

**COVID-19 VERİLERİ İÇİN
BAYES AĞLARI İLE MAKİNE ÖĞRENMESİ**

**MACHINE LEARNING WITH BAYESIAN NETWORKS
FOR COVID-19 DATA**

HÜSEYİN CAN YILMAZ

PROF. DR SERPİL AKTAŞ ALTUNAY

Tez Danışmanı

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

İstatistik Anabilim Dalı için Öngördüğü

YÜKSEK LİSANS TEZİ olarak hazırlanmıştır

2022

ÖZET

Covid-19 Verileri İçin Bayes Ağları İle Makine Öğrenmesi

Hüseyin Can YILMAZ

Tez Danışmanı:

Prof. Dr. Serpil AKTAŞ ALTUNAY

Mayıs 2022, 51 sayfa

Covid-19 pandemisi, 17 Kasım 2019 tarihinde Çin'in Vuhan Eyaleti'nde ilk defa görülmüştür. Küresel pandemi ilk başta Vuhan'daki deniz mahsülleri ve hayvan satışı yapılan yerlerde görülmüştür. Sonra insanlar arasında da yayılışını devam ettirerek Vuhan başta olmak üzere Çin'in diğer eyaletindeki bölgelere ve diğer dünya ülkelerine yayılmıştır. 11 Mayıs 2022 tarihi itibariyle dünyada 519.682.000 vaka meydana gelmiştir ve 6.266.278 hasta ölmüştür.

Ülkemizde ve dünya genelinde Covid-19 pandemisinin etkilerini gösteren birçok araştırma ve analiz çalışmaları yapılmıştır. Bu tez çalışmasında 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 çalışılmıştır.

Bu çalışmada, 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: Covid-19, Bayes Ağları, Makine Öğrenmesi, Pandemi

ABSTRACT

Machine Learning with Bayesian Networks for Covid-19 Data

Hüseyin Can YILMAZ

Supervisor:

Prof. Dr. Serpil AKTAŞ ALTUNAY

May 2022, 51 pages

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 May 11, 2022, 519.682.000 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 thesis study, 215,968 worldwide cases from 104 countries around the world were analyzed and the patients were tried to be classified using Bayesian Networks and machine learning techniques.

In this study, 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: Covid-19, Bayesian Networks, Machine Learning, Pandemic

TEŞEKKÜR

Tez çalışmam boyunca her türlü destekçim ve yol göstericim, pozitif tavırlarıyla öğrencilerini daima güçlü hissettiren, bilgi birikimiyle tezime farklı yönlerden bakmamı sağlayan daima öğrencisi olmaktan gurur duyduğum değerli akademik olarak yol göstericim Prof. Dr. Serpil AKTAŞ ALTUNAY'a ve engin bilgilerini benimle paylaşan saygı değer hocam Doç. Dr. Derya ERSEL'e ve her zaman yanımda olan, beni her daim destekleyen sevgili eşime sonsuz teşekkürlerimi sunarım.



İÇİNDEKİLER

ÖZET	i
ABSTRACT.....	ii
TEŞEKKÜRLER.....	iii
İÇİNDEKİLER.....	iv
ÇİZELGELER.....	v
ŞEKİLLER.....	vi
1. GİRİŞ.....	1
2. BAYES AĞLARI	8
2.1. Grafikselsel Modeller ve Döngü.....	9
2.2. Bayes Ağları'nın Matematiksel Gösterimi.....	16
2.3. Bayes Ağları'nın Koşullu Olasılık Gösterimi	19
2.4. Bağımsız Olaylar	22
2.5. Zincir Kuralı.....	22
2.5.1. Markov Özelliği.....	23
2.5.2. Tan Özelliği.....	23
2.6. Bayes Ağları'nda Öğrenme	24
2.6.1. Olasılık Temelli Çıkarımlar.....	24
2.6.2. Bayes Ağları'nda Öğrenme Algoritmaları.....	25
2.7. Bayes Ağları Sınıflandırıcıları	26
3. COVID-19 VERİLERİNİN ÇÖZÜMLENMESİ	28
3.1. Veri Kümesinin Tanımı.....	28
3.2. Betimleyici İstatistikler.....	30
3.3. COVID-19 Verileri İçin Bayes Ağları Çözümlemesi.....	31
4. SONUÇ VE TARTIŞMA.....	42
KAYNAKLAR.....	45
ÖZGEÇMİŞ.....	51

ÇİZELGELER

Sayfa

Çizelge 2. 1. Aile ebeveyn, çocuk ve torun dışı çizelgesi	17
Çizelge 3. 1. Veri kümesine genel bakış	29
Çizelge 3. 2. Verinin içerdği değişkenler ve tanımları	29
Çizelge 3. 3. Tan modeli ve olasılıkları	32
Çizelge 3. 4. Markov modeli ve olasılıkları	32
Çizelge 3. 5. Hastalığa hangi ayda yakalandığına ait koşullu olasılık tablosu.....	34
Çizelge 3. 6. Yaşa ait koşullu olasılık tablosu.....	35
Çizelge 3. 7. Hastanın cinsiyete ait koşullu olasılık tablosu	35
Çizelge 3. 8. Hastanın seyahat geçmişine ait koşullu olasılık tablosu	36
Çizelge 3. 9. Sonuç değişkenine ait koşullu olasılık tablosu	37
Çizelge 3. 10. Tan modeline göre verilerin sınıflandırılması	38
Çizelge 3. 11. Markov modeline göre verilerin sınıflandırılması.....	39
Çizelge 3. 12. Test ve deneme modeline göre verilerin sınıflandırılması	40
Çizelge 3. 13. Yeni hastanın özelliklerine göre Markov modelinde sınıflanması	41

ŞEKİLLER

Sayfa

Şekil 1. 1. Koronavirüsün şematik yapısı.....	3
Şekil 2. 1. Bir Bayes Ağı örneği.....	10
Şekil 2. 2. Bayes Ağı olmayan döngüsel bir ağ.....	11
Şekil 2. 3. Dört düğümden oluşan yön verilmiş ve yön verilmemiş grafik örnekleri	11
Şekil 2. 4. Yakınsak bağlantı, seri bağlantı ve ayrık bağlantı için grafiksel gösterim.....	12
Şekil 2. 5. Bir Bayes Ağı'nda serisel, ıraksayan ve yakınsayan bağlantılar	13
Şekil 2. 6. DAG örneği	14
Şekil 2. 7. Döngüsel olmayan bir Bayes Ağı ve döngüsel Ağ örneği.....	15
Şekil 2. 8. Aile ebeveyn, çocuk ve torun dışı düğümler için verilen Bayes Ağı örneği	17
Şekil 2. 9. Döngüsel bir ağ örneği ve olasılıkları.....	18
Şekil 2. 10. EF= Taralı alan, EFC = Noktalı alan gösterimi.....	20
Şekil 2. 11. Tipik bir TAN yapısı ağı.....	24
Şekil 3. 1. Değişkenlere ait betimleyici istatistikler.....	30
Şekil 3. 2. Hastaların yaş dağılımı.....	31
Şekil 3. 3. IBM SPSS Modeler programı ile Bayes Ağları kesiti.....	31
Şekil 3. 4. Tan ve Markov modelleri için karşılaştırma oranları	33
Şekil 3. 5. Markov modeli için değişkenlerin anlamlılık düzeyleri	34
Şekil 3. 6. Test ve deneme modelleri için karşılaştırma oranları	41

1. GİRİŞ

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 Pazarı'ndan (Huanan Seafood and Live Animal Market) ortaya çıktığı düşünülen bir hastalık kısa sürede salgına dönüştü. Bugüne kadar yapılan çalışmalarda hastalığı geçiren bireylerin büyük çoğunluğunun bu hayvan ticarethanelerini ziyaret ettikleri ve burası ile teması olduğu gözlemlendi. Bu ticarethanede yaras, kümes hayvanları, fare ve birçok çeşit yılan türü gibi hayvanların satıldığı bilinmektedir. İlk zamanlarda salgının görüldüğü hastalardan yedi tanesinin durumu ciddiye, bir hasta yaşamını yitirmişti. Yaklaşık dokuz gün içerisinde doktorlar, 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 bunun da yeni Coronavirüs, Covid-19 şeklinde tanımlandığını duyurdular.

Koronavirüs aileleri (CoV), toplumda sıklıkla rastlanan soğuk algınlığı gibi, kendini kısıtlayan güçsüz enfeksiyondan, Orta Doğu Solunum Sendromu (Middle East Respiratory Syndrome, MERS) ve Ağır Akut Solunum Sendromu (Severe Acute Respiratory Syndrome, SARS) gibi daha güçlü enfeksiyonlara neden olabilen büyük ve bilindik bir virüs ailesidir. Koronavirüslerin insanlarda bulunan ve insanlar arasında oldukça bulaşıcı çeşitli türleri (HCoV-229E, HCoV-OC43, HCoV-NL63 ve HKU1-CoV) bu güne kadar gözlemlenmiştir.

Bu virüsler soğuk algınlığına sebep olan virüslerdir. Ayrıca hayvanlarda görülen birçok koronavirüs çeşidi vardır ve bazılarının hayvanlardan insanlara geçerek insanlarda ciddi hastalıklara neden olabilmektedir. Kapsamlı araştırmalar, SARSCoV'un misk türü kedilerden ve MERS-CoV'un bir hörgüçlü deve türlerinden insanlara bulaştığını göstermiştir[1].

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.

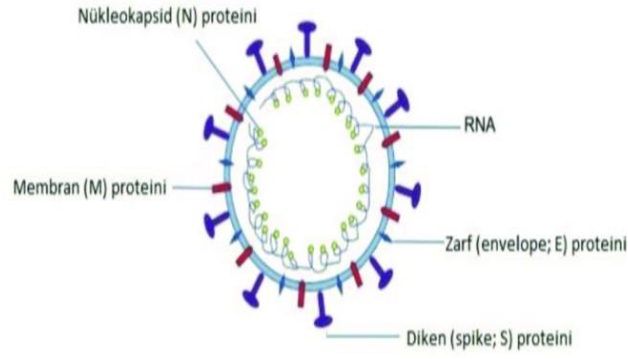
Dünya Sağlık Örgütü (DSÖ) 31 Aralık 2019’da, Çin’in Hubei bölgesindeki Wuhan kentinde etiyolojisi bilinmeyen pnömoni vakalarını kayıt altına almıştır. 7 Ocak 2020’de vaka öncesinde insanlarda gözlemlenmemiş olan yeni bir koronavirüs (2019- nCoV) olarak kayıt altına alınmıştır. Sonrasında 2019-nCoV salgının 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’ta “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.

Covid-19’un ilk ortaya çıktığı anda vaka sayısı özelinde bir çok yorum yapıldı. Bunlardan birine göre hasta sayısı kırk iki [2], diğer bir kaynağa göre hasta sayısı kırk bir [3] şeklindedir. Bu hastalardan Wuhan’dan Tayland’a gidenler hariç, neredeyse tamamının Çin vatandaşı olduğu gözlemlenmiş ve salgının bu kadar hızlı yayılmasına bu kişilerin sebep olabileceği belirtilmiştir [4][5][6].

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 savaşırsınız. 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 [7].

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. [8]



Şekil 1. 1. Koronavirüsün şematik yapısı

Koronavirüsün bulaşıcılığı hakkında, semptomu olmayan vakalar olabilirken, semptomsuz vakaların oranı tam olarak tespit edilememektedir. En yaygın semptom yüksek ateş, kuru öksürük ve nefeste daralmalardır. Ağır geçirilen vakalarda zatürre, solunum ve böbrek yetmezliği ve nihayetinde ölüm bile gelişebilmektedir[9].

Covid-19, hastaların öksürük yolu ile havaya yayılan damlacıkların solunması yolu ile bulaşabilmektedir. Hastaların solunum partikülleri ile pis yüzeylere dokunduktan sonra ellerini yıkamadan yüz, göz, burun veya ağız gibi açık yüzeylere götürülmesi de Covid-19 virüsünün yayılmasını neden olabilmektedir. Elleri iyi bir şekilde yıkamadan göz, burun veya ağıza götürmek Covid-19 bulaşma riskini artırmaktadır.

Covid-19 pandemisi ile ilgili bugüne kadar toplanan bilgiler ışığında, kimi insanların hastalıklarını ağır geçirdikleri ve ciddi belirtiler geliştirme riski altında olduğunu görülmüştür.

Türkiye'deki vakalara bakıldığında, hastaların yüzde 82'si hastalığı rahat geçirmekte, hastaların %20'si hastanede tedavisi yapılmaktadır. Covid-19'un genellikle 60 yaş ve üzerindeki kişiler üzerinde etkisi daha büyük olduğu bilinmektedir.

Covid-19 genellikle 60 yaşın üstündeki kalp hastalığı olan kişileri, diyabet hastalığı olan kişileri, hipertansiyon hastalarını, kanser hastalığı olan kişileri ve kronik solunum yolu hastalığı olan kişileri daha çok etiler.

Yeni Koronavirüs tanısı için gerekli testler ülkemizde mevcuttur.

Covid-19'den korunmak için pandeminin başında bazı önlemler alınması gerekmiş ve bunlar halka anlatılmıştır. Örneğin eller minimum 20 saniye süresince su ve sabun ile iyice yıkanmalı, su ve sabunun mevcut olmadığı ortamlarda ise alkol bazlı antiseptik solüsyon kullanılmalıdır gibi önlemler vurgulanmıştır. Antiseptik veya

antibakteriyel barındıran sabun bulundurmaya ihtiyaç yoktur, evlerde kullanılan normal sabun tedbir almak için uygundur. Buruna, gözlere ve ağıza elleri yıkadıktan sonra dokunulmalıdır. Semptom belirtisi gösteren kişilerle temas edilmemelidir. İnsanlarla olan sosyal mesafe korunmalı eğer mümkünse 1.5 metre aralıklarla iletişime geçilmelidir. Semptom gösteren kişilerle ya da bu kişilerin çevrelerindeki kişilerle direkt temas edilmiş ise eller bol su ve sabun ile çok iyi temizlenmelidir.

