T.C.
İNÖNÜ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

VERİLER ARASINDAKİ İLİŞKİLERİN BELİRLENMESİ VE BAYES AĞININ OLUŞTURULMASI

YÜKSEK LİSANS TEZİ

Elif Aslı OYMAK

Bilgisayar Bilimleri Anabilim Dalı

Tez Danışmanı: Prof. Dr. Ali KARCI

OCAK 2021

T.C.
İNÖNÜ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

VERİLER ARASINDAKİ İLİŞKİLERİN BELİRLENMESİ VE BAYES AĞININ OLUŞTURULMASI

YÜKSEK LİSANS TEZİ

Elif Aslı OYMAK
(36173619012)

Bilgisayar Bilimleri Anabilim Dalı

Tez Danışmanı: Prof. Dr. Ali KARCI

OCAK 2021

TEŞEKKÜR VE ÖNSÖZ

Yüksek lisans tez sürecimin başlangıcından bitişine kadar her aşamasında destek ve önerilerini benimle paylaşan, bana güç veren değerli danışman hocam Sayın Prof. Dr. Ali Karcı' ya;

Bölüm başbakanımız Sayın Prof.Dr. Celalleddin Yeroğlu'na ve tüm bölüm çalışanlarına;

Çalışmalarım boyunca yanımda olan Ayşe Danışmanoğlu, Arş. Gör. Sara Altun, Arş. Gör. Oya Köksal, Arş. Gör. Zeynep Özdemir ve Arş. Gör. Fırat Orhan Bulucu' ya;

Ayrıca tüm hayatım boyunca olduğu gibi tez çalışmalarım süresince de benden desteklerini esirgemeyen Ailem'e;

teşekkür ederim.

ONUR SÖZÜ

Yüksek Lisans Tezi olarak sunduğum ” Veriler Arasındaki İlişkilerin Belirlenmesi ve Bayes Ağının Oluşturulması “ başlıklı bu çalışmanın bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın tarafımdan yazıldığını ve yararlandığım bütün kaynakların, hem metin içinde hem de kaynakçada yöntemine uygun biçimde gösterilenlerden oluştuğunu belirtir, bunu onurumla doğrularım.

Elif Aslı OYMAK

İÇİNDEKİLER

ÇİZELGELER DİZİNİ

ŞEKİLLER DİZİNİ

ÖZET

Yüksek Lisans Tezi

Veriler Arasındaki İlişkilerin Belirlenmesi ve Bayes Ağının Oluşturulması

Elif Aslı Oymak

İnönü Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Bilimleri Anabilim Dalı

49+ vii sayfa

2020

Danışman : Prof. Dr. Ali Karcı

Bu tez çalışmasında, veriler arasındaki ilişkiler istenilen şartlara göre filtreleme yapılarak belirlenmiş, Bayes Ağı oluşturulmuş, sonuçların doğruluk oranları hesaplanmıştır. Veriler arasındaki ilişkiler belirlenirken ve Bayes Ağı oluşturulurken Naive Bayes fonksiyonu ve Tepe Tırmanma Algoritması kullanılmıştır. Ardından veri setinden istatistiki bilgiler elde edilmiştir. Bir verinin özelliklerinden yola çıkarak sonucun gerçekleşmesinin olasılıksal hesabı , verilerin birbirleriyle aralarındaki koşullu olasılıkları açıklanmıştır. İleriki bölümlerde gereken teknikler ve kullanılan yöntemler açıklanmış, veriler arasındaki ilişkilerin olasılık hesabı Tepe Tırmanma Algoritması ile bulunmuş ve ilişkilere ait Bayes Ağ yapısı graf olarak çizdirilmiştir. Matematiksel hesaplamalardan oluşan bu teknikler ve yöntemler RStudio çalışma ortamı ve R dili ile yazılan kod satırları ile gösterilmiştir. Oluşturulacak graf da RStudio kütüphanesinden faydalanılarak kod satırları ile gösterilmiştir. Böylece veriler arasındaki ilişkiler sayısal ve görsel olarak gösterilmiştir. Özelliklerin her birinin koşullu olasılıkları çizelgeler şeklinde gösterilecektir. Bir başka deyişle, özelliklerin bağlı olduğu niteliğe göre gerçekleşme ihtimali hesaplanmıştır ve bu hesaplanan değer sınıf olarak adlandırılan niteliğe göre elde edilmiştir. Sonrasında özelliklerin birbiri ile olan koşullu olasılık değerleri hesaplanmış ve sonuçları gösterilmiştir. Sonuç olarak, örnek bir veri setiyle Bayes Ağının oluşturulması bulgularıyla anlatılmıştır.

ANAHTAR KELİMELER: Veri Madenciliği, Bayes Ağları, Tepe Tırmanma Algoritması

ABSTRACT

Master Thesis

Determining Relationships Between Data and Creating Bayesian Network

Elif Aslı Oymak

Inonu University
Institute of Science
Department of Computer Science

49+ vii pages

2020

Supervisor : Prof. Dr. Ali Karcı

In this thesis, the relationships between the data were determined by filtering according to the desired conditions, a Bayesian Network was created, and the accuracy of the results was calculated. Naive Bayes function and Hill Climb Algorithm were used while determining the relationships between the data and creating the Bayes Network. Then, statistical information was obtained from the data set. The probabilistic calculation of the realization of the result based on the properties of a data, and the conditional probabilities of the data with each other are explained. In the following chapters, the required techniques and the methods used are explained, the probability calculation of the relationships between the data is found with the Hill Climb Algorithm and the Bayes Network structure of the relations is graphed. These techniques and methods consisting of mathematical calculations are shown with the RStudio working environment and lines of code written in R language. The graph to be created is shown with lines of code using the RStudio library. Thus, the relationships between data are shown numerically and visually. The conditional probabilities of each of the properties will be shown in tabular form. In other words, the probability of occurrence of the features according to the attribute to which they depend is calculated and this calculated value is obtained according to the quality called class. Then, conditional probability values of the properties with each other were calculated and the results were shown. As a result, the creation of the Bayes Network with an example data set is explained with its findings.

KEYWORDS: Data Mining, Bayesian Networks, Hill Climb Algorithm.

