri kümeleri

Veriler genellikle bir veri tabanı ve veri ambarı tarafında düzenlenip yönetilmektedir. Kurumsal olan yapılarda veriler genellikle veritabanı ve veri ambarlarında saklanmaktadır. Fakat bireysel olarak çalışan araştırmacılar çoğunlukla veri kaynağı olarak veri kümeleri kullanırlar. Veri kümeleri de genellikle excel formatında dosyalar olmaktadır. Veri kümelerindeki veriler yapısal formda saklanmış olmayabilir. Bu yüzden, veri madenciliği uygulamalarında kullanmak için belirli süreçlerden geçirilmesi gerekebilmektedir.

2.1.3. Veritabanı sistemleri

Veritabanı sistemleri veri kümelerinin sistematik bir şekilde tutulduğu ortamlardır. Bu veriler çeşitli yazılımlar aracılığıyla yönetilmektedir. Geleneksel dosya sistemlerinin karmaşık yapısı ve veriye ulaşmada yaşanan sıkıntılar nedeniyle veritabanı yönetim sistemleri kullanılmaya başlanmıştır. Veritabanı sistemleri birbirleriyle ilişkili olan veri ve programlardan oluşmaktadır [16].

Şekil 2.1’de veritabanı ve kullanıcı arasında süreç gösterilmiştir.



Şekil 2.1. Veritabanı ve kullanıcı arasındaki süreç

Veritabanlarında veri ekleme, silme, güncelleme gibi işlemler yapılabilir. Fakat veri ambarlarında bu mümkün değildir. Veri ambarları veri tabanlarından daha basit yapıdadırlar. Veri ambarları karar destek sistemlerine yardımcı olmak üzere oluşturulmuş yapılardır [16].

2.2. Veri Madenciliği Uygulama Alanları

Veri madenciliği günümüzde sağlık, bankacılık, ticaret, pazarlama, sosyal medya gibi birçok sektörde kullanılabilmektedir [16].

Sağlık:

- Hastalıkların teşhisi
- DNA sıra analizi
- İlaç analizleri
- Risk analizi

Bankacılık:

- Kredi başvurularının değerlendirilmesi
- Müşteri kayıp analizleri
- Riskli müşteri profillerinin belirlenmesi
- Dolandırıcılık tespiti

Ticaret:

- Müşteri profillerinin belirlenmesi
- Memnuniyet analizleri
- Satış eğilimleri
- Yeni müşteri eğilimlerinin belirlenmesi
- Satış tahmini

Sosyal Medya:

- Duygu tespit ve analizleri
- Anket analizleri
- Okuyucu analizleri
- Popüler sanatçı, yayın, gazete araştırmaları

Günümüzde görüldüğü üzere pek çok alanda veri madenciliği çalışmaları yapılmaktadır. Veri madenciliği birbirinden bağımsız pek çok alanda kullanılabilmektedir. Veri madenciliğin bu kadar her alana yayılmasının sebebi hem veri madenciliği hem yapay zekâ tekniklerinin benzer algoritmaları kullanmasıdır [17]. Her iki alan birbirine önemli faydalar sağlamaktadır.

2.3. Veri Madenciliği Yöntemleri

Veri madenciliği yöntemleri temel olarak üç grupta incelenmektedir. Bunlar sınıflandırma, kümeleme ve birliktelik analizi yöntemleridir. Bu yöntemlerinin her birinin kendine ait algoritmaları vardır.

2.3.1. Sınıflandırma

Sınıflandırma en çok kullanılan ve bilinen veri madenciliği tekniklerinden birisidir. Sınıflandırmada kullanılan başlıca yöntemler karar ağaçları, yapay sinir ağları, mesafeye dayalı algoritmalar, istatistiğe dayalı algoritmalar olarak dört ana gruba ayrılmaktadır [17].

Sınıflandırma teknikleri, örüntü tanıma, hastalık tanı tahminleri, dolandırıcılık tespiti, hava tahmini, risk tespiti vb. konularda kullanılmaktadır.

Örnek: Bir banka kredi verdiği müşterilerinin risk durumlarını tespit etmek istemektedir. Risk durumlarını tespit ederek yeni bir kredi talebi geldiğinde bu müşteriye kredi verip vermeyeceğini değerlendirecektir. Bunu yaparken karar ağaçları yöntemi kullanılmıştır. Kullanılan eğitim verisinde müşterilerin borç durumu yüksek-düşük, gelir durumu yüksek-düşük, statü durumu işveren-ücretli, risk durumu ise iyi-kötü şeklinde tutulmuştur. Bu veriler karar ağacı algoritmalarından C4.5 algoritması kullanılarak eğitilmiş ve oluşan karar ağacından karar kuralları elde edilmiştir. Oluşan karar ağacı kurallarından birine örnek vermek gerekirse, “borç durumu ve gelir durumu düşük olan ücretli statüsünde çalışanların kredi risk durumu iyidir” şeklindedir. Oluşan bu kurallara dayanarak yeni gelen müşterilerin risk durumları hakkında çıkarımlar yapılabilmektedir [16].

Kullanılan algoritmalar: Bayesyen sınıflandırma, en yakın komşu, en küçük mesafe sınıflandırıcısı, karar ağaçları, yapay sinir ağları, genetik algoritmalar [17].

2.3.2. Kümeleme

Kümeleme yöntemi nesnelere inceleyerek birbirleriyle ortak özellikleri olanları aynı gruba, ortak özellikleri bulunmayanları ayrı bir gruba toplamaktadır. Verilerin arasındaki benzerlikler verilerin birbirleri arasındaki uzaklıklarının ölçülmesiyle bulunur. Veriler arasındaki uzaklıklar Öklid, Manhattan ve Minkowski uzaklık bağıntı hesaplama yöntemleri kullanılarak bulunur.

Kümeleme yöntemi ses tanıma, resim tanıma, istatistiksel tahmin, DNA analizi, coğrafi bilgi sistemleri vb. birçok alanda kullanılmaktadır.

Örnek: Ülkelerin ormanlık alan oranı, fosil enerji tüketim enerji tüketim oranı ve karbondioksit salınım oranı verileri kullanılarak insan sağlığı için riskli ya da risksiz bölgeler, çevre için önlem alınması ya da önleme gerek olmayan bölgeler belirlenmiştir. Bu işlemi yaparken k-ortalamlar algoritması kullanılmıştır. Bu algortmada k parametresinin belirlenmesi gerekmektedir. K parametresi küme sayısı demektir. K parametresi 10'dan küçük seçilirse daha keskin ayrımlar, 10'dan büyük seçilirse daha ince ayrımlar yapılabilmektedir. Algoritma uygulandıktan sonra fosil enerji tüketim miktarı fazla olan kümeden yola çıkılarak riskli ya da risksiz bölgeler belirlenip, gerekli önlemlerin alınması konusunda yol gösterici sonuçlar elde edilmiştir.

Kullanılan algortmalar: K-ortalama, BIRCH, PAM, dalga kümeleme, yapay sinir ağları, genetik algortmalar [17].

2.3.3. Birliktelik Analizi

Veritabanı içinde bulunan kayıtların birbirleri ile olan ilişkileri incelenerek, olayların eş zamanlı meydana gelme olasılıklarının belirlenmesinde kullanılan yöntem birliktelik analizi yöntemidir.

Birliktelik analizi, önceki yıllarda süpermarket verilerinin analizi üzerine yapılan çalışmalarda sıklıkla kullanıldığından market sepet, pazar sepet analizi olarak da anılmaktadır. Bu çalışmalarda müşteri bir ürün aldığında, bu ürün ile hangi ürünü veya ürünleri sepetine attığının olasılığı ortaya konulmaktadır. Bu olasılıklar belirlendiğinde marketlerdeki, mağazalardaki raflar müşteri alışkanlıklarına göre düzenlenebilmektedir. Böylelikle müşteriler ürünlere daha kolay erişebilmektedir.

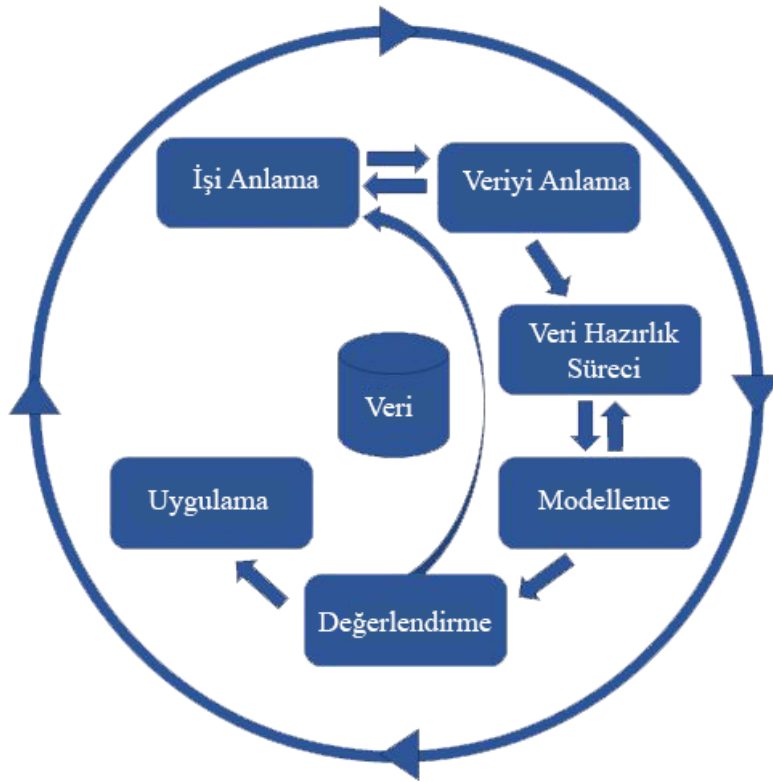
Örnek: Bir marketten alışveriş yapan müşterilerin sepetlerinde bulunan ürünler inceleyerek birliktelik analiz çalışması yapılmak istenmektedir. Bu çalışma sonucunda müşterilerin ürünleri birlikte alma olasılıkları belirlenecektir. Birliktelik analizi çalışması yapılırken apriori algortmasından yararlanılacaktır. Aşağıda müşterilerin sepetindeki ürünlere ait tablo verilmiştir.

Kullanılan algortmalar: Apriori, Apriori TID, SETM [17].

2.4. Veri Madenciliği Süreci

Bir veri projesine başlanırken nereden başlanacağı, hangi yolların izleneceği, hangi adımlarda neler yapılacağı, proje aşamalarının çıktıları CRISP-DM adı verilen süreç ile yönetilebilmektedir [18]. CRISP-DM (Cross-Industry Standard Process for Data Mining), veri madenciliği süreci ile ilgili en yaygın kullanılan metodolojilerden birisidir. CRISP-DM kelimesinin türkçeye çevrilmiş hali Çapraz Endüstri Veri Madenciliği Standart Süreci olarak adlandırılmaktadır. CRISP-DM metodolojisi genel olarak işi anlama, veriyi anlama, veri hazırlık süreci, modelleme, değerlendirme ve uygulama adımları olmak üzere altı adımdan oluşmaktadır.

CRISP-DM süreci Şekil 2.2' de görsel olarak verilmiştir.



Şekil 2.2. CRISP-DM süreci [19].

2.4.1. İş anlama

İş anlama adımı, CRISP-DM sürecinin ilk adımıdır. Bu adımda en önemli olan şey veri madenciliği hedefinin belirlenmesidir [20]. İş anlama evresinde öncelikle işin amacı ve kapsamı belirlenmektedir. Hangi yöntemler kullanılacağı, projenin çıktıları ve projeden elde edilmesi beklenen sonuçlar açık bir dille ortaya konulmaktadır.

2.4.2. Veriyi anlama

Veriyi anlama adımı CRISP-DM sürecinin ikinci adımıdır. Bu evrede probleme uygun olan veri toplanmakta ya da probleme uygun, elimizde bulunan veri üzerinde çalışılmaya başlanmaktadır. Veriyi anlamak için verinin içindeki niteliklerin tespit edilip doğru anlaşılması gerekmektedir. Doğru anlaşılmayan bir veri üzerinde doğru çıkarımlar yapmak mümkün olmamaktadır. Veriyi anlama noktasında bulunan eksiklikleri gidermek için veriyi elde ettiğimiz kaynaklara başvurmak doğru bir tercih olacaktır.

