



Covid-19 Verileri için Bayes Ağları ile Makine Öğrenmesi

Hüseyin Can Yılmaz¹, Serpil Aktaş Altunay^{2*}

¹Gelir İdaresi Başkanlığı, Ankara, Türkiye

²Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara, Türkiye

Makale Tarihiçesi

Gönderim: 16.08.2022

Kabul: 13.10.2022

Yayın: 05.03.2023

Araştırma Makalesi

Öz – Covid-19 pandemisi, 17 Kasım 2019 tarihinde Çin'in Wuhan Eyaleti'nde ilk defa görülmüştür. Küresel pandemi ilk başta Wuhan'daki deniz mahsülleri ve hayvan satışı yapılan yerlerde görülmüştür. Sonra insanlar arasında da yayılımı devam ettirerek ilk olarak Wuhan ve Çin'in diğer eyaletindeki bölgelere ve dünya üzerinde diğer ülkelere de yayılmıştır. 14 Ağustos 2022 tarihi itibarıyla dünyada 590.624.000 vaka meydana gelmiştir ve 6.431.291 hasta ölmüştür. Ülkemizde ve dünya genelinde Covid-19 pandemisinin etkilerini gösteren birçok araştırma ve analiz çalışmaları yapılmıştır. Bu çalışmada dünya genelinde 104 ülkeden oluşan 215.968 adet dünya çapında meydana gelen vaka analiz edilmiştir ve Bayes Ağları (Bayesian Networks) ile makine öğrenimi tekniği kullanılarak hastalar sınıflandırılmaya çalışılmış ve dokuz adet değişkenle Covid-19 virüsüne yakalanan hastaların hayatta kalıp kalmayacağını araştırılmıştır. Böylelikle hangi hastaya öncelik verip tedavi edilmesi gerektiği veya gözlem altında tutulması gerektiği belirlenecektir. Sonuç olarak bu çalışmayla dünya genelindeki Covid-19 pandemisinden kaynaklı ölüm oranlarının düşürülmesi hedeflenmektedir.

Anahtar Kelimeler – Bayes ağları, Covid-19, makine öğrenmesi, pandemi

Machine Learning with Bayesian Networks for Covid-19 Data

¹Revenue Administration, Ankara, Turkey

²Hacettepe University, Faculty of Science, Department of Statistics, Ankara, Turkey

Article History

Received: 16.08.2022

Accepted: 13.10.2022

Published: 05.03.2023

Research Article

Abstract – The Covid-19 pandemic emerged on November 17, 2019, in Wuhan Province of China. The outbreak was initially detected in those found in the seafood and animal market in this region. Later, it spread from person to person and spread to other cities in Hubei province, especially in Wuhan, other provinces of China, and other world countries. Until Aug 14, 2022, 590.624.000 cases occurred globally and 6.266.278 patients died from Covid-19. Many research and analysis studies have been conducted in our country and around the world showing the effects of the Covid-19 pandemic. In this study, 215,968 worldwide cases from 104 countries around the world were analysed and the patients were tried to be classified using Bayesian Networks and machine learning techniques. It was investigated whether the patients who caught the Covid-19 virus would survive using nine variables. In this way, it will be determined which patient should be given priority and treated or kept under observation. Thus, this study aims to reduce the death rates due to the Covid-19 pandemic worldwide.

Keywords – Bayesian networks, Covid-19, machine learning, pandemic

¹ canyilmaz.hun@gmail.com

² spxl@hacettepe.edu.tr

*Sorumlu Yazar

1. Giriş

İlk defa 15 Aralık 2019'da Çin'in Hubei bölgesindeki Wuhan şehrinde, canlı hayvan ticareti olarak adlandırılan Huanan Deniz Ürünleri ve Canlı Hayvan Toptan Satış Pazar'ında ortaya çıktığı düşünülen bir hastalık kısa sürede salgına dönüşmüştür. Bu pazarda vahşi hayvanların satıldığı bilinmektedir. Virüse rastlandıktan kısa süre içinde bilim insanları ve epidemiyolojistlerden oluşan uzman bir ekip; bulaşıcı patojenlerle oluşan (Severe Acute Respiratory Syndrome) SARS-CoV; 2002, (Middle East Respiratory Syndrome) MERS-CoV; 2012, influenza kuş (H5N1 Avian Influenza) gribi adı verilen binlerce insanı öldüren solunum yolları hastalıklarına etken olan virüsleri, gelişen bu yeni tablodan dışarda tutarak farklı bir hastalığın meydana geldiğini ve bununda yeni Coronavirüs, Covid-19 şeklinde tanımlandığını duyurmuşlardır. Bu virüsler soğuk algınlığına sebep olan virüslerdir. Ayrıca hayvanlarda görülen birçok koronavirüs çeşidi vardır ve bazılarının hayvanlardan insanlara geçerek insanlarda ciddi hastalıklara neden olabilmektedir. SARS-CoV, 21. Yüzyılda, ilk milletler arasındaki acil sağlık durumu 2003 yılında ortaya çıkmıştır. Önceden hiç duyulmamış bir virüs olarak meydana gelen ve binlerce kişinin hayatını yitirmesine sebep olan SARS-CoV yaklaşık on sene sonra koronavirüs ailesinden, öncesinde hiç insan ya da hayvanlarda ortaya çıkmamış olan MERS-CoV ortaya çıkmıştır, Eylül 2012'de ilk defa Suudi Arabistan'daki kişilerde görüldüğü sanılmıştır ancak daha sonradan yapılan araştırmalarda ilk vakaların Nisan 2012'de Ürdün Zarqa'daki bir hastanede görüldüğü kayıtlara geçirilmiştir. Tek zincirli, zarflı ve pozitif polariteli virüslere RNA virüsleri denir. Pozitif polariteli bu virüsler RNA'ya bağımlı RNA polimeraz enzimi içermezler fakat RNA polimeraz enzimini genomlarında kodlarlar. Üst katmanlarında çubuksu çıkıntıları vardır. Bu uzantılara latince "corona", yani "taç" denilmektedir. Bu tanımdan yola çıkılarak bu virüslere Coronavirus (taçlı virüs) denilmiştir (Zhou, 2019). Dünya Sağlık Örgütü (DSÖ) 31 Aralık 2019'da, Çin'in Hubei bölgesindeki Wuhan kentinde etiyolojisi bilinmeyen pnömoni vakalarını kayıt altına almıştır. 7 Ocak 2020'de vaka öncesinde insanlarda gözlemlenmemiş olan yeni bir koronavirüs (2019- nCoV) olarak kayıt altına alınmıştır. Sonrasında 2019-nCoV salgının adı Covid-19 olarak kabul görmüş ve virüs SARS CoV'e çok benzerlik gösterdiğinden SARS-CoV-2 olarak nitelendirilmiştir. DSÖ, Covid-19 küresel pandemisini 30 Ocak 2020'de "uluslararası boyutta halk sağlığı acil durumu" ile sınıflandırmış ve salgının başladığı Çin'in Hubei bölgesi dışında 113 ülkede Covid-19 hastalığının görülmesi, virüsün salgını ve ağır geçirilmesi nedeniyle 11 Mart'ta küresel salgın (pandemi) olarak tanımlamıştır. Aynı zamanda bu virüstün halk sağlığı üzerinde yaratacağı etkilerde tartışılmıştır (Hui, 2020). Türkiye'de Covid-19 ile ilgili araştırmalar 10 Ocak'ta başlamış ve 22 Ocak'ta T.C. Sağlık Bakanlığı Bilimsel Danışma Kurulu ilk toplanmasını yapmıştır, Covid-19'a yönelik almış olduğumuz önlemlere rağmen ilk Covid-19 vakası 11 Mart'ta görülmüştür. DSÖ 13 Mart 2020'de hastalığın seyrini değiştirebilecek ve insanların kafasındaki soru işaretlerini gidermekte yardımcı olacak basın açıklamasında "Nerede olduğunu bilmediğiniz bir virüsle savaşamazsınız. Covid-19 yayılım zincirini kırmak için tespit edin, izole edin, test edin ve tedavi edin! Bulduğumuz ve tedavi ettiğimiz her olgu hastalığın yayılımını kısıtlayacak." açıklamalarında bulundu. Covid-19 verileri ile ilgili çok sayıda araştırma yapılmıştır (Şencan, 2020).

Bu çalışmada, Türkiye'de 11 Mart 2020 – 19 Mayıs 2020 tarihleri arasında ortaya çıkmış olan toplam Covid-19 vaka sayıları için Lojistik Büyüme ve Üstel Büyüme modellerini kullanarak ileriki dönemler için vaka sayılarını tahmin etmişlerdir. Covid-19 teşhisinde Yapay Sinir Ağları yöntemleri kullanılarak sınıflama yapan çalışmalar da vardır (Tang, 2021; Karacan & Eryılmaz, 2021). Yaptıkları çalışmada bazı matematiksel modeller kullanarak salgının gidişatı hakkında tahminler yapmışlardır (Adiga, 2020). Bu modellerden en çok bilineni SIR modelidir. ARIMA modelini kullanarak İtalya Covid-19 verilerini analiz etmiştir (Perone, 2020). Covid-19 verileri için matematiksel modeller, istatistiksel modeller ve derin öğrenme yöntemlerini karşılaştıran bir çalışma yapmışlardır (Masum, 2022). Derin öğrenme yöntemleri ile Covid-19 verilerini analiz etmişlerdir (Zeroual, 2020). Zaman serisi modelleri ile Covid-19 verilerini analiz etmişlerdir (Maleki, 2020).