Sağlık merkezlerine zorunda olmadıkça gidilmemelidir çünkü semptom gösteren kişiler çoğunlukla sağlık merkezlerini kullanmaktadır. Eğer zorunda kalınıp gidildi ise semptom gösteren kişilerle temastan olabildiğince kaçınılmalıdır. Eğer semptom gösteren kişi sizseniz hapşıрма ve öksürme yoluyla damlacıkların havaya karışmaması için kağıt mendil kullanılmalıdır. Yoğun ortamlara girmek zorunda kalmanız durumunda ise burun ve ağız tıbbi maske ile kapatılmalıdır.

Et ürünleri yerken iyi pişmiş bir şekilde yenmelidir. Enfeksiyon kapma ihtimali yüksek olan canlı hayvan pazarı, çiftlikler, mezbaha gibi yerlerden uzak durulmalıdır.

Eğer seyahat edilecekse maske, mesafe ve el temizliği tedbirleri unutulmamalıdır. Seyahat sonrasında kişi kendini 14 gün izlemeli ve bir semptomu varsa sağlık merkezlerine hemen başvurmalıdır [10].

Son zamanda tecrübe edilmiş olan Ebola, MERS ve SARS gibi ciddi geçirilen salgınlardan sonra insanlar artık böyle salgınlarla mücadele etmek adına bu virüslerin yapısını anlamak, nerede ve nasıl ortaya çıkacağını tahmin etmek böyle salgınlarla mücadele etmek adına oldukça önemlidir. Bu kapsamda devletler, sivil toplum kuruluşları, akademisyenler ve alanında uzman kişiler bu salgınların seyrini değiştirmek amacı ile araştırmalar yapmaya başlamıştır.

Bu çalışmalara, Sierra Leone bölgesindeki Ebola salgını için salgınını izlemek amacı ile oluşturduğu ve bölgesel olarak IBM tarafından hayata geçirilen bu çalışma örnek teşkil etmektedir [16]. Çalışma sayesinde kişiler gördükleri ve tespit ettikleri hasta kişileri hemen iletebileceği ve salgının seyrini değiştirebilecek hızlı bir müdahale sistem geliştirilmiştir.

Bu kapsamda Bayes Ağları birçok modelin sınıflandırılmasında işimize yaramaktadır, Bayes Ağları belirsiz ilişkileri ortaya çıkarmak üzere matematiksel bir model ortaya koymaktadır. Bayes Ağları değişkenler arası ilişkileri ortaya çıkarmak adına diyagramsal modelleri kullanmaktadır.

Bayes Ağları son zamanlarda psikoloji, biyoloji, sosyal bilimler tıp vb. birçok alanda kullanılmaktadır. En yaygın kullanım alanlarından biride risk tahminleridir [51].

Ha ve Seong (2004) araştırmalarında meydana getirdikleri modelde koşullu olasılık yorumlamalarının uzman bilgisini gerektirdiğini dile getirerek, Bayes Ağları'nın etkin uygulanması için koşullu olasılıkların uzmanlara danışılması gerektiğini söylemiştir [31]. Diğer bir araştırmada, uzmanlar tarafından oluşturulmuş bir ekip Bayes Ağı modeli için hayati önemi sahip olan koşullu olasılık değerlerini ekip içinde tartışarak özen ile doldurmuşlardır [32]. Bayes Ağları'nın modellenmesinde düğümlerin oluşturulması ve koşullu olasılık değişkenlerinin yorumlanması açısından uzman görüşleri oldukça önem arz etmektedir.

Bayes Ağları kompleks sistemler veya süreçlerde meydana gelen olayların anlamlı çözümlere ulaşması için çoğunlukla uygulanır. Karışık sistemlerde bu konularda yetkin olan uzman bilgisine ihtiyaç duymaktadır [30].

Uzmanlar, De Melo ve Sanchez'in [34] açıklamalarına göre, tüm değişkenlerin olasılık dağılımına ek olarak, tüm koşullu dağılımların diğer koşullu dağılımlarla etkileşimini de değerlendirmelidir. Böylece koşullu dağılımları çaprazlayarak, bir olasılık dağılım matrisini meydana getirebilir fakat uzmanlarca doldurulan koşullu olasılık matrisi bu kişilere ekstra sorumluluk da getirmektedir [35].

Bayes Ağlarının avantajlarından biri de verinin olmadığı aşamalarda, veri yerine uzman görüşlerinin kullanılabilmesine olanak sağlamaktadır [36]. Verilerin elde edilebildiği aşamalarda ise Bayes Ağlarının avantajı, modelde objektif bir biçimde yapılan öngörülerin modelde daha sonradan yeniden değiştirilebilmesidir [33].

Bayes Ağlarının, bileşik olasılık dağılımlarına bağlı olarak elde edilen diğer modellere göre daha etkili olduğu değerlendirilmektedir. Yani Bayes Ağları, direkt olarak çarpım kuralının yapılabilmesi sayesinde, koşullu olasılık etkenlerinin üstel artışına neden olmaksızın daha fazla değişkeni içerebilmektedir [49].

Bayes Ağlarının bir diğer avantajına bakılacak olursa, koşulu olasılıkların uzman birikimi ve bilgisi ile elde edilen veriler ile geçmişe dayalı önsel bilgilerden veya önceki çalışmalardan elde edilen bilgileri sentezleyebiliyor olmasıdır [50]. Bayes Ağları, değişkenler arasındaki olasılıksal süreç ilişkileri temsil eden grafiklere dayalı modeldir [44]. Bayes Ağlarının oluşturulmasında grafik teorisinden yararlanılır. Değişkenler hakkındaki olasılık bilgilerinin grafiksel olarak gösterimini sağlamak, değişkenlerin ortak olasılık fonksiyonunu görsel olarak incelemek, gözlemlenen

verilerden en etkili sonuçları çıkarmak grafiksel modellerin en önemli özelliklerindendir [37].

Grafiksel modeller ile verilerin analiz edilmesinin bazı avantajları vardır. Bunların ilki modelin tüm değişkenlerden yararlanması dolayısıyla bazı verilerin eksik olduğu durumlarda da kolaylıkla kullanılmasıdır [38]. İkincisi, nedensel ilişkileri ortaya çıkarmayı, bu ilişkilere dayalı olarak problemi daha iyi anlamayı ve değişkenlerin etkileşiminin sonuçlarını tahmin etmeyi sağlar [39]. Üçüncüsü, veriden elde edilen bilgi ile değişkenler hakkındaki önsel bilginin Bayes yaklaşımı ile birleştirilerek gösterilmesini sağladığından güçlü bir yöntemdir [46]. Yani gerçek bilgi ile öznel bilginin birlikte kullanımına olanak sağlar [40]. Dördüncüsü, Bayesci yöntemleri Bayes Ağları ile birlikte kullandığında modellemeye etkili ve prensipli bir yaklaşım sunarak verilerin modele aşırı uyumunun önüne geçer [47]. Yani modeli düzgünleştirmek için uygun olan bazı verileri ayırıp onları kullanmak yerine tüm uygun verileri kullanır[62].

İnsan davranışlarının belirsizliğini diğer yöntemlerle modellemek neredeyse imkânsızdır. Oysa Bayes Ağları ile X özelliği müşterinin bu özelliğin kullanışlılık ve yararlılığına olan inancının derecesi olarak alınıp bu derecenin olasılığı modelde gösterilebilir [48].

Bayes Ağlarında bir değişkenin olasılık fonksiyonu ağda bu değişkeni temsil eden düğümün ebeveyn düğümlerine bağlıdır[41]. Bu durum çok sayıda değişkenin bulunduğu ağlarda önemli bir avantaj sağlar [63]. Bayes Ağlarının nedensel semantik yapısı nedensel önsellerin daha basit anlaşılmasına imkân sağlamaktadır. Bu özellik veriler karmaşık olduğunda kullanıcının ön bilgisini kullanarak BN oluşturmaya yardımcı olmaktadır [42].

Bayes Ağları kesin olmayan karmaşık bilgiler içeren problemlerin modellenmesinde oldukça kullanışlı bir yöntemdir [64]. Modelde verilen düğümlerin koşullu olasılıklarının ortaya çıkarılmasında birçok farklı deneysel ve istatistiksel teknik mevcuttur [65]. Bununla birlikte, incelenen problem için model tutarlılığını garanti eden Bayes Ağı modeli oluşturmak ve geliştirmek için özel bir yöntem veya kılavuz yoktur. Bunun için araştırmacılar, geliştirilen modellerin veriler üzerinde

sağlamlıklarını tespit etmek ve duyarlılık analizini doğrulamak için bazı yöntemler geliştirmişlerdir. Model geçerliliği, belirtilen yöntemlerin bir parçasıdır [66]. Bayes Ağlarının avantajlarından en iyisi kanıt yayılımıdır. Bahsedilen bu karakteristik özellik ile her bir düğümün olasılıklarını bir düğümden iki yönlü olarak yeni bir bilgi ile bütün yapı boyunca güncellemeye izin verir [67].

Bu çalışmanın amacı, Covid-19'a yakalanmış hastalar için bazı değişkenler yardımı ile bir sınıflama oluşturmak ve risk faktörlerine göre hastaların hayatta kalma ve ölüm risklerini Bayes Ağları ile hesaplamaktır. Bu çalışma aynı zamanda Türkiye için veriler yayınlandığında aynı modelin Türkiye verilerine uygulanması için bir altyapı oluşturmayı hedeflemektedir.

Bu çalışmada, koronavirüse yakalanan hastaların yaş, cinsiyet, ülkesi, kronik rahatsızlığı, seyahat geçmişi, semptomu ve hastalığa yakalandığı tarihi gibi bağımsız değişkenleri kullanarak hangi değişkenlerin koronavirüs sonucunda hastanın hayatta kalıp kalamayacağını en çok etkileyeceğini tespit etmek ve Bayes Ağları kullanarak bu hastanın hayatta kalıp kalamayacağını sınıflandırmaktır.

Bayes Ağları yönteminin Markov modeli kullanarak yapılan bu sınıflama hastaları %99,16 doğru sınıflandırmıştır. Yani modele yeni girilen ve hastalıktan sonra hayatta kalıp kalmayacağı bilinmeyen bir kişinin sonucunu çok yüksek bir olasılıkla doğru tahmin edecektir.

Bu sonuçlar göz önünde bulundurularak koronavirüs pandemisinin hangi kişilerde daha fazla ölüm riski içerdiğini tespit edip bu kişiler için ayrıca önlemler alınabilmesi gereklidir.

2. BAYES AĞLARI

Veri madenciliği, Bayes Ağları denilince akla gelen ilk şeylerden biri olur. Sadece ham bir veriyi kullanarak, amaçladığımız sürece etki edecek kararları almak mümkün değildir. Sürece yön verecek etkili kararlar almak isteniliyorsa ham veriyi işleyerek bilgi elde edilmelidir. Ham veriden bilgiye ulaşma sürecine ise veri madenciliği denir. Veri madenciliği adımlara sahip olan süreçler sistemidir. Bu sistem bazen bilgisayar ortamında otomatik bir biçimde bazen de bir uzman tarafından çeşitli algoritmalar kullanılarak yarı otomatik bir şekilde uygulanır.

Veri madenciliği, zor olan ve önceden keşfedilmemiş örüntü ve dizilerin ortaya çıkarılması[15] ve yararlı olan bilgileri süzüp veri kümesinden çıkarılması işlemidir [12].

Aynı zamanda veri madenciliği, büyük veriden (Big data) ileriye tahmin etmek için önemli ve faydalı bilgi ve verilerin çeşitli yazılım programları ile ortaya çıkarılması ve analiz edilmesi işlemidir[13]. Ayrıca veri madenciliği, büyük veri içindeki değişkenlerin ilişkileri araştırarak aralarındaki bağlantıyı keşfetmeye çalışan teknikler bütünüdür[11].

Bu süreç, bilgiyi getirdiği gibi zaman, emek ve maliyet de getirmektedir. Büyük veri incelendiğinde, tüm veriden bilgiyi sentezlemek ana amaç olsa da yapılan araştırmalar ve eski çalışmalar tüm verinin sadece %20-30'undan anlamlı sonuçlar ve bilgi elde edildiğini göstermektedir [43]. Kalan verinin %70-80'i zaman, emek ve maliyet açılarından bakıldığında bazıları için riskli bir durum oluşturabilir çünkü verinin bu kısmından anlamlı bir bilgiye ulaşılmamaktadır[14].

Veri madenciliğinin veri özetleme, veriden farklı örüntüler çıkarma ve tahmin amaçları vardır fakat analize uygun model bilinmediğinden veri madenciliğinde araştırmanın amacı doğru modeli keşfetmektir [45].

Veri madenciliğinin önemli sınıflandırma modellerinden biri ise Bayes ağlarıdır. Bayes Ağları, olasılığı belirsizliğin bir ölçüsü olarak kullanan olasılıksal uzman sistemlerin (probabilistic expert systems) bir türüdür [59].

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.

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.

Diğer bir ifade ile Bayes Ağları tablodaki rastgele değişkenler arasındaki bağımlılık ilişkilerinin anlaşılmasını ve görselleştirilmesini sağlayan grafiksel olasılık yapısı ortaya çıkaran istatistiksel bir yöntemdir [60]. Aynı zamanda bu ağ koşullu olasılıklar için çıkarsama yapılmasını sağlar. Bayes Teoremi yardımıyla araştırmacıların eldeki verilerin yanında uzman bilgisini de ağa katmasını sağlar [61]. Bayes Ağlarının oluşturulmasında grafik teorisinden yararlanılır. Değişkenler hakkındaki olasılık bilgilerinin grafiksel olarak gösterimini sağlamak, değişkenlerin ortak olasılık fonksiyonunu görsel olarak incelemek, gözlemlenen verilerden en etkili sonuçları çıkarmak grafiksel modellerin en önemli özelliklerindendir.