2.4.3. Veri hazırlık süresi

Veri hazırlık süreci CRISP-DM sürecinin üçüncü adımıdır. Kullanılmak verilerin mevcut haliyle projede kullanılması sağlıklı çıktılar elde etmek için yeterli olmayabilir. Sonuçların doğruluğundan emin olmak için verilerin bir standartta olması gerekmektedir. Bu standarda ulaşmak için verilerin çeşitli işlemlerden geçmesi gerekmektedir. Çalışmalarda en zaman alıcı kısım veri hazırlık süreci olmaktadır. Veritabanındaki satırlardaki eksik veriler, tutarsız girilen veriler, birden fazla aynı anlama gelen sütunlar, tekrarlanan veriler gibi sorunları düzeltilmesi gerekmektedir. Örneğin hem doğum tarihi bilgisi hem de yaş bilgisi tutmak aynı anlama geldiği için gereksizdir. Bunlardan birinin kaldırılması doğru bir yaklaşım olacaktır. Veri hazırlık aşaması verilerin temizlenmesi ve verilerin tekrar dönüştürülmesi olmak üzere iki başlık şeklinde incelenebilmektedir [17].

2.4.3.1. Verilerin temizlenmesi

Verilerin temizlenmesi aşamasında yapılacak işlemler kayıp verilerin ortaya çıkaracağı sonuçların giderilmesi ve verilerdeki gürültüleri yok etme işlemleridir.

Kayıp veriler: Kayıp verilerin oluşturacağı karmaşıklıkları ortadan kaldırmak için belirli teknikler bulunmaktadır. Bu teknikler maddeler halinde aşağıda verilecektir [17].

- Kayıp verilerin sayısı toplam veri kaydının içinde küçük bir yüzdeye sahip ise bu verilerin bulunduğu satırlar veritabanından çıkarılabilmektedir.
- Küçük bir veritabanı ise ve verilere ulaşmak kolay olacaksa kayıp olan veriler elle doldurulabilmektedir.

- Tüm kayıp verilere aynı bilgiyi girmek bir çözüm olabilir fakat ortaya çıkan sonuçlar her zaman sağlıklı olmayacaktır.
- Kayıp olan verilere bulunan diğer verilerin ortalaması olan bir değer verilerek kayıp veri sorunu ortadan kaldırılabilir.
- Eksik olmayan veriler kullanılarak regresyon denklemi ve regresyon katsayısı elde edilerek kayıp verilerin tahmin edilmesi yolu izlenebilir. Ayrıca zaman seri analizi, bayesyen sınıflandırma, karar ağaçları gibi diğer yöntemlerde kayıp veri tahmininde kullanılabilir.

Verilerdeki gürültüyü ortadan kaldırma: Gürültülü veri veritabanına yanlış girilmiş ya da diğer tüm veriler içinde değeri uçlarda olan veriler olarak tanımlanabilir [17]. Çalışmanın düzgün sonuçlar vermesi için bu verilerin temizlenmesi gerekmektedir. Gürültülü verilerin temizlenmesinde kullanılan yöntemler; kutulama yöntemiyle gürültünün temizlenmesi, sınırlar yardımı ile düzgünleştirme yapılarak gürültünün temizlenmesi ve kümeleme yöntemiyle gürültünün temizlenmesi olarak açıklanabilmektedir.

2.4.3.2. Verilerin tekrar yapılandırılması

Veri madenciliği uygulamasında kullanılan yöntemler her zaman aynı tür veriler ile çalışmamaktadır. Bazı yöntemler sadece sayısal değerler alırken bazıları metin şeklinde değer isteyebilmektedir. Bazı yöntemlerde ise sadece 0 ve 1'lerden oluşan değerler ile işlem yapılmaktadır [17]. Böyle durumlarda elimizdeki verileri uygun formlara getirmemiz gerekmektedir. Verileri uygun formlara getirmek için aşağıda bazı yöntemlerden bahsedilmiştir.

Normalizasyon: Normalizasyon işlemi eldeki verileri 0.0 ve 1.0 gibi veri aralığına getirme işlemidir. Normalizasyon işlemi yapabilmek için çeşitli yöntemler kullanılır. Min-max normalizasyonu, sıfır ortalama normalizasyonu ve ondalıklı normalizasyon yöntemleri normalizasyon işlemi için kullanılan yöntemlerden bazılarıdır. Min-max normalizasyon yönteminde min verinin alacağı en küçük değeri, max ise en büyük değeri ifade eder. Normalizasyon işleminin formülü aşağıdaki gibidir [17]. Burada V , verinin orijinal değerini, V' ise verinin normalize edildikten sonra aldığı değeri ifade etmektedir. Aşağıdaki denklem veriyi 0-1 aralığında normalize etmektedir.

$$V' = \frac{V - \text{minVal}}{\text{maxValue} - \text{minVal}} \quad (2.1)$$

Ancak veri sadece 0-1 aralığında değil örneğin 1-5 aralığında normalize etmek gerekebilmektedir. Böyle durumlarda 2. Denklem kullanılmaktadır. Buradaki newMin normalize etmek istediğimiz aralıktaki küçük olan değer, newMax ise normalize etmek istediğimiz aralıktaki büyük olan değeri ifade etmektedir.

$$V' = \frac{V - \text{minVal}}{\text{maxValue} - \text{minVal}} * (\text{newMax} - \text{newMin}) + \text{newMin} \quad (2.2)$$

Örneğin yaş verisinin tutulduğu bir sütunda en küçük değer 18, en büyük değer 80 olsun. 30 verisini 0-1 aralığında normalize etmek istersek denklem 2.1'de V yerine 30 değerini, minVal yerine 18 değerini, maxVal yerine ise 80 değerini koyarak V' değerini 0,193 buluruz. Bu değer 30 verisinin normalize edilmiş değeridir. Aynı şekilde 30 verisini 1-5 aralığında normalize etmek istersek bu sefer denklem 2.2'yi kullanmamız gerekmektedir. Burada V yerine 30, minVal yerine 18, maxVal yerine 8*, newMax yerine 5, newMin yerine 1 koyarak V' değerini 1,774 buluruz. Bu değer 30 verisinin 1-5 aralığında normalize edilmiş değeridir.

Sıfır-ortalama normalizasyon yönteminde ise normalizasyon yapılmak istenen sütunun standart sapmasına ve ortalamasına ihtiyaç duyulmaktadır. Sıfır-ortalama normalizasyon yönteminin denklemi aşağıdaki gibidir.

$$V' = \frac{V - \text{ortalama}}{\text{standartSapma}} \quad (2.3)$$

Boyut indirgeme: Normalizasyon işleminin dışında verilerin sayılarının azaltılması yani indirgenmesi gerekebilmektedir. Yani verideki değişken miktarının azaltılması demektir. Boyut indirgeme işleminde birden fazla değişken birleştirilerek tek bir değişkenle ifade edilmektedir. Boyut indirgeme yöntemi olarak dalga dönüşümü ve temel bileşenler analizi yöntemleri kullanılmaktadır [17].

Görsel etiketleme: Bu yöntemdeki verilerin histogram grafiği çıkarılarak verilerin arasında keskin farklar olan yerler belirlenmektedir. Bu farkların olduğu yerler sınır noktalar olarak kabul edilmektedir. Örneğin yaş verisinin tutulduğu bir sütunda veriler histogram grafiğinden yola çıkılarak aşağıdaki gibi etiketlenebilmektedir.

18-29 yaş aralığı: 1. Grup

30-37 yaş aralığı: 2. Grup

38-45 yaş aralığı: 3. Grup

46-60 yaş aralığı: 6. Grup

60+ yaş aralığı: 5. Grup

Etiketlere verilen isim yapılacak olan çalışmanın amacına ve konusuna göre değişebilmektedir.

2.4.4. Modelleme

Modelleme adımı CRISP-DM sürecinin dördüncü adımıdır. Modelleme adımında uygulanacak olan veri madenciliği yöntemi belirlenmektedir. Bu aşamada tüm veri madenciliği yöntemleri kullanılabilir. Eldeki verilere, belirlenmiş olan iş tanımına ve kriterlerine uygun olan modelin seçilmesi bu aşamada önemlidir. Doğru modeli bulana kadar birçok model denemek gerekebilir. Düzenlenmiş olan verilerde seçilen modele göre veri dönüşüm ve veri temizleme işlemlerinin yapılması gerekebilir.

2.4.5. Değerlendirme

Değerlendirme adımı CRISP-DM sürecinin beşinci adımıdır. Bu adımda elde edilen sonuçlar ve süreç genel olarak gözden geçirilmektedir. Elde edilen sonuçların sürecin başında tanımlanan iş tanımına ve ulaşılmak istenen kriterleri ne kadar sağladığına bakılmaktadır.

2.4.6. Uygulama

Uygulama adımı CRISP-DM sürecinin son adımıdır. Bu aşamada oluşturulan modelden elde edilen çıktılar ile süreç tamamlanmış olmaktadır. Elde edilen sonuçlar değerlendirilerek ilk aşamada ortaya konan problem tanımına ne kadar yaklaşıldığı test edilir. Oluşturulan model, bağımsız bir uygulama olabileceği gibi başka uygulamalar içinde kullanılacak bir parçaya olabilmektedir. Örneğin bir e-ticaret sitesi için çalışıldıysa uygun iş akışı belirlenerek platformun kodlanmasına geçilebilir. Bir reklam firması için çalışıldıysa elde edilen sonuçlar ile uygun olan reklamlar müşteriye gösterilebilmektedir [18].

3. BİRLİKTELİK ANALİZİ

Zaman ilerledikçe kurumlarda kayıt altına alınan veri miktarı da artmaktadır. Veri miktarı arttıkça bu verilerin arasındaki ilişkiyi de ortaya koyma ihtiyacı doğmuştur. Verilerin arasındaki ilişkinin ortaya çıkması kurumlarda karar almada ve işleyişte önemli bir yere sahiptir.

Birliktelik analizi, olayların aynı anda meydana gelme olasılıklarını ortaya çıkaran veri madenciliği yöntemidir. Birliktelik analizi uygulamalarının en yaygın örnekleri pazar sepet analizi, market sepet analizi olarak adlandırılan uygulamalardır.

Bu uygulamalarda müşterinin ürün satın alma tercihleri belirlenmektedir. Müşterinin bir ürün aldığı anda yanında hangi ürünü hangi olasılıkla tercih ettiği, en çok hangi ürünün satıldığı, en az hangi ürünün satıldığı ortaya koymaktadır. Bu olasılıkların hesaplanması sonucunda, birlikte satın alınan ürünleri mağazalarda yakın raflara koyarak, satışı az olan ürünler için kampanyalar düzenleyerek mağazaların satış grafikleri artırılabilir.

Birliktelik analizi çalışmalarının yapılması için çeşitli algoritmalar geliştirilmiştir. Bu algoritmalarından en çok tercih Apriori Algoritması, AIS ve SETM algoritmalarıdır.

3.1. Destek ve Güven Değeri

Birliktelik analizi uygulamalarında birliktelik kuralları çıkarılırken destek (support) ve güven (confidence) değeri olarak isimlendirilen iki değerden yararlanılmaktadır.

Destek değeri, bir verinin veri kümesi içinde ne sıklıkla geçtiğini belirtmektedir. Güven değeri ise bir verinin diğer veriler ile veri kümesi içinde ne sıklıkla geçtiğini belirtmektedir.

X ve Y farklı iki veri olmak üzere X ürünü için destek değeri aşağıdaki gibi hesaplanmaktadır.

$$\text{Destek}(X) = \frac{\text{Sayı}(X)}{N} \quad (3.1)$$

Yukarıdaki denklemde Sayı(X), tüm veri seti içinde X verisinin kaç kez bulunduğunu belirtmektedir. N değeri veri setinin boyutudur. Destek(X) ise tüm veri seti içinde X verisinin bulunma olasılığını bize vermektedir.

X ve Y verisinin destek değeri ise aşağıdaki denklem ile hesaplanmaktadır.

$$\text{Destek}(X, Y) = \frac{\text{Sayı}(X, Y)}{N} \quad (3.2)$$

Yukarıdaki denklemde Sayı (X, Y), tüm veri seti içinde X ve Y verisinin kaç kez bir arada bulunduğunu belirtmektedir. N değeri veri setinin boyutudur. Destek (X, Y) ise tüm veri seti içinde X ve Y verisinin tüm veri seti içerisinde birlikte bulunduğu durumların olasılığını bize vermektedir.