Diğer çalışmalardan farklı olarak bu çalışmadaki amaç, Bayes Ağları ile COVID-19 hastalarının yaşama ve ölüm tahminlerini yapmaktır. Bayes ağları son zamanlarda psikoloji, biyoloji, sosyal bilimler vb. birçok alanda kullanılmaktadır (Gemela, 2001). En yaygın kullanım alanlarından biride risk tahminleridir.

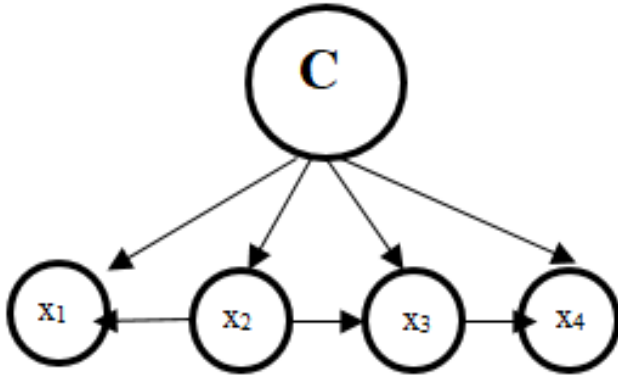
2. Materyal ve Yöntem

Veri madenciliğinin önemli sınıflandırma modellerinden biri Bayes ağlarıdır. Bayes Ağları, olasılığı belirsizliğin bir ölçüsü olarak kullanan olasılıksal uzman sistemlerin (probabilistic expert systems) bir türüdür (Spiegelhalter et.al, 1993). Bayes Ağları aynı zamanda Bayesçi İnanç Ağları (Bayesian Belief Networks), Nedensel Olasılık Ağları (Causal Probabilistic Networks), Olasılıksal Sebep - Sonuç Modelleri (Probabilistic Cause - Effect Models), Olasılıksal Etki Diyagramları (Probabilistic Influence Diagrams) olarak da isimlendirilmektedir (Spiegelhalter et.al., 1993). Yapay zekâ ve makine öğrenmesi konularındaki birçok çalışmada Bayes Ağları sıklıkla kullanılmaktadır. Özellikle karmaşık koşullu ilişki yapılarının ortaya çıkarılmasında Bayes Ağları tercih edilmektedir.

2.1. Bayes Ağları

Veri madenciliğinin önemli sınıflandırma modellerinden biri Bayes ağlarıdır. Bayes Ağları, olasılığı belirsizliğin bir ölçüsü olarak kullanan olasılıksal uzman sistemlerin (probabilistic expert systems) bir türüdür (Spiegelhalter et.al, 1993). Bayes Ağları aynı zamanda Bayesçi İnanç Ağları (Bayesian Belief Networks), Nedensel Olasılık Ağları (Causal Probabilistic Networks), Olasılıksal Sebep - Sonuç Modelleri (Probabilistic Cause - Effect Models), Olasılıksal Etki Diyagramları (Probabilistic Influence Diagrams) olarak da isimlendirilmektedir (Spiegelhalter et.al., 1993). Yapay zekâ ve makine öğrenmesi konularındaki birçok çalışmada Bayes Ağları sıklıkla kullanılmaktadır. Özellikle karmaşık koşullu ilişki yapılarının ortaya çıkarılmasında Bayes Ağları tercih edilmektedir.

Bu teknikler değişkenlerin koşullu olasılıkları ile temsil edilen anlaşılması zor ilişkilerin matematiksel modelini yönlendirilmiş döngüsüz diyagram (DAG =Directed Acyclic Graph) ile görselleştirir. Bayes Ağlarının yapısı “Grafiksel Kısım” ile gösterilirken ilişkili değişkenler arasındaki koşullu olasılıklar “Koşullu Olasılıklar Tablosu” nda gösterilmektedir. Bayes Ağları tablodaki rastgele değişkenler arasındaki bağımlılık ilişkilerinin anlaşılmasını ve görselleştirilmesini sağlayan grafiksel olasılık yapısı ortaya çıkaran istatistiksel bir yöntemdir (Jensen & Nielsen, 2007). Aynı zamanda bu ağ koşullu olasılıklar için çıkarsama yapılmasını sağlar. Bayes Teoremi yardımıyla araştırmacıların eldeki verilerin yanında uzman bilgisini de ağa katmasını sağlar (Atılğan & Ersel, 2017). TAN (Tree augmented naive) Bayes ağı son zamanlarda, şu andaki makine öğrenmesi ve veri madenciliği yöntemlerinden daha çok ilgi görmektedir. TAN modelleri bir ağaç oluşumuna katkı sağlayan Naive Bayes ağlarına kadar uzanır.



Şekil 1. Tipik bir TAN yapısı ağı

Şekil 1’deki TAN yapısındaki “C” bir düğümü, X1, X2, X3, X4 ise çocukları temsil etmektedir. $G=(V,A)$ grafiği ve P bileşik olasılık dağılımı verilen bir Bayes Ağı verildiğinde grafikte yer alan herhangi bir X düğümünün ebeveynleri verildiğinde X düğümü ebeveyni ya da torunu olmayan herhangi bir düğümden yönsel ayrılmıştır. Yani ağda bulunan tüm $v \in V$ düğümleri için $v \perp \text{Nondesc}(v) \setminus \text{Pa}(v)$ ise, Bayes Ağı Markov koşulunu sağlamaktadır. Markov koşulu dikkate alındığında yönsel ayrılma ve koşullu bağımsızlık arasında güçlü bir ilişki vardır (Heckerman, 2008).

$G=(V,A)$ yönlü döngüsel olmayan grafikte herhangi X,Y,Z düğümleri ve P olasılık fonksiyonları için:

G: düğüm, **P:** olasılık

1) G ve P her zaman tutarlı ise $(X \perp Y | Z) \Rightarrow (X \perp Y | Z)P$

2) $(X \perp Y | Z)P$ G ile tutarlı tüm dağılımları sağlıyorsa $(X \perp Y | Z)G$ olacaktır (Jensen, 2007; Nielsen, 2007).

Bayes Ağları oluşturdukları koşullu olasılıklarla, olasılığa dayalı çıkarım yapabilen bir sistem gibi düşünülebilir. Olasılığı kullanan çıkarımsal sistemlerde, düğümler bütünü için bazı düğümlere ilişkin bilgiler sunulduğunda, olasılıkların çıkarılması öğrenen sistemlerde ana görevini oluşturmaktadır. Olasılık çıkarma süreci ağ güncellemesi diye de nitelendirilmektedir. Bayes Ağları'nda bir düğüm için bilgi girdisi olduğunda diğer bütün düğümlerin yeniden revize edilmesinden dolayı çıkarımlar geniş bir yelpazeye aittir (Korb & Nicholson, 2011). X değişkenler vektöründe tanımlı olan Bayes Ağları, X vektöründeki değişkenlere ait bileşik olasılık dağılımlarını ortaya koyduğu için Bayes Ağları'ndan faydalanarak bu vektörler kümesindeki tüm olasılıklar hesaplanır (Heckerman, 2008; Lauria, 2006; Duchessi, 2006).

Bayes Ağları'nın yapısal özelliklerini meydana getiren olmazsa olmazlarından biri de öğrenme kısmıdır. Pratikte çoğunlukla araştırılan konu veya soruna ait Bayes Ağı yapısının çok iyi bilinmemesinden kaynaklı mevcut veriler üzerinden modelin öğrenilmesi gerekir (Olmuş ve Erbaş, 2012). "Bayes Ağları Öğrenme" (Bayesian Network Learning) ile tanımlanan bu öğrenme süreci, Bayes Ağı'ndaki DAG ve bileşik olasılık dağılımlarının parametrelerinin tahmin edilmesi olarak nitelendirilebilmektedir. Bayes Ağları'nda yapısal öğrenme ve parametre öğrenme olmak üzere iki çeşit öğrenme tanımı olduğu söylenebilir. Bu öğrenme modelleri oluşturulacak olan modelin yapısını, modelin düğümlerini ve ilişkilerini belirleme amacıyla etkin bir şekilde kullanılmaktadır. Yapısal öğrenme algoritmaları, Bayes Ağı'nın diyagramsal yapısını kendi kendine oluşturmak için uygulanan metotlardır. Yapısal öğrenme algoritmaları hakkındaki ilk araştırma (Pearl & Verma, 1990) neden sonuç araştırması üzerinde olmuştur. Yapısal öğrenme algoritmaları koşullu bağımsızlık testlerini uygulayarak Bayes Ağları'nın yapısını öğrenmek adına bir temel oluşturmaktadır (Scutari, 2011).

Bir Bayes Ağı bir dizi ayrık değişkenler üzerinde ortak bir olasılık dağılımı sunar. Bayes Ağları sınıflandırıcıları, sınıflama problemi için tasarlanan Bayes Ağları'nın özel bir çeşididir. Denetimli sınıflandırma, bir dizi değişken veya özellikleri tahmin etmek için etiketleme veya kategorilere ayırmayı amaçlamaktadır. Pratikte basit bir Bayes ağ sınıflandırıcısı olan Navie Bayes sınıflandırıcıları genellikle şaşırtıcı derecede iyi performans gösterir. Bu sınıflandırıcılar temel olarak her bir Xi değişkenin koşullu olasılıklarından öğrenir.