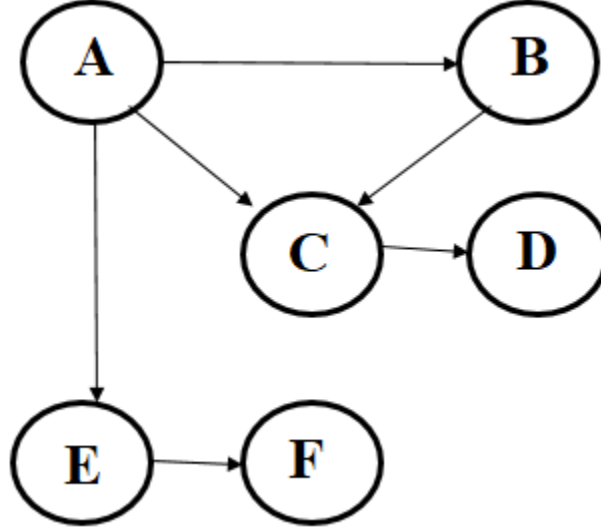
Bayes Ağları sonsuz olmayan sayıdaki vakalara sahip değişkenler arasındaki koşullu bağımlılıkları belirten oklardan ve bu değişkenleri gösteren düğümlerden oluşan DAG’lardır.

Düğümünden çıkan okların belirttiği düğüme Çocuk Düğüm (Child Nodes), düğümünden çıkan okların çıktığı düğüme Ebeveyn Düğüm (Parent Nodes) ve düğümünden çıkan okların hiç biri çıktığı düğümü belirtmezse o düğüme Kök Düğüm (Root Nodes) ismi verilir [17]. Çocuk Düğümünden eğer ok çıkmaz ise o düğüme Yaprak Düğüm (Leaf Nodes) denir [18].

2.1. Grafiksel Modeller ve Döngü

Bir grafiksel modelde $V=\{V_1, V_2, \dots, V_n\}$ değişkenler kümesine sahip olduğu varsayılın. V kümesi, değişkenlerin bir kümesidir ve her bir değişken grafiksel modelde bir düğüme karşılık gelir. Ayrıca bu düğümler ok şeklindeki bağlantılar ile birbirlerine bağlanırlar. V_i ve V_j düğümleri arasında bir bağlantı varsa bu bağlantı A_{ij} ile gösterilecektir. Ağda verilen tüm bağlantıların kümesi $A=\{A_{ij}: V_i \text{ düğümü } V_j \text{ düğümü ile bağlantılıdır}\}$ şeklinde gösterilir. Yani, V ve A kümeleri, tam bir grafiği tanımlar. Bu grafik $G=(V,A)$ ile verilmektedir. $G=(V,A)$ grafiği V sonlu düğümler

kümesi ve bu düğümlerin arasındaki bağlantıların sonlu kümesi A'dan oluşan bir yapı olarak tanımlanır. A kümesi, V'den alınan rastgele değişken çiftlerinden oluşur. Yani A kümesi (V_i, V_j) düzenlenmiş çiftler kümesidir. Burada verilen her (V_i, V_j) , V'nin elemanlarıdır. Bir grafikte, düğümler rastgele değişkenleri ve oklar rastgele değişken çiftlerinde var olan ilişkiyi temsil etmektedirler [70].

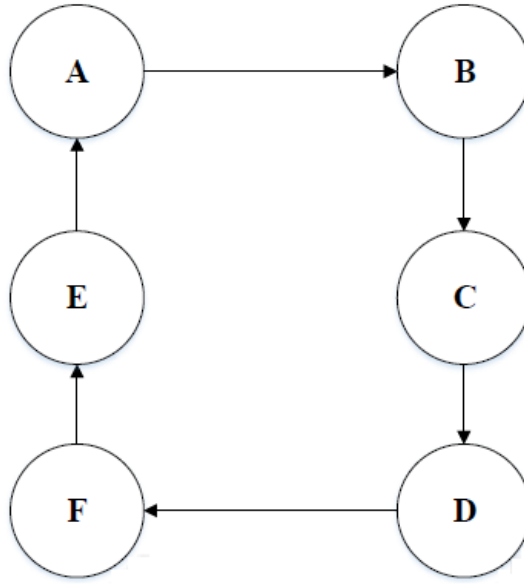


Şekil 2. 1. Bir Bayes Ağı örneği

Bir Bayes Ağı modelinde $(V_1, V_2) \in V$ şeklindeki iki düğüm arasında bulunan düğümlerin farklı sıralamaları ile oluşturulan bağlantılı yapıya yol denir. Şekil 2.1'de verilen ağ örneğinde $A \rightarrow E \rightarrow F$ veya $A \rightarrow B \rightarrow C \rightarrow D$, $A \rightarrow C \rightarrow D$ yolları görülmektedir. Ağda verilen bir yolun uzunluğu yolda bulunan düğüm sayısının bir eksiği olacaktır. $A \rightarrow B \rightarrow C \rightarrow D$ yolunda 4 düğüm vardır ve yolun uzunluğu 3'tür. Aynı düğüm ile başlayan ve biten yola kapalı yol denir. Aynı grafikte A ile C düğümleri arasındaki bağlantının yönü ters çevrilecek olursa uzunluğu 3 olan kapalı bir yol elde edilmiş olacaktır.

Yani Bayes Ağları'nda kapalı yol olmayacağı için Şekil 2.1.'da gösterilen şekil de bir Bayes Ağı olmayacaktır.

Bir grafikte aynı düğüm ile başlayıp aynı düğüm ile biten yola döngü denir. Şekil 2.2'de verilen örnek ağda $A \rightarrow B \rightarrow C \rightarrow D \rightarrow F \rightarrow E \rightarrow A$ yolu döngüsel bir yolu göstermektedir.

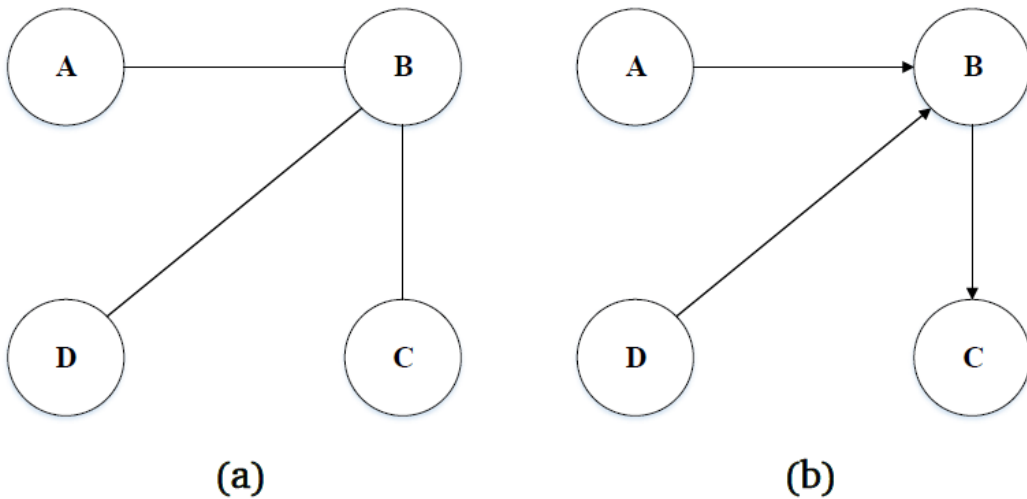


Şekil 2. 2. Bayes Ağı olmayan döngüsel bir ağ

$G=(V,A)$ grafiğinde A 'daki oklar arasında sadece yön verilmemiş oklar varsa, G grafiğine yön verilmemiş grafik denir. A 'da sadece yön verilmiş oklar varsa G grafiğine yön verilmiş grafik denir. Verilen bir $G=(V,A)$ grafiğinde, (V_i, V_j) ve (V_j, V_i) 'nin her ikisi de A 'nın elemanı ise, V_i 'den V_j 'ye olan bir oka yön verilmemiş ok veya yön verilmemiş bağlantı denir. Bu bağlantı $V_i - V_j$ ile gösterilir.

Verilen grafikte $(V_i, V_j) \in V$ ancak $(V_j, V_i) \notin V$ ise, V_i 'den V_j 'ye doğru olan bir oka yön verilmiş ok veya yön verilmiş bağlantı denir ve $V_i \rightarrow V_j$ şeklinde gösterilir.

Aşağıdaki şekilde yön verilmiş ve yön verilmemiş grafik örnekleri verilmiştir [71].

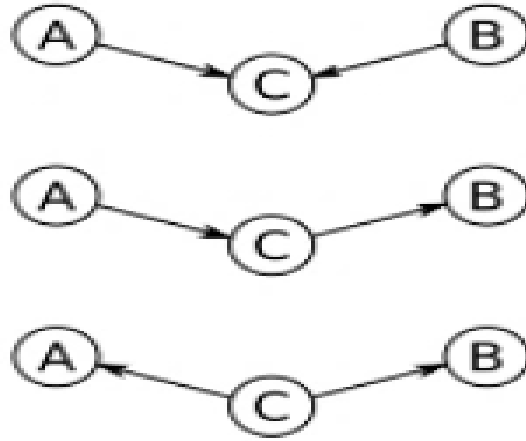


Şekil 2. 3. Dört düğümden oluşan yön verilmiş ve yön verilmemiş grafik örnekleri

Yukarıda verilen Şekil 2.3 a ve b grafiklerinin aynı düğümler arası oklara sahip olduğu görülmektedir. Verilen grafiklerden “a” grafiği yön verilmemiş oklara

sahiptir ve bu grafik $V=\{A,B,C,D\}$ düğümler kümesine ve $A=\{(A,B),(B,A),(B,C),(C,B),(B,D),(D,B)\}$ oklar kümesine sahiptir. Diğer grafik olan “b” grafiği ise $V=\{A,B,C,D\}$ düğümler kümesine sahiptir ancak $A=\{(A,B),(B,C),(D,B)\}$ yön verilmiş oklar kümesine sahiptir. Bir grafik eğer hem döngü olmama hem de yön verilmiş olma özelliklerini aynı anda sağlarsa bu grafiğe yön verilmiş döngüsel olmayan grafik denir.

Bayes Ağları, çok değişkenli olasılık tablolarının grafiksel olarak gösterilmesini sağlayan grafiksel modellerin bir çeşididir. Bu ağlar iki temel kısımdan oluşmaktadır. Bunlardan ilki $V=\{V_1, V_2, \dots, V_p\}$ olan ve değişkenlerin özelliklerini tanımlayan olasılık tablolarıdır. X değişkeninin çok değişkenli olasılık dağılımı global dağılım, her bir V_i değişkenlerinin dağılımları ise yerel dağılımlar olarak adlandırılmaktadır. İkincisi ise $G = (V, A)$ olan ve DAG olarak adlandırılan grafiksel yapıdır. Her bir $v \in V$ düğümü bir V_i değişkeni ile ilişkilendirilir. Grafikte verilen her bir $a \in A$ ise düğümleri birbirine bağlayan yönlendirilmiş oklardır. Bu oklar değişkenler arasındaki direkt olasılıksal ilişkiyi göstermektedir. Eğer değişkenler arasında ok yoksa bu değişkenlerin olasılıksal bağımsızlığını veya koşullu olasılıksal bağımsızlığını ortaya koymaktadır [72].



Şekil 2. 4. Yakınsak bağlantı, seri bağlantı ve ayırık bağlantı için grafiksel gösterim

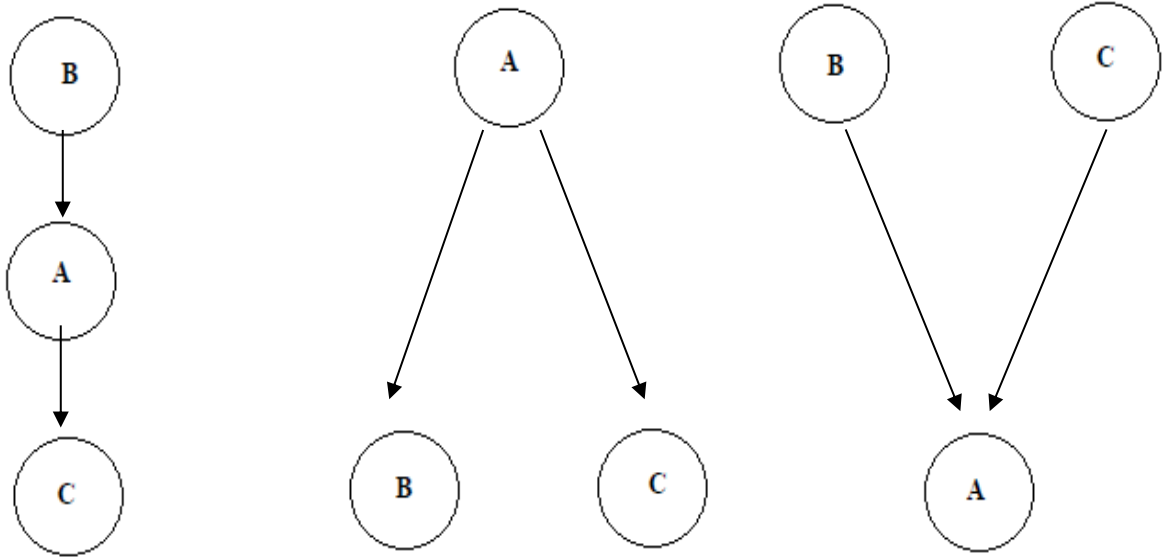
Bayes Ağlarda temel olarak Yakınsak, Iraksak ve Serisel bağlantı olmak üzere 3 çeşit bağlantı vardır. Serisel ve Iraksayan bağlantılarda bir olay ile ilgili kuvvetli bilgi var ise koşullu bağımsızlık koşulu sağlanacaktır. Ancak yakınsayan bağlantılarda koşullu bağımsızlığın sağlanması için kuvvetli ya da zayıf bir olay ile koşullu bağımsızlık sağlanacaktır.