1. GİRİŞ

Günümüzde Dünya nüfusunun yarısından çoğu internet ve araçlarını kimi zaman bilgi edinmek, kimi zaman mobil veya web platformlarda sunulan uygulamaları kullanmak ve faydalanmak gibi sebeplerle kullanmaktadır. Bu aşamada işlemlerin hatasız şekilde uygulanması için veriler istenilmektedir. Bu veri setleri kullanıcıları analiz etmek için depolanmaktadır. Dolayısıyla verilerin boyutu ciddi bir şekilde artmaya devam etmektedir. Veri madenciliği, büyük miktarda veriden faydalı desenler bulan bir süreçtir [1]. Veri madenciliği birçok sınıflandırma yöntemine sahiptir. Bu tez çalışmasında bu yöntemlerden olasılık tabanlı olan, istatiksel işlemler yapacağımız Bayes sınıflandırıcıları ve ağları kullanılacaktır. İstenilen formata dönüştürdüğümüz veri seti üzerinde Bayes teoremine dayanarak koşullu olasılıklar hesaplanması, Bayes ağının oluşturulması ve sonuçların doğruluk oranının hesaplanması amaçlanmaktadır. Bu kapsamda R Studio ortamında R dili ile sınıflandırma yöntemlerinin verilerin birbiriyle ilişkilerinin bulunabileceği ve de sonuç çıkarımının inceleneceği bir uygulama geliştirilmiştir.

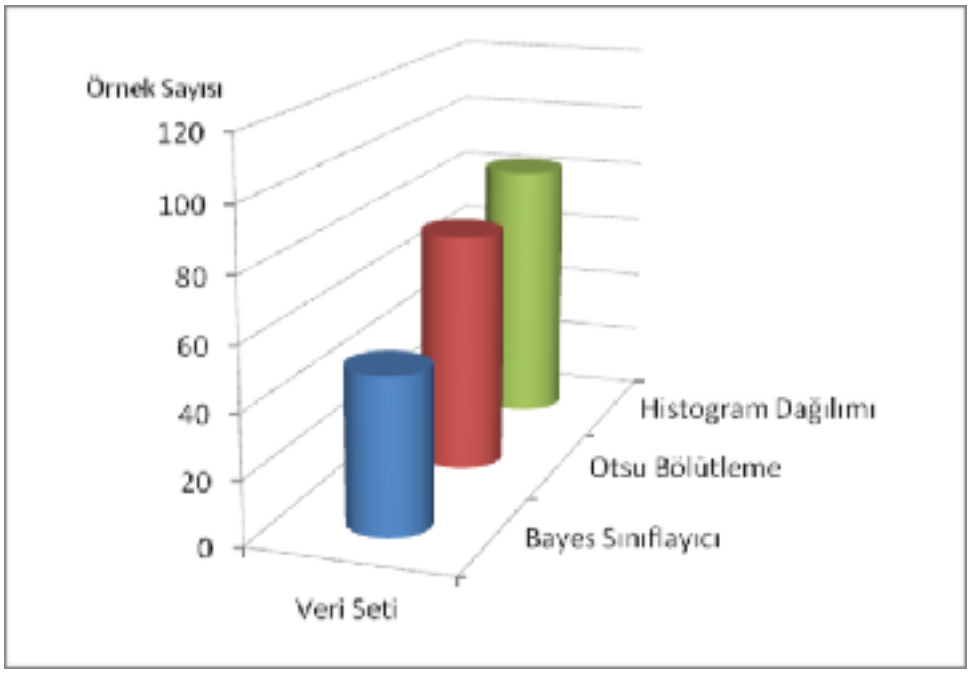
Tez içeriğiyle ilgili literatür taraması yapılmış, önceki çalışmalarda kullanılan teknik, yöntem ve kavramlar hakkındaki incelemeler tezin 2. bölümünde paylaşılmıştır. Uygulama için kullanılan teknik ve yöntemler literatürden elde edilen bilgiler aracılığıyla tezin 3. bölümünde paylaşılmıştır. Geliştirilen uygulama ve uygulamanın sonuçları 4. ve 5. bölümde kendi ana başlıkları altında paylaşılmıştır. Çalışmanın sonucunda elde edilen bulgular ve sonuçlar 6. bölümde incelenmiştir.

Naive Bayes ve Bayes Ağları verilerin hangi sınıflara ait olabilecekleri olasılığını tahmin eden sınıflandırıcılardır [2]. Kaynaklarda Bayes Ağları basit bir sınıflandırıcı olarak tanımlansa da bunun aksine çok etkilidir [3].

N.B. Sebik ve H.İ. Bülbül, akciğer kanseri veri seti üzerinde veri madenciliği modellerinin başarılarının analiz edilmesi üzerine bir çalışma yapmıştır. Yapılan çalışmada akciğer kanseri teşhisinde literatüre katkı sağlayacak bir veri seti toplanmıştır. Elde edilen veri setine çeşitli algoritmalar WEKA yazılım ortamında

uygulanmıştır. Çalışmada veriler ayrıntılı olarak kontrol edilip standart bir hale dönüştürülmüştür. Ardından ön işleme süreçleri tamamlanmış ve WEKA kullanılarak veri setine farklı algoritmalar uygulanıp modeller çıkartılmıştır. Sonuç olarak en etkili algoritma Naive Bayes algoritması olarak tespit edilmiştir [4].

B. Kır Savaş v.d. çalışmasında önce yapılmış çalışmalardaki öneriler üzerine gölge tespit yöntemlerinden Bayes Sınıflandırma Yöntemi, Otsu Bölütleme Yöntemi ve Histogram Dağılımı Yöntemini inceleyerek görüntü seti üzerinde test etmiştir. Çalışmada tüm uygulamalar için elde edilen test sonuçları karşılaştırılarak 3 algoritmanın da gölge tespitindeki başarımları sunulmuştur. Kullanılan veri seti üzerinde Bayes Sınıflandırma Yöntemi ile bulunan başarım oranı % 49, Otsu Bölütleme Yöntemi ile bu başarım oranı % 75 ve Histogram Dağılımı Yöntemi ile ise % 83' tür. Başarım grafiği kullanılan yöntemlere göre Şekil 1.1.' deki grafikte gösterilmektedir [5].



Şekil 1.1. Uygulanan yöntemlerin başarım grafiği [5].

M.O. Olgun ve G. Özdemir, Kontrol Grafiklerinde Örüntü Tanıma üzerine İstatiksel Özellik Temelli Bayes Sınıflandırıcı kullanarak çalışma yapmıştır. Dolayısıyla, sınıflandırıcıların test ve performans özelliklerini ölçmek için 5 farklı(5x900) örnek kümesi oluşturulmuştur. Ham veri ve eşitliklerden oluşturulan Bayes Örüntü Sınıflandırıcılarının eğitim ve test durumlarındaki sınıflandırma oranları sonucu istatiksel özellikler kullanılarak elde edilip Çizelge 1.1.' de verilmiştir. Çalışma çıktılarına göre, Bayes sınıflandırıcının iyi bir performans sergilemesinden dolayı, gerçek zamanlı örüntü tanıma çalışmalarında bu sınıflandırıcı tavsiye edilmektedir. Yine çıktılara göre; Bayes Örüntü Tanıyıcı, Yapay Sinir Ağlarına kıyasla sınıflandırma performansında daha başarılıdır. Bu tür gerçek zamanlı kontrol grafikleri çalışmalarında Bayes Sınıflandırıcısının örüntü tanıma hedefli kullanılabileceği sonucu çıkarılmıştır [6].