Naive Bayes ve TAN birbirine bağlı çeşitli türler arasında birçok belirsizlik içeren büyük veri kümelerini modellemek için kullanılan olasılığa dayalı grafiksel modellerdir. Bu modellerin, resim parçalama, tıbbi teşhis, diğer çeşitli veri sınıflandırmaları gibi yaygın uygulamaları mevcuttur. Bir sınıflandırma problemi, bu tür çeşitli örneklerin analiziyle elde edilen önceki bilgilere dayanarak belirli bir örneğin hangi kategoriye ait olduğunun belirlenmesiyle ilgilenir.

Naive Bayes modeli, bir örneğin tüm niteliklerinin, o örneğin sınıfı verildiğinde birbirinden bağımsız olduğunu varsayar fakat bu modelde yapılan bağımsızlık varsayımı doğru ve gerçekçi değildir.

TAN modeli, sistemin öznelikleri arasına bir etkileşim düzeyi daha ekleyerek Naive Bayes modelini geliştirir. Tan modelinde her nitelik sınıflara ve diğer niteliklere bağlıdır. Bu model, nitelikler arasındaki bağımlılıkları içerdiğinden, Naive Bayes modelinden daha gerçekçidir. TAN süreci Chow and Liu tarafından önerilen eski bir algoritmanın temeline dayanır. Kısıtlı olmayan Bayes Ağı sınıflandırıcı ise Navie Bayes ve TAN Ağlarından farklıdır. Kısıtlı olmayan Bayes Ağı modelinin DAG'ı değişkenler arasında (marjinal ve koşullu) bağımsız ilişkiler kurar. Sınırsız bir olasılıksal ağın yapısını öğrenmek için birkaç farklı yaklaşım mevcuttur. Bu çalışma için Bayes yapısal öğrenme yaklaşımından yararlanılmıştır.

2.1.1. Veriler

Bu çalışmanın amacı, Covid-19 geçiren kişiler için bazı değişkenleri kullanarak bir risk tahmini yapmak ve bu kişilerin hayatta kalma olasılıklarını hesaplamaktır. Bu nedenle github kaynağında 104 ülke için verilmiş olan yaklaşık 10 milyon Covid-19 hastası içeren veriye her değişken için bir değer olacak şekilde temizleme

işlemleri yapılarak yaklaşık 215 bin veri elde edilmiştir. Bu veriye Bayes Ağları ile çözümlenmesi yapılarak hayatta kalıp kalmayacaklarına dair olasılık hesaplamaları ve sınıflama sonuçları verilmiştir.

Veriler github üzerinden “https://github.com/beoutbreakprepared/nCoV2019/blob/master/latest_data/latestdata.tar.gz” linkinden çekilmiştir.

Analiz Sonuçları

Bu çalışma sayesinde koronavirüse yakalanan kişilerin, Bayes Ağları yöntemi kullanılarak hayatta kalıp kalamayacağı sınıflandırılıp tahmin edilecektir. Öte yandan hastalığa yakalanan kişiler ya da diğer ilgili kişiler yapılan modele o hastanın verilerini girerek ölüm riskini hesaplayabileceklerdir. Sonuç olarak dünya genelinde oluşacak olan farkındalık sayesinde ölüm riski yüksek olan kişiler tespit edilecek ve ölüm oranları en aza indirilmesi sağlanacaktır.

Bu çalışmanın amacı, Covid-19 geçiren kişiler için bazı değişkenleri kullanarak bir risk tahmini yapmak ve bu kişilerin hayatta kalma olasılıklarını hesaplamaktır kullanılmamalıdır.

Tablo 1

Veri kümesine genel bakış

YAS	CINSİYET	ULKE	KRONIK	SONUC	SEYAHAT	HASTALIK TARİHİ	SEMPATOM	HASTALIK AYI
35	Erkek	Afganistan	Yok	Hayatta	Var	24.02.2020	Yok	Şubat
35	Erkek	Afganistan	Yok	Hayatta	Var	24.02.2020	Yok	Şubat
53	Kadın	Cezayir	Yok	Hayatta	Var	2.03.2020	Yok	Mart
24	Kadın	Cezayir	Yok	Hayatta	Var	2.03.2020	Yok	Mart
53	Kadın	Cezayir	Yok	Hayatta	Yok	2.03.2020	Yok	Mart
24	Kadın	Cezayir	Yok	Hayatta	Yok	2.03.2020	Yok	Mart
78	Kadın	Cezayir	Yok	Hayatta	Yok	5.03.2020	Yok	Mart
62	Erkek	Cezayir	Yok	Öldü	Yok	5.03.2020	Yok	Mart
54	Kadın	Cezayir	Yok	Hayatta	Yok	5.03.2020	Yok	Mart
84	Kadın	Cezayir	Yok	Öldü	Yok	15.03.2020	Yok	Mart
77	Kadın	Cezayir	Yok	Öldü	Yok	16.03.2020	Yok	Mart
50	Erkek	Cezayir	Yok	Hayatta	Yok	17.03.2020	Yok	Mart
42	Erkek	Cezayir	Yok	Öldü	Yok	17.03.2020	Yok	Mart
83	Erkek	Cezayir	Yok	Hayatta	Var	19.03.2020	Yok	Mart
75	Erkek	Cezayir	Yok	Hayatta	Var	19.03.2020	Yok	Mart
78	Erkek	Cezayir	Yok	Hayatta	Var	20.03.2020	Yok	Mart
49	Erkek	Cezayir	Yok	Hayatta	Var	20.03.2020	Yok	Mart
21	Erkek	Cezayir	Yok	Hayatta	Yok	20.03.2020	Yok	Mart
85	Erkek	Cezayir	Yok	Öldü	Yok	21.03.2020	Yok	Mart
64	Erkek	Cezayir	Yok	Hayatta	Yok	25.03.2020	Yok	Mart

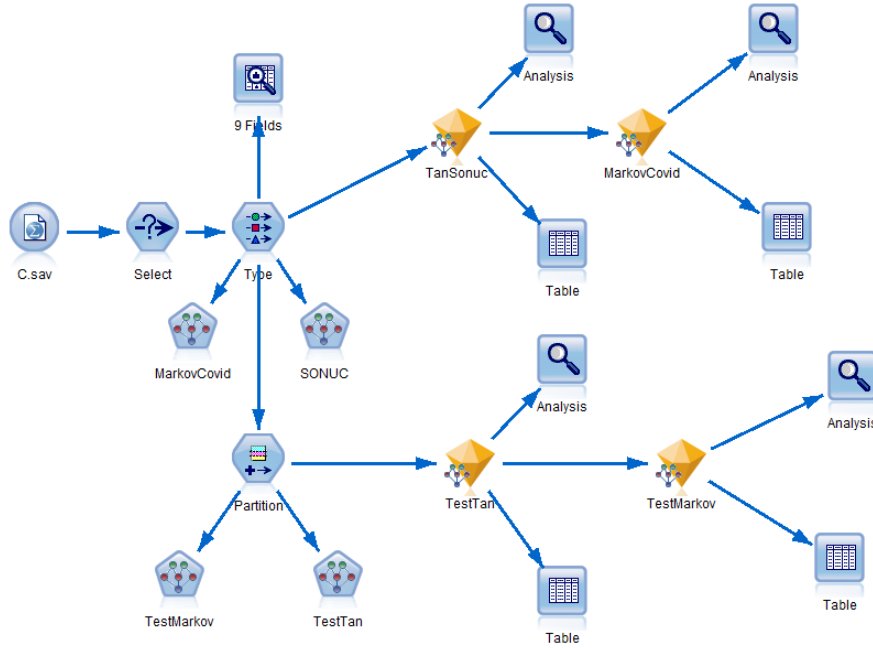
Tablo 1’de görüldüğü üzere tüm verinin 20 satırlık bir kesiti verilmiştir. Veriler 10.309.980 hastanın içinden seçilen 215.968 hastadan oluşan ve dokuz değişken içeren 104 ülkeye ait bir veri kümesidir. Seçilen hastaların yaşları 1’den 105’e kadar değişmektedir. Ek olarak hastaların kronik rahatsızlığı olup olmaması, son üç ay içinde yurtdışı seyahatinde bulunup bulunmadığı, cinsiyeti ve hangi ayda hastalığa yakalandığı bilgileri mevcuttur. Bağımlı değişken ise hastanın koronavirüs sonucunda hayatta kalıp kalamadığıdır.

Tablo 2

Verinin içerdiği değişkenler ve tanımları

Değişken Adı	Tanımı
YAS	Hastaların hastalığa yakalandığı tarihteki yaşları
CINSİYET	Hastanın cinsiyeti
ULKE	Hastanın hangi ülkede yaşadığı
KRONIK	Hastanın kronik rahatsızlığı olup olmadığı
SONUC	Hastanın koronavirüs sonucu hayati durumu
SEYAHAT	Hastanın son 3 ay içinde yurtdışı seyahati durumu
HASTALIKTARIHI	Hastanın hastalığa yakalandığı tarih
SEMPATOM	Hastalığa yakalandığında semptom gösterme durumu
HASTALIKAYI	Hastanın hastalığa yakalandığı ay

Tablo 2’de veri kümesindeki dokuz değişkenin tanımları verilmiştir. Verilerin analizi IBM SPSS Modeler 18.0 programı ile yapılmıştır (IBM Corp., 2016) IBM SPSS Modeler uygulaması ile “Makine Öğreniminin Bayes Ağları” modeli kullanılmıştır (Wendler & Sören, 2016). Tan ve Markov modelleri denenmiş ve birbirleri ile karşılaştırılmıştır. Test modülü çalıştırılarak gözelerin %20’si tahmin edilmiştir.