Bayes ağın içerisinde A, B düğümleri ile bu düğümlerin arasındaki tüm yollar için X orta değişkeni bulunduğunu varsayalım. Eğer X ;

1. X düğümünün durumu biliniyor ve serisel ya da ıraksayan okları varsa,
2. Yakınsayan oklar var ve ne X ne de X' in torunları olayda bulunmuyor ise

X, A' dan B' yi yönsel ayırmaktadır denir.

Eğer düğümün minimum iki adet ebeveyni mevcutsa, o düğümün yakınsayan okları bulunmaktadır.



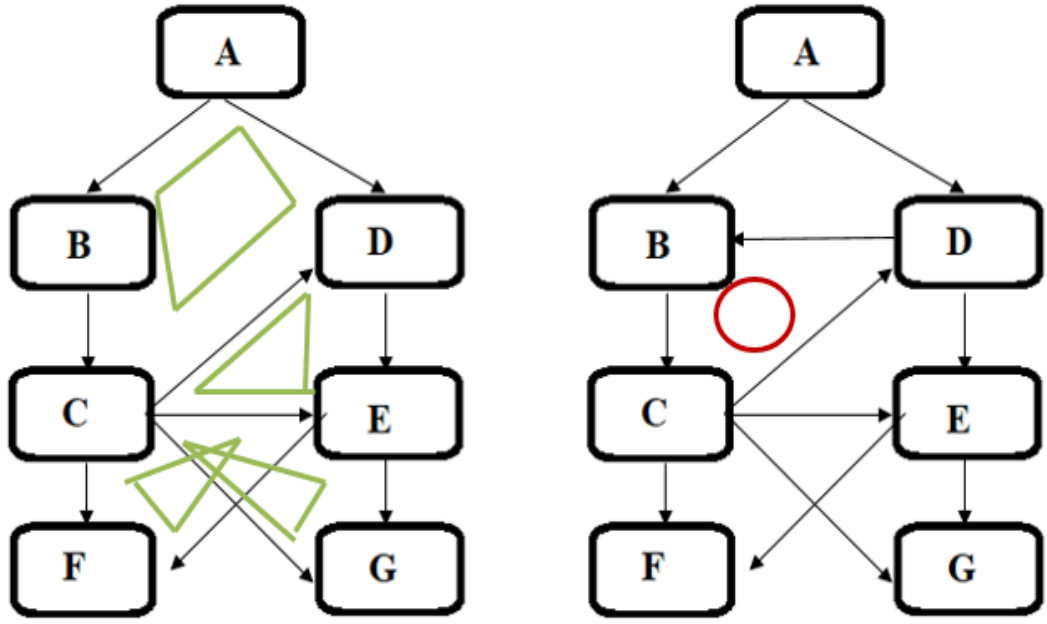
Şekil 2. 5. Bir Bayes Ağı'nda serisel, ıraksayan ve yakınsayan bağlantılar

Şekil 2.5'ta bağlantı çeşitleri verilmiştir. Buna göre birinci grafikte A düğümü bilindiğinde B ve C düğümleri koşullu bağımsızdır. İki numaralı grafikte A verildiğinde B ve C düğümleri koşullu bağımsızdır. Üçüncü grafikte ise A düğümü bilindiğinde B ve C düğümleri koşullu bağımlı olacaktır.

Birinci ve ikinci grafiklerde B ve C düğümleri arasındaki oklarda A düğümü yakınsamadığı için $B \perp C | A$ koşullu bağımsızlık ilişkisi vardır. Ancak üçüncü grafikte A düğümüne doğru yakınsayan iki ok olduğu için koşullu bağımlılık ilişkisi bulunmaktadır.

2006 senesindeki Lauria ve Duchessi'nin araştırmalarına göre Bayes Ağları, merak edilen değişkenler içindeki neden-sonuç ilişki kümesi, birbiri ile bağlantılı bileşik olasılıklarını ve koşullu bağımsızlık varsayımlarını nitelendirmektedir.

Değişkenler arası bu neden-sonuç ilişki kümeleri, hem birbiri ile bağlantılı düğümleri hem de bağlantısı olmayan düğümleri DAG ile (Şekil 2.6.) meydana getirir [19].



Şekil 2. 7. Döngüsel olmayan bir Bayes Ağı ve döngüsel Ağ örneği

Bayes Ağları, çoğunlukla değişkenler arasındaki ilişkileri diyagram yardımı ile ortaya koyan ve koşullu olasılık değerlerinden faydalanarak değişkenler arasındaki ilişkileri ortaya çıkarmaktadır [22]. Bayes Ağları, değişkenler arasındaki neden sonuç ilişkisini ve modeldeki diyagramdan yararlanarak belirsizliği çözmeyi hedefleyen DAG'dır. [23].

Bayes Ağları'nda bileşik olasılık dağılımı ve koşullu olasılıkları ortaya çıkarmak için olasılık teorisinden sıklıkla yararlanır. Ebeveyn Düğümleri bilinen düğümler için koşullu olasılıklar hesaplanır. Bayes Ağları'nda ortak olasılık dağılımları, düğümlerin olasılık dağılımlarından elde edilmektedir[24].

İlişki yapıları için diyagramları içeren modellere başvurulur. Diyagram içeren modeller yoluyla düğümler arası ilişkilere ilişkin ve koşullu bağımsızlık durumlarına ilişkin yorumlar yapılabilir.

Bayes Ağı'nın yapısal öğrenme sürecinde çeşitli programlardan yararlanılmaktadır. Bu programlara en çok öğrenme algoritmalarının yazılımı yapılırken ihtiyaç duyulmaktadır. Bu da istatistik teorisinin Bayes Ağları'nı meydana getirmede hayati önem taşımakta olduğunu göstermektedir.

Bayes Ağı'ndaki düğümler arası ilişkilerin ortaya çıkarılmasında farklı istatistiksel değerlendirmelerden yararlanır buna ek olarak parametrelerin öğrenilmesi aşamasında koşullu olasılık ve Bayes yöntemlerinden yararlanılmaktadır .

Bayes ağları için yapısal öğrenme süreci, DAG yapısını verilerden öğrenmek anlamına gelir. Yapısal öğrenme için puana dayalı yaklaşım (score-based approach), kısıtlamaya dayalı yaklaşım (constraint-based approach) ve hibrit yaklaşım (hybrid methods) olmak üzere üç adet ana yaklaşım vardır.

Skora dayalı yaklaşım algoritmaları Bayes Ağı'nı, çok fazla düğümden oluşan veri kümesi için nedensellik yapısının doğruluğunu gösteren skor fonksiyonunu en büyük yapacak şekilde oluşturur. önce Bayes Ağı'nın verilere ne kadar iyi uyduğunu değerlendirmek için bir kriter tanımlar, ardından maksimum puanı elde eden bir yapı için DAG'lerin alanını araştırır. Skora dayalı yaklaşım temelde iki bölümden oluşan bir arama problemidir: bir skor metriğinin tanımı ve arama algoritması [29].

Kısıtlamaya dayalı yöntem, grafik için bir dizi ok kısıtlamasını belirlemek için bağımsızlık değerlendirmeleri kullanır ve ardından kısıtlamaları karşılayan optimal DAG'yi bulur. Koşullu bağımsızlık değerlendirmelerinden faydalanarak değişkenlerin ilişkilerini en iyi şekilde temsil edebilecek bir yapı oluşturmayı hedefler. Bu yaklaşım, diğer bazı (uzman) yapı bilgisi ile iyi çalışır ancak test gücünü garanti etmek için çok sayıda veri örneği gerektirir. Bu nedenle, örnek sayısı az olduğunda daha az güvenilirdir [22].

Hibrit yaklaşıma dayalı yöntem, kısıtlama yapıları algoritmalar barındıran koşullu bağımsızlık değerlendirmelerini ve puanlama yapıları algoritmaların puan ölçeklerinden faydalanarak Bayes Ağı yapısını meydana getirir. Kısıtlama yapıları algoritmalarından faydalanarak bir yönsüz grafik meydana getirmek ve sonrasında puan yapıları algoritmaları çalıştırarak okların yönlerini belirlemek bu algoritmaların temel amacıdır.

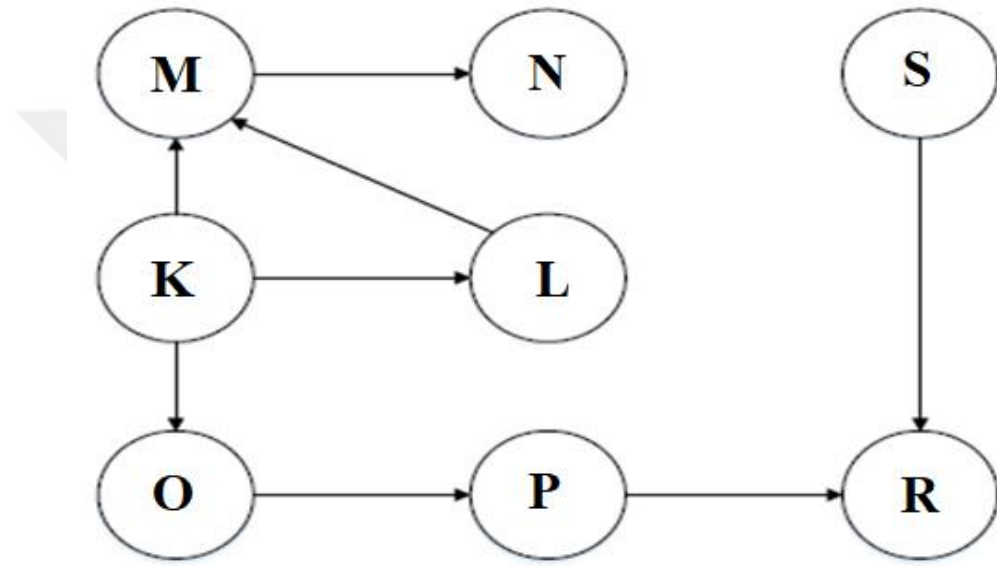
Grafiksel modellerin en önemli özelliklerinden birisi “Koşullu Olasılık”tır. Grafikte yer alan düğümlerin birbirine bağlayan okların meydana getirilmesi ve değişkenlerin ilişkilerini açığa çıkarması sürecinde koşullu olasılık ilişkileri hayati önem taşımaktadır [24].

2.2. Bayes Ağları'nın Matematiksel Gösterimi

Bayes Ağları'nın matematiksel gösterimi denilince akla ilk Bayes Teoremi oluşturmak gelmektedir. Bayes Ağları'nın olasılık teorisinden faydalanarak olasılık temelli karar destek sistemi olduğu bilinmektedir. Olasılık teorisi, matematiksel

olarak var olan ağlar üzerinde, sonradan gelen olayların sisteme eklendikten sonra farklılıklar yapılabileceğini belirtmektedir[68][69].

Bayes Ağları'nın öğrenme tekniklerini kavrayabilmek için, Bayes'in olasılık ve istatistik yaklaşımını grafik gösteriminde kavramak önemlidir [62]. Bu kapsamda Şekil 2.8'de gösterilen aile ebeveyn, çocuk ve torun dışı düğümler için verilen Bayes Ağı örneğinde, L düğümüne K'nın çocuğu denir yani $Desc(K)=\{L\}$ (descendant) şeklinde ifade edilir. Eğer P düğümü K düğümünün ebeveyni veya çocuğu değilse $Nondesc(K)=\{P\}$ şeklinde gösterilir ve torun dışı düğüm (nondescendant) olarak adlandırılır [73]. Çocuğu olmayan düğüme kısır düğüm denir.

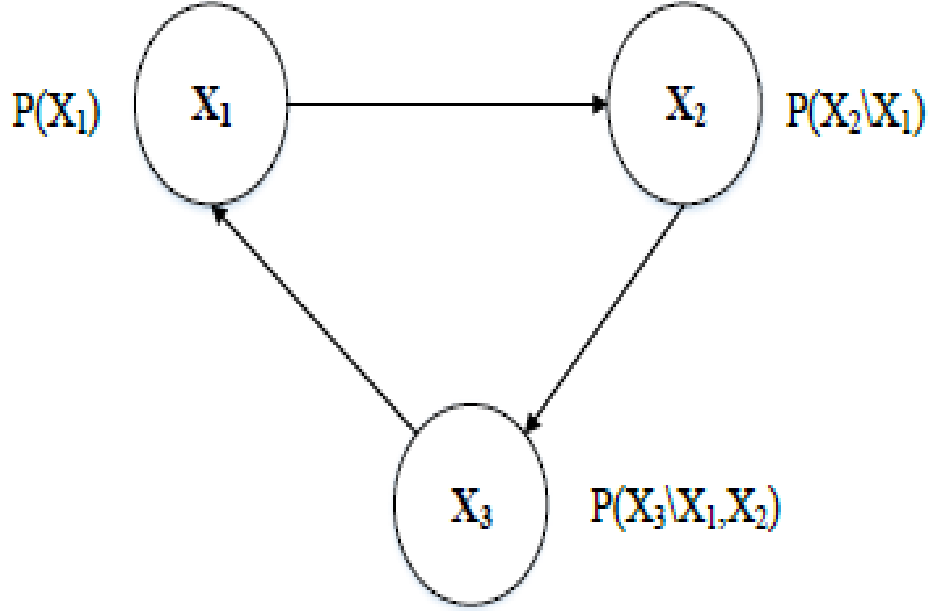


Şekil 2. 8. Aile ebeveyn, çocuk ve torun dışı düğümler için verilen Bayes Ağı örneği
Bu ağ için her bir düğüm ile ilgili Çizelge 2.1 ortaya çıkarılmıştır.