Y ürününün hangi olasılıkla X ürünü birlikte bulunduğunu hesaplamak için ise aşağıdaki denklem kullanılmaktadır.

$$\text{Güven}(X, Y) = \frac{\text{Sayı}(X, Y)}{\text{Sayı}(X)} \quad (3.3)$$

Yukarıdaki denklemde Sayı (X, Y), tüm veri seti içinde X ve Y verisinin kaç kez bir arada bulunduğunu belirtmektedir. Sayı (X), tüm veri içinde X verisinin kaç kez bulunduğunu belirtmektedir. Güven (X, Y) ise tüm veri seti içinde Y verisinin hangi olasılıkla X verisi ile birlikte bulunduğunu bize vermektedir.

Destek ve güven değerlerini hesapladıktan sonra bunları karşılaştırabilmemiz için destek eşik ve güven eşik değerlerine ihtiyacımız olmaktadır [16]. Hesapladığımız destek ve güven değerlerinin belirlediğimiz eşik değerlerinden büyük olması beklenmektedir. Hesaplanan değerler, belirlenen eşik değerinden ne kadar büyük ise oluşturulan birliktelik kurallarının doğruluğu o kadar güçlü kabul edilmektedir.

Örneğin; 100 müşteriden 30'u peynir, ekmek ve yumurtayı birlikte almış ise destek değeri hesabı aşağıdaki gibi olmaktadır. (Peynir: P, Ekmek: E, Yumurta: Y) [21].

$$\text{Destek}(P, E \rightarrow Y) = \frac{\text{Sayı}(P, E, Y)}{\text{Toplam müşteri}} = \frac{30}{100} = 0,3$$

Yani peynir ve ekmek alan müşterilerin bu ürünlerin yanında yumurta da alma olasılığı 0,3'tür.

Peynir ve ekmeđi birlikte alan müşterilerin sayısını 40 olduđu varsayarsak güven deđeri hesabı ařađıdaki gibi olmaktadır.

$$\text{Güven}(P, E \rightarrow Y) = \frac{\text{Sayı}(P, E, Y)}{\text{Sayı}(P, E)} = \frac{30}{40} = 0,75$$

3.2. Lift Deđeri

Birliktelik analizi uygulamalarında destek ve güven deđeri dıřında önemsedimiz bir diđer deđer de lif deđeridir. Lift deđeri ilginçlik, farklılık deđer olarak yorumlanmaktadır. Lift deđer 1'den ne kadar büyük çıkarsa o kuralın farklılığı o kadar yüksek kabul edilmektedir.

$$\text{Lift}(X, Y) = \frac{\text{Güven}(X, Y)}{\text{Destek}(X)} \quad (3.4)$$

Örneđin bir marketten alınan patlıcan, sođan ve sarımsak ürünleri birlikte incelendiđinde destek, güven ve lift deđerleri ařađıdaki gibi hesaplanmıřtır [21].

$$\text{Destek}(\text{Su}) = 3$$

$$\text{Güven}(\text{Ekmek}, \text{Su}) = \%75$$

$$\text{Lift}(\text{Ekmek}, \text{Su}) = \frac{\%75}{3} = \%25$$

$$\text{Destek}(\text{Kola}) = 5$$

$$\text{Güven}(\text{Ekmek}, \text{Kola}) = \%75$$

$$\text{Lift}(\text{Ekmek}, \text{Kola}) = \frac{\%75}{5} = \%15$$

Yukarıdaki lift deđerleri incelendiđinde Lift(Ekmek, Su) deđer Lift(Ekmek, Sođan) deđerinde daha büyük çıkmıřtır.

Buda bize ekmek ürününün su ürünü üzerindeki etkisinin kola ürününden üzerindeki etkisinden daha fazla olduđu göstermektedir.

3.3. Birliktelik Analizi Algoritmaları

3.3.1. Apriori algoritması

Apriori algoritması Agrawal ve Srikant tarafından 1994'te ortaya çıkarılan bir algoritmadır [24]. Birliktelik analizi uygulamalarında en çok tercih edilen algoritmadır. Nesne kümeleri arasındaki bağıntıları çıkarmak için kullanılmaktadır. Algoritma adımları aşağıdaki gibi sıralanabilmektedir.

- Destek ve güven değerlerini karşılaştırmak üzere eşik değerler belirlenir.
- Her bir ürün için destek değerleri hesaplanır.
- Hesaplanan destek değerleri eşik destek değeri ile karşılaştırılır.
- Eşik destek değerinden küçük olan değerlere sahip satırlar tablodan çıkarılır.
- Tabloda kalan ürünler ikişerli gruplandırılarak, grupların destek değerleri elde edilir.
- Eşik destek değerinden küçük olan satırlar tablodan çıkarılır.
- Üçerli, dörderli gruplandırmalar da yapılarak aynı işlemler tekrarlanır.
- Ürün grubu belirlendikten sonra kural destek değerine bakılarak birliktelik kuralları oluşturulur.
- Oluşturulan kuralların güven değerleri hesaplanır.

Algoritmanın anlaşılabilmesi için aşağıdaki örnek bir çalışmaya yer verilmiştir. Bu örnekte bir polikliniğe başvuran hastaların şikayetleri incelenmiştir. Minimum destek eşik değeri %60, güven eşik değeri ise %75 olarak belirlenmiştir. Tablo 3.1.'de hasta şikayetleri verilmiştir.

Tablo 3.1. Hasta şikâyet tablosu.

Hasta	Şikâyet
1	Ateş, Baş Ağrısı, Vücut Ağrısı
2	Baş Ağrısı, Kusma, İshal, Öksürük
3	Ateş, Kusma, Boğaz Ağrısı, Baş Ağrısı, Öksürük
4	Baş Ağrısı, Kusma, Vücut Ağrısı, Öksürük
5	Kusma, Öksürük, Ateş, Tansiyon

Burada 5 hasta olduğuna göre ve destek eşik değeri 0.60 olduğuna göre destek sayısı 3'tür.

Tablo 3.2. Destek değerleri

Şikâyet	Sayı
Ateş	3
Vücut Ağrısı	2
Baş Ağrısı	4
Kusma	4
Öksürük	4
İshal	1
Boğaz Ağrısı	1
Tansiyon	1

Destek eşik sayısı 3 olduğuna göre değeri 3'ten küçük olanların tablodan çıkartılması gerekmektedir.

Tablo 3.3. Eşik destek değerine eşit ve büyük olan şikâyetler

Şikâyet	Sayı
Ateş	3
Baş Ağrısı	4
Kusma	4
Öksürük	4

Şikâyetler belirlendikten sonra Tablo 3.1'de yer alan şikâyet gruplarından ikili şikâyet grupları oluşturulmaktadır. Daha sonra oluşturulan ikili şikâyet gruplarının destek değerleri hesaplanacaktır.

Tablo 3.4. İkili şikâyet grupları ve destek değerleri

Şikâyet	Sayı
Ateş, Baş Ağrısı	2
Ateş, Kusma	2
Ateş, Öksürük	2
Baş Ağrısı, Kusma	3
Baş Ağrısı, Öksürük	3
Kusma, Öksürük	4

İkili şikâyet grupları ve destek değerleri hesaplandıktan sonra bu tablodan eşik destek sayısı 3'ten küçük olan şikâyetler çıkarılacaktır. İkili şikâyet grupları belirlendikten sonra üçlü şikâyet grupları belirlenecektir.

Tablo 3.5. Eşik destek değerine eşit ve büyük olan ikili şikâyet grupları

Şikâyet	Sayı
Baş Ağrısı, Kusma	3
Baş Ağrısı, Öksürük	3
Kusma, Öksürük	4

Tablo 3.6. Eşik destek değerine eşit ve büyük olan üçlü şikâyet grupları

Şikâyet	Sayı
Baş Ağrısı, Kusma, Öksürük	3

Üçlü şikâyet grupları da tespit edildikten sonra birliktelik kuralları oluşturulabilir duruma gelmiş demektir. {Baş Ağrısı, Kusma, Öksürük} kümesi için kural destek değeri hesaplaması aşağıdaki gibi yapılmaktadır.

$$\text{Destek}(A \rightarrow B) = \frac{\text{Sayı}(\text{Baş Ağrısı, Kusma, Öksürük})}{\text{Hasta Sayısı}} = \frac{3}{5} = 0,6$$

Bulunan değer belirlenen destek eşik değerinden küçük olmadığına göre bu kural geçerli bir kuraldır.

Tablo 3.7. Oluşturulan birliktelik kuralları

Birliktelik Kuralı	Açıklama	Güven Değeri
Baş Ağrısı, Kusma →Öksürük	Baş ağrısı ve kusma şikâyeti bulunan hastalarda öksürük şikâyetinin de bulunma olasılığı	3/3=% 100
Baş Ağrısı → Öksürük, Kusma	Baş ağrısı şikâyetinin yanında öksürük ve kusma şikâyetinin birlikte bulunma olasılığı	3/4=% 75
Öksürük → Baş Ağrısı, Kusma	Öksürük şikâyetinin yanında baş ağrısı ve kusma şikâyetinin birlikte bulunma olasılığı	3/4=% 75
Kusma → Baş Ağrısı, Öksürük	Kusma şikâyetinin yanında baş ağrısı ve öksürük şikâyetinin birlikte bulunma olasılığı	3/4=% 75

Tablodaki en yüksek güven değerine sahip kural şu şekilde yorumlanmaktadır. Baş ağrısı ve kusma şikâyetinin birlikte bulunduğu hastalarda öksürük şikâyetinin de bulunma olasılığı %100'dür.

4. UYGULAMA

Bu çalışmada Covid-19 yatan hasta verileri kullanılmıştır. Çalışma sonucunda, Covid-19 nedeniyle hastanede tedavi olmak zorunda kalan hastaların ortak özellikleri belirlenecektir.

4.1. Veriyi Anlama ve Hazırlık Süreci

Bu çalışmada veri seti olarak Geyve Devlet Hastanesi Covid-19 servisinde 2 Nisan 2020 ile 30 Eylül 2021 tarihleri arasında yatan hastaların verileri kullanılmıştır. Pandemi döneminde Geyve Devlet Hastanesinde hastaneye ilk yatış 2 Nisan 2020 tarihinde olmuştur. Hastane verileri bir excel dosyası olarak teslim etmiştir. Bu dosyada her satırda hastaya ait veriler tam şekilde doldurulmamış olup, eksik olan veriler hastane sisteminden temin edilip tek tek işlenmiştir. Veri setinde bulunan aynı kayıtlar tespit edilip veri setinden kaldırılmıştır. Birden fazla kez hastaneye yatan hastaların kayıtları birleştirilmiştir. Tüm bu işlemlerin sonunda çalışmamızda kullanılmayacak olan sütunlar kaldırılmıştır. Tüm sütunlardaki veriler anlaşıldıktan sonra, elimizde 8 adet sütun yani nitelik kalmıştır. Bunlar aşı türü, cinsiyeti, yaşı, hastaneye yatmadan önce aşısı varsa bu aşının türü, hastaneden çıktıktan sonra aşı olduysa bu aşının türü, hastanede kalma süresi, hastaneden çıkış türü ve tanı nitelikleridir.

Hastaların aşı bilgileri 1 Ekim 2021 tarihinde hastane sisteminden tek tek işlenmiştir. Hastaların o tarihteki aşı bilgileri dosyaya girilmiştir. Türkiye’de aşılama süreci sağlık bakanlığı tarafından 14 Ocak 2021 tarihinde Sinovac firmasının CoronaVac aşısı ile başlamıştır. Çalışmada bundan sonra CoronaVac aşısı Sinovac olarak anılacaktır. Türkiye 12 Nisan 2021 tarihinden itibaren Biontech aşısının kullanıma başlamıştır. 30 Haziran 2021 tarihinden sonra 50 yaş ve üzerindeki ve sağlık personellerinin 3.doz aşı olması konusunda çağrıda bulunulmuştur. 16 Ağustos 2021’de bazı ülkelerin, ülkelerine gelenlerden iki Biontech aşısı istemesi üzerine iki Sinovac ve bir Biontech aşısı yaptıranlara 4. Doz aşı olma hakkı tanınmıştır [3]. Çalışmada kullanılan aşı durumları ve açıklamaları Tablo 4.1.’de verilmiştir.