Şekil 2. IBM SPSS Modeler programı ile Bayes Ağları kesiti

Şekil 2’de (IBM Corp., 2016) IBM SPSS Modeler uygulaması ile “Makine Öğreniminin Bayes Ağları” modeli kullanılarak yapılan analizin görsel bir şeması verilmiştir.

Tablo 3
Tan modeli ve olasılıkları

YAŞ	CINSİYET	ULKE	KRONİK	SONUC	SEYAHAT	HASTALIK TARIHI	SEPTOM	HASTALIK AYI	TAN- SONUC	TAN-P- SONUC
39	Kadın	Avustralya	Yok	Hayatta	Var	27.01.2020	Yok	Ocak	Hayatta	0,87
29	Kadın	Avustralya	Yok	Hayatta	Var	27.01.2020	Yok	Ocak	Hayatta	0,87
21	Kadın	Avustralya	Yok	Hayatta	Var	27.01.2020	Yok	Ocak	Hayatta	1,00
21	Kadın	Avustralya	Yok	Hayatta	Var	27.01.2020	Yok	Ocak	Hayatta	1,00
79	Erkek	Avustralya	Yok	Hayatta	Var	27.01.2020	Yok	Ocak	Hayatta	0,85
79	Erkek	Avustralya	Yok	Hayatta	Yok	27.01.2020	Yok	Ocak	Hayatta	0,99
29	Kadın	Avustralya	Yok	Hayatta	Yok	27.01.2020	Yok	Ocak	Hayatta	1,00
44	Erkek	Avustralya	Yok	Hayatta	Var	29.01.2020	Yok	Ocak	Hayatta	0,84
44	Erkek	Avustralya	Yok	Hayatta	Var	29.01.2020	Yok	Ocak	Hayatta	0,84
42	Kadın	Avustralya	Yok	Öldü	Var	30.01.2020	Yok	Ocak	Hayatta	0,87

Yapılan analiz sonucunda Tablo 3'e bakıldığında Tan Modeli için 10 kişiye ait tahminler ve hastanın hayatta kalıp kalamayacağı görülmektedir. Örneğin 29 yaşındaki bir kadın hasta için gerçekleşen sonuç "Hayatta" iken Tan Modeli tarafından "Hayatta" şeklinde %87 olasılıkla doğru tahmin edilmiştir. 42 yaşındaki kadın hasta için gerçekleşen sonuç "Öldü" iken Tan Modeli tarafından "Hayatta" şeklinde %87 olasılıkla yanlış tahmin edilmiştir.

Tablo 4
Markov modeli ve olasılıkları

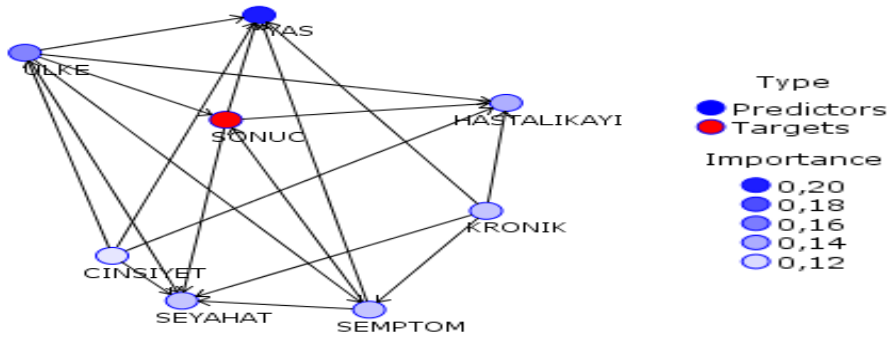
YAŞ	CINSİYET	ULKE	KRONİK	SONUC	SEYAHAT	HASTALIK- TARIHI	SEPTOM	HASTALIK AYI	MARKOV- SONUC	MARKOV- P-SONUC
20	Erkek	Avustralya	Yok	Öldü	Var	25.02.2020	Yok	Şubat	Öldü	0,52
20	Erkek	Avustralya	Yok	Öldü	Var	25.02.2020	Yok	Şubat	Öldü	0,52
79	Erkek	Avustralya	Yok	Hayatta	Var	27.02.2020	Yok	Şubat	Öldü	0,59
79	Erkek	Avustralya	Yok	Hayatta	Var	27.02.2020	Yok	Şubat	Öldü	0,59
69	Kadın	Avustralya	Yok	Hayatta	Var	28.02.2020	Yok	Şubat	Hayatta	0,57
59	Erkek	Avustralya	Yok	Hayatta	Var	28.02.2020	Yok	Şubat	Hayatta	1,00
69	Kadın	Avustralya	Yok	Hayatta	Yok	28.02.2020	Yok	Şubat	Hayatta	0,94
59	Erkek	Avustralya	Yok	Hayatta	Yok	28.02.2020	Yok	Şubat	Hayatta	1,00
59	Kadın	Avustralya	Yok	Hayatta	Var	29.02.2020	Yok	Şubat	Hayatta	1,00
59	Kadın	Avustralya	Yok	Hayatta	Var	29.02.2020	Yok	Şubat	Hayatta	1,00

Yapılan analiz sonucunda Tablo 4'e bakıldığında Markov modeli için 10 örnek için tahmin oranları ve hastanın hayatta kalıp kalamayacağı görülmektedir. 69 yaşındaki kadın hasta için gerçekleşen sonuç "Hayatta" iken Markov modeli tarafından "Hayatta" şeklinde %94 olasılıkla doğru tahmin edilmiştir. 79 yaşındaki erkek hasta için gerçekleşen sonuç "Hayatta" iken Markov modeli tarafından "Öldü" şeklinde %59 olasılıkla yanlış tahmin edilmiştir.

Tablo 5
Tan ve Markov modelleri için karşılaştırma oranları

Modelleri Oransal Karşılaştırma	Doğru Sayısı	Oran	Yanlış Sayısı	Oran
Tan Modeli	213.768	%98,98	2.200	%1,02
Toplam		215.968		
Markov Modeli	214.149	%99,16	1.819	%0,84
Toplam		215.968		
Tan-Markov Modeli Birlikte	215.253	%99,67	715	%0,33
Toplam		215.968		
Modelin Aynı Tahmin Ettiği	213.601	%99,23	1.652	%0,77
Toplam		215.253		

Yapılan analiz sonucunda Tablo 5’te görüldüğü üzere Tan modeli verinin %98,98’ini doğru sınıflarken, %1,02’sini yanlış sınıflamıştır. Markov modeli verinin %99,16 doğru sınıflarken, %0,84’ünü yanlış sınıflamıştır. Her iki modelin aynı sonucu verdiği oran %99,67’dir. İki model de birbiri ile uyumludur denilebilir. Her iki modelin aynı sonucu verdiği 215.253 hastadan 213.601 tanesi ham verideki sonuçla aynı iken 1.652 tanesi farklı sonuç vermiştir. Bu olasılıklar dikkate alındığında Markov modeli seçilmiş ve analizlere Markov modeli üzerinden devam edilmiştir.



Şekil 3. Markov modeli için değişkenlerin anlamlılık düzeyleri

Seçilen Markov modeli üzerinden değişkenlerin anlamlılık düzeyi Şekil 3’te görüldüğü gibi, hastanın koronavirüsten sonra yaşayıp yaşayamayacağını etkileyen en önemli değişken, %20’lik bir oranla “Yaş” faktörüdür. İkinci önemli değişken yaklaşık %16 ile hastanın hastalığa yakalandığı ülkedir. Üçüncü ve dördüncü önemli değişkenler ise sırası ile %14’lük bir oranla hastanın hastalığa yakalandığı ay ile %13’lük bir oranla hastanın kronik rahatsızlığı olup olmadığı gelmektedir. Yapılan araştırmada; hastanın cinsiyeti ve son üç ay içinde yurt dışına çıkıp çıkmaması değişkenleri bilinen algının aksine hastanın hayatta kalıp kalamayacağına en az etkili değişkenlerdir.

Tablo 6
Hastalığa hangi ayda yakalandığına ait koşullu olasılık tablosu

Sıra	Aile					Olasılık		
	Cinsiyet	Ülke	Kronik	Sonuç	Semptom	Mart	Nisan	Mayıs
1	Erkek	Ekvator Ginesi	Yok	Hayatta	Var	0	0,333	0,666
2	Erkek	Ekvator Ginesi	Yok	Hayatta	Yok	0	0	1
3	Erkek	Ermenistan	Yok	Hayatta	Yok	0	0	1
4	Erkek	Estonya	Yok	Hayatta	Yok	0,717	0,281	0,001
5	Erkek	Esvatini	Yok	Hayatta	Yok	0	0,333	0,666
6	Erkek	Etiyopya	Yok	Hayatta	Yok	0	0,116	0,848
7	Erkek	Etiyopya	Yok	Öldü	Yok	0	0,5	0,5
8	Erkek	Fildişi Sahili	Yok	Hayatta	Yok	0	0,333	0,666
9	Erkek	Filipinler	Var	Hayatta	Var	0	1	0
10	Erkek	Filipinler	Var	Hayatta	Yok	0	0,833	0,166
11	Erkek	Filipinler	Var	Öldü	Var	0	0,938	0,061
12	Erkek	Filipinler	Var	Öldü	Yok	0	1	0
13	Erkek	Filipinler	Yok	Hayatta	Var	0	0,4	0,6
14	Erkek	Filipinler	Yok	Hayatta	Yok	0	0,304	0,646
15	Erkek	Filipinler	Yok	Öldü	Var	0	0,714	0,285

Hastalığa hangi ayda yakalandığına ait koşullu olasılık tablosuna (Tablo 6) bakıldığında, hastanın cinsiyeti “Erkek, ülkesinin “Etiyopya”, kronik rahatsızlığının olmadığı, semptomunun olmadığı ve hastalık sonucu öldüğü bilindiğinde bu hasta %50 ihtimalle ya Mayıs ayında ya da Nisan ayında hastalığa yakalanmıştır. Hastanın cinsiyetinin “Erkek”, ülkesinin “Filipinler”, kronik rahatsızlığının olduğu, semptomunun olmadığı ve hastalık sonucu öldüğü bilindiğinde bu hastanın nisan ayında hastalığa yakalanma olasılığı yaklaşık %94’tür.