Çizelge 2. 1. Aile ebeveyn, çocuk ve torun dışı çizelgesi

Düğüm	Ebeveyn	Çocuk	Torun dışı	Aile
K	$\emptyset$	L,M,O	N, P, R, S	K, L, M, O
L	K	M	N, O, P, R, S	K,L,M
M	K,L	N	O, P, R, S	K,L,M,N
N	M	$\emptyset$	K, L, O, P, R, S	M,N
O	K	P	L, M, N, R, S	K,O,P
P	O	R	K, L, M, N, S	O,P,R
R	P,S	$\emptyset$	K, L, M, N, O	P,S,R
S	$\emptyset$	R	K, L, M, N, O, P	R,S

Şekil 2.9’ de gösterilen ağda X_1 değişkeni X_2 ve X_3 değişkenlerinin ebeveyn düğümüdür. X_1 değişkeni ile X_2 değişkenleri de X_3 değişkeninin ebeveyn düğümleridir. Görüldüğü gibi X_2 değişkeni hem ebeveyn hem de çocuk düğümü olarak ağda yer almıştır.



Şekil 2. 9. Döngüsel bir ağ örneği ve olasılıkları

Eğer ağda herhangi bir düğüme okla bağlanmamış düğümler varsa bu düğüm ile diğer düğümlerin olasılıksal olarak bağımsız olduğunu ve ağda marjinal olasılıkları ile bulunduğunu gösterir.

Bu ağın birleşik olasılıkları, ağda yer alan düğümlerin koşullu olasılıklarının çarpımıdır. X_i çocuk ve X_j de ebeveyn düğümü olmak üzere ağın olasılığı verilen formül ile hesaplanır.

$$P(X_1, X_2, \dots, X_N) = \prod_{i=1}^N P(X_i | X_j)$$

(2.1)

Burada N değişken yani düğüm sayısıdır. Bayes Ağları, DAG yapısı ve Koşullu olasılıklar tablosu olmak üzere iki kısımdan oluşmaktadır. Bu tabloda ebeveyn düğümler ile her bir çocuk düğümün alınan değerlere göre koşullu olasılıkları bulunmaktadır.

2.3. Bayes Ağları'nın Koşullu Olasılık Gösterimi

E ve F'nin olgu olduğu varsayalım. F olgusunun olduğu bilindiği takdirde E olgusunun olma “koşullu olasılığı” $P(E|F)$ belirtilir ve aşağıdaki eşitlikte gösterilir:

$$P(E|F) = \frac{P(EF)}{P(F)} \quad (2.2)$$

E ve F olgularının aynı zamanda olma olasılığı $P(EF)$ ve F olgusunun olma olasılığı $P(F)$ 'dir. Bu durum $P(F) > 0$ kısıtı altında için denklemin iki tarafıda $P(F)$ ile çarpılarak Newbold'a göre (1995) aşağıdaki gibi yazılabilir:

$$P(EF) = P(F)P(E|F) \quad (2.3)$$

Eşitlik 2.3'e bakıldığında, E ve F olgularının aynı zamanda olma olasılığı, F olgusunun olma olasılığı ile F olgusunun olduğu bilindiği takdirde E olgusunun koşullu gerçekleşme olasılığının çarpımı ile elde edilir. Buradan esinlenerek, bu denklemin biraz genelleştirilirse, olasılığın “çarpım kuralı” (*Multiplication Rule*) olarak tanımlanabilmektedir:

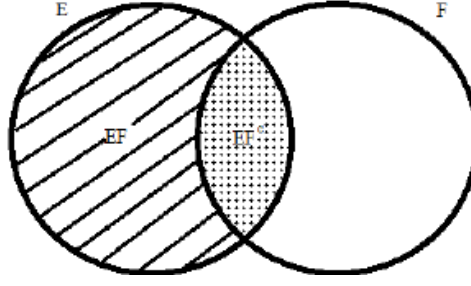
$$P(E_1 E_2 \dots E_n) = P(E_1)P(E_2|E_1) \dots P(E_n|E_1 \dots E_{n-1}) \quad (2.4)$$

Burada n farklı olgunun aynı zamanda olma olasılığı, zincirleme bir şekilde koşullu olasılıklarının çarpımı ile elde edilir.

Aynı E ve F olguları hesaplanırken E olgusunun tanımı Eşitlik 2.5'te gösterilmiştir:

$$E = EF \cup EF^c \quad (2.5)$$

Burada F^c , F olgusunun olmamasını göstermektedir aynı şekilde 'F' sembolü de kullanılır. Yani E kümesi içindeki bir nokta; ya E kümesi ile F kümesinin kesişimine aittir yada E kümesine ait iken F kümesine ait olmamalıdır (Şekil 2.10.). Bu tanıma göre EF^c ve EF ayrık olgulardır.



Şekil 2. 10. EF= Taralı alan, EFC = Noktalı alan gösterimi

Eşitlik 2.3, 2.4 ve 2.5'ten esinlenerek ve EF^c ve EF ayrık olgular olmasından dolayı;

$$P(E) = P(EF) + P(EF^c)$$

$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$$

$$P(E) = P(E|F)P(F) + P(E|F^c)[1 - P(F)]$$

(2.6)

Eşitlik 2.6 ortaya çıkarılır. Bu eşitlikte, E olgusunun olma olasılığının, F olayı olduğunda E olgusunun olması koşullu olasılığı ile F olgusu olmadığında E olgusunun olma olasılığının ağırlıklı ortalaması olduğu saptanmıştır. Eşitlik 2.6, bir olgunun olma olasılığını, ikinci bir olgunun olma veya olmama durumları üzerine “koşullandırarak” saptamayı gerçekleştirdiği için çok etkili bir eşitliktir. Koşullandırma metodu genellikle, olma olasılığı direkt olarak hesaplanamayan fakat ikinci olgunun olması veya olmaması bilindiği durumlarda çok kullanışlı bir metottur.

Yeni kanıtların elde edilmesiyle olasılıkların yeniden değerlendirilmesi yöntemi, koşullu olasılıklar için önemli bir başka durumdur. Bu yöneme göre, sisteme yeni bir bilgi girdiğinde, bir varsayımın olasılığındaki değişim kısaca varsayımın Bahis Oranı” (Odds Ratio) olarak tanımlanabilir. Yani bir A olgusunun Bahis Oranı aşağıdaki eşitlikte verilmiştir.

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

(2.7)

Eşitlik 2.7'e göre, A olgusunun bahis oranı, A olgusunun olmasının, olmamasına oranla kaç kat daha yüksek veya alçak olduğunu saptar. Bir örnekle anlatılacak olursa, $P(A) = 3/4$ ise, $P(A) = 3P(A^c)$ hesaplanacaktır. Yani, A olgusunun bahis oranı 3 olarak hesaplanacaktır.

Olma olasılığı $P(H)$ ile gösterilen H varsayımı sonrasında yeni bir E bilgisinin meydana geldiği bilindiğinde, H varsayımının yeni bahis oranı Eşitlik 2.8 ile gösterilmiştir:

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)}{P(H^c)} \frac{P(E|H)}{P(E|H^c)}$$

(2.8)

Eşitlik 2.8'e göre, H varsayımının önceki bahis oranının, varsayımın doğru olduğu bilindiğinde E bilgisinin olmasının koşullu olasılığı ile varsayımın yanlış olduğu bilindiğinde E bilgisinin olmasının koşullu olasılığı oranının çarpımına eşit olur. Bu durumda, yeni bir E bilgisinin meydana geldiğinde, H varsayımının da doğru olma durumu, doğru olmadığı durumuna oranla fazlaysa bahis oranı ve H varsayımının doğru olma olasılığı yükselmektedir. Diğer bir taraftan E bilgisi meydana geldiğinde, H varsayımının yanlış olduğu durumda, doğru olduğu duruma oranla fazlaysa bahis oranı ile H hipotezinin doğru olma olasılığı düşecektir. Bu varsayımlardan yola çıkarak, Eşitlik 2.8'i biraz daha genişletebiliriz. Yani $F_1, F_2, \dots, F_n$ bütün örneklerin uzayı olan E 'yi veren ayrık olaylar olduğu kabul edilsin. Her bir $i=1, \dots, n$ için EF_i olayları da ayrık olaylar olarak tanımlanabilmektedir. Bu bilgiler ışığında:

$$P(E) = \sum_{i=1}^n P(EF_i)$$

$$P(E) = \sum_{i=1}^n P(E|F_i)P(F_i)$$

(2.9)

denklemleri oluşturulur. Eşitlik 2.9'a göre $F_1, F_2, \dots, F_n$ gibi sadece bir tanesinin ortaya çıkması zorunda olan olaylar elde edildiğinde, bütün örneklem uzayının gerçekleşme olasılığı olan $P(E)$, ilk olarak ortaya çıkan F_i örneklemi üzerinde koşullandırılarak oluşturulur. Eşitlik 2.9'da $P(E)$ 'nin, her $P(E|F_i)$ olasılığı koşullandırılmasının kullanıldığı ilgili olayın olasılığı ile ağırlıklandırma yöntemi ile elde edilen ağırlıklı

ortalamaya eşit olacağı saptanmıştır. Bu durumda, E olayının ortaya çıkması halinde F_j olaylarından hangisinin ortaya çıkacağını belirlemek için aşağıdaki eşitlik kullanılır:

$$P(F_j|E) = \frac{P(EF_j)}{P(E)}$$

$$P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)}$$

(2.10)

Eşitlik 2.10'da Bayes formülünün denklemi yer almaktadır. Bu denklemin isim babası İngiliz filozof Thomas Bayes'tir (1702-1761). Eşitlik 2.10'a göre F_j olayları bir deneyin konusu üzerindeki olası varsayımlar olarak düşünülürse; Bayes formülü yapılacak çalışmanın öncesinde meydana gelen varsayımlara ilişkin bilgilerin (denklemdaki $P(F_j)$ olasılığı) deneyin bilgilerine göre nasıl yön verilmesi gerektiğine ilişkin değerlendirmeler yapılabilir.

2.4. Bağımsız Olaylar

E ve F olguları için $P(EF) = P(E)P(F)$ ise E ve F olguları bağımsızdır denilir. Başka bir deyişle iki ayrı olaya ilişkin ortak sonuçlar yok ise bu olaylara bağımsız olaylar denir [27]. Bağımsızlık durumunu koşullu olasılıklar açısından incelersek; $P(E|F)=P(E)$ veya $P(F|E)=P(F)$ eşitlikleri sağlanması durumunda F ve E olguları için bağımsızdır diyebiliriz. Yani bu açıklamadan yola çıkarak E durumunun biliniyor olması F olayının olması olasılığını etkilemeyecektir. Bu bilgiler ışığında, $E_1...E_n$ olaylarının herhangi bir alt kümesi $E_{i1}...E_{ir}$ için $P(E_{i1}...E_{ir})=P(E_{i1})...P(E_{ir})$ bu eşitlik sağlanıyorsa bağımsız olduğunu söyleyebiliriz.

2.5. Zincir Kuralı

Bayes Ağlarının temelinde değişkenlerin bileşik olasılık dağılımlarını inceleyen zincir kuralı (Chain Rule) vardır. Bu kurala göre, ağdaki değişkenlerin bileşik olasılık dağılımı tüm değişkenlerin koşullu olasılıklarından ya da marjinal olasılıklarından faydalanarak bulunur. Zincir kuralında ilgili değişkenin ebeveyni yoksa koşullu olasılık yerine marjinal olasılık kullanılır.

$X_1,...,X_d$ Eğer değişkenleri temsil ederse, X_i değişkeninin bileşik olasılığı gösterimi Eşitlik 2.11'de verilir:

$$P(X_1 \dots X_d) = \prod_{i=1}^d P(X_i | \text{ebeveynler}(X_i))$$

(2.11)

Dışarıdan, bir değişkene veya değişkenler kümesine ait bilgiler elde edildiği takdirde değişkenlerin koşullu ve marjinal olasılıkları, formülde marjinalleştirme yöntemiyle saptanabilir. Değişkene ait bazı bilgiler verilmişse, olayların ortaya çıkma olasılığı Eşitlik 2.12'deki gibi hesaplanır.

$$P(U|e) = \frac{P(U, e)}{P(e)}$$

(2.12)

U , X_i değişkenlerinin evrensel kümesi olarak tanımlanır. Koşullu olasılıklara dayanan Bayes teoremi, araştırılan sorunu, değişkenlerin birbirleri ile ilişkilerini kullanarak modeller sonrasında ise yeni bulguların ortaya çıkması ile birlikte modelin kendini yenilemesine fırsat tanır [28].

2.5.1. Markov Özelliği

$G=(V,A)$ 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.

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

G : düğüm, P : olasılık

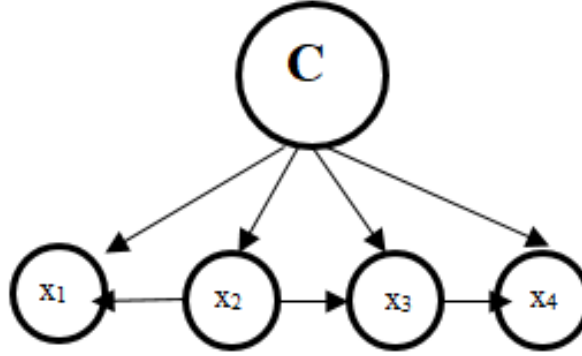
1) G ve P her zaman tutarlı ise $(X \perp Y \setminus Z)G \Rightarrow (X \perp Y \setminus Z)P$

2) $(X \perp Y \setminus Z)P$ G ile tutarlı tüm dağılımları sağlıyorsa $(X \perp Y \setminus Z)G$ olacaktır [64].