Çizelge 1.1. Ham veri ve istatiksel özellik kullanarak elde edilen Bayes eğitim ve test performansları.

Ham Veri Bayes Örüntü Tanıyıcı		
Sınıflandırıcı Numarası	Eğitim Sınıflandırma Yüzdesi	Test Sınıflandırma Yüzdesi
3,1	100	90,22
3,2	100	91,56
3,3	100	92,89
3,4	100	90,89
3,5	100	92,22
Ortalama	100	91,56
Standart Sapma	0	1,05
Özellik-Temelli Bayes Örüntü Tanıyıcı		
Sınıflandırıcı Numarası	Eğitim Sınıflandırma Yüzdesi	Test Sınıflandırma Yüzdesi
4,1	99,78	99,33
4,2	99,78	99,11
4,3	99,56	98,67
4,4	99,56	99,11
4,5	99,56	98,89
Ortalama	99,65	99,02
Standart Sapma	0,12	0,25

R.Solmaz v.d. , Fonksiyonel Tiroit Hastalığı teşhisinde Naive Bayes Sınıflandırıcının kullanılması üzerine çalışma yapmıştır. Yapılan çalışmada, Naive Bayes Sınıflandırıcı, kan değerleri tabanlı iki veri setine uygulanmıştır. Sınıflama doğruluğu önerilen teknikle veri setleri % 97,20 ve % 95,04 oranında sınıflandırılmıştır. Kazanılan sonuçlara göre önerilen sınıflama tekniği kan değerleri temelli tiroit tanılama sistemi için kullanılabilmektedir. Ayrıca Naive Bayes Sınıflandırıcının tiroit hastalığı teşhisinde % 95' ten daha başarılı olduğu ve hazırlanacak karar destek sistemine entegre edilebileceği sonucuna varılmıştır [8].

Bu çalışmada, literatürde belirlenen özelliklere göre; tedavi yöntemlerinden olan immunotherapy yönteminin, hastaya uygulanıp uygulanmaması konusunda veri madenciliği yöntemleri ile ön bir değerlendirme yapılmış ve değerlendirme başarı oranının artırılması sağlanmıştır. Böylece hekime tedavi yöntemini seçerken, immunotherapy yöntemini seçip seçmeme konusunda daha doğru karar vermesi için yardımcı olunabilecektir. Başarı oranının artırabilmek için veri seti üzerinde birçok yöntem denenmiştir. Gözlemlenen sonuçlara göre en yüksek başarı oranı, Bayes net ile yapılan sınıflandırmada %85.55 olarak görülmüştür. Yapılan çalışma ile en iyi tedavi yöntemini seçmede hekimlere yardımcı olmanın yanı sıra hastalara zaman kazandırmak, tedavi maliyetini düşürmek ve tedavi kalitesini iyileştirmek gibi birçok fayda sağlanacaktır [8] .

Naive Bayes sınıflandırması, Bayes teoreminden geliştirilmiş bir yöntem olup Thomas Bayes toplam olasılık formülünün tersini alıp hesaplayarak oluşturduğu formül, Bayesci yaklaşımın zeminini oluşturmuştur [9, 10]. NB sınıflandırıcı çoğunluk olarak tıbbi teşhis ve metin belgelerinin sınıflandırılması için kullanılmaktadır [11]. Bayes teoremini temel alan ve büyük veri setleri için kullanışlı olan istatistik tabanlı Naive Bayes sınıflandırma algoritmasının uygulanabilmesi için tahmin ediciler birbiriyle bağımlı olmamalıdır [12]. Naive Bayes sınıflandırma algoritmasının eğitim verisi üzerinde yapılan olasılık hesaplamalarıyla test edilecek verilerin hangi sınıf içine dahil olacağı bulunmaya çalışılmaktadır. Eğitim için kullanılacak veri ne kadar fazla ise test verisinin ait olduğu sınıfı bulma olasılığı artmaktadır [13].

Felsefi olarak çeşitli olasılık değerlerinin objektif bir nitelik değil, gözlemci tarafından meydana çıkarılan subjektif bir değer olarak kabul edilen subjektivist olasılık düşünürlerinin görüşüne göre Bayesian teoremi, yeni bilgiler aracılığıyla olasılık değeri ile ilgili subjektif inanışların güncelleştirilip değiştirilmesine olanak veren temel bir gereçtir; dolayısıyla sonsal bir yaklaşımın temeli olduğu ifade edilmektedir. Naive Bayesian, tahminci ve tanımlayıcı bir sınıflama algoritması olup hedef değişkenle bağımsız değişkenler arasındaki bağlantıyı analiz etmektedir[14].

Bayes karar verme kuralı öznitelikler arasında bulunan bazı ilişkiler ve bağımlılıklar gösterilememiş olmasına rağmen birçok sınıflandırma probleminde oldukça etkili sonuçlar vermiştir[15].

Naive Bayes, bir modeli yani veri setini öğrenirken, öğrenme kümesinde her çıktının kaç defa tekrarlandığını hesaplar. Hesap sonucu elde edilen bu değer, öncelikli olasılık olarak isimlendirilmiştir. Örnek verirsek; bir banka kredi kartı başvurularını "iyi" ve "kötü" olmak üzere iki sınıf şeklinde gruplandırmak istemektedir. İyi sınıf çıktısı toplam 10 vaka içinde 4 kere tekrarlandıysa iyi sınıf çıktısı için öncelikli olasılık 0,4'tür. Bunun sonucunda, "Kredi kartı başvurusu yapan bir kişi ile ilgili hiçbir şey bilinmiyorsa, bu kişi 0,4 olasılıkla iyi sınıf grubundadır" olarak ifade edilir. Ayrıca Naive Bayes her bağımsız değişken ve bağımlı değişken kombinasyonunun gerçekleşme sıklığını bulur. Bulunan sıklıklar öncelikli olasılıklarla birlikte tahminler için kullanılır [14].

Naive Bayesi bir kez daha açıklayacak olursak; genel olarak sonrasal olasılıkları hesaplamak için kullanılan ve rastgele seçilen iki olayın koşullu ve marjinal olasılıklarını ilişkilendiren bir teoremdir. Ayrıca Maksimum Olabilirlik ilkesini temel alan bir teoremdir. Bu durumda Bayes Teoremi, mevcut olasılıkların doğruluk oranını hesaplamak için kullanılabilir [16]. Bu da günlük hayatta birçok alanda seçimler yapmadan önce Naive Bayes teoremine yer verebileceğimizi göstermektedir.