Tablo 4.1. Aşı durumları.

Aşı Durumu	Açıklama
1BIONTECH	1 tane Biontech aşısı olanlar
1SINOVAC	1 tane Sinovac aşısı olanlar
1SINOVAC+1BIONTECH	1 tane Sinovac aşısı ve 1 tane Biontech aşısı olanlar
1SINOVAC+2BIONTECH	1 tane Sinovac aşısı ve 2 tane Biontech aşısı olanlar
2BIONTECH	2 tane Biontech aşısı olanlar
2SINOVAC	2 tane Sinovac aşısı olanlar
2SINOVAC+1BIONTECH	2 tane Sinovac aşısı ve 1 tane Biontech aşısı olanlar
2SINOVAC+2BIONTECH	2 tane Sinovac aşısı ve 2 tane Biontech aşısı olanlar
3SINOVAC	3 tane Sinovac aşısı olanlar
ASISIZ	Hiç aşısı olmayanlar

Yaş verisi, hastanın doğum tarihi verisinden hesaplanarak ayrı bir sütun oluşturulmuştur. Daha sonra yaş verisi kategorize edilerek düzenlenmiştir. Tablo 4.2.'de yaş sütununun detay bilgileri verilmiştir.

Tablo 4.2. Yaş kategori tablosu.

Yaş	Kategori
15 – 29 yaş aralığı	0
30 – 44 yaş aralığı	1

Tablo 4.2. (Devamı) Yaş kategori tablosu.

Yaş	Kategori
45 – 59 yaş aralığı	2
60 – 74 yaş aralığı	3
75 – 89 yaş aralığı	4
90+ yaş	5

Hastanede kalma süresi verisi, hastanın hastaneye giriş yaptığı gün kaydı ile hastaneden çıkış yaptığı gün kaydının farkı alınarak, hastanede kalma süresi hesaplanmıştır. Hastanede kalma süreleri kategorize edilerek düzenlenip hastanede kaldığı gün aralığı olarak ayrı bir sütun haline getirilmiştir. Tablo 4.3.'te hastanede kalma sürelerinin detay bilgileri verilmiştir.

Tablo 4.3. Yatış süresi kategori tablosu.

Yatış süresi	Kategori
0 – 7 gün aralığı	0
8 – 15 yaş aralığı	1
16 – 30 gün aralığı	2
30+ gün	3

Tanı sütunu sadece Covid-19 tanısı ve Covid-19 tanısıyla birlikte ek hastalık kaydı girilenler olarak ikiye ayrı şekilde girilmiştir. Tablo 4.4'te tanı detay bilgileri verilmiştir.

Tablo 4.4. Tanı tablosu.

Tanı	Kategori
Sadece Covid-19 tanısı	C0

Tablo 4.4. (Devamı) Tanı tablosu.

Tanı	Kategori
Covid-19 + Ek Hastalık	C1

Hastaneden çıkış türü verisi olarak hastaneden taburcu, başka servise geçiş ve sevk olmak üzere üç farklı şekilde girilmiştir.

Cinsiyet sütunu ise kadın ve erkek olarak girilmiştir.

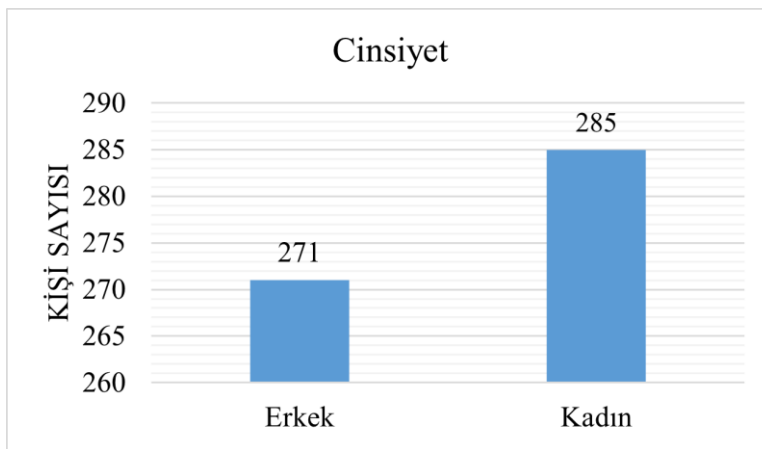
Hastaların aşı olduğu tarih ve hastaneye yattığı tarih bilgisinden çıkarımlar yapılarak, hastalanmadan önceki ve sonraki aşı tercihlerinin ne olduğunu gösteren sütunlar eklenmiştir. Hastanın hiç aşısı yok ise aşısız olarak kayıt girilmiştir. Aşısı olanlarda ise hastalanmadan önce veya sonra aşısı yok ise yok şeklinde kayıt girilmiştir.

4.2. Analiz Sonuçları

Veriler düzenlenip işlenebilir hale getirildikten sonra veri üzerinde analizler yapılmıştır. Bu başlık altında veriler üzerinden ulaşılan sonuçlar grafiksel olarak yorumlanacaktır.

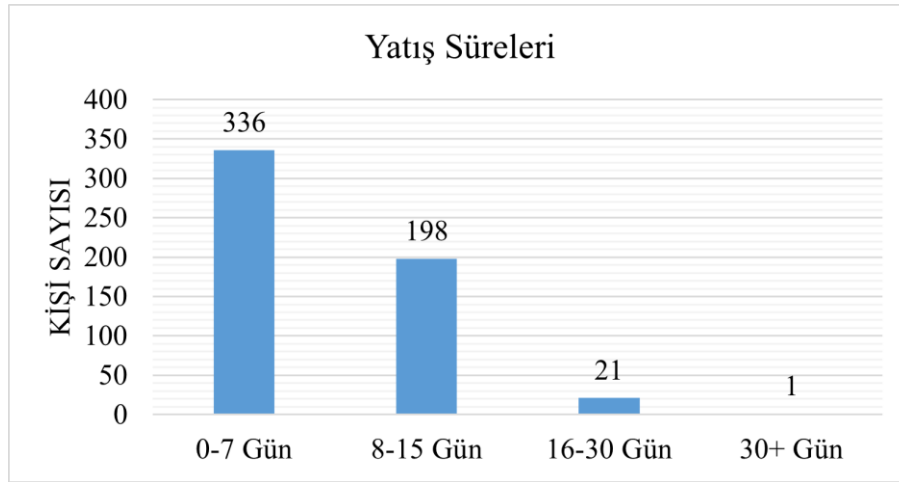
Yapılan analizler sonucunda aşağıda yatan hastalara ait grafikler sunulmuştur. Toplam 556 yatan hasta kaydı bulunmaktadır.

Hastaların cinsiyete göre dağılımı 271 kişinin cinsiyeti erkek, 285 kişinin cinsiyeti kadın şeklindedir. Cinsiyet dağılım grafiği Şekil 4.1.'de verilmiştir.



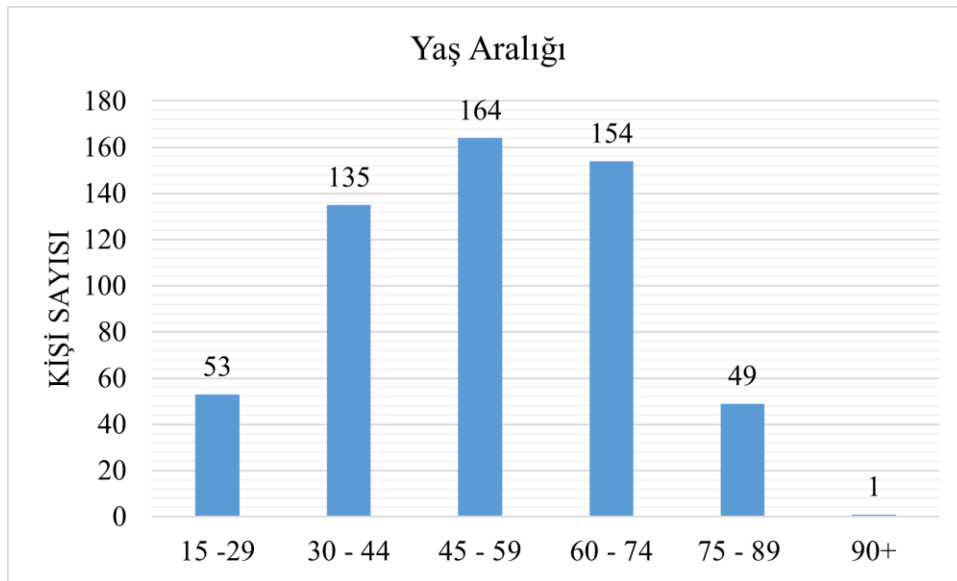
Şekil 4.1. Cinsiyet dağılım grafiği.

Hastaların hastaneye yatış süresi olarak dağılım grafiği Şekil 4.2.'de verilmiştir. 336 kişi ile yatılan gün sayısı 0-7 gün aralığında yoğunlaşmıştır.



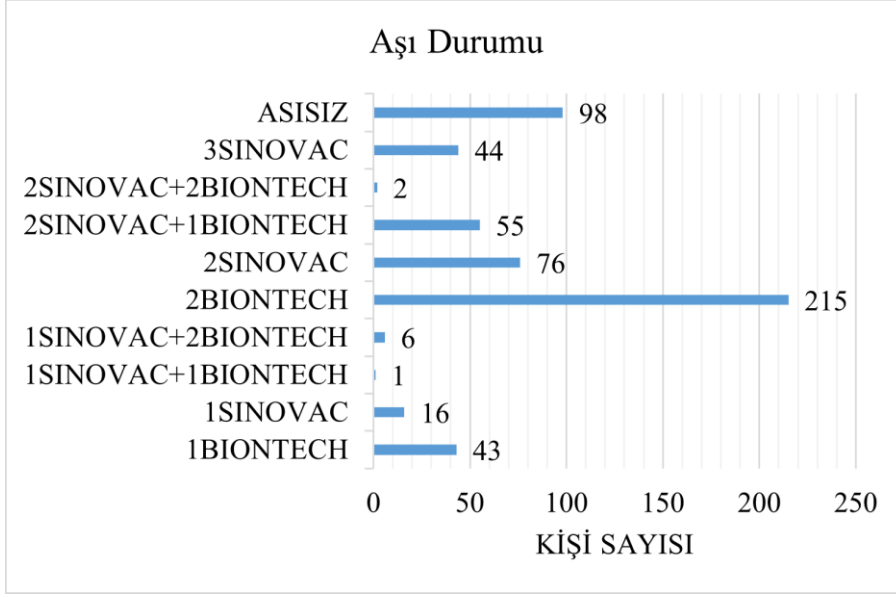
Şekil 4.2. Yatış süresi dağılım grafiği.

Hastaneye yatış yapan hastaların yaş dağılım grafiği Şekil 4.3.'te verilmiştir. 45 – 59 yaş aralığında 164 kişi hastaneye yatmıştır. En çok birikim bu yaş aralığında olmuştur. Ardından 154 kişi ile en çok yatış 60 – 74 yaş aralığında olmuştur.



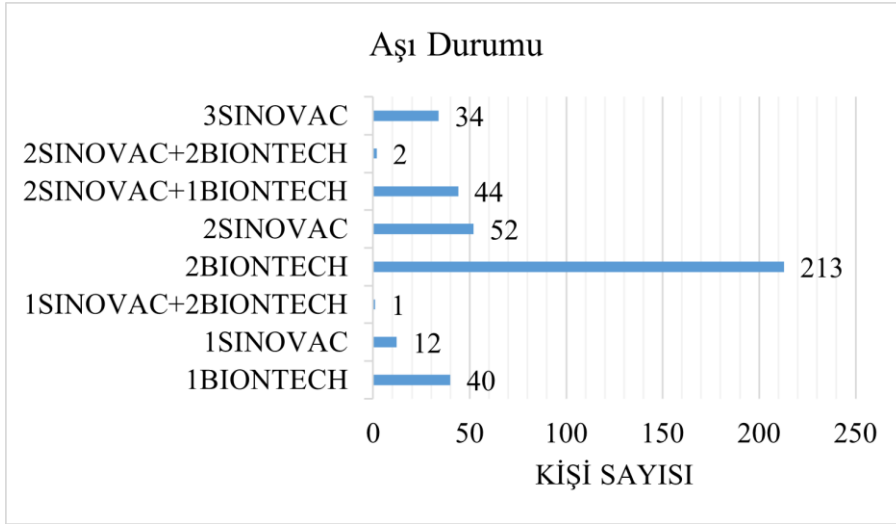
Şekil 4.3. Yaş dağılım grafiği.