2.5.2. Tan Özelliği

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 2.11’deki TAN yapısındaki “C” bir düğümü, X_1 , X_2 , X_3 , X_4 ise çocukları temsil etmektedir.



Şekil 2. 11. Tipik bir TAN yapısı ağı

2.6. Bayes Ağları’nda Öğrenme

Bayes Ağları oluşturdıkları 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 [25]. Öğrenme çok canlı bir araştırma alanıdır ve bu alanda birçok yöntem geliştirilmiştir. Bu yöntemlerin çoğu her tür veri için uygun değildir. Bu nedenle uygulamada bazı problemler yaşanabilir. Ağ yapısının öğrenilmesi parametrelerin öğrenilmesinden daha zor bir süreçtir. Ayrıca gizli düğümler yada kayıp veri gibi kısmi gözlenebilirlik durumu söz konusu olduğunda başka zorluklarda ortaya çıkabilir.

Çoğunlukla farklı öğrenme yöntemlerinin önerdiği dört farklı öğrenme yöntemi vardır. Bunlar, En çok olabilirlik tahmini, EM ya da MCMC, Model uzayının araştırılması, EM+Model uzayının araştırılması yöntemleridir [74].

2.6.1. Olasılık Temelli Çıkarımlar

Bayes Ağları modeli algoritmaları kullanarak meydana getirilmesinin ardından, modele ilişkin olasılıklar ek bilgilerle, modele ait verilerle veya çeşitli birleşimler ile hesaplanmaktadır. Bir olasılık hesabı “olasılığa dayalı çıkarım” diye nitelendirilebilir. 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 olasılıkların hesaplanması olağandır [29].

Farklı düğümleri ve bu düğümlerin bir biri ile bağlantılarını içeren bir model olan Bayes Ağları'ndan, düğümlerle ilgili tüm olasılığa dayalı sorguların ilişkilerini çözmek adına faydalanabilir.

Bayes Ağları'nda olasılık temelli çıkarımlarda kullanılan yöntemlerin “kesin” ve “yaklaşık” yöntemler olmak üzere ikiye ayrıldığı bilinmektedir. Değişken eleme (Variable Elimination), Ve/Veya Araması (AND/OR Search), Sınıf Ağacı Yayımı (Clique Tree Propagation) ve Tekrarlayan Koşullandırma (Recursive Conditioning) gibi metotlar kesin çıkarım yöntemlerine örnek olarak verilebilir. Bütün bu metotların kompleksliği Bayes Ağı'nın esnekliği ile alakalıdır.

Döngüsel Ağ Yayılımı (Loopy Belief Propagation), Markov Zinciri Monte Carlo Simülasyonu (Stochastic Markov Chain Monte Carlo Simulation), Öne göre Örnekleme (Importance Sampling), Genelleştirilmiş Ağ Yayılımı (Generalized Belief Propagation) ve Mini-Kova Elemesi (Mini-Bucket Elimination) gibi metotlar yaklaşık çıkarım yöntemlerine örnek olarak verilebilir.

2.6.2. Bayes Ağları'nda Öğrenme Algoritmaları

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. “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 [52] neden sonuç araştırması üzerinde olmuştur. Yapısal öğrenme

algoritmaları koşullu bağımsızlık testlerini uygulayarak Bayes Ağları'nın yapısını öğrenmek adına bir temel oluşturmaktadır [54].

Tümevarımsal neden – sonuç algoritmasını kolay uygulayabilmesine olanak tanıyan PC algoritması kısıtlama tabanlı algoritmalarla yönelik tasarlanmıştır [53]. Markov örtülerinden sağlanan birçok kısıtlama tabanlı algoritmalar oluşturulmuştur [55].

Sonrasında Markov örtülerine dayalı BüyümeDaralma adında algoritma geliştirilmiştir[56]. İki aşamadan oluşan Aralıklı Artımsal İlişki Markov Örtüsü ve Artımsal İlişki Markov Örtüsü algoritmalarını oluşturulmuştur [57].

Koşullu bağımsızlı ilişkilerini istatistiksel testler uygulayarak belirleyen Hızlı Artımsal İlişki Markov Örtüsü algoritması kullanılmıştır [56]. Çocuklar ve aday ebeveyn adını verdikleri örneklemden faydalanarak, iki aşamadan oluşan Çocuklar ve En Çok En Az Ebeveynler algoritması kullanılmıştır.

Bayes Ağı kurulurken parametresi ve yapısı olmak üzere iki ana bölüm oluşturulmalıdır [58]. Yapı öğrenmesi, çok değişkenli veri kümesinden faydalanarak düğümler arası bağlantıları belirten DAG'dır.

Parametre öğrenme, ebeveynleri olan düğümlerin koşullu olasılık değerlerinin belirlenmeye çalışılmasıdır.

Sonuç olarak Bayes Ağı yapısının öğrenilme sürecinin parametrelerin öğrenilmesi sürecine oranla çok daha karmaşık bir sorun olduğu gözlemlenmiştir [26].

2.7. Bayes Ağları Sınıflandırıcıları

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 X_i 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 olduğunu 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 öz nitelikleri 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 Naive 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.

3. COVID-19 VERİLERİNİN ÇÖZÜMLENMESİ

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 https://github.com/beoutbreakprepared/nCoV2019/blob/master/latest_data/latestdata.tar.gz linkinden alınmıştır.

3.1. Veri Kümesinin Tanımı

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.

Bu çalışma koronavirüse yakalanan kişilerin, hastalıktan sonra hayatta kalıp kalamayacağını Bayes Ağları yöntemi kullanarak hayatta kalıp kalamayacağını sınıflandırmak ve tahmin etmektir.

Bu çalışma ile hastalığa yakalanan kişiler ya da diğer ilgili kişiler yapılan modele o hastanın verilerini girerek ölüm riskini hesaplayabileceklerdir. Dünya genelinde oluşacak olan bu farkındalık sayesinde ölüm riski yüksek olan kişiler tespit edilerek ölüm oranlarını en aza indirmek hedeflenmektedir.

Aşağıdaki Çizelge 3.1.'de görüldüğü üzere tüm verinin 20 satırlık bir kesiti ve Çizelge 3.2.'de değişkenlerin adları ve tanımları verilmiştir.

Çizelge 3. 1. Veri kümesine genel bakış

YAŞ	CINSİYET	ULKE	KRONİK	SONUC	SEYAHAT	HASTALIK TARİHİ	SEMPTOM	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

Çizelge 3. 2. Verinin içerdiği değişkenler ve tanımları

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

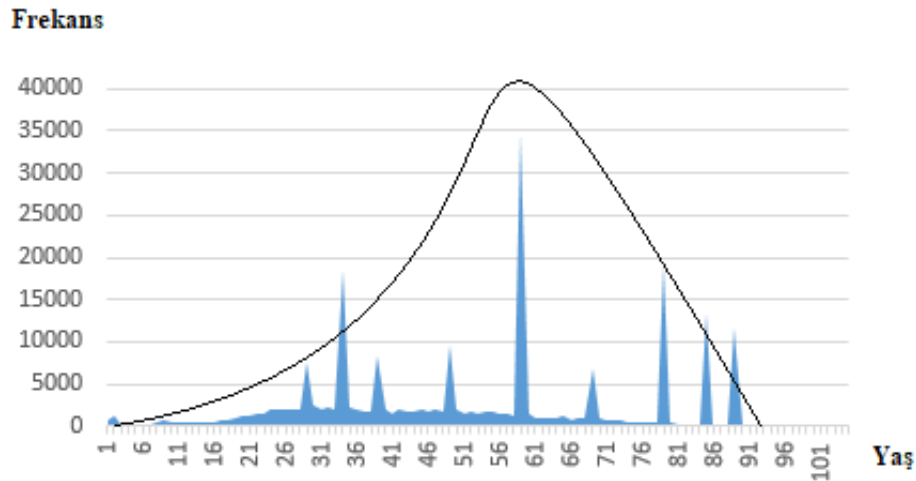
3.2. Betimleyici İstatistikler

Verinin betimleyici özelliklerine göre, 215.968 hastanın yaş ortalaması 52’dir, hastalar en düşük 1, en yüksek 105 yaşındadır. Araştırma hastaların hastalığa yakalandığı 104 ülkeyi içermektedir. Hastaların seyahat geçmişinin, semptomunun ve kronik rahatsızlığının olup olmadığına bakılmıştır. Hastaların hastalandığı tarih, Kasım 2019’dan Kasım 2020’ye kadar uzanmaktadır. Hastaların hastalığa yakalandıkları ay ise veri içinde 8 farklı aydan oluşmaktadır.

Field	Sample Graph	Type	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
YAS		Range	1	105	52.845	21.502	0.010	--	215968
CINSIYET		Set	--	--	--	--	--	2	215968
ULKE	Too many values	Set	--	--	--	--	--	104	215968
KRONIK		Set	--	--	--	--	--	2	215968
SONUC		Set	--	--	--	--	--	2	215968
SEYAHAT		Set	--	--	--	--	--	2	215968
HASTALIKTA...		Range	2019-11-12	2020-11-26	--	--	--	--	215968
SEMPATOM		Set	--	--	--	--	--	2	215968
HASTALIKAYI		Set	--	--	--	--	--	8	215968

Şekil 3. 1. Değişkenlere ait betimleyici istatistikler

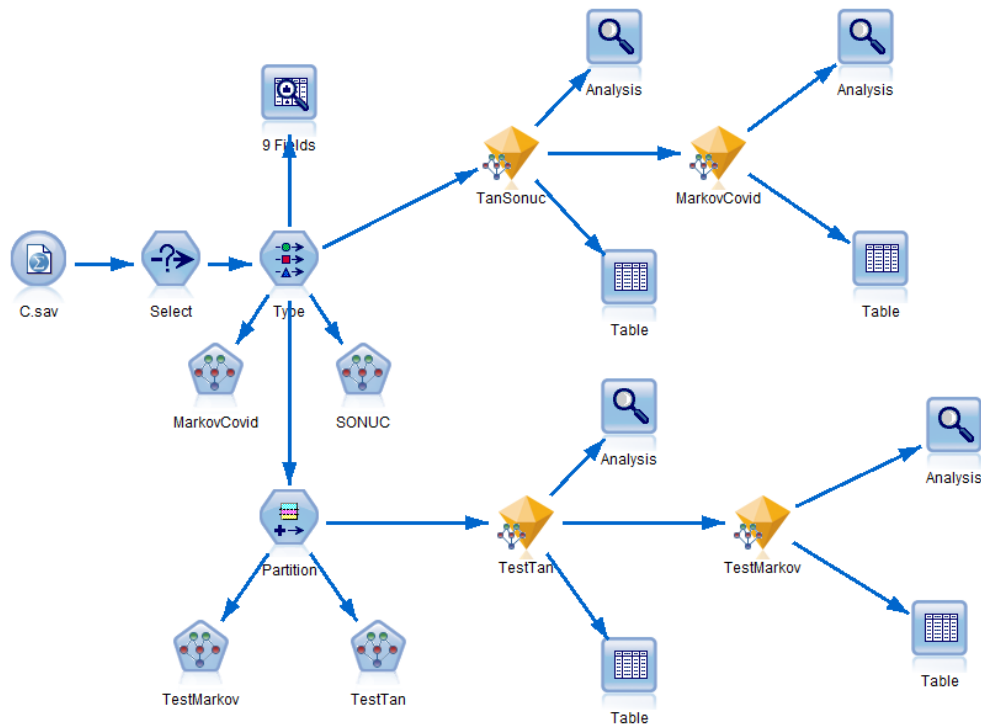
Şekil 3.2’deki grafikten de anlaşılacağı üzere araştırmadaki hastaların yaşlarına ilişkin grafik sola çarpıktır. Veride yaş dağılımının ortalaması 52,845’tir.



Şekil 3. 2. Hastaların yaş dağılımı

3.3. Covid-19 Verileri İçin Bayes Ağları Çözümlemesi

Verilerin analizi IBM SPSS Modeler 18.0 programı ile yapılmıştır. IBM SPSS Modeler uygulaması ile “Makine Öğreniminin Bayes Ağları” modeli kullanılmıştır [75,76]. Tan ve Markov modelleri denenmiş ve birbirleri ile karşılaştırılmıştır. Test modülü çalıştırılarak gözelerin %20’si tahmin edilmiştir.



Şekil 3. 3. IBM SPSS Modeler programı ile Bayes Ağları kesiti

Yapılan analiz sonucunda Çizelge 3.3’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.

Çizelge 3. 3. Tan modeli ve olasılıkları

YAŞ	CİNSİYET	ULKE	KRONİK	SONUC	SEYAHAT	HASTALIK TARİHİ	SEMPTOM	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 Çizelge 3.4’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.