Tablo 7
Yaşa ait koşullu olasılık tablosu

Sıra	Aile			Olasılık				
	Cinsiyet	Ülke	Sonuç	<=21.8	21.8-42.6	42.6-63.4	63.4-84.2	>84.2
1	Erkek	Afganistan	Hayatta	0	1	0	0	0
2	Erkek	Almanya	Hayatta	0.016	0.282	0.391	0.211	0.098
3	Erkek	Almanya	Öldü	0.2	0.6	0.2	0	0
4	Erkek	ABD	Hayatta	0.049	0.289	0.357	0.276	0.026
5	Erkek	ABD	Öldü	0	0.228	0.228	0.473	0.070
6	Erkek	Andora	Öldü	1	0	0	0	0
7	Erkek	Angola	Hayatta	0	1	0	0	0
8	Erkek	Arjantin	Hayatta	0	0.15	0.55	0.3	0
9	Erkek	Avustralya	Hayatta	0.064	0.273	0.319	0.326	0.015
10	Erkek	Avustralya	Öldü	0.12	0.08	0	0.8	0
11	Erkek	Bahamalar	Hayatta	0	0.185	0.518	0.296	0
12	Erkek	Belçika	Hayatta	0.010	0.149	0.279	0.304	0.255
13	Erkek	Benin	Hayatta	0.333	0.333	0.333	0	0
14	Erkek	BAE	Hayatta	0	0	0	1	0
15	Erkek	Bolivya	Hayatta	0	0.227	0.5	0.227	0.045

Hastanın yaşlarına ait koşullu olasılık tablosuna bakıldığında, hastanın cinsiyetinin erkek, ülkesinin Bahamalar ve koronavirüs sonrası hayatta kaldığı bilindiğine göre bu kişinin 42,6 ile 63,4 yaşları arasında olması olasılığı %51,8’dir. Hastanın cinsiyetinin “Erkek”, ülkesinin “Avustralya” ve koronavirüs sonrası öldüğü bilindiğine göre bu kişinin 63,4 ile 84,2 yaşları arasında olması olasılığı %80’dir.

Tablo 8
Hastanın cinsiyete ait koşullu olasılık tablosu

Sıra	Aile		Olasılık	
	Ülke	Erkek	Kadın	
1	Burkina Faso	0.8	0.2	
2	Butan	1	0	
3	Cezayir	0.590	0.409	
4	Ekvator Ginesi	0.454	0.545	
5	Ermenistan	1	0	
6	Estonya	0.441	0.558	
7	Esvatini	0.6	0.4	
8	Etiyopya	0.692	0.307	
9	Fildişi Sahili	0.666	0.333	
10	Filipinler	0.558	0.441	
11	Finlandiya	0.2	0.8	
12	Fransa	0.857	0.142	
13	Gabon	0.6	0.4	
14	Gambiya	0.666	0.333	
15	Gana	0.615	0.384	

Hastanın cinsiyetine ait koşullu olasılık tablosuna bakıldığında, Burkina Faso'da koronavirüs hastalarının %80'i erkek ve %20'si kadındır. Finlandiya'da koronavirüs hastalarının %80'i kadın ve %20'si erkektir.

Tablo 9
Hastanın seyahat geçmişine ait koşullu olasılık tablosu

Sıra	Cinsiyet	Aile			Olasılık	
		Ülke	Kronik	Sonuç	Var	Yok
1	Erkek	Afganistan	Yok	Hayatta	1	0
2	Erkek	Almanya	Var	Hayatta	0	1
3	Erkek	Almanya	Yok	Hayatta	0.001	0.998
4	Erkek	Almanya	Yok	Öldü	1	0
5	Erkek	ABD	Var	Hayatta	0.333	0.666
6	Erkek	ABD	Yok	Hayatta	0.130	0.869
7	Erkek	ABD	Yok	Öldü	0.807	0.192
8	Erkek	Andora	Yok	Öldü	0	1
9	Erkek	Angola	Yok	Hayatta	0	1
10	Erkek	Arjantin	Yok	Hayatta	0.75	0.25
11	Erkek	Avustralya	Yok	Hayatta	0.543	0.456

12	Erkek	Avustralya	Yok	Öldü	1	0
13	Erkek	Bahamalar	Yok	Hayatta	1	0
14	Erkek	Belçika	Yok	Hayatta	0	1
15	Erkek	Benin	Yok	Hayatta	0	1
16	Erkek	BAE	Yok	Hayatta	1	0
17	Erkek	Bolivya	Var	Hayatta	0	1
18	Erkek	Bolivya	Yok	Hayatta	0	1
19	Erkek	Bolivya	Yok	Öldü	0.5	0.5
20	Erkek	Brezilya	Var	Hayatta	0	1
21	Erkek	Brezilya	Yok	Hayatta	0.152	0.847
22	Erkek	Brezilya	Yok	Öldü	0.444	0.555
23	Erkek	Bulgaristan	Yok	Hayatta	0	1
24	Erkek	Burkina Faso	Yok	Hayatta	0.5	0.5
25	Erkek	Butan	Yok	Hayatta	0.5	0.5
26	Erkek	Cezayir	Yok	Hayatta	0.4	0.6
27	Erkek	Cezayir	Yok	Öldü	0	1
28	Erkek	Ekvator Ginesi	Yok	Hayatta	0.4	0.6
29	Erkek	Ermenistan	Yok	Hayatta	0.5	0.5
30	Erkek	Estonya	Yok	Hayatta	0.001	0.998
31	Erkek	Esvatini	Yok	Hayatta	0.666	0.333
32	Erkek	Etiyopya	Yok	Hayatta	0.441	0.558
33	Erkek	Etiyopya	Yok	Öldü	0	1
34	Erkek	Fildişi Sahili	Yok	Hayatta	0.666	0.333
35	Erkek	Filipinler	Var	Hayatta	0.9	0.1

Hastanın seyahat geçmişine ait koşullu olasılık tablosuna bakıldığında, hastanın cinsiyetinin erkek, Brezilya'da yaşadığı, kronik rahatsızlığının olmadığı ve koronavirüs sonucu hayatta kaldığı bilindiğinde bu kişinin son üç ay içinde seyahat geçmişinin olmama olasılığı %84,7'dir.

Hastanın cinsiyetinin erkek, Amerika Birleşik Devletleri'nde yaşadığı, kronik rahatsızlığı olmadığı ve koronavirüs sonucu öldüğü bilindiğinde bu kişinin son üç ay içinde seyahat geçmişinin olma olasılığı %80,7'dir.

Tablo 10
Sonuç değişkenine ait koşullu olasılık tablosu

Sıra	Aile		Olasılık	
	Ülke	Semptom	Hayatta	Öldü
1	Afganistan	Yok	1	0
2	Almanya	Var	1	0
3	Almanya	Yok	0.999	0
4	ABD	Var	0.75	0.25
5	ABD	Yok	0.980	0.019
6	Andora	Var	0	1
7	Angola	Yok	1	0
8	Arjantin	Var	1	0
9	Arjantin	Yok	1	0
10	Avustralya	Var	1	0
11	Avustralya	Yok	0.929	0.070
12	Bahamalar	Var	0.5	0.5
13	Bahamalar	Yok	1	0
14	Belçika	Yok	1	0
15	Benin	Var	1	0
16	BAE	Yok	1	0
17	Birleşik Krallık	Yok	1	0
18	Bolivya	Var	0.666	0.333
19	Bolivya	Yok	0.957	0.042
20	Brezilya	Var	0.844	0.155
21	Brezilya	Yok	0.972	0.027

Sonuç değişkenine ait koşullu olasılık tablosuna bakıldığında, hastanın ülkesinin Avustralya olduğu ve semptomunun olmadığı bilindiğinde bu hastanın hayatta kalma olasılığı %92,9'dur. Hastanın ülkesinin Amerika Birleşik Devletleri olduğu ve semptomu olduğu bilindiğinde bu hastanın koronavirüs hastalığından ölme olasılığı %25'tir. Hastanın ülkesinin Bolivya olduğu ve herhangi bir semptomunun olmadığı bilindiğinde bu hastanın koronavirüs hastalığından sonra hayatta kalma olasılığı %95,7'dir. Hastanın ülkesinin Brezilya olduğu ve herhangi bir semptomunun olmadığı bilindiğinde bu hastanın koronavirüs hastalığından sonra hayatta kalma olasılığı %97,2'dir. Hastanın ülkesinin Bahamalar olduğu ve herhangi bir semptomunun olmadığı bilindiğinde bu hastanın koronavirüs hastalığından sonra hayatta kalma olasılığı birbirine eşittir.