Hastaneye yatan hastaların aşı tercihleri Şekil 4.4.'te verilmiştir. 215 kişi en çok tercih edilen aşı kombinasyonu iki Biontech aşısı olmuştur. Ardından en çok tercih edilen aşı kombinasyonu 76 kişi işe iki Sinovac aşısı olmuştur.



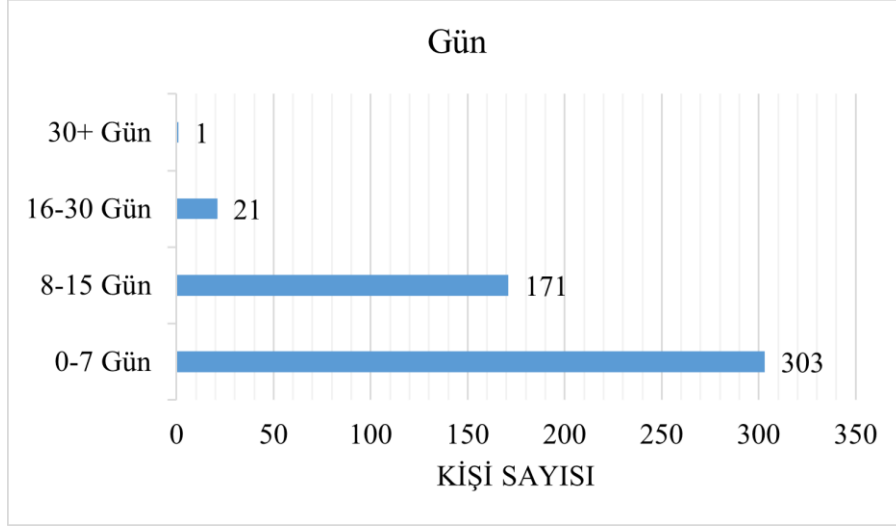
Şekil 4.4. Aşı tercihi dağılım grafiği.

Hastaların hastaneden çıktıktan sonraki aşı tercihleri Şekil 4.5.'te ki gibi olmuştur. 556 yatan hastadan 398'i hastaneden çıktıktan sonra aşı olmayı tercih etmiştir. Bu 398 kişinin hastaneye yatmadan önce hiç aşı kayıt bilgisi bulunmamaktadır. Hiç aşı olmayan 398 kişinin 213'ü hastalığı atlattıktan sonra iki Biontech aşısı olmayı tercih etmiştir.



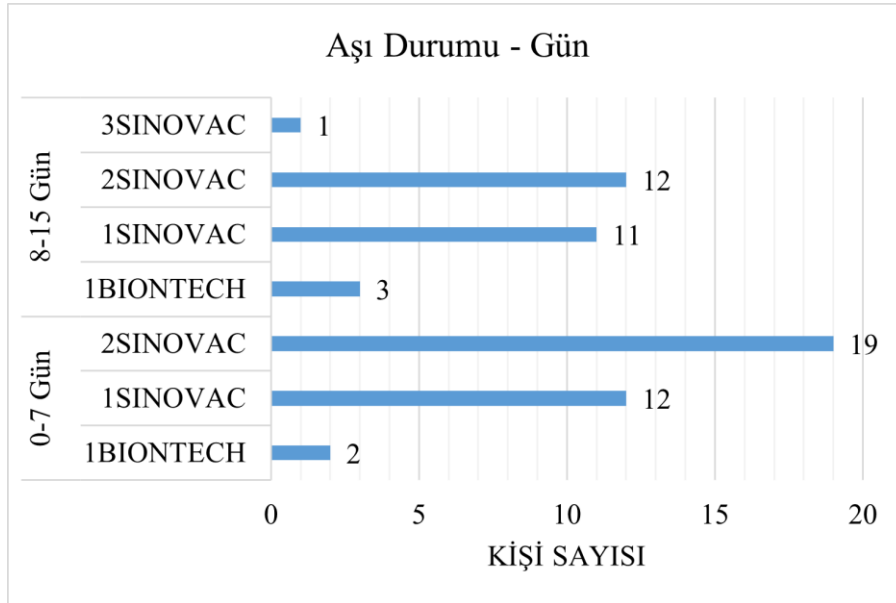
Şekil 4.5. Hastaneden çıktıktan sonraki aşı tercihleri.

Hastalanmadan önce aşı kaydı olmayan 496 hastanın hastanede yatış süresi dağılım grafiği Şekil 4.6.'da verilmiştir. 303 kişi ile en çok yığılma 0 – 7 gün aralığında olmuştur.



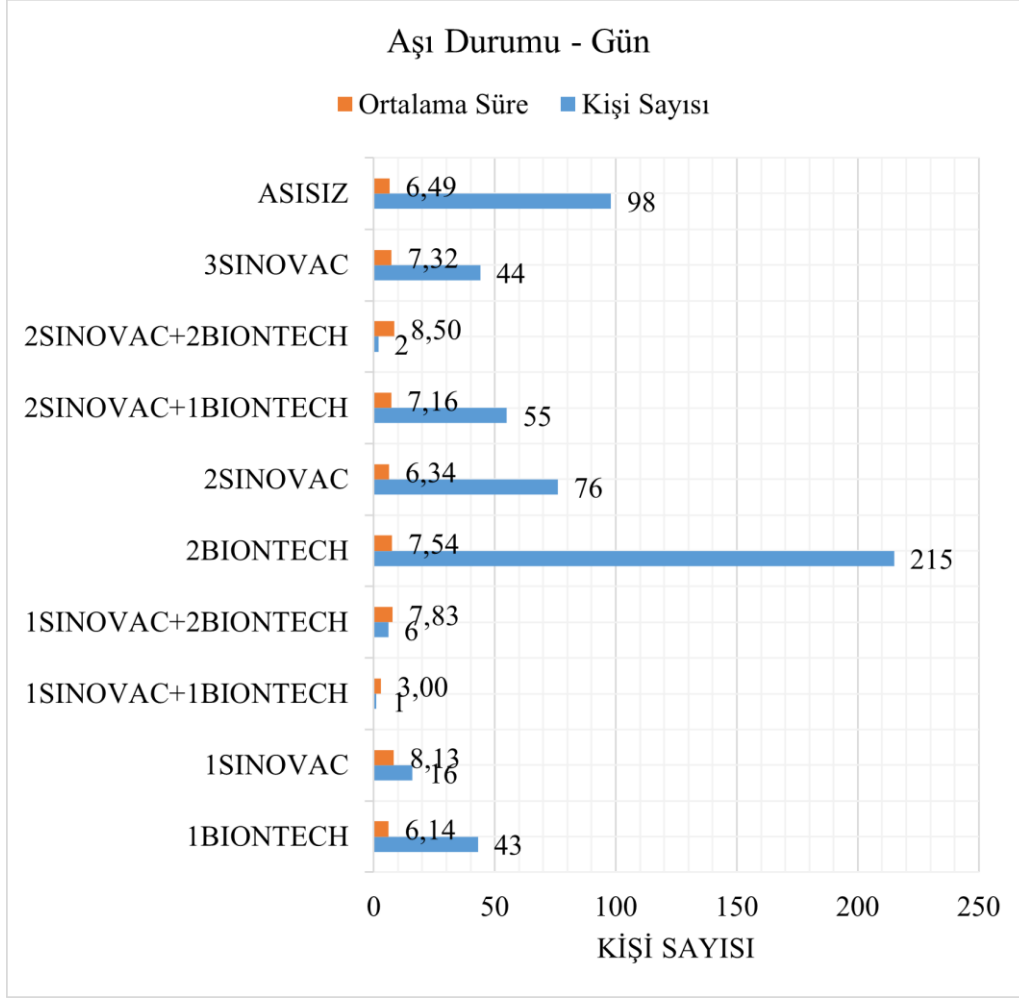
Şekil 4.6. Aşısı olmayanların hastanede yatış süresi.

Hastalanmadan önce aşısı olan 60 hastanın hastanede yatış süresi dağılım ve aşı türü grafiği Şekil 4.7.'de ki gibi olmuştur. 60 kişiden 33'ü 0 – 7 gün aralığında hastanede yatarken, 27'si 8 – 15 gün aralığında hastanede yatmıştır. Aşısı olup hastanede yatan 60 kişiden 19'unun hastalığa yakalandığında iki Sinovac aşısı vardır.



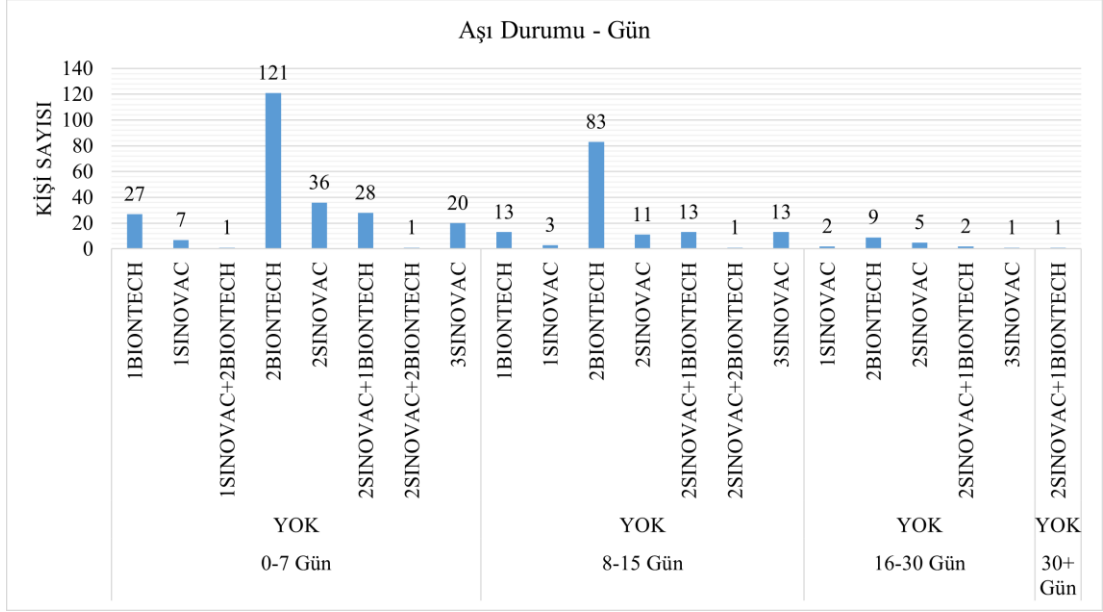
Şekil 4.7. Aşısı olanların hastanede yatış süresi.

Aşı türlerine göre ortalama hastanede yatış süreleri Şekil 4.8.'de verilmiştir. İki Biontech aşısı olmayı tercih eden 215 kişinin hastanede ortalama yatış süresi 7,54 gün olmuştur. Aşısı olmayan 98 kişinin ortalama yatış süresi ise 6,49 gündür.



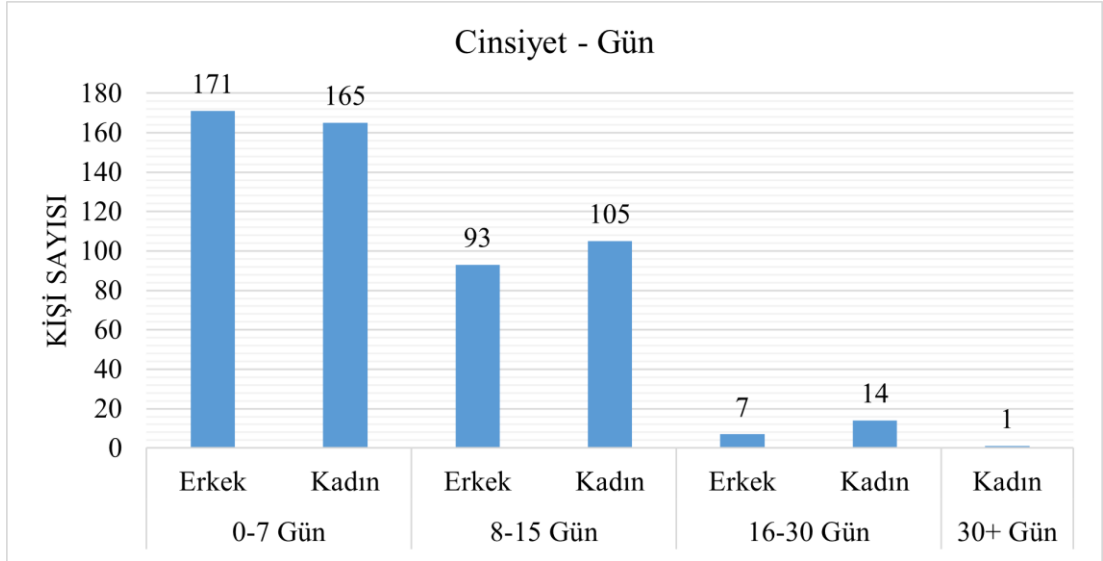
Şekil 4.8. Aşı türlerine göre ortalama hastanede yatış süresi.