Koşullu olasılık bilgisi ile Bayes formülü oluşturulmaktadır. Ek koşullarla örneklem uzayından ayrılan alt dallardaki olaylara ilişkin olasılıklardır [17]. İki olayın kesişim olasılıklarının marjinal olasılık değerine bölünmesi koşullu olasılığın matematiksel ifadesidir [10]. Bayes formülünde önsel olasılık P(Ci) şeklinde gösterilmektedir ve sınıflandırma öncesi değeri elde edilmiş, bilinen sınıfların olasılığıdır. Sonsal olasılık ise P(Xj/Ci) ile gösterilir ve sınıf bilgisi bilinmesi durumundaki koşullu olasılığı ifade etmektedir [17].

Bayes Teoremi : $$P(C_i/X) = \frac{(P(X_1 \cap X_2 \cap \ldots \cap X_p/C_i)*P(C_i)}{P(X_1 \cap X_2 \cap \ldots \cap X_p)}$$ 1.1

Naive Bayes : $$P(C_i/X_1, X_2, \ldots, X_p) = \prod_{j=1}^{p} P(X_j/C_i)*P(C_i)$$ 1.2

Koşullu olasılık üzerinde durmak gerekirse:

P(A\B), A'nın B ile olan koşullu olasılığı olarak ifade edilmektedir. Yani B olayı bilindiği takdirde A olayının gerçekleşme olasılığıdır. Denklem 3.11 ile A'nın B'ye koşullu olasılığı ifade edilmektedir [18] :

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$ 1.3

Denklem 1.1'deki eşitlik göz önüne alınarak ($A \cap B$)' nin olasılığı eşitlik 1.2'deki gibi bulunur:

$$P(A \cap B) = P(A\backslash B).\ P(B)$$ 1.4

Eşitlik 1.1'de verilen koşullu olasılık ifadesi göz önüne alınarak B'nin A'ya koşullu olasılığı eşitlik 1.3'te verilmiştir:

$$P(B\backslash A) = P(B \cap A) / P(A)$$ 1.5

Eşitlik 1.3'te $P(A \cap B)$ yerine eşitlik 1.2'deki eşitlik uygulandığı takdirde eşitlik 1.4'e ulaşılır:

$$P(B\backslash A) = P(A\backslash B).\ P(B) / P(A)$$ 1.6

Veri madenciliğinde bağımsız değişken sayısını p ile ifade edersek p arttıkça sınıf koşullu kesişim olasılıklarının hepsine ulaşmak zor bir hal alacağı için çözümü karmaşık bir hal almaktadır. Naif Bayes metodunda işlemleri kolaylaştırmak amacı ile her bir sınıftaki değişkenlerin yani sınıf koşullu değişkenlerin birbirinden

bağımsız olduğu kabul edilir. Bu kabul çoğu kaynakta "koşullu bağımsızlık" olarak yer almaktadır [19].

Koşullu bağımsızlık kabulü, p(A∩B)=P(A)*P(B) ile ifade edilen temel olasılık kuralı ile Bayes formülünü, sınıf koşullu olasılıkların ifadesi olan P(Xj/Ci) ve önsel olasılıkların ifadesi olan P(Ci)'nin çarpımı olarak basit hale getirmektedir [19]. Paydadaki P(X1∩X2∩…∩Xp) ifadesi, sabit bir değer olduğu için sınıf belirlemede herhangi bir fark yaratmayacağı için göz ardı edilmektedir [20].

İki sınıftan oluşan bir veri setinde sınıf tespiti için aşağıda belirtilen olasılık verilerini ele alalım. Önsel olasılıkları Şekil 1.2 ve Şekil.1.3' te gösterirsek:

P(Riskli grup = Evet) = 2/ 10= 0, 2
P(Riskli grup = Hayır) = 8/ 10= 0, 8

Şekil 1.2. Önsel olasılıklar.

P(Yaş = <30 / Riskli grup = Evet) = 3/ 7= 0, 42
P(Yaş = <30 / Riskli grup = Hayır) = 2/ 3= 0, 66
P(Gelir düzeyi = orta / Riskli grup = Evet) = 1/ 5= 0, 2
P(Gelir düzeyi = orta / Riskli grup = Hayır) = 2/ 5= 0, 4
P(Cinsiyet = Kadın / Riskli grup = Evet) = 3/ 6= 0, 5
P(Cinsiyet = Kadın / Riskli grup = Hayır) = 3/ 4= 0, 75
P(Sigara = Evet / Riskli grup = Evet) = 4/ 7= 0, 57
P(Sigara = Evet / Riskli grup = Hayır) = 1/ 3= 0, 33

Şekil 1.3. Sınıf koşullu olasılıklar.

Yukarıda hesaplanan olasılıklar kullanılarak, 30 yaşın altında, orta gelir düzeyinde, sigara tüketen bir kadın hastanın hastalık riski içerisinde olma ve olmama olasılıkları Şekil 1.4'teki gibi ayrı ayrı hesaplanabilmektedir:

P(X / Riskli grup = Evet) = 0, 42 x 0, 2 x 0, 5 x 0, 57= 0, 023

P(X / Riskli grup = Hayır) = 0, 66 x 0, 4 x 0, 75 x 0, 33= 0, 065

Şekil 1.4. Şartlı olasılık hesabı.

Yeni bir gözlemin sınıfı bilinmiyorsa en ideal sınıfı belirlerken en yüksek olasılık değeri göz önüne alınır. Naive Bayes formülüyle her bir sınıf için bulunan olasılıklar içerisinden en yüksek olasılık değerine sahip olan sınıf, yeni gözlemin ait olduğu sınıf olur [21,10]. Yukarıdaki örnekte hesaplamalar sonucu hastanın hastalık riski altında olmadığına karar verilir.

Naive Bayes sınıflandırma yönteminin yaygın olarak kullanılmasına neden olan bazı avantajları bulunmaktadır. Bu avantajlar aşağıda maddeler halinde belirtilmiştir.

- Anlaşılması kolay ve uygulanması basit bir yöntemdir.
- Oldukça hızlı eğitilir.
- İkili veya çoklu sınıflamalar için kullanılması uygundur.
- Özelliklerden ilişkisiz olanları ortadan kaldırarak sınıflandırma performansını etkili biçimde artırır.
- Olasılık tahmin hesaplaması yaparken örnekten vazgeçip kayıp değeri değerlendirmek için mücadele eder [11].
- Hesaplama süresi kısa olduğu için gayet hızlı çalışır.