Çizelge 3. 4. Markov modeli ve olasılıkları

YAŞ	CİNSİYET	ULKE	KRONİK	SONUC	SEYAHAT	HASTALIK TARİHİ	SEMPTOM	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

Çizelge 3.4. Devam ediyor

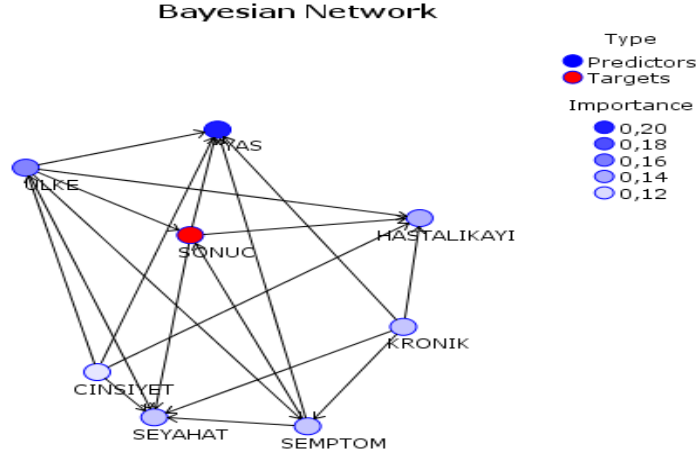
59	Erkek	Avustralya	Yok	Hayatta	Var	28.02.2020	Yok	Şubat	Hayatta	1,00
69	Kadın	Avustralya	Yok	Hayatta	Yok	28.02.2020	Yok	Şubat	Hayatta	0,94
59	Erkek	Avustralya	Yok	Hayatta	Yok	28.02.2020	Yok	Şubat	Hayatta	1,00
59	Kadın	Avustralya	Yok	Hayatta	Var	29.02.2020	Yok	Şubat	Hayatta	1,00
59	Kadın	Avustralya	Yok	Hayatta	Var	29.02.2020	Yok	Şubat	Hayatta	1,00

Yapılan analiz sonucunda Şekil 3.4’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.

Results for output field SONUC			
Individual Models			
Comparing \$B-SONUC with SONUC			
Correct	213.768	98,98%	
Wrong	2.200	1,02%	
Total	215.968		
Comparing \$B1-SONUC with SONUC			
Correct	214.149	99,16%	
Wrong	1.819	0,84%	
Total	215.968		
Agreement between \$B-SONUC \$B1-SONUC			
Agree	215.253	99,67%	
Disagree	715	0,33%	
Total	215.968		
Comparing Agreement with SONUC			
Correct	213.601	99,23%	
Wrong	1.652	0,77%	
Total	215.253		

Şekil 3. 4. Tan ve Markov modelleri için karşılaştırma oranları

Seçilen Markov modeli üzerinden değişkenlerin anlamlılık düzeyi Şekil 3.5’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.



Şekil 3. 5. Markov modeli için değişkenlerin anlamlılık düzeyleri

Hastalığa hangi ayda yakalandığına ait koşullu olasılık tablosuna 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.

Çizelge 3. 5. Hastalığa hangi ayda yakalandığına ait koşullu olasılık tablosu

Aile						Olasılık							
Sıra	Cinsiyet	Ülke	Kronik	Sonuç	Semptom	Aralık	Haziran	Kasım	Mart	Mayıs	Nisan	Ocak	Şubat
1	Erkek	Ekvator Ginesi	Yok	Hayatta	Var	0	0	0	0	0.666	0.333	0	0
2	Erkek	Ekvator Ginesi	Yok	Hayatta	Yok	0	0	0	0	1	0	0	0
3	Erkek	Ermenistan	Yok	Hayatta	Yok	0	0	0	0	1	0	0	0
4	Erkek	Estonya	Yok	Hayatta	Yok	0	0	0	0.717	0.001	0.281	0	0
5	Erkek	Esvatini	Yok	Hayatta	Yok	0	0	0	0	0.666	0.333	0	0
6	Erkek	Etiyopya	Yok	Hayatta	Yok	0	0.034	0	0	0.848	0.116	0	0
7	Erkek	Etiyopya	Yok	Öldü	Yok	0	0	0	0	0.5	0.5	0	0
8	Erkek	Fildişi Sahili	Yok	Hayatta	Yok	0	0	0	0	0.666	0.333	0	0
9	Erkek	Filipinler	Var	Hayatta	Var	0	0	0	0	0	1	0	0
10	Erkek	Filipinler	Var	Hayatta	Yok	0	0	0	0	0.166	0.833	0	0
11	Erkek	Filipinler	Var	Öldü	Var	0	0	0	0	0.061	0.938	0	0
12	Erkek	Filipinler	Var	Öldü	Yok	0	0	0	0	0	1	0	0
13	Erkek	Filipinler	Yok	Hayatta	Var	0	0	0	0	0.6	0.4	0	0
14	Erkek	Filipinler	Yok	Hayatta	Yok	0	0.049	0	0	0.646	0.304	0	0
15	Erkek	Filipinler	Yok	Öldü	Var	0	0	0	0	0.285	0.714	0	0

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.

Çizelge 3. 6. Yaşa ait koşullu olasılık tablosu

Aile				Olasılık				
Sıra	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 cinsiyetine ait koşullu olasılık tablosuna bakıldığında, Burkina Faso’da koronavirüs hastalarının %80’i erkek ve %20’si kadındır.

Finlandiya’da koranavirüs hastalarının %80’i kadın ve %20’si erkektir.

Çizelge 3. 7. Hastanın cinsiyete ait koşullu olasılık tablosu

Aile		Olasılık	
Sıra	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 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.

Çizelge 3. 8. Hastanın seyahat geçmişine ait koşullu olasılık tablosu

Sıra	Aile				Olasılık	
	Cinsiyet	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

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.

Çizelge 3. 9. Sonuç değişkenine ait koşullu olasılık tablosu

Aile			Olasılık	
Sıra	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

Çizelge 3.10'a bakılırsa, Tan modelinin sınıflandırması ve sınıflandırma olasılıklarına bakıldığında, tan modelinde aşağıdaki çizelgede görüldüğü üzere ilk 25 hastaya baktığımızda 19 hastanın hayatta kalıp kalamayacağını doğru sınıflanmış, geri kalan 6 hastayı yanlış sınıflamıştır. Yani 18. Gözeye baktığımızda 85 yaşındaki Cezayir'de yaşayan, kronik rahatsızlığı olmayan erkek hastanın sonucu koranavirüs hastalığından sonra öleceği yönünde olmasına rağmen, tan modeli bu hastanın %60 olasılıkla hastalıktan sonra hayatta kalacağını yanlış olarak sınıflandırmıştır.

Çizelge 3. 10. 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

Çizelge 3.11’de, Markov modelinin sınıflandırması ve sınıflandırma olasılıklarına bakıldığında, Markov modelinde aşağıdaki çizelgede görüldüğü üzere ilk 25 hastaya baktığımızda 23 hastanın hayatta kalıp kalamayacağını doğru sınıflanmış, geri kalan 2 hastayı yanlış sınıflamıştır. Örneğin, 84 yaşındaki Cezayir’de yaşayan, kronik rahatsızlığı olmayan kadın hastanın koronavirus hastalığından sonra hayatını kaybeden kadını, markov modeli %63 olasılıkla hastalıktan sonra öleceği yönünde doğru bir sınıflama yapmıştır.

Çizelge 3. 11. Markov modeline göre verilerin sınıflandırılması

SIRA	YAŞ	CİNSİYET	ULKE	KRONİK	SONUC	SEYAHAT	HASTALIK TARİHİ	SEMPTOM	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

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 Çizelge 3.12.'de verilmiştir. Tan modeli bu 17 hastanın 15'ini doğru sınıflandırırken 2 hastayı yanlış sınıflandırmıştır. Markov modeli bu 17 hastanın tümünü doğru sınıflandırmıştır.

Çizelge 3. 12. 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

Yapılan analiz sonucunda Şekil 3.6'da görüldüğü üzere Tan Modeli test verilerinin %98,97'sini doğru sınıflarken, %1,03'ünü yanlış sınıflamıştır. Markov Modeli test verinin %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 Markov Modeli üzerinden yapılmıştır.

Results for output field SONUC

Individual Models

Comparing \$B-SONUC with SONUC

'Partition'	Testing		Training	
Correct	43.008	98,97%	170.741	98,97%
Wrong	446	1,03%	1.773	1,03%
Total	43.454		172.514	

Comparing \$B1-SONUC with SONUC

'Partition'	Testing		Training	
Correct	43.039	99,04%	171.043	99,15%
Wrong	415	0,96%	1.471	0,85%
Total	43.454		172.514	

Agreement between \$B-SONUC \$B1-SONUC

'Partition'	Testing		Training	
Agree	43.261	99,56%	171.926	99,66%
Disagree	193	0,44%	588	0,34%
Total	43.454		172.514	

Comparing Agreement with SONUC

'Partition'	Testing		Training	
Correct	42.939	99,26%	170.598	99,23%
Wrong	322	0,74%	1.328	0,77%
Total	43.261		171.926	

Şekil 3. 6. Test ve deneme modelleri için karşılaştırma oranları

Çizelge 3.13'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.

Çizelge 3. 13. Yeni hastanın özelliklerine göre Markov modelinde sınıflandırması

SIRA	YAS	CINSİYET	ULKE	KRONİK	SONUC	SEYAHAT	HASTALIK TARİHİ	SEMPTOM	HASTALIK 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

4. SONUÇ VE TARTIŞMA

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 sađlık idaresi gibi önemli kuruluřların ve kiřilerin lkemizde yařanan Covid-19 ile savařta ayrıca desteklenmelidir. alıřılan veri kmesi Covid-19'un farklı yař aralıklarını farklı etkilemekte olduđunu gstermektedir. Bu saptamadan yola ıkarak halkı yařa gre sınıflandırarak sađlık politikaları hem maliyetleri azaltacak hem de tıbbi mdahalenin daha dođru uygulamasını sađlayacaktır.

Bu alıřmaya ek olarak, daha ok veriye ulařılıp alıřma detaylandırılabilir. Trkiye verisine ulařıp daha ok deđiřken ekleyerek daha verimli sonulara ulařılabilir. Bu alıřmayı herkes tarafından kolayca ulařılabilecek bir uygulama haline dnřtrp insanların olası bir hastalık durumunda riski llebilir. Koronavirs hayatımızda ne kadar zaman daha kalacađı belirli deđilken bu alıřmada hastalıkla karřılařan bireylerin hastalık sonucu hayatta kalıp kalamayacađı ile ilgili Trkiye verileri ile alıřılabilir.

Bayes Ađlarının sađlık alanında kullanımı yaygındır. Bir kiřide ortaya ıkan semptomlara bakılarak bu kiřinin herhangi bir hastalıđa sahip olma olasılıđı hesaplanabilir. Buna benzer olarak neden-sonu iliřkisi olan birok olayın olasılıđı bu modelleme ile grselleřtirilebilir. Covid-19 ile ilgili verilere de Bayes Ađları uygulaması yapılmaktadır. Bayes Ađları ile riskli hastalar belirlenir ve tanı testleri ilk etapta riskli hastalar zerinde yapılabilir. Buradan elde edilen model sonuları ile kiři kendi verilerini girerek olası bir hastalık durumunda kendinin hayatta kalıp kalamayacađını hesaplayabilir ve bu sayede ek nlemler alabilir.

Bu tez alıřmasında 'https://github.com/beoutbreakprepared/nCoV2019/blob/master/latest_data/latestdata.tar.gz' adresinden alınan 19 lkeye ait 10.309.980 Covid-19 hastası iinden seilen 215.968 hastadan oluřan 9 deđiřkenli veri kmesi kullanılmıřtır. Bađımlı deđiřken ise hastanın koronavirs'e yakalandıktan sonra hayatta kalıp kalmadıđı alınmıřtır. Ek olarak hastanın kronik rahatsızlıđı olup olmaması, son  ay iinde yurtdıřı seyahatinde bulunup bulunmadıđı, cinsiyeti ve hangi ayda hastalıđa yakalandıđı bilgileride kullanılmıřtır. Bu deđiřkenlere iin Markov ve Tan modelleri uygulanmıř; Tan Modeli test verilerinin %98,97'sini dođru sınıflarken, Markov Modeli test verinin %99,04'n dođru sınıflamıřtır. Buradan elde edilen sonular ile sisteme yeni giren bir hastanın hayatta kalma olasılıkları hesaplanmıřtır.

Bu analiz yurtdışındaki benzer verilerle çalışılmış olup %92,3 doğru sınıflama oranı yakalanmıştır. Türkiye içinde benzer veriler elde edildiği taktire, 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 taktirde de uygulaması yapılarak elde edilen sonuçlar karar vericiler için yararlı bilgi niteliğinde olacaktır.



KAYNAKLAR

- [1] Afyon Kocatepe Üniversitesi, Covid-19 (Yeni Koronavirüs Hastalığı) Nedir? Bilgilendirme Sayfası, <https://aku.edu.tr/covid19/>, (Erişim Tarihi: 29.12.2021).
- [2] Lu H. Stratton CW. Tang YW., Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle, 92:401–402, **2020**.
- [3] Zhao S. Musa SS. Lin Q. Ran J. Yang G. Wang W. Et al., Estimating the Unreported Number of Novel Coronavirus (COVID- 19) Cases in China in the First Half of January, 9(2), 388, **2020**.
- [4] Hui DS. Azhar EI. Madani TA. Ntoumi F. Kock R. Dar O. et al. 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, **2020**.
- [5] Yeni Koronavirüs “COVID-19” Nedir? <https://www.yeditepe.edu.tr/tr/duyuru/yeni-koronavirus-COVID-19-nedir> (Erişim Tarihi: 26.03.2021).
- [6] Tesini BL. Coronaviruses and Acute Respiratory Syndromes (COVID-19, MERS, and SARS) Coronaviruses and Acute Respiratory Syndromes (COVID-19, MERS, and SARS). <https://www.msmanuals.com/professional/infectiousdiseases/respiratoryviruses/coronaviruses-and-acute-respiratory-syndromes-COVID-19,-mers,-and-sars> (Erişim Tarihi: 23.03.2021).
- [7] Koronavirüs neden hızlı yayıldı? Corona Virüsü. <https://www.epsikiyatri.com/koronavirus-neden-hizli-yayildi-corona-virusu> (Erişim Tarihi: 27.03.2020)
- [8] Zhou Y, Yang Y, Huang J, Jiang S, Du L. Advances in MERS-CoV Vaccines and Therapeutics Based on the Receptor-Binding Domain. Viruses. 11(1), 60, **2019**.
- [9] Ceric, Mustafa. Kovid-19 (Koronavirüs) Salgınının Ekonomik Etkileri. Politika Notu, Ankara Yıldırım Beyazıt Üniversitesi, **2020**.
- [10] Sağlık Bakanlığı, Covid-19 Rehberi. Bilgilendirme Rehberi, T.C. Sağlık Bakanlığı, (Erişim Tarihi: 29.01.2022).
- [11] Kalıkov, Anaberk, Veri Madenciliği ve Bir E-Ticaret Uygulaması, Yüksek Lisans Tezi, Gazi Üniversitesi, **2006**.
- [12] Frawley, W.J., Piatetsky-Shapiro ve Matheus C.J., Knowledge Discovery in Databases: An Overview, AI Magazine Vol 13(3), 57, **1992**.