Hastaneye yatmadan önce hiç aşısı olmayan 398 kişinin hastaneye yattığı süre ve aşı tercihi grafiği Şekil 4.9.'da verilmiştir. 0 – 7 gün aralığında hastanede yatan 121 kişi, sonrasında iki Biontech aşısı olmayı tercih etmiştir. 8 – 15 gün aralığında hastanede yatan 83 kişi, sonrasında iki Biontech aşısı olmayı tercih etmiştir.



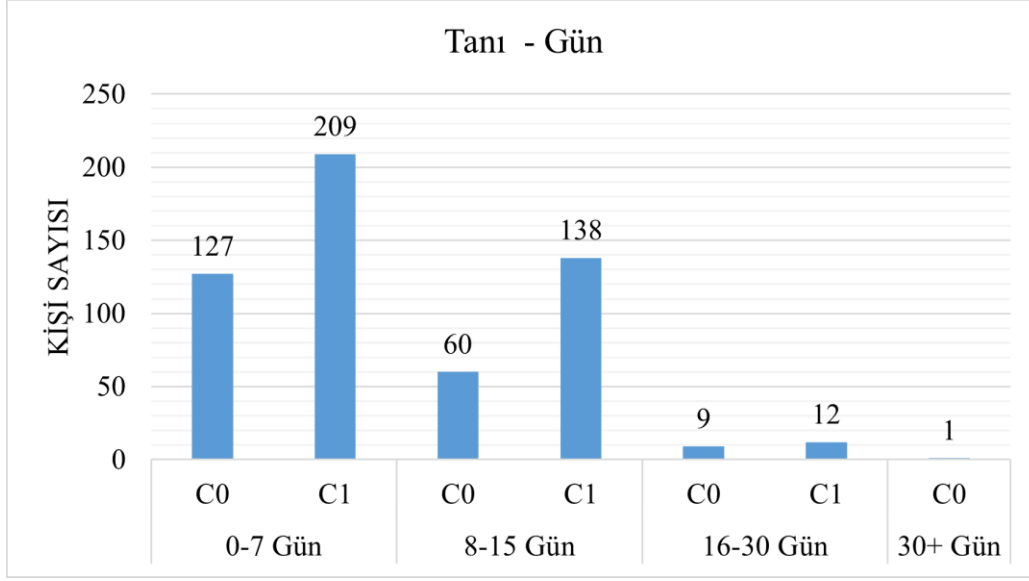
Şekil 4.9. Hastalanmadan önce aşısı olmayan hastaların yatış süresi ve aşı tercihi grafiği.

Hastanede yatış süresi ve cinsiyet dağılım grafiği Şekil 4.10.'da verilmiştir.



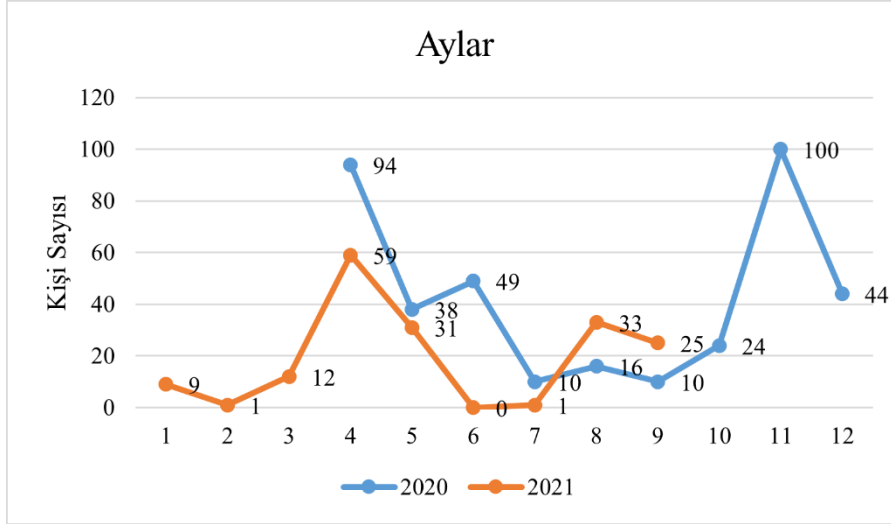
Şekil 4.10. Hastanede yatış süresi ve cinsiyet dağılım grafiği.

Hastanede yatış süresi ve tanı dağılım grafiği Şekil 4.11.'de verilmiştir. Hastanede yatış yapan 556 kişiden 336 kişinin hastanede yatış süresi 0 – 7 gün aralığındadır. Bunlardan 127 kişi sadece covid-19 tanısı ile hastaneye yatış yapmıştır. 209 kişi ise Covid-19 ile ek hastalığı olup hastaneye yatış yapmıştır.



Şekil 4.11. Hastanede yatış süresi ve tanı dağılım grafiği.

Hastanede yatan hastaların aylara göre dağılım grafiği Şekil 4.12.'de ki gibidir. 2020 yılında en yoğun hastaneye yatış nisan ve kasım aylarında olmuştur. 2021 yılında en yoğun yatış nisan ayında olmuştur. 2021 yılında yatışın en az olduğu ay haziran ayıdır. Bu ayda hastaneye hiç kimse yatış yapmamıştır.



Şekil 4.12. Aylara göre dağılım grafiği.

4.3. Apriori Algoritmasının Uygulanması

4.3.1. Birliktelik kuralları

Bu çalışmada Sakarya Geyve Devlet Hastanesi Covid-19 yatan hasta verileri kullanılmıştır. Covid-19 servisinde 2 Nisan 2020 ile 30 Eylül 2021 tarihleri arasında

yatan hastaların verileri kullanılmıştır. Geyve Devlet hastanesinde pandemi döneminde ilk hastaneye yatış 2 Nisan 2020 tarihinde olmuştur. Alınan bu veriler, veri temizleme ve hazırlık aşamalarından geçirildikten sonra analiz yapmaya uygun hale getirilmiştir. Temizlenen ve çalışmaya uygun hale getirilen verilere Rapidminer uygulaması kullanılarak apriori algoritması uygulanmıştır. Veri kümesi içinde sıklıkla birlikte geçen kümeleri bulmak için FP-Growth operatörü kullanılmıştır. FP-Growth algoritmasından çıkan sonuçlar Create Association Rules operatörüne verilerek birliktelik kuralları elde edilmiştir. Güven eşik değeri %80 olarak seçilmiştir. Çıkan sonuçlarda lift değeri en yüksek olan kurallar dikkate alınmıştır.

Cinsiyeti erkek olan hastaların verileri ile algoritma çalıştırıldığında ulaşılan kurallar aşağıdaki gibi olmuştur.

- Hastanede yattığı gün aralığı 0-7 gün olan ve aşı durumu 2BIONTECH olan erkek hastaların %100'ünün hastalanmadan önce aşısı yoktur ve 2BIONTECH aşısını da hastaneden çıktıktan sonra olmuşlardır. Bu kuralın lift değeri 2.765'tir.
- Hastanede yattığı gün aralığı 0-7 gün olan ve aşı durumu 2BIONTECH olan hastaların %100'ü erkektir, hastalanmadan önce aşısı yoktur ve 2BIONTECH aşısını da hastaneden çıktıktan sonra olmuşlardır. Bu kuralın lift değeri 2.765'tir.
- Aşı durumu 2BIONTECH olan erkek hastaların %100'ünün hastalanmadan önce aşısı yoktur ve 2BIONTECH aşısını da hastaneden taburcu olarak çıktıktan sonra olmuştur. Bu kuralın lift değeri 2.765'tir.
- Hastaneden taburcu olarak çıkan ve hastanede yattığı gün aralığı 0 – 7 gün olan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden erkek hastaların %100'ünün hastalanmadan önce aşısı yoktur ve aşı durumu 2BIONTECH'tir. Bu kuralın lift değeri 2.765'tir.
- Aşı durumu 2BIONTECH olan erkek hastaların %100'ü hastanede taburcu olarak çıkmıştır, hastalanmadan önce aşısı yoktur ve aşılarını hastalandıktan sonra olmayı tercih etmişlerdir. Bu kuralın lift değeri 2.765'tir.

Cinsiyeti kadın olan hastaların verileri ile algoritma çalıştırıldığında ulaşılan kurallar aşağıdaki gibi olmuştur.

- Aşı durumu 2BIONTECH olan kadın hastaların %92'sinin hastalanmadan önce aşısı yoktur ve 2BIONTECH aşısını da hastaneden taburcu olarak çıktıktan sonra olmuştur. Bu kuralın lift değeri 2.457'dir.
- Aşı durumu 2BIONTECH olan kadın hastaların %99'unun hastalanmadan önce aşısı yoktur ve 2BIONTECH aşısını da hastalandıktan sonra olmuşlardır. Bu kuralın lift değeri 2.457'dir.
- Hastalanmadan önce aşısı olmayan ve 2BIONTECH aşısını hastalandıktan sonra olan kadın hastaların %100'ünün aşı durumu 2BIONTECH'tir. Bu kuralın lift değeri 2.457'dir.
- Hastaneden taburcu olarak çıkan, hastalanmadan önce aşısı olmayan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %100'ünün aşı durumu 2BIONTECH'tir. Bu kuralın lift değeri 2.457'dir.
- Hastalanmadan önce aşısı olmayan, hastanede yattığı gün aralığı 0-7 gün arasında olan kadın hastaların %100'ünün aşı durumu 2BIONTECH'tir. Bu kuralın lift değeri 2.457'dir.
- Hastalanmadan önce aşısı olmayan ve 2BIONTECH aşısını hastalandıktan sonra olmayı tercih eden kadın hastaların %93'ü hastaneden taburcu olarak çıkmıştır ve aşı durumu 2BIONTECH'tir. Bu kuralın lift değeri 2.455'tir.
- Hastaneden taburcu olarak çıkan ve aşı durumu 2BIONTECH olan kadın hastaların %99' u aşılarının hastaneden çıktıktan sonra olmuştur. Bu kuralın lift değeri 2.455'tir.
- Hastanede yattığı gün aralığı 0-7 gün olan ve aşı durumu 2BIONTECH olan kadın hastaların %98'inin hastalanmadan öce aşısı yoktur ve aşılarını hastalandıktan sonra olmayı tercih etmişlerdir. Bu kuralın lift değeri 2.438'dir.
- Hastanede yattığı gün aralığı 0 – 7 gün olan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %96'sının hastalanmadan öce aşısı yoktur ve aşı durumu 2BIONTECH'tir. Bu kuralın lift değeri 2.398'dir.
- Hastalandıktan sonra 2BIONTECH aşısı olan kadın hastaların %89'u hastaneden taburcu olarak çıkmış ve hastalanmadan önce aşıları yoktur ve aşı durumu 2BIONTECH'tir. Bu kuralın lift değeri 2.395'tir.

Cinsiyet = Erkek ve Tanı = C1 olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Covid-19 + Ek hastalık teşhisi ile hastaneye yatan ve hastalandıktan sonra 2BIONTECH aşısı olan erkek hastaların %100'ünün hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.368'tir.
- Covid-19 + Ek hastalık teşhisi ile hastaneye yatan, hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden ve hastaneden taburcu olarak çıkan erkek hastaların %100'ünün hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.368'tir.
- Covid-19 + Ek hastalık teşhisi ile hastaneye yatan, hastanede yattığı gün aralığı 0 – 7 gün olan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden erkek hastaların %100'ünün hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.368'tir.