Tablo 11
Tan modeline göre verilerin sınıflandırılması

SIRA	YAS	CINSİYET	ULKE	KRONİK	SONUC	SEYAHAT	HASTALIK TARİHİ	SEMPTOM	HASTALIK AYI	TAN-SONUC	TAN-P-SONUC
1	35	Erkek	Afganistan	Yok	Hayatta	Var	24.02.2020	Yok	Şubat	Hayatta	1,00
2	53	Kadın	Cezayir	Yok	Hayatta	Var	2.03.2020	Yok	Mart	Hayatta	1,00
3	24	Kadın	Cezayir	Yok	Hayatta	Var	2.03.2020	Yok	Mart	Hayatta	1,00
4	53	Kadın	Cezayir	Yok	Hayatta	Yok	2.03.2020	Yok	Mart	Hayatta	0,76
5	24	Kadın	Cezayir	Yok	Hayatta	Yok	2.03.2020	Yok	Mart	Hayatta	0,81
6	78	Kadın	Cezayir	Yok	Hayatta	Yok	5.03.2020	Yok	Mart	Hayatta	0,54
7	62	Erkek	Cezayir	Yok	Öldü	Yok	5.03.2020	Yok	Mart	Hayatta	0,66
8	54	Kadın	Cezayir	Yok	Hayatta	Yok	5.03.2020	Yok	Mart	Hayatta	0,76
9	84	Kadın	Cezayir	Yok	Öldü	Yok	15.03.2020	Yok	Mart	Hayatta	0,54
10	77	Kadın	Cezayir	Yok	Öldü	Yok	16.03.2020	Yok	Mart	Hayatta	0,54
11	50	Erkek	Cezayir	Yok	Hayatta	Yok	17.03.2020	Yok	Mart	Hayatta	0,66
12	42	Erkek	Cezayir	Yok	Öldü	Yok	17.03.2020	Yok	Mart	Hayatta	0,80
13	83	Erkek	Cezayir	Yok	Hayatta	Var	19.03.2020	Yok	Mart	Hayatta	1,00
14	75	Erkek	Cezayir	Yok	Hayatta	Var	19.03.2020	Yok	Mart	Hayatta	1,00
15	78	Erkek	Cezayir	Yok	Hayatta	Var	20.03.2020	Yok	Mart	Hayatta	1,00
16	49	Erkek	Cezayir	Yok	Hayatta	Var	20.03.2020	Yok	Mart	Hayatta	1,00
17	21	Erkek	Cezayir	Yok	Hayatta	Yok	20.03.2020	Yok	Mart	Hayatta	0,55
18	85	Erkek	Cezayir	Yok	Öldü	Yok	21.03.2020	Yok	Mart	Hayatta	0,60
19	64	Erkek	Cezayir	Yok	Hayatta	Yok	25.03.2020	Yok	Mart	Öldü	0,55
20	84	Kadın	Cezayir	Yok	Hayatta	Yok	1.04.2020	Yok	Nisan	Hayatta	1,00
21	64	Erkek	Cezayir	Yok	Hayatta	Yok	1.04.2020	Yok	Nisan	Hayatta	1,00
22	51	Erkek	Cezayir	Yok	Hayatta	Yok	1.04.2020	Yok	Nisan	Hayatta	1,00
23	49	Erkek	Cezayir	Yok	Hayatta	Yok	1.04.2020	Yok	Nisan	Hayatta	1,00
24	20	Erkek	Andora	Yok	Öldü	Yok	2.03.2020	Var	Mart	Öldü	1,00
25	38	Erkek	Angola	Yok	Hayatta	Yok	19.03.2020	Yok	Mart	Hayatta	1,00

Tablo 11’de verilen Tan modelinin sınıflandırması ve sınıflandırma olasılıklarına bakıldığında, tan modelinde aşağıdaki tabloda görüldüğü üzere ilk 25 hastaya baktığımızda 19 hastanın hayatta kalıp kalamayacağını doğru sınıflanmış, geri kalan 6 hastayı yanlış sınıflanmıştır. Örneğin 18. satırdaki gözleme baktığımızda 85 yaşındaki Cezayir’de yaşayan, kronik rahatsızlığı olmayan erkek hastanın sonucu koronavirüs hastalığından sonra öleceği yönünde olmasına rağmen, tan modeli bu hastanın %60 olasılıkla hastalıktan sonra hayatta kalacağını yanlış olarak sınıflandırmıştır.

Tablo 12
Markov modeline göre verilerin sınıflandırılması

SIRA	YAŞ	CINSİYET	ULKE	KRONİK	SONUC	SEYAHAT	HASTALIK TARİHİ	SEMPYOM	HASTALIK AYI	\$MARKOV-SONUC	\$MARKOV-P-SONUC
1	35	Erkek	Afganistan	Yok	Hayatta	Var	24.02.2020	Yok	Şubat	Hayatta	1,00
2	53	Kadın	Cezayir	Yok	Hayatta	Var	2.03.2020	Yok	Mart	Hayatta	1,00
3	24	Kadın	Cezayir	Yok	Hayatta	Var	2.03.2020	Yok	Mart	Hayatta	1,00
4	53	Kadın	Cezayir	Yok	Hayatta	Yok	2.03.2020	Yok	Mart	Hayatta	1,00
5	24	Kadın	Cezayir	Yok	Hayatta	Yok	2.03.2020	Yok	Mart	Hayatta	1,00
6	78	Kadın	Cezayir	Yok	Hayatta	Yok	5.03.2020	Yok	Mart	Öldü	0,63
7	62	Erkek	Cezayir	Yok	Öldü	Yok	5.03.2020	Yok	Mart	Hayatta	0,63
8	54	Kadın	Cezayir	Yok	Hayatta	Yok	5.03.2020	Yok	Mart	Hayatta	1,00
9	84	Kadın	Cezayir	Yok	Öldü	Yok	15.03.2020	Yok	Mart	Öldü	0,63
10	77	Kadın	Cezayir	Yok	Öldü	Yok	16.03.2020	Yok	Mart	Öldü	0,63
11	50	Erkek	Cezayir	Yok	Hayatta	Yok	17.03.2020	Yok	Mart	Hayatta	0,63
12	42	Erkek	Cezayir	Yok	Öldü	Yok	17.03.2020	Yok	Mart	Öldü	1,00
13	83	Erkek	Cezayir	Yok	Hayatta	Var	19.03.2020	Yok	Mart	Hayatta	1,00
14	75	Erkek	Cezayir	Yok	Hayatta	Var	19.03.2020	Yok	Mart	Hayatta	1,00
15	78	Erkek	Cezayir	Yok	Hayatta	Var	20.03.2020	Yok	Mart	Hayatta	1,00
16	49	Erkek	Cezayir	Yok	Hayatta	Var	20.03.2020	Yok	Mart	Hayatta	1,00
17	21	Erkek	Cezayir	Yok	Hayatta	Yok	20.03.2020	Yok	Mart	Hayatta	1,00
18	85	Erkek	Cezayir	Yok	Öldü	Yok	21.03.2020	Yok	Mart	Öldü	1,00
19	64	Erkek	Cezayir	Yok	Hayatta	Yok	25.03.2020	Yok	Mart	Hayatta	1,00
20	84	Kadın	Cezayir	Yok	Hayatta	Yok	1.04.2020	Yok	Nisan	Hayatta	1,00
21	64	Erkek	Cezayir	Yok	Hayatta	Yok	1.04.2020	Yok	Nisan	Hayatta	1,00
22	51	Erkek	Cezayir	Yok	Hayatta	Yok	1.04.2020	Yok	Nisan	Hayatta	1,00
23	49	Erkek	Cezayir	Yok	Hayatta	Yok	1.04.2020	Yok	Nisan	Hayatta	1,00
24	20	Erkek	Andora	Yok	Öldü	Yok	2.03.2020	Var	Mart	Öldü	1,00
25	38	Erkek	Angola	Yok	Hayatta	Yok	19.03.2020	Yok	Mart	Hayatta	1,00

Tablo 12’de verilen Markov modelinin sınıflandırması ve sınıflandırma olasılıklarına bakıldığında, Markov modelinde aşağıdaki tabloda görüldüğü üzere ilk 25 hastaya baktığımızda 23 hastanın hayatta kalıp kalamayacağını doğru sınıflanmış, geri kalan 2 hastayı yanlış sınıflanmıştır. Örneğin, 84 yaşındaki Cezayir’de yaşayan, kronik rahatsızlığı olmayan kadın hastanın koronavirüs hastalığından sonra hayatını kaybeden kadını, markov modeli %63 olasılıkla hastalıktan sonra öleceği yönünde doğru bir sınıflama yapmıştır.