Bu avantajların yanı sıra bir takım dezavantajlar da bulunmaktadır. Bu dezavantajlar aşağıda belirtilmiştir.

- İyi sonuçlar elde etmek için büyük verilerden oluşan veri setine ihtiyaç duymaktadır.
- Verilen eğitim verilerinin hepsini sakladıkları için tembeldir [22].

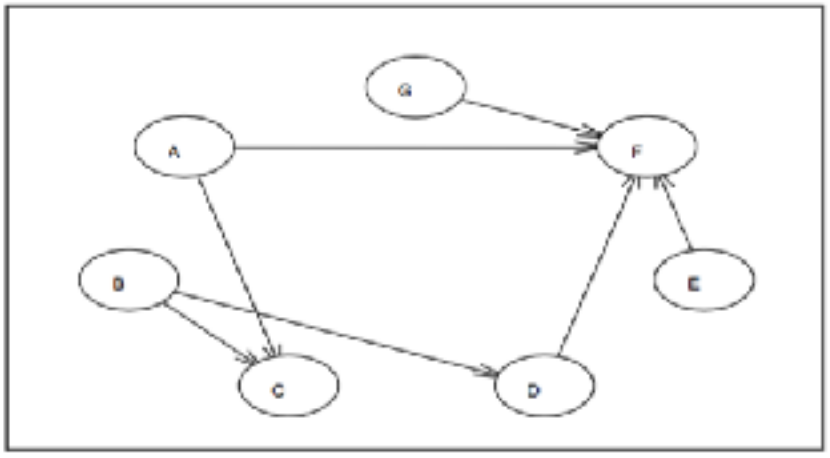
Bayes ağları, birçok değişkene sahip veri kümesi için değişkenler arası nedensellik ve koşullu bağımsızlık ilişkilerini ifade eden grafiksel modellerdir. Bir Bayes ağının oluşması için üç temel bileşen gerekmektedir:

1. X={ $X1$, $X2$,..., Xn } şeklindeki değişkenler kümesi,

2. G=(V,E) şeklinde gösterilen yönlü döngüsel olmayan bir grafik,

3. Yerel olasılık dağılımlarının çarpımsal olarak ifade edildiği P ile gösterilen bir ortak olasılık dağılımı.

Bayes ağını oluşturacak veri seti içerisindeki her değişken bir düğüm olarak ifade edilir. Yönsüz bir kenar ile bağlı iki değişken yalnızca değişkenler arasında ilişki bulunduğunu ifade eder. Bayes ağı üç temel kavramdan oluşur, bunlar ebeveyn, torun ve torun dışı kavramlarıdır. Bir düğümden diğer düğüme bağlantı yapılırken yönlü kenarın başladığı nokta ebeveyn düğümü , bittiği nokta çocuk düğüm olarak adlandırılır. Bayes ağında A ve B düğümleri olduğunu varsayalım, eğer A'dan B'ye doğru yönlü bir kenar var ise A düğümü, B düğümünün ebeveynidir, Pa(B)={A} şeklinde gösterilir. Buna göre B değişkeninin gerçekleşmesinde ebeveyni olan A'nın etkisi vardır ve A'da herhangi bir değişiklik olacak olursa B değişkeni de bu değişimden etkilenecektir. B değişkeni A değişkeninin çocuk düğümüdür ve Desc(A)={B} olarak gösterilir. Eğer B düğümü A'nın ebeveyni veya çocuğu değilse torun dışı olarak isimlendirilir ve Nondesc(A)= {B} olarak gösterilir. Şekil 1.5'te bu yapı gösterilmektedir.



Şekil 1.5. Düğüm, ebeveyn, çocuk ve torun dışı kavramlarının Bayes Ağ yapısında gösterimi [23].

Şekil 1.5' teki Bayes ağında ebeveyn, torun ve torun dışı düğümler Çizelge 1.2'de gösterilmektedir.

Çizelge 1.2. Ebeveyn, çocuk ve torun dışı düğümler [23].

Düğüm	Ebeveyn	Çocuk	Torun dışı
A	Ø	C,F	B,D,E,G
B	Ø	C,D	A,E,F,G
C	A,B	Ø	D,E,F,G
D	B	F	A,C,E,G
E	Ø	F	A,B,C,D,G
F	A,D,E,G	Ø	A,B,C
G	Ø	F	A,B,C,D,E

M. Karabıyık ve B. Yet, çalışmalarında Türkiye'deki futbol ligleri için kendilerinin geliştirdiği bir Bayes Ağ modeli önermektedir. Bu model futbol yarışmalarına katılan takımların stratejilerini gözlemleyerek maçın sonucu hakkında çıkarımda bulunmayı hedeflemektedir. FutBA Türkiye spor ligleri için üretilen ilk Bayes ağı modeli olması, tamamen özgün Bayes ağı yapısına sahip olması, uzman bilgisi, geçmiş maç verisi veya ikisinin karışımı ile tahmin üretme esnekliğine sahip olması gibi birçok yenilik sunmaktadır. Modelin geçmişe ve geleceğe yönelik performansları daha önceki futbol modelleri düşünüldüğünde başarılıdır. Model, tüm Bayes ağları gibi, eksik girdilerle tahmin üretebilmesine karşın böyle tahminlerin doğruluğunun daha az olması beklenmektedir. Dolayısıyla, FutBA modelinde girdiler için harcanacak efor ile tahminlerin doğruluğu arasında ödünleşim vardır [24].

M.Atalay v. d. Trafik Kazaları Analizi için Bayes Ağları Modeli üzerine çalışmıştır. Çalışılan modelde trafikte meydana gelen kazalar kazalara sebep olan unsurlar Bayes ağları yardımıyla incelenmektedir. Bayes ağlarının önemli bir grafiksel model olduğu belirtilmiştir. Koşullu bağımlılık ilişkileri hakkında bilgi vermekte, gözlemler sonucunda çıkarımlar yapılıp insanların faydalanması için

kullanılabilmektedir. Anlatılan çalışmada Silivri Bölge Trafik Şube Müdürlüğü ve İlçe Jandarma Trafik Tim Komutanlığı' ndan elde edilen maddi hasarlı trafik kaza tespit tutanakları ve trafik kaza tespit tutanaklarının içerdiği bilgilere göre oluşturulan veri setinden ilgili Bayes Ağı oluşturulmuştur. Oluşturulan Bayes Ağı' nın hatasız tahmin üretme bilgisi test verisi olarak kullanılarak denenmiş ve kullanılan model, model için elde edilmiş logskorun marjinal modelin logskoru ile kıyaslanması ile doğrulanmıştır. Önerilen çalışma, trafikte meydana gelen kazalara sebep olan unsurların birbirleri ve kaza sonuçları ile bağlantılarını tespit edebilen faydalı bir model oluşturmuştur [25].