Cinsiyet = Kadın ve Tanı = C1 olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Sadece Covid-19 teşhisi ile hastaneye yatan, hastanede yattığı gün aralığı 0 – 7 gün olan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %97'sinin hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.402'dir.
- Sadece Covid-19 teşhisi ile hastaneye yatan, hastanede yattığı gün aralığı 0 – 7 gün olan ve hastaneden taburcu olarak çıkan ve hastaneden çıktıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %97'sinin hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.400'dür.
- Sadece Covid-19 teşhisi ile hastaneye yatan, hastanede yattığı gün aralığı 0 – 7 gün olan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %92'si hastaneden taburcu olarak çıkmıştır ve hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.396'dır.
- Sadece Covid-19 teşhisi ile hastaneye yatan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %91'i hastaneden taburcu olarak çıkmış ve hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.387'dir.
- Sadece Covid-19 teşhisi ile hastaneye yatan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %95'inin hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.379'dur.

- Sadece Covid-19 teşhisi ile hastaneye yatan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %95'inin hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.376'dır.
- Sadece Covid-19 teşhisi ile hastaneye yatan, hastanede yattığı gün aralığı 8-15 gün olan kadın hastaların %98'i hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.051'dir.
- Sadece Covid-19 teşhisi ile hastaneye yatan, hastanede yattığı gün aralığı 8-15 gün olan ve hastalanmadan önce aşısı olmayan kadın hastaların %97'si hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.043'tür.
- Sadece Covid-19 teşhisi ile hastaneye yatan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %95'i hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.022'dir.
- Sadece Covid-19 teşhisi ile hastaneye yatan, hastalanmadan önce aşısı olmayan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %95'i hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.020'dir.

Cinsiyet = Erkek ve Tanı = C0 olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Sadece Covid-19 teşhisi ile hastaneye yatan, hastaneden taburcu olarak çıkan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden hastaların %96'sı erkektir ve hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.305'tir.
- Sadece Covid-19 teşhisi ile hastaneye yatan, hastanede yattığı gün aralığı 0-7 gün olan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden hastaların %95'i erkektir ve hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.284'tür.
- Hastanede yattığı gün aralığı 0-7 gün olan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden hastaların %95'i sadece Covid-19 teşhisi almış erkek hastalardır ve hastalanmadan önce aşıları yoktur. Bu kuralın lift değeri 1.284'tür.
- Sadece Covid-19 teşhisi ile hastaneye yatan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden hastaların %94'ü erkektir ve hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.271'dir.

- Sadece Covid-19 teşhisi ile hastaneye yatan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden hastaların %85'i erkektir, hastaneden taburcu olarak çıkmıştır ve hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.251'dir.

Cinsiyet = Kadın ve Tanı = C0 olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Sadece Covid-19 teşhisi ile hastaneye yatan, hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden ve hastanede 8-15 gün aralığında yatan kadın hastaların %100'ünün hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.421'dir.
- Hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden ve hastanede 8-15 gün aralığında yatan kadın hastaların %100'ü sadece Covid-19 tanısı almıştır ve hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.421'dir.
- Sadece Covid-19 teşhisi ile hastaneye yatan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %87'si hastaneden taburcu olarak çıkmıştır ve hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.406'dır.
- Sadece Covid-19 teşhisi ile hastaneye yatan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %97'sinin hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.391'dir.
- Sadece Covid-19 teşhisi ile hastaneye yatan, hastaneden taburcu olarak çıkan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %97'sinin hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.387'dir.

Aşı durumu = 2BIONTECH olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Aşı durumu 2BIONTECH olan ve yaş aralığı 30-44 yaş olan hastaların %98'inin hastalanmadan önce aşısı yoktur ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.068'dir.
- Aşı durumu 2BIONTECH olan ve yaş aralığı 30-44 yaş olan hastaların %98'i hastalandıktan sonra 2BIONTECH aşısı olmayı tercih etmiştir ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.068'dir.

- Yaş aralığı 30-44 yaş olan hastaların %98'inin aşı durumu 2BIONTECH'tir, hastalanmadan önce aşıları yoktur ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.068'dir.
- Yaş aralığı 30-44 yaş olan hastaların %98'inin aşı durumu 2BIONTECH'tir, hastalandıktan sonra 2BIONTECH aşısı olmayı tercih etmiştir ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.068'dir.
- Hastalanmadan önce aşısı olmayan ve yaş aralığı 30-44 yaş olan hastaların %98'i hastalandıktan sonra 2BIONTECH olmayı tercih etmiş ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.068'dir.
- Yaş aralığı 30-44 yaş olan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden hastaların %98'inin hastalanmadan önce aşısı yoktur ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.068'dir.

Aşı durumu = 2BIONTECH ve Tanı = C0 olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Aşı durumu 2BIONTECH olan ve sadece Covid-19 tanısı ile hastaneye yatan ve yaş aralığı 30-44 yaş olan hastaların %100'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.113'tür.
- Aşı durumu 2BIONTECH olan ve sadece Covid-19 tanısı ile hastaneye kasım ayında yatan ve hastanede yattığı gün aralığı 8-15 gün olan hastaların %100'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.113'tür.

Aşı durumu = 2SINOVAC olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Aşı durumu 2SINOVAC olan ve yaş aralığı 30-44 yaş olan hastaların %85'inin hastanede yattığı gün aralığı 0-7 gündür. Bu kuralın lift değeri 1.292'dir.
- Aşı durumu 2SINOVAC olan ve yaş aralığı 45-59 yaş olan hastaların %100'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.134'tür.
- Aşı durumu 2SINOVAC olan ve hastanede yattığı gün aralığı 8-15 gün olan hastaların %100'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.134'tür.
- Yaş aralığı 45-59 yaş olan hastaların %100'ünün aşı durumu 2SINOVAC'tır ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.134'tür.

- Aşı durumu 2SINOVAC olan ve hastaneye yattığı ay nisan olan hastaların %99'u hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.083'tür.

Aşı durumu = 2SINOVAC + 1BIONTECH olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Aşı durumu 2SINOVAC + 1BIONTECH olan ve hastanede yattığı gün aralığı 8-15 gün olan hastaların %93'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.250'dir.
- Aşı durumu 2SINOVAC + 1BIONTECH olan ve yaş aralığı 45-59 yaş olan hastaların %93'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.250'dir.
- Aşı durumu 2SINOVAC + 1BIONTECH olan, yaş aralığı 45-59 yaş olan ve hastanede yattığı gün aralığı 8-15 gün olan hastaların %100'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.250'dir.

Aşı durumu = ASISIZ olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Aşı durumu ASISIZ olan ve yaş aralığı 30-44 yaş olan erkek hastaların %100'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.045'tir.
- Aşı durumu ASISIZ olan ve hastanede yattığı gün aralığı 8-15 gün olan erkek hastaların %100'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.045'tir.
- Aşı durumu ASISIZ olan ve hastanede yattığı ay ağustos olan erkek hastaların %100'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.045'tir.
- Aşı durumu ASISIZ olan ve hastanede yattığı gün aralığı 8-15 gün olan kadın hastaların %100'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.182'dir.

Aşı durumu = 1SINOVAC + 2BIONTECH olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Aşı durumu 1SINOVAC + 2BIONTECH olan ve hastaneye yattığı ay mayıs olan erkek hastaların %100'ünün hastanede yattığı gün aralığı 0-7 gün olmuştur. Bu kuralın lift değeri 2.0'dir.

- Aşı durumu 1SINOVAC + 2BIONTECH olan ve hastaneye yattığı ay mayıs olan erkek hastaların %100'ünün hastanede yaş aralığı 60-74 yaş olmuştur. Bu kuralın lift değeri 2.0'dır.
- Aşı durumu 1SINOVAC + 2BIONTECH olan ve hastalanmadan önce 1SINOVAC olan kadın hastaların %100'ü hastaneden çıktıktan sonra 2BIONTECH olmayı tercih etmiştir. Bu kuralın lift değeri 1.333'tür.
- Aşı durumu 1SINOVAC + 2BIONTECH olan ve hastaneden çıktıktan sonra 2BIONTECH olan kadın hastaların %100'ü hastalanmadan önce 1SINOVAC olmuştur.

Hastalanmadan önceki aşı durumu = Yok ve Tanı = C0 olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Hastaneden çıktıktan 2BIONTECH aşısı olmayı tercih eden ve hastanede yattığı gün aralığı 8-15 gün olan hastaların %97'si hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.077'dir.
- Hastaneden çıktıktan 2BIONTECH aşısı olmayı tercih eden ve hastanede yattığı gün aralığı 8-15 gün olan hastaların %97'sinin hastalanmadan önce aşısı yoktur ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.077'dir.
- Hastaneden çıktıktan 2BIONTECH aşısı olmayı tercih eden ve hastanede yattığı gün aralığı 8-15 gün olan ve sadece Covid-19 tanısı almış hastaların %97'sinin hastalanmadan önce aşısı yoktur ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.077'dir.

Aşı durumu = 2SINOVAC, Tanı = C0 ve Cinsiyet = Kadın olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Hastanede yattığı gün aralığı 8 – 15 gün olan hastaların %100'ünün yaş aralığı 45-59 yaş aralığındadır. Bu kuralın lift değeri 2.750'dir
- Hastanede yattığı gün aralığı 8 – 15 gün olan hastaların %100'ünün yaş aralığı 45-59 yaş aralığındadır ve aşı durumu 2SINOVAC'tır. Bu kuralın lift değeri 2.750'dir.
- Aşı durumu 2SINOVAC olan ve hastanede yattığı gün aralığı 8 – 15 gün olan hastaların %100'ünün yaş aralığı 45-59 yaş aralığındadır. Bu kuralın lift değeri 2.750'dir.

- Hastanede yattığı gün aralığı 8-15 gün olan hastaların %100'ü sadece Covid-19 tanısı almıştır ve 45-59 yaş aralığındadır. Bu kuralın lift değeri 2.750'dir.
- Hastaneden taburcu olarak çıkan ve hastanede 8-15 gün aralığında yatan hastaların %100'ü 45-59 yaş aralığındadır. Bu kuralın lift değeri 2.750'dir.
- Sadece Covid-19 tanısı almış ve hastanede yattığı gün aralığı 8-15 gün hastaların %100'ünün aşı durumu 2SINOVAC'tır ve yaş aralığı 45-59'dur. Bu kuralın lift değeri 2.750'dir.

Aşı durumu = 2SINOVAC, Tanı = C0 ve Cinsiyet = Erkek olarak algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Yaş aralığı 45 – 59 yaş aralığında olan hastaların %100'ü hastanede 0-7 gün aralığında yatmış ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.727'dir.
- Yaş aralığı 45-59 yaş aralığında olan hastaların %100'ünün aşı durumu 2SINOVAC'tır, hastanede 0-7 gün aralığında yatmış ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.727'dir.
- Aşı durumu 2SINOVAC olan ve 45-59 yaş aralığında olan hastaların %100'ü hastanede 0-7 gün aralığında yatmış ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.727'dir.
- Yaş aralığı 45-59 yaş aralığında olan hastaların %100'ü erkektir, hastanede 0-7 gün aralığında yatmış ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.727'dir.
- Yaş aralığı 45-59 yaş aralığında olan erkek hastaların %100'ü hastanede 0-7 gün aralığında yatmış ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.727'dir.

Hastalanmadan önce aşısı olmayanların verileri ile algoritma çalıştırıldığında elde edilen kurallar aşağıdaki gibi olmuştur.

- Hastalandıktan sonra 2BIONTECH aşısı olan ve hastanede yattığı gün aralığı 8-15 gün olan hastaların %97'si hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.050'dir.
- Hastalandıktan sonra 2BIONTECH aşısı olan ve hastanede yattığı gün aralığı 8-15 gün olan hastaların %97'sinin hastalanmadan önce aşısı yoktur ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.050'dir.