Tablo 13

Test ve deneme modeline göre verilerin sınıflandırılması

SIRA	YAŞ	CINSİYET	ULKE	KRONİK	SONUC	SEYAHAT	HASTALIK TARİHİ	SEMPTOM	HASTALIK AYI	TEST KÜMESİ	TAN-SONUC	TAN-P-SONUC	MARKOV-SONUC	MARKOV-P-SONUC
1	57	Kadın	Avustralya	Yok	Hayatta	Yok	24.02.2020	Yok	Şubat	Training	Hayatta	0,99	Hayatta	1,00
2	49	Erkek	Avustralya	Yok	Hayatta	Yok	24.02.2020	Yok	Şubat	Training	Hayatta	0,99	Hayatta	1,00
3	24	Kadın	Avustralya	Yok	Hayatta	Yok	24.02.2020	Yok	Şubat	Training	Hayatta	0,99	Hayatta	0,82
4	20	Erkek	Avustralya	Yok	Öldü	Var	25.02.2020	Yok	Şubat	Training	Hayatta	0,83	Öldü	0,64
5	20	Erkek	Avustralya	Yok	Öldü	Var	25.02.2020	Yok	Şubat	Training	Hayatta	0,83	Öldü	0,64
6	79	Erkek	Avustralya	Yok	Hayatta	Var	27.02.2020	Yok	Şubat	Training	Hayatta	0,70	Hayatta	1,00
7	79	Erkek	Avustralya	Yok	Hayatta	Var	27.02.2020	Yok	Şubat	Training	Hayatta	0,70	Hayatta	1,00
8	69	Kadın	Avustralya	Yok	Hayatta	Var	28.02.2020	Yok	Şubat	Testing	Hayatta	0,77	Hayatta	1,00
9	59	Erkek	Avustralya	Yok	Hayatta	Var	28.02.2020	Yok	Şubat	Training	Hayatta	0,75	Hayatta	1,00
10	69	Kadın	Avustralya	Yok	Hayatta	Yok	28.02.2020	Yok	Şubat	Training	Hayatta	0,99	Hayatta	1,00
11	59	Erkek	Avustralya	Yok	Hayatta	Yok	28.02.2020	Yok	Şubat	Testing	Hayatta	0,99	Hayatta	1,00
12	59	Kadın	Avustralya	Yok	Hayatta	Var	29.02.2020	Yok	Şubat	Training	Hayatta	0,83	Hayatta	1,00
13	59	Kadın	Avustralya	Yok	Hayatta	Var	29.02.2020	Yok	Şubat	Training	Hayatta	0,83	Hayatta	1,00
14	63	Kadın	Avustralya	Yok	Hayatta	Var	29.02.2020	Yok	Şubat	Training	Hayatta	0,83	Hayatta	1,00
15	63	Kadın	Avustralya	Yok	Hayatta	Yok	29.02.2020	Yok	Şubat	Training	Hayatta	0,99	Hayatta	1,00
16	29	Kadın	Avustralya	Yok	Hayatta	Yok	29.02.2020	Yok	Şubat	Testing	Hayatta	0,99	Hayatta	0,82
17	29	Kadın	Avustralya	Yok	Hayatta	Yok	29.02.2020	Yok	Şubat	Training	Hayatta	0,99	Hayatta	0,82

Veri kümesinin %80'i ile deneme yapıp %20'si ile test yapılmıştır. Test ve deneme modeline göre rastgele 17 hastanın tan ve markov modellerinin sınıflandırması ve sınıflandırma olasılıkları Tablo 13'te verilmiştir. Tan modeli bu 17 hastanın 15'ini doğru sınıflandırırken 2 hastayı yanlış sınıflandırmıştır. Markov modeli bu 17 hastanın tümünü doğru sınıflandırmıştır.

Tablo 14

Test ve deneme modelleri için karşılaştırma oranları

Modelleri Oransal Karşı- laştırma	Test				Deneme			
	Doğru	Oran	Yanlış	Oran	Doğru	Oran	Yanlış	Oran
Tan Modeli	43.008	%99,97	446	%1,03	170.741	%98,97	1.773	%1,03
Toplam		43.454				172.514		
Markov Modeli	43.039	%99,04	415	%0,96	171.043	%99,04	1.471	%0,85
Toplam		43.454				172.514		
Tan-Markov Modeli Birlikte	43.261	%99,56	193	%0,44	171.926	%99,66	588	%0,34
Toplam		43.454				172.514		
Modelin Aynı Tahmin Ettiği	42.939	%99,26	322	%0,74	170.598	%99,23	1.328	%0,77
Toplam		43.261				171.926		

Yapılan analiz sonucunda Tablo 14'te görüldüğü üzere Tan Modeli test verilerinin %98,97'sini doğru sınıflarken, %1,03'ünü yanlış sınıflamıştır. Markov Modeli test verininin %99,04'ünü doğru sınıflarken, %0,96'sını yanlış sınıflamıştır. Her iki modelin test verilerinde aynı sonucu verdiği oran %99,56'dır. İki model de birbiri ile uyumludur denilebilir. Her iki modelin test verilerinde aynı sonucu verdiği 43.261 hastadan 42.939 tanesi ham verideki sonuçla aynı iken 322 tanesi farklı sonuç vermiştir. Bu olasılıklar dikkate alındığında test verisi için Markov modeli seçilmiş ve analizler yapılmıştır.

Tablo 15

Yeni hastanın özelliklerine göre Markov modelinde sınıflandırılması

SIRA	YAS	CINSİYET	ULKE	KRONİK	SONUC	SEYAHAT	HASTA-LIK TA-RIHI	SEMP-TOM	HASTA-LIK AYI	MARKOV-SONUC	MARKOV-P-SONUC
1	59	Kadın	Avustralya	Yok	Hayatta	Var	29.02.2020	Yok	Şubat	Hayatta	1,00
2	63	Kadın	Avustralya	Yok	Hayatta	Var	29.02.2020	Yok	Şubat	Hayatta	1,00
3	63	Kadın	Avustralya	Yok	Hayatta	Yok	29.02.2020	Yok	Şubat	Hayatta	1,00
4	29	Kadın	Avustralya	Yok	Hayatta	Yok	29.02.2020	Yok	Şubat	Hayatta	0,82
5	29	Kadın	Avustralya	Yok	Hayatta	Yok	29.02.2020	Yok	Şubat	Hayatta	0,82
6	79	Erkek	Avustralya	Yok		Var	1.03.2020	Yok	Mart	Hayatta	0,68

Tablo 15’te görüldüğü gibi yeni girilen hastanın hayatta kalıp kalamayacağı tahminine göre Markov modeli, 79 yaşında, Avustralya’da yaşayan, kronik rahatsızlığı olmayan, son üç ayda seyahat geçmişi olan ve mart ayında ilk hastalığa yakalanmış olan erkek bir hastanın %68,3 olasılıkla hayatta kalacağı olasılığını vermektedir.

Sonuç ve Tartışma

Bu çalışmanın amacı, diğer ülkeler için elde edilmiş verilerden yararlanarak bir ön çalışma yaparak Türkiye verisi içinde bir hazırlık yapmak ve altyapı oluşturmaktır. Bu çalışmada Bayes Ağları ile bazı değişkenler kullanılarak hastaların hayatta kalıp kalmayacakları tahmin edilmiştir. Türkiye verisi elde edildiğinde Bayes Ağları kullanılarak Covid-19 geçiren hastalar için hayatta kalma ile ilgili bir risk hesaplaması yapıldığında, bu hem sağlık ile ilgili plan ve programların oluşturulması; ölüm riski yüksek hastalar için ek tedbirler elde edilerek bu hastaların ölüm oranlarının düşürülmesi mümkün olacaktır. Salgının başından beri 12 Mayıs 2022 itibarı ile Türkiye’de Covid-19’dan 98.870 kişinin vefat ettiği düşünülürse bu tür modellerin önemi daha da ortaya çıkmaktadır. Koronavirüse karşı yapılacak analizler ve veri izlemeye ilişkin, salgınla etkin mücadele için ileriye dönük bir planla veriler hayata geçirilmeli, salgının zamanı, seyri ve coğrafi/idari ilgililer açısından izlenmesi gerekmektedir. Çözümleme sürecinin etkin bir şekilde ilerleyebilmesi ve konu ile ilgili eksik bilgi birikimlerinin giderilmesi açısından veri kümesinin tüm alanlardan uzman kişilerle paylaşılması çözümleme sürecine katkı sağlayacaktır. Koronavirüsün ekonomiye olası etkilerine bakılırsa, küresel pandemiden kaynaklı kısa dönem artan ekonomik faaliyetlerin %12 azalacağını düşünürsek, bir ayda neredeyse 31,76 milyar TL’lik bir gayri safi yurt içi hâsıla (GSYH) eksiği oluşabilir. Başka bir kısa dönem elektrik kullanımındaki düşüşü temel alan bir çalışmaya göre, GSMH’de %3,20’lik bir azalma meydana gelebilir. Salgının uzun dönemli izleri göz önünde bulundurulduğunda, ekonomik kayıp tahminlere göre çok daha fazla çıkabilir. COVID 19’dan kaynaklı olarak büyüme oranında düşüş, işsizlik oranlarında artış, enflasyon oranında artış, dış finansmanda zorluklar ve bütçe ve ödemeler dengesinde bozulma gerçekleşebilecektir. Salgının küresel çapta uzun dönem etkileri dikkate alındığında bu tür kötü etkileri teğet geçebilmek adına ekonomiye ihracatlar yoluyla gelir aktarımı yapılmalı ve parasal kolaylaştırmaya gidilmelidir. Borç ertelemeleri, doğrudan fonlama, vergi/kredi kolaylıkları gibi tedbirlerin iyice artırılması gerektiği görülmektedir. Şehir hastaneleri, tecrübeli ve özverili sağlık çalışanları, tarım/gıda ve imalat sektörü, organize sağlık idaresi gibi önemli kuruluşların ve kişilerin ülkemizde yaşanan Covid-19 ile savaşta ayrıca desteklenmelidir. Çalışılan veri kümesi Covid-19’un farklı yaş aralıklarını farklı etkilemekte olduğunu göstermektedir. Bu saptamadan yola çıkarak halkı yaşa göre sınıflandırarak sağlık politikaları hem maliyetleri azaltacak hem de tıbbi müdahalenin daha doğru uygulamasını sağlayacaktır.