Hipokrat-I: Bayes Ağı Tabanlı Tıbbi Teşhis Destek Sistemi olarak çalışılmış tezde Bayes Ağ yapısı kullanılarak teşhis destek sistemi sunulmuştur. Sunulan sistem tiroit hastalıkları için geliştirilmiştir ve tiroit türlerini tespit edebilmektedir. Elektronik ve elektronik olmayan hasta kayıtlarından yararlanılarak sistem oluşturulmuştur. Ek olarak, belirtilen hastalığın tespiti için uygulanan testler, konsültasyon seçimine bulguların tanıya ne kadar katkı sağladığı alanındaki uzmanlarca tespit edilmiştir [26].

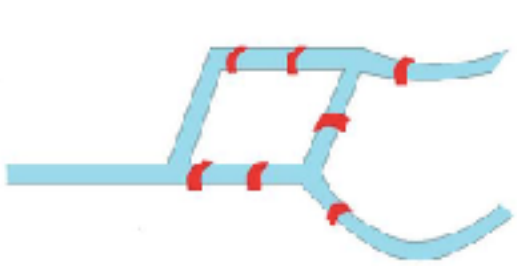
Z.D. Akşehir v.d., İş Sağlığı ve Güvenliği Sektöründe Bayes Ağları Uygulaması ile ilgili çalışma yapmıştır. Günümüzde inşaat sektöründeki gelişmeyle beraber iş kazalarının da sayısı artmıştır. Teknolojinin gelişimi, iş güvenliğindeki önlemlerde eksiklikler ve çalışanların eğitimsiz oluşu bu iş kazalarındaki ana nedenlerdir. Sunulan çalışmada, kullanılan iş kazaları verileri ilk olarak veri ön işleme aşamasına tâbi tutulup ardından elde edilmiş verilere tek değişkenli sıklık ve çapraz tablolama çözümlemesi uygulanmıştır. Çözümlemelerden edinilen sonuçlardan iş kazalarının oluşmasında güçlü risk oluşturan nicelikler belirlenmiştir. Ardından bu değişkenlerin iş kazasına etkileri Bayes ağları ile analiz edilmiştir. Bayes ağı, değişkenler arasındaki koşullu bağımlılık ilişkilerini ve tek bir bağımsız değişkene bağımlı olmadıklarını yansıtmaktadır. Bayes ağı, uluslararası bir inşaat firmasından bir veri kümesi üzerinde uygulanmıştır. Kurulan Bayes ağının doğruluk oranı ve diğer performans ölçütleri analiz edilmiş ve yapılan modelin etkinliği yorumlanmıştır. Deneysel sonuçlara göre, bazı iş kazası vakalarının makine öğrenme tekniklerini

kullanarak yüksek doğruluk oranları ile önceden tahmin edilebileceği gösterilmiştir [27].

Graf teorisi ise ilk olarak 1736'da Leonhard Euler katkısıyla literatüre kazandırılmıştır. Euler bu teorisini Königsberg köprü problemini üzerindeki çalışması ile sunmuştur [28]. Şekil 1.3.a'da görülen Pregel nehrinde yedi adet köprü bulunmaktadır. Euler'e göre bu yedi köprünün oluşturduğu kapalı döngüde her bir köprüyü sadece bir kere kullanılmak şartıyla başlanılan noktaya varmak imkansızdır. Problemin Euler tarafından Şekil 1.3.b'deki çizimi graf teorisinin de temellerinin atılmasını sağlamıştır. Eğer bir grafta her bir ayrıttan sadece bir kere geçilerek tüm ayrıtları gezilerek başlangıç noktasına dönülüyorsa bu graf Eularian'dır denir [33].

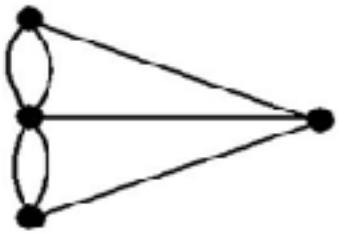
Eğer bir graftaki tüm düğümler çift dereceli düğüm ise bu graf net olarak Eularian graftır [29,30]. Çünkü bir nokta çift dereceye sahip ise, o noktaya gidilen her bir ayrıt için çıkılacak bir ayrıt olduğu garantilenmiştir.

Şekil 1.6.1'de gösterilen alanda, düşük maliyetle gezebilmek için bu alandaki tüm ayrıtlardan en az bir kere geçilmesi ve bu alanı tam olarak gezebilmesi için geçtiği ayrıtlardan tekrar geçme davranışını da minimum düzeyde gerçekleştirebilmesi gerekmektedir [31].



Şekil 1.6.1. Könisberg köprülerinin bir şeması [32].

Şekil 1.6.2’ de Könisberg köprüleri problemine matematiksel bakış gösterilmektedir



Şekil 1.6.2. Könisberg köprüleri problemine matematiksel bakış [32].

2. TEPE TIRMANMA ALGORİTMASI, BAYES AĞI VE NAIVE BAYES SINIFLANDIRICISI

2.1. Tepe Tırmanma Algoritması

Tepe Tırmanma metodu basit yapısı ve hızı sebebiyle optimizasyon için kullanılan rastlantısal bir iteratif yerel arama yöntemidir. Bu metodun esasında, tanımlanan birtakım kurallara göre bir çözümden bir diğer komşu çözüme erişimi vardır. Ayrıca metodun uygulamasında mutlak açıdan en iyi olmasa da iyi bir komşuluk yapısı seçiminin önemi büyüktür. Metodun zayıf olduğu taraf yerel ve genel en iyi arasında var olan ayrımı yapamaması sonucu yerel optimumdan kaçamamasıdır. Özetle T-T Algoritmasının adımları aşağıdaki gibi ifade edilebilir [34]:

1. Başlangıç çözümü verilir; x_0 : Eldeki çözüm ve $x_0 \in$ R

2. Aşağıdaki adımları tekrarlanır:

a) N(x_n) komşu seti içerisinden en iyi x′ komşusu seçilir.

b) f(x′) ≤ f(x_n) ise x′ çözümü yeni mevcut çözüm olarak atanır: x_{n+1} = x′

c) Aksi halde durulur [34].

Tepe Tırmanma yöntemi daima mevcut çözümü en fazla geliştiren yönde ilerleme prensibine dayanır ve hafıza gereksinimi oldukça düşüktür. Şekil 2.1'de yönteme ait akış diyagramı verilmiştir [35].