- Hastalanmadan önce aşısı olmayan, hastalandıktan sonra 2BIONTECH aşısı olan ve hastanede yattığı gün aralığı 8-15 gün olan hastaların %97'si hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.050'dir.
- Hastanede yattığı gün aralığı 8-15 gün olan hastaların %97'si hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.044'tür.
- Hastanede yattığı gün aralığı 8-15 gün olan hastaların %97'sinin hastalanmadan önce aşısı yoktur ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.044'tür.
- Hastalanmadan önce aşısı olmayan ve hastanede yattığı gün aralığı 8-15 gün olan hastaların %97'si hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.044'tür.
- Yaş aralığı 30-44 yaş olan hastaların %96'si hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.033'tür.
- Yaş aralığı 30-44 yaş olan hastaların %96'sinin hastalanmadan önce aşısı yoktur ve hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.033'tür.
- Hastalanmadan önce aşısı olmayan ve yaş aralığı 30-44 yaş olan hastaların %96'si hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.033'tür.

5. SONUÇ VE ÖNERİLER

Dünyada ve ülkemizde her geçen gün artan veri hacmi ile birlikte veri madenciliği konusuna olan önem ve ilgi gitgide artmaktadır. Veri madenciliği alanı birçok sektöre katkı sağladığı gibi sağlık alanına da katkı sağlamaktadır. Bu çalışmada da yatan hasta verileri veri madenciliği çerçevesinde incelenmiştir.

Ülkemizde Covid-19 vaka sayılarının yüksek olduğu dönemlerde hastaneye yatış oranları artabilmektedir. Hasta profillerinin incelenip hastaların hastanede ne kadar kalacağını öngörmek yatan hasta servislerindeki planlama açısından önemli olabilmektedir. Bu çalışmadan Sakarya Geyve Devlet Hastanesi Covid-19 yatan hasta servisinden alınan 556 hastanın verileri birliktelik kural analizi yöntemlerinden olan apriori algoritması ile incelenmiş ve birliktelik kuralları çıkarılmıştır. Veriler parçalı olarak apriori algoritması incelenmiş lift değeri yüksek olan kurallar çalışmada listelenmiştir. Listelenen kurallardan bazıları şunlardır.

Hastanede yattığı gün aralığı 0-7 gün olan ve aşı durumu 2BIONTECH olan erkek hastaların %100'ünün hastalanmadan önce aşısı yoktur ve 2BIONTECH aşısını da hastaneden çıktıktan sonra olmuşlardır. Bu kuralın lift değeri 2.765'tir.

Aşı durumu 2BIONTECH olan kadın hastaların %92'sinin hastalanmadan önce aşısı yoktur ve 2BIONTECH aşısını da hastaneden taburcu olarak çıktıktan sonra olmuştur. Bu kuralın lift değeri 2.457'dir.

Sadece Covid-19 teşhisi ile hastaneye yatan, hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden ve hastanede 8 – 15 gün aralığında yatan kadın hastaların %100'ünün hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.421'dir.

Sadece Covid-19 teşhisi ile hastaneye yatan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %87'si hastaneden taburcu olarak çıkmıştır ve hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.406'dır.

Sadece Covid-19 teşhisi ile hastaneye yatan, hastanede yattığı gün aralığı 0-7 gün olan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden kadın hastaların %97'sinin hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.402'dir.

Covid-19 + Ek hastalık teşhisi ile hastaneye yatan ve hastalandıktan sonra 2BIONTECH aşısı olan erkek hastaların %100'ünün hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.368'tir.

Sadece Covid-19 teşhisi ile hastaneye yatan, hastaneden taburcu olarak çıkan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden hastaların %96'sı erkektir ve hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.305'tir.

Aşı durumu 2SINOVAC olan ve yaş aralığı 30-44 yaş olan hastaların %85'inin hastanede yattığı gün aralığı 0 -7 gündür. Bu kuralın lift değeri 1.292'dir.

Sadece Covid-19 teşhisi ile hastaneye yatan, hastanede yattığı gün aralığı 0-7 gün olan ve hastalandıktan sonra 2BIONTECH aşısı olmayı tercih eden hastaların %95'i erkektir ve hastalanmadan önce aşısı yoktur. Bu kuralın lift değeri 1.284'tür.

Aşı durumu 2SINOVAC + 1BIONTECH olan ve hastanede yattığı gün aralığı 8-15 gün olan hastaların %93'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.250'dir.

Aşı durumu 2BIONTECH olan ve sadece Covid-19 tanısı ile hastaneye yatan ve yaş aralığı 30-44 yaş olan hastaların %100'ü hastaneden taburcu olarak çıkmıştır. Bu kuralın lift değeri 1.113'tür.

Yapılan analizler sonucunda yapılan çıkarımlar şunlardır.

Hastaneye yatan 556 hastadan 506'sı hastaneden sağlıklı bir şekilde taburcu edilmiştir.

Hastalığa yakalandığında aşısı olmayan 398 hastadan 198'i hastaneden taburcu olduktan sonra iki Biontech aşısı olmayı tercih etmiştir.

Hastaneye yatan 556 hastanın 336'sı, 0-7 gün hastanede yatmıştır. Bunlardan 241 kişinin hastalanmadan önce aşı kayıt bilgisi yoktur ve sonrasında iki Biontech aşısı olmayı tercih etmişlerdir.

Hastanede yatan hastalardan sadece 1 kişi 30 günden fazla hastane yatmıştır. Bu kişinin cinsiyeti kadındır, 60-74 yaş aralığındadır ve hastaneden çıktıktan sonra iki Sinovac ve bir Biontech aşısı olmayı tercih etmiştir.

Hastanede yatış yapan hastalar 164 kişi ile en çok 45-74 yaş aralığında kümelenmiştir. Bu 164 kişiden 126'sının hastalanmadan önce aşı bilgisi yoktur.

Hastanede yatan 556 hastadan 60'ının hastalanmadan önce aşı kayıt bilgisi vardır ve 33 kişi iki Sinovac aşısı olmayı tercih etmiştir ve bu 33 kişinin 19'u 0-7 gün aralığında hastanede yatmıştır

Yapılan çalışma sonucunda elde edilen bulgular Covid-19 salgın döneminde hastaneye yatan kişilerde bir arada bulunan özellikleri ortaya koymuştur. Bu çalışmanın sağlık alanında yapılacak olan diğer çalışmalara ve doktorlara fayda sağlayacağı düşünülmektedir.

KAYNAKLAR

- [1] T.C. Sağlık Bakanlığı COVID-19 Bilgilendirme Platformu (2022, 23 Ağustos). COVID-19 Nedir?. <https://covid19.saglik.gov.tr/TR66300/covid-19-nedir-.html>
- [2] COVID-19 Aşısı Bilgilendirme Platformu (2022, 23 Ağustos). <https://covid19asi.saglik.gov.tr>
- [3] Özgür Ansiklopedi Vikipedi (2022, 23 Ağustos).. Türkiye'de COVID-19 aşılması. https://tr.wikipedia.org/wiki/T%C3%BCrkiye%27de_COVID19_a%C5%9F%C4%B1lamas%C4%B1
- [4] Yağar, D. S. (2021), Sağlık Hizmetlerinde Veri Madenciliğinin Kullanımı: Türkiye’de Lisansüstü Tezlerinin İncelenmesi. Sağlık Hizmetlerinde Kuram ve Uygulama Dergisi, 1(2)
- [5] Tomar, D., Agarwal, S. (2013). A survey on data mining approaches for healthcare. *International Journal of Bio-Science and Bio-Technology*, 5(5), 241-266. <http://dx.doi.org/10.14257/ijbsbt.2013.5.5.25>
- [6] Silahtaroglu, G., Şahinbaş, K. (2022). Segmentation of Covid-19 Positive Patients Regarding Symptoms and Complaints. *Eskisehir Technical University Journal of Science and Technology A- Applied Sciences and Engineering*. 23(1), 37-47. <https://doi.org/10.18038/estubtda.877029>
- [7] Çelik, A. (2020). Using Apriori Data Mining Method in COVID-19 Diagnosis. *Journal of Engineering Technology and Applied Sciences*, 5(3), 121-131. <https://doi.org/10.30931/jetas.857665>
- [8] Pinter, G., Felde, I., Mosavi, A., Ghamisi, P., & Gloaguen, R. (2020). Covid-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach. *Mathematics*, 8(6). <https://doi.org/10.3390/math8060890>
- [9] Kalezhi, J., Chibuluma, M., Chembe, C., Chama, V., Lungo, F., & Kunda, D. (2022). Modelling Covid-19 Infections in Zambia Using Data Mining Techniques. *Results in Engineering*, 13. <https://doi.org/10.1016/j.rineng.2022.100363>
- [10] Shahin, M., Inoubli, W., Shah, S. A., Yahia, S. B., & Draheim, D. (2021) Distributed Scalable Association Rule Mining Over Covid-19 Data. *International Conference on Future Data and Security Engineering*, 39(52). DOI: 10.1007/978-3-030-91387-8_3
- [11] Bozkurt, A. (2021). Veri Madenciliği Yöntemleri ile Tam Kan Sayımı Sonuçlarından Covid-19 Test Sonuçlarının Tahmini. [Yüksek Lisans Tezi]. Necmettin Erbakan Üniversitesi
- [12] Ahoz, F., Golabpour, A. (2021). Predicting the Incidence of Covid-19 Using Data Mining. *BMC Public Health*, 21(1). DOI:10.1186/s12889-021-11058-3
- [13] Tandan M, Acharya Y, Pokharel S, Timilsina M. “Discovering Syntom Patterns of Covid-19 Patients Using Association Rule Mining”. *Computer in Biology and Medicine*, 131. <https://doi.org/10.1016/j.combiomed.2021.104249>

- [14] Gök, E. C. (2021). Kan Değerleri ile Covid-19 Enfekte Düzeyinin Rassal Orman Sınıflandırıcı ile Tahmin Edilmesi. [Yüksek Lisans Tezi]. Süleyman Demirel Üniversitesi
- [15] Oğuz, Ç. (2021). Covid-19 Olası Vakalarının Derin Öğrenme Teknikleri Kullanarak Tespiti. [Yüksek Lisans Tezi]. Atatürk Üniversitesi
- [16] Özkan, Y. (2016). Veri Madenciliği Yöntemleri (Üçüncü baskı). Papatya Yayıncılık Eğitim.
- [17] Silahtaroglu, G. (2016). Veri Madenciliği Kavram ve Algoritmaları (Üçüncü baskı). Papatya Yayıncılık Eğitim.
- [18] Şeker, S. E. (2018). CRISP-DM: Endüstriler Arası Standard İşleme – Veri Madenciliği için (Cross Industry Standard Processing – Data Mining). YBS Ansiklopedi. <http://ybsansiklopedi.com/wp-content/uploads/2018/08/crisp-dm.pdf>
- [19] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reeinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0- Step-by-step data mining guide.SPSS
- [20] Schröer, C., Kruse, F., & Marx, G. J. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181. <https://doi.org/10.1016/j.procs.2021.01.199>
- [21] Çelik, U., Akçetin, E. Ve Gök, M. (2017). Rapidminer ile Uygulamalı Veri Madenciliği (Birinci baskı). Pusula 20 Teknoloji ve Yayıncılık A.Ş.
- [22] Agrawal, R., Imielinski, T. ve Swami, A., (1993). Mining Association Rules Between Sets of Items in Large Databases, In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 22 (207-216). <https://doi.org/10.1145/170036.170072>
- [23] Houtsma, M., Swami, A., (1995). Set-Oriented Mining for Association Rules in Relational Databases, *Proceedings of the 11th IEEE International Conference on Data Engineering*, 25-34. DOI: 10.1109/ICDE.1995.380413
- [24] Agrawal, R., Srikant, R. (1994). Fast Algorithms for Mining Association Rules in Large Databases. *Proceedings of the 20th International Conference on Very Large Databases (VLDB)*, 487-499

ÖZGEÇMİŞ

Ad-Soyad : Seda UÇAR

ÖĞRENİM DURUMU:

- **Lisans** : 2013, Sakarya Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2014-2022 yılları arasında Arvasis Bilişim ve Danışmanlık firmasında yazılım uzmanı olarak çalıştı.
- 2022 yılında Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi Bilgisayar Mühendisliği araştırma görevlisi kadrosuna atandı.

TEZDEN TÜRETİLEN ESERLER:

- Uçar, S., Yurtay, N. (2022, 7-11, Eylül). Veri Madenciliği Yöntemleriyle Covid-19 Yatan Hasta Verilerinin İncelenmesi. *International Conference on Data Science and Applications*, Mugla, Turkey.