- [13] “What is Big Data?”, Oracle, <https://www.oracle.com/tr/big-data/guide/what-is-big-data.html>, (Erişim tarihi: 27 Mart 2020).
- [14] “The Role of Big Data in Global Epidemics”, University of Nevada, <https://onlinedegrees.unr.edu/blog/the-role-of-big-data-in-global-epidemics>, (Erişim tarihi: 30 Mart 2020).
- [15] Mikael Hagstrom, “From SARS to Ebola, How Big Data Fight Disease”, World Economic Forum, 2 Aralık 2014, <https://www.weforum.org/agenda/2014/12/from-sars-to-ebola-how-big-data-fights-disease>, (Erişim tarihi: 30 Mart 2020).
- [16] “Ebola Tracking System for Sierra Leone Offered by IBM”, BBC, (Erişim tarihi: 27 Kasım 2021).
- [17] Trucco, P., Cagno, E., Ruggeri, F. ve Grande, O., A Bayesian Belief Network Modelling of Organisational Factors in Risk Analysis: A Case Study in Maritime Transportation, Reliability Engineering and System Safety, 93(6), 845-856, **2008**.
- [18] Norrington, L., Quigley, J., Russel, A. ve Van der Meer, R., Modelling the Reliability of Search and Rescue Operations with Bayesian Belief Networks, Reliability Engineering and System Safety, 93(7), 940-949, **2008**.
- [19] Lauria, E. J. M. ve Duchessi, P. J., A Bayesian Belief Network for IT Implementation Decision Support, Decision Support Systems, 42(3), 738-742, **2006**.
- [20] Lee, E., Park, Y. ve Shin, J. G., Large Engineering Project Risk Management Using a Bayesian Belief Network, Expert Systems with Applications, 36(3), 5880-5887, **2009**.
- [21] Korb, K. B. ve Nicholson A. E., Bayesian Artificial Intelligence, Chapman & Hall/CRC, 23-25 Blades Court, London, UK, **2004**.
- [22] Sigurdsson, J. H., Walls, L. A. ve Quigley, J. L., Bayesian Belief Nets for Managing Expert Judgement and Modelling Reliability, Quality and Reliability Engineering International, 17, 181-190, **2001**.
- [23] Mahadevan, S., Zhang, R. ve Smith, R., Bayesian Networks for System Reliability Reassessment, Structural Safety, 23(3), 231-251, **2001**.
- [24] Newbold, P., İşletme ve İktisat İçin İstatistik, (U. Senesen, Çevir.) Prentice Hall, Inc. , NJ,USA, **1995**.
- [25] Korb, Kevin B.; Nicholson, Ann E. Bayesian artificial intelligence (Computer science and data analysis series). **2011**.
- [26] Ben-Gal, Irad, et al. Encyclopedia of statistics in quality and reliability. chapter Bayesian Networks, Wiley & Sons, **2007**.

- [27] Newbold, Paul. İşletme ve İktisat İçin İstatistik,[Çev Şenesen Ü.], 4. Basım, Literatür Ya., İstanbul, **2008**.
- [28] Cinar, D. ve Kayakutlu, G., Scenario Analysis Using Bayesian Networks: A Case Study in Energy Sector, Knowledge-Based Systems, 22(3), 267-276, **2010**.
- [29] Heckerman, D., A Tutorial on Learning with Bayesian Networks, Microsoft Corporation, Redmond, pp, 33-82, **1996**.
- [30] Trucco, P., Cagno, E., Ruggeri, F. ve Grande, O., A Bayesian Belief Network Modelling of Organisational Factors in Risk Analysis: A Case Study in Maritime Transportation, Reliability Engineering and System Safety, 93(6), 845-856, **2008**.
- [31] Ha, J. S. ve Seong, P. Hoang, A Method for Risk-Informed Safety Significance Categorization Using the Analytic Hierarchy Process and Bayesian Belief Networks, Reliability Engineering and System Safety, 83(1), 1-15, **2004**.
- [32] Gran, B. A., Assessment of Programmable Systems Using Bayesian Belief Nets, Safety Science, 40(9), 797-812, **2002**.
- [33] Jones, B., Jenkinson, I., Yang, Z. ve Wang, J., The Use of Bayesian Network Modelling for Maintenance Planning in a Manufacturing Industry, Reliability Engineering and System Safety, 95(3), 267-277, **2010**.
- [34] De Melo, A. C. V. ve Sanchez, A. J., Software Maintenance Project Delays Prediction Using Bayesian Networks, Expert Systems with Applications, 34(2), 908-919, **2008**.
- [35] Mazumdar, S., Seybold, D., Kritikos, K., & Verginadis, Y., A survey on data storage and placement methodologies for cloud-big data ecosystem. Journal of Big Data, 6(1), 15, **2019**.
- [36] Siddiqa, A., Karim, A., & Gani, A., Big data storage technologies: A survey. Frontiers of Information Technology & Electronic Engineering, 18(8), 1040-1070, **2017**.
- [37] GoogleCloud , Büyük veri nedir. Erişim Adresi: <https://cloud.google.com/what-is-big-data?hl=tr>, (Erişim Tarihi: 27.04.2022)
- [38] Wu, J., Wang, J., Nicholas, S., Maitland, E., & Fan, Q., Application of big data technology for covid-19 prevention and control in china: Lessons and recommendations. Journal of Medical Internet Research, 22(10), e21980, **2020**.
- [39] Mariani, M., Big data and analytics in tourism and hospitality: A perspective article. Tourism Review, 75(1), 299-303, **2019**.
- [40] Phillips-Wren, G., Ai tools in decision making support systems: A review. International Journal on Artificial Intelligence Tools, 21(2), e1240005, **2012**.

- [41] Liu, S., Artificial intelligence (ai) worldwide - statistics & facts. Eriřim Adresi: <https://www.statista.com/topics/3104/artificial-intelligence-ai-worldwide/>, (Eriřim Tarihi: 23.02.2022)
- [42] Oracle., Make ai collaboration your next b2b marketing move, Eriřim Adresi: <https://blogs.oracle.com/cx/make-your-next-b2b-marketing-move-ai-collaboration>, (Eriřim Tarihi: 02.12.2021)
- [43] Seagate., Rethink data: Put more of your business data to work from edge to cloud.
EriřimAdresi:https://www.seagate.com/files/wwwcontent/ourstory/rethink%20data/files/Rethink_Data_Report_2020.pdf, (Eriřim Tarihi: 23.11.2021)
- [44] Jeffcock, P., What's the difference between ai, machine learning, and deep learning? Eriřim Adresi:<https://blogs.oracle.com/bigdata/differenceaimachinelearningdeeplearning#:~:text=Machine%20learning%20is%20a%20subset,to%20solve%20more%20complex%20problems.>, (Eriřim Tarihi: 17.08.2021).
- [45] Shapoval, V., Wang, M. C., Hara, T., & Shioya, H., Data mining in tourism data analysis: Inbound visitors to japan. Journal of Travel Research, 57(3), 310-323, **2018**.
- [46] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y., Deep learning, (1), MIT press Cambridge, **2016**.
- [47] Wolfewicz, A., Deep learning vs. Machine learning – what's the difference? Eriřim Adresi: <https://www.levity.ai/blog/difference-machine-learning-deep-learning>, (Eriřim Tarihi: 15.03.2022)
- [48] Acumen., Deep learning market global industry analysis, market size, opportunities and forecast, 2019 - 2026. Acumen Research and Consulting. Eriřim Adresi: <https://www.acumenresearchandconsulting.com/deep-learning-market>, (Eriřim Tarihi: 11.05.2021).
- [49] Lauria, E. J. M. ve Duchessi, P. J., A Bayesian Belief Network for IT Implementation Decision Support, Decision Support Systems, 42(3), 1573-1588, **2006**.
- [50] Hanea, D. ve Ale, B., Risk of Human Fatality in Building Fires: A Decision Tool Using Bayesian Networks, Fire Safety Journal, 44(5), 704-710, **2009**.
- [51] Gemela J., Financial analysis using Bayesian Networks, Applied Stochastic Models in Business and Industry, 17(1), 57-67, **2001**.

- [52] Verma, T., and J. Pearl. Equivalence and synthesis of causal models, *Uncertainty in Artificial Intelligence*, Vol. 6, P. Bonissone, M. Henrion, LN Kanal and JF Lemmer., 255-268, **1990**.
- [53] Spirtes, P., Glymour, C., and Scheines, R., *Causation, Prediction, and Search*. IT Press., **2000**.
- [54] Scutari, M. Measures of Variability for Graphical Models. PhD thesis, *Universita degli Studi di Padova, Dipartimento di Scienze Statistiche*, 10(2), 872-873, **2011**.
- [55] Margaritis, D., *Learning Bayesian Network Model Structure from Data*. PhD thesis, School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA. CMU-CS 03-153, **2003**.
- [56] Tsamardinos, Ioannis, et al. Algorithms for large scale Markov blanket discovery. In: *FLAIRS conference*. 2003. p. 376-380, **2003**.
- [57] Yaramakala, S., Margaritis, D., *Speculative Markov Blanket Discovery for Optimal Feature Selection*, In *Proceedings of the 5th IEEE International Conference on Data Mining*, pages 809-812, **2005**.
- [58] Neapolitan, Richard E. *Learning bayesian networks*. Vol. 38. Upper Saddle River: Pearson Prentice Hall, **2004**.
- [59] D. J. Spiegelhalter, A. P. Dawid, S. L. Lauritzen, ve R. G. Cowell, *Bayesian analysis in expert systems*”, *Statistical Science*, 8(3), 219-247, **1993**.
- [60] Jensen, Finn V., and Thomas Dyhre Nielsen. *Bayesian networks and decision graphs*. Vol. 2. New York: Springer, **2007**.
- [61] Galapero, J., Fernández, S., Pérez, C. J., Calle-Alonso, F., Rey, J., & Gómez, L. Identifying risk factors for ovine respiratory processes by using Bayesian networks. *Small Ruminant Research*, 136, 113-120, **2016**.
- [62] Heckerman, David. A tutorial on learning with Bayesian networks. *Innovations in Bayesian networks*, 33-82, **2008**.
- [63] J. Lu, C. Bai, ve G. Zhang, “Cost-benefit factor analysis in e-services using bayesian networks”, *Expert Systems with Applications*, 36(3), 4617-4625, **2009**.
- [64] Pearl, Judea. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann, **1988**.
- [65] Shenoy, Catherine, and Prakash P. Shenoy. "Bayesian network models of portfolio risk and return." *The MIT Press*, **2000**.

- [66] C. A. Pollino, O. Woodberry, A. Nicholson, K. Korb, ve B. T. Hart, "Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment", *Environmental Modelling & Software*, 22(8), 1140-1152, **2007**.
- [67] Heckerman, D. A tutorial on learning with Bayesian networks. *Innovations in Bayesian networks*, 13-18, **2008**.
- [68] Scutari, M., and S. Lebre. "Bayesian networks in R: with applications in systems biology." **2013**.
- [69] P. Antal, G. Fannes, D. Timmerman, Y. Moreau, ve B. De Moor, "Using literature and data to learn Bayesian networks as clinical models of ovarian tumors", *Artificial Intelligence in Medicine*, 30(3), 257-281, **2004**.
- [70] P. P. Rodrigues, D. Ferreira-Santos, A. Silva, J. Polónia, ve I. Ribeiro-Vaz, "Causality assessment of adverse drug reaction reports using an expert-defined Bayesian network", *Artificial Intelligence in Medicine*, 91(12), 22, **2018**.
- [71] H. Olmuş, "Bayes Ağlar ve Markov Ağları Kullanan Kümeleme Yönteminin İncelenmesi ve Bir Uygulama", Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Gazi Üniversitesi, Ankara, Türkiye, **2007**.
- [72] Scutari, Marco; DENIS, Jean-Baptiste. *Bayesian networks: with examples in R*. Chapman and Hall/CRC, **2014**.
- [73] E. Dünder, "Bayesçi Ağlarda Öğrenme Algoritmalarının Karşılaştırılması", Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Ondokuz Mayıs Üniversitesi, Samsun, Türkiye, **2013**.
- [74] D.Ersel, "Birliktelik Analizinde Özgün Bir Birleşik İlginçlik Ölçümü", Doktora Tezi, Fen Bilimleri Enstitüsü, Hacettepe Üniversitesi, Ankara, Türkiye, **2012**.
- [75] Wendler, Tilo, and Sören Gröttrup. 2016. *Data Mining with SPSS Models Theory, Exercises and Solutions*, Cham Springer International Publishing, <https://doi.org/10.1007/978-3-319-28709-6>, **2016**.
- [76] IBM Corp, *IBM SPSS Statistics for Windows, Version 23.0*. Armonk, New York, U.S.A, **2016**.