İlk olarak başlangıç çözümü, var olan probleme uygun olarak elde edilir. Komşu üretebilen bir algoritma ile mevcut çözüme benzeyen bir miktar rastgele komşu çözüm üretilerek uygunluk değeri en iyi olan komşu çözüm alınır. Bu çözüm mevcut çözüme göre daha iyiyse mevcut çözümle yer değiştirir ve sonraki iterasyon uygulanır. Belirli bir iterasyon miktarına, uygunluk değerine veya bu değerin iterasyona bağlı değişimine varıldığında algoritma son bulur. Algoritma basit bir yapıya sahiptir ve hızlıdır. Ancak adımları dolayısıyla optimizasyonu birinci çözüme göre sonuçlanır ve yerel en iyi çözümlerde takılıp kalma ihtimali vardır [35]. Tepe tırmanma yöntemine ait akış diyagramı Şekil 2.1' de gösterilmektedir.

Şekil 2.1. Tepe tırmanma yöntemine ait akış diyagramı [35].

Tepe tırmanma algoritması, arama algoritmaları arasındaki en iyi sonucu veren algoritma olmamasına rağmen kodlanması ve tasarımının basit olmasından dolayı sık sık kullanılır.

2.1.1. Tepe tırmanma algoritması çeşitleri

Mevcut algoritma üzerine birtakım düzeltmeler yapılarak daha iyi sonuçlar elde edilmeye çalışılmıştır. Literatürde sıklıkla bahsedilen bir kaç tepe tırmanma algoritması burada farkları ile birlikte kısaca açıklanacaktır.

Standart olarak tepe tırmanma algoritması bir başlangıç noktası seçer ve buradan komşu notaları gezerek sonuç bulmaya çalışır. Bir grafik üzerinde rastgele seçilecek bir nokta için 3 ihtimal bulunmaktadır:

1. Seçilen noktanın bir tarafında problem iyiye giderken diğer tarafında kötüye gitmektedir. Dolayısıyla tırmanma algoritmamız iyi yönde gezinmeye devam eder.

2. Seçilen noktanın iki tarafında da problem sonucu kötüye gitmektedir. Dolayısıyla bulunduğumuz nokta problem için en iyi noktalardandır çıkarımını yaparız. Bulunan

nokta en iyi sonuçtur diyemeyiz yani bu sonuçtan daha iyi sonuçlar olabilir fakat klasik tepe tırmanma algoritması artık arama yapmaz ve bulunduğu noktada kalır.

3. Seçilen noktanın iki tarafında da problem iyiye gitmektedir. Yani bulunan nokta aslında problem için erişilebilecek en kötü noktalardandır çıkarımını yaparız. Dolayısıyla tepe tırmanma algoritması iki yönden birisini seçecek ve tırmanmaya devam eder. Ayrıca her iki yöne de tırmanan farklı bir algoritma da bulunmaktadır.

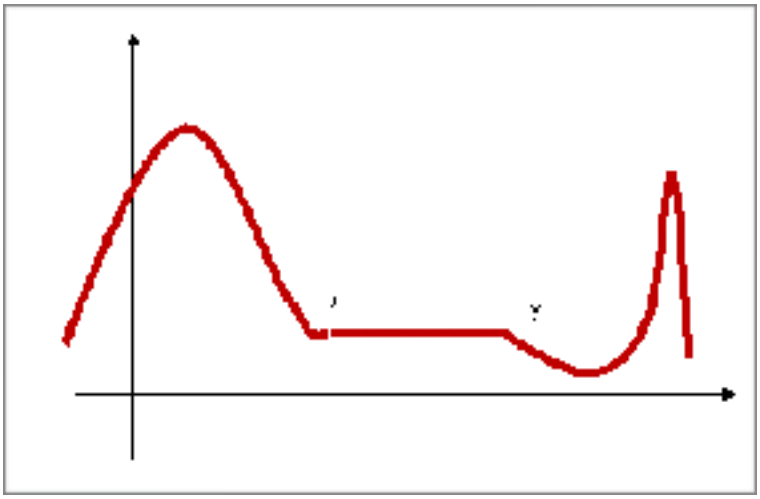
Örneğin; steepest ascent tepe tırmanma algoritmasında klasik tepe tırmanma algoritmasından farklı olarak, bulunabilen tüm çıktılar arasından bir seçim yapılır. Bu algoritmada da klasik tepe tırmanma algoritmasında olduğu gibi sorun aynıdır. Eğer arama işlemi yapıldığı sırada bir yerel çukura yani en düşük çukura rastlanılırsa algoritma bu durumdan kurtulamayıp en doğru sonucu bulamayabilir.

Diğer bir algoritma olan olasılıksal tepe tırmanma algoritmasında tüm komşuların aranması ve de komşulardan alınan sonuca göre davranmak yerine, rastgele bir şekilde komşu nokta seçilmektedir. Eğer seçilen komşu iyi yönde gitmiyorsa aynı doğrultu üzerinde arama yapmaya devam edilir.Eğer arama sonucunda iyileştirme elde edilemiyorsa, farklı bir komşu nokta seçilerek aramaya devam edilir.

Açıklanan tepe tırmanma algoritmalarına ek olarak rastgele başlangıç tepe tırmanma algoritması şaşırtıcı bir şekilde iyi sonuç vermektedir. Algoritma çalışma mantığında şöyle bir yol izlemektedir: bir x durumunu başlangıç olarak kabul eder , ardından daha iyi bir başlangıç durumu bulursa bu noktaya kayar. Yani bu algoritma iyi durum bulduğu sürece başlangıç durumu değiştirilir ama bulamadığı durumlarda da aramaya devam etmektedir.

Bazı kaynaklar rastgele başlangıç tepe tırmanma algoritmasını pompalı tüfek tepe tırmanma algoritması olarak adlandırabilmektedir.

Tepe tırmanma algoritmalarının ortak zayıflığı yerel optimum noktasında kalmalarıdır, bunun sonucunda daha iyi noktalar atlanabilmektedir. Şekil 1.2.1’ de tepe tırmanma algoritmasının çalışma grafiği gösterilmektedir.



Şekil 2.1.2. Tepe Tırmanma Algoritması muhtemel çalışma grafiği [36].

Yukarıdaki şekli incelersek x ve y noktaları arasındaki doğrusal nokta olduğunu görmekteyiz. Eğer algoritma bu doğru üzerinden bir nokta seçerse komşularda da iyiye veya kötüye gitmekte olan bir sonuç bulamayacağından karar aşamasında hata elde edilebilmektedir.

Tepe tırmanma algoritmaları az rastlanan bir durum da olsa aynı sonuç elde edilince hatalı sonuçlar verebilmektedir. Bu durum grafikte görülen düzlüklerin olduğu yerlerde oluşur ve gittiği hiç bir yön iyi bir sonuç vermez. Bunun sonucunda da hata meydana gelmektedir [36].