Bu çalışmaya ek olarak, daha çok veriye ulaşıp çalışma detaylandırılabilir. Türkiye verisine ulaşp daha çok değişken ekleyerek daha verimli sonuçlara ulaşılabilir. Bu çalışmayı herkes tarafından kolayca ulaşılabilir

bir uygulama haline dönüştürüp insanların olası bir hastalık durumunda riski ölçülebilir. Koronavirüs hayatımızda ne kadar zaman daha kalacağı belirli değilken bu çalışmada hastalıkla karşılaşan bireylerin hastalık sonucu hayatta kalıp kalamayacağı ile ilgili Türkiye verileri ile çalışılabilir. Bayes Ağlarının sağlık alanında kullanımı yaygındır. Bir kişide ortaya çıkan semptomlara bakılarak bu kişinin herhangi bir hastalığa sahip olma olasılığı hesaplanabilir. Benzer şekilde aralarında neden-sonuç ilişkisi olan birçok olayın olasılığı bu modelleme ile ortaya çıkarılabilir. Covid-19 ile ilgili verilere de Bayes Ağları uygulaması yapılmaktadır. Bayes Ağları ile riskli hastalar belirlenir ve tanı testleri ilk etapta riskli hastalar üzerinde yapılabilir. Buradan elde edilen model sonuçları ile kişi kendi verilerini girerek olası bir hastalık durumunda kendinin hayatta kalıp kalamayacağını hesaplayabilir ve bu sayede ek önlemler alabilir.

Bu çalışmada “https://github.com/beoutbreakprepared/nCoV2019/blob-/master/latest_data/latestdata.tar.gz” adresinden elde edilen 19 ülkeye ait 10.309.980 Covid-19 hastası içinden seçilen 215.968 hastadan oluşan 9 değişkenli veri kümesi kullanılmıştır. Bağımlı değişken ise hastanın koronavirüs’e yakalandıktan sonra hayatta kalıp kalmadığı alınmıştır. Ek olarak hastanın kronik rahatsızlığı olup olmaması, son üç ay içinde yurtdışı seyahatinde bulunup bulunmadığı, cinsiyeti ve hangi ayda hastalığa yakalandığı bilgileri de kullanılmıştır. Bu değişkenlere için Markov ve Tan modelleri uygulanmış; Tan Modeli test verilerinin %98,97’sini doğru sınıflarken, Markov Modeli test verininin %99,04’ünü doğru sınıflamıştır. Buradan elde edilen sonuçlar ile sisteme yeni giren bir hastanın hayatta kalma olasılıkları hesaplanmıştır.

Türkiye için benzer veriler elde edildiği takdirde bu gibi çalışmalarla gelecekte meydana gelebilecek pandemilerde de ölüm oranlarında ve artan vaka sayılarında önemli bir azalma gözlemlenebilir ve buna ilişkin önlemler alınabilir. Gelecek çalışmalarda Bayes Ağları yönteminin bu veri kümesi üzerinden diğer Makine Öğrenmesi algoritmaları ile karşılaştırılması yapılabilir. Türkiye verileri yayınlandığı takdirde de uygulaması yapılarak elde edilen sonuçlar karar vericiler için yararlı bilgi niteliğinde olacaktır (Yılmaz, 2022).

Yazar Katkıları

Hüseyin Can Yılmaz: Konunun saptanması, Çalışmada kullanılan verileri toplama, analiz etme ve sonuçları yorumlama.

Serpil Aktaş Altunay: Çalışmanın planlanması, Analiz sonuçlarının yorumlanması ve makale yazımı.

Çıkar Çatışması

Yazarlar çıkar çatışması bildirmemişlerdir.

Kaynaklar

- Adiga, A., Dubhashi, D., Lewis, B. et al (2020). *Mathematical Models for COVID-19 Pandemic: A Comparative Analysis*. J Indian Inst Sci 100, 793–807. Doi:<https://doi.org/10.1007/s41745-020-00200-6>
- Atılğan K.Y. & Ersel D., (2017). *Bayesci ağ yapısının öğrenilmesinde grafiksel bir yaklaşım*, Journal of Statisticians: Statistics and Actuarial Sciences, 1, 1-10.
- Chen, J. M. (2022). *Novel statistics predict the COVID-19 pandemic could terminate in 2022*. Journal of Medical Virology, 94(6), 2845-2848.
- Heckerman, D. (2008). A Tutorial on learning with Bayesian networks. Innovations in Bayesian networks, 33-82. Doi:https://doi.org/10.1007/978-3-540-85066-3_3
- Gemela J. (2001). *Financial analysis using Bayesian Networks*, Applied Stochastic Models in Business and Industry, 17(1), 57-67. Doi:<https://doi.org/10.1002/asmb.422>
- Hui D.S. Azhar E.I., Madani, T.A., Ntoumi F., Kock R. Dar O. et al (2020). *The continuing COVID-19 epidemic threat of novel coronaviruses to global health — The latest 2019 novel coronavirus outbreak in Wuhan, China*. International Journal of Infectious Diseases, 91; 264-266. Doi:<https://doi.org/10.1016/j.ijid.2020.01.009>
- IBM Corp (2016). *IBM SPSS Statistics for Windows*, Version 23.0. Armonk, New York, U.S.A. Jensen, V. & Nielsen T.D. (2007). Bayesian networks and decision graphs. Vol. 2. New York: Springer. Doi:https://doi.org/10.1007/978-3-540-85066-3_3
- Karacan, H. & Eryılmaz, F. (2021). *Covid-19 Detection from Chest X-Ray Images and Hybrid Model*

- Recommendation with Convolutional Neural Networks* . Journal of Advanced Research in Natural and Applied Sciences , 7 (4) , 486-503. Doi:<https://doi.org/10.28979/jarnas.952700>
- Korb, K.B.& Nicholson, A. E. (2011). *Bayesian artificial intelligence* (Computer science and data analysis series) CRC press. Doi:<https://doi.org/10.1201/9780203491294>
- Lauria, E. J. M. & Duchessi, P. J.(2006). *A Bayesian Belief Network for IT Implementation Decision Support, Decision Support Systems*, 42(3), 738-742, Doi:<https://doi.org/10.1016/j.dss.2006.01.003>
- Masum, M., Masud, M. A., Adnan, M. I., Shahriar, H., & Kim, S. (2022). *Comparative study of a mathematical epidemic model, statistical modeling, and deep learning for COVID-19 forecasting and management*. SocioEconomic Planning Sciences, 80, 101249.
- Maleki, M., Mahmoudi, M. R., Wraith, D., & Pho, K. H. (2020). *Time series modelling to forecast the confirmed and recovered cases of COVID-19*. Travel medicine and infectious disease, 37, 101742.
- Olmuş, H., & Erbaş, S. (2012). *Bayes Ağlarda Kümeleme Metotunu Kullanarak Meme Kanseri Tanısının Modellenmesi*. Türkiye Klinikleri Journal of Biostatistics, 4(1),10-19.
- Pearl, J. & Verma, T.S. (1990) *Equivalence and Synthesis of Causal Models*. Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence, Cambridge, 27-29 July, 220-227.
- Perone, G. (2020) *An ARIMA model to forecast the spread and the final size of Covid-2019 epidemic in Italy (first version on SSRN 31 march)*. SSRN Electron J. Doi:<https://doi.org/10.2139/ssrn.3564865>
- Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L. & Cowell R. G. (1993). *Bayesian analysis in expert systems*, Statistical Science, 8(3), 219-247. Doi:<https://doi.org/10.1214/ss/1177010888>
- Scutari, M. (2011). *Measures of Variability for Graphical Models*. (Yayımlanmamış doktora tezi), Università degli Studi di Padova, Dipartimento di Scienze Statistiche.
- Şencan, Ş. N. , Şencan, B. , Borazan Çelikbıçak, M. , Arslan, D. , Özkan, E. S. , Gökçen, A. Ş. , Çiftçi, R. B. , Arıkan, İ. , Uğur, B. , Şahin, H. , Coşkun, A. E. , Konşuk Ünlü, H. & Aktaş, S. (2020). *Lojistik Büyüme ve Üstel Büyüme Modelleri ile Türkiye 'de Covid-19 Modellemesi* . Nicel Bilimler Dergisi , 2 (1) , 1-18 . <https://dergipark.org.tr/tr/pub/nicel/issue/55149/748068>
- Tang, S., Wang, C., Nie, J., Kumar, N., Zhang, Y., Xiong, Z. & Barnawi, A. (2021). *EDL-COVID: Ensemble Deep Learning for COVID-19 Case Detection From Chest X-Ray Images*. *IEEE Transactions on Industrial Informatics*, 17, 6539-6549. doi: 10.1109/TII.2021.3057683
- Wendler, T. & Sören G. (2016). *Data Mining with SPSS Modeler Theory, Exercises and Solutions*, Cham Springer International Publishing, Springer. Doi:<https://doi.org/10.1007/978-3-319-28709-6>
- Yılmaz, H. C. (2022). *Covid-19 Verileri İçin Bayes Ağları İle Makine Öğrenmesi*, (Yayımlanmamış yüksek lisans tezi), Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, Türkiye.
- Zhou Y, Yang Y, Huang J, Jiang S. & Du L (2019). *Advances in MERS-CoV Vaccines and Therapeutics Based on the Receptor-Binding Domain*. *Viruses*. 11(1), 60. <https://doi.org/10.3390/v11010060>
- Zeroual A., Harrou F., Dairi A. & Sun Y. (2020). *Deep Learning Methods for Forecasting Covid-19 Time-Series Data: A comparative study*. *Chaos, Solit Fractals*, 140:110121. Doi:<https://doi.org/10.1016/j.chaos.2020.110121>