2.2. Bayes Ağı

Bayes ağları, şekilsel olarak yönlü graflara benzerler. Fakat burada her ayrıtın olasılık değeri söz konusu olduğundan Bayes ağları graf yapısından biraz ayrılmış olarak karşımıza çıkmaktadır. Bayes ağı için bahsedebileceğimiz çeşitli durumlardan aşağıda bahsedilecektir [37].

1) A durumundan B durumuna geçebiliyorsak Şekil 2.2.3' teki gibi gösterilir:

Şekil 2.2.3. İki olay arasındaki geçiş.

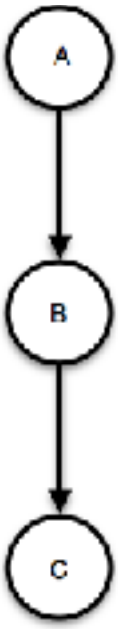
α bu durumun şartlı olasılığı olarak kabul edilirse;

α = P(B |A) = P(A ∩ B) / P(A)

P(B |A) = B' nin A' ya bağlı olasılığıdır ve P(A) = Evrensel kümedir.

α bu iki olayın ardışık bir biçimde meydana gelme durumudur.

2) Olaylar sadece tek bir olaya bağlı olmayıp birden fazla olaya da bağlı olabilmektedir. Eğer üç olay meydana gelmiş olsaydı Şekil 2.2.4' teki gibi gösterilir:



Şekil 2.2.4. Üç olay arasındaki geçiş.

A'nın olasılığı P(A)' dır. Çünkü A herhangi bir olaya bağlı değildir.

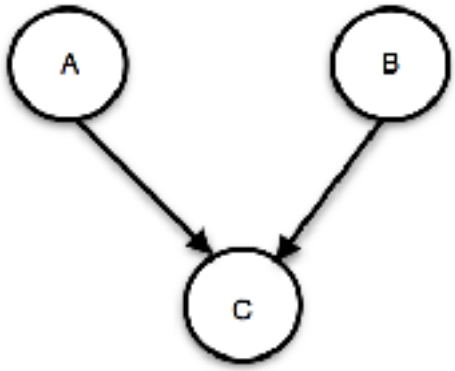
$P(B|A) = P(A \cap B) / P(A)$

C, hem A' ya hem B'ye bağlı bir değişken durumunda olduğu için;

$P(C| A, B) = P(C| B, A)$ şeklinde ifade edilir. C, A' dan bağımsız olabilir bu durumda şartlı bağımsızlık denilecektir. Çünkü eğer B olayı gerçekleştiyse C için A olayının gerçekleşip gerçekleşmeyeceği durumuna bakılmayacaktır.

Bileşke olasılığı $P(A, B, C) = P(A). P(B|A). P(C|B)$ olacaktır.

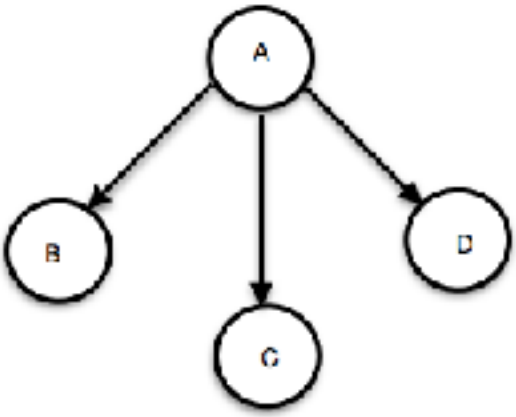
3) C değişkeninin hem A hem B olayına bağlılığı söz konusu olduğunda bileşke olasılığı $P(A, B, C) = P(A). P(B). P(C|A, B)$ olacaktır. İki olaya bağlı tek olay arasındaki geçiş Şekil 2.2.5' te gösterilmektedir.



Şekil 2.2.5. İki olaya bağlı tek olay arasındaki geçiş.

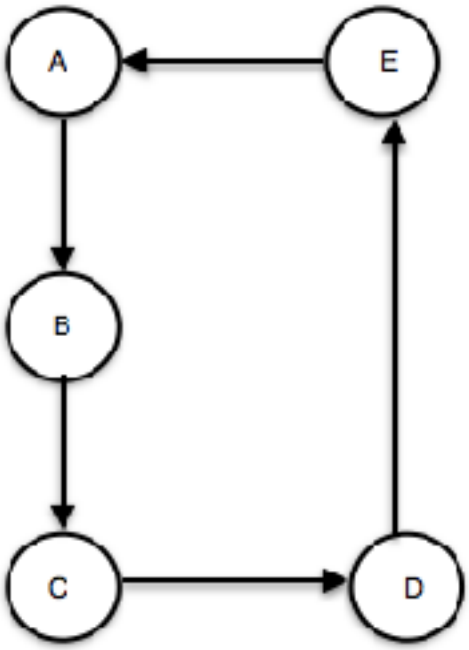
4) Birden fazla olay tek olaya bağımlı olabilir. Burada B, C ve D kendi başına A' ya bağlıdır.

Olasılıklar P(D|A), P(C|A), P(B|A) şeklinde ayrı ayrı yazılmaktadır. Üç olaya bağlı tek olay arasındaki geçiş Şekil 2.2.6' daki gibi gösterilmektedir.



Şekil 2.2.6. Üç olaya bağlı tek olay arasındaki geçiş.

5) Bayes ağı, her bir ayrıtın birden fazla olasılık katmanı olacağı ve nereden başlanacağı bilinemeyeceği için kesinlikle döngü içermemelidir. Döngü içeren olay ağı Şekil 2.2.7' deki gibi gösterilmektedir.



Şekil 2.2.7. Döngü içeren olay ağı [37].

Bir örnek ile Bayes ağ yapısı ile ilgili bazı hesaplamaları yapıp açıklarsak;

H : Hava kirliliği

S : Sigara içme

N : Nefes darlığı

K : Kanser

Y : Yüksek

D : Düşük

E : Evet

H : Hayır

P : Pozitif

Hava kirliliği ve sigara içme parametrelerinin akciğer kanseri üzerindeki etkisi bilinmektedir. Yine kanser olan bir kişinin röntgen sonucunun pozitif çıkma ve nefes darlığı çekme durumu söz konusudur [37]. Kanser için hava kirliliği ve sigara içme durumuna göre olasılıkları Çizelge 2.2.1'de gösterilmektedir [37].

P(H=D) = 0.9

P(S=E)=0.3

Çizelge 2.2.1. Kanser için hava kirliliği ve sigara içme durumu.

H	S	P(K=E I H,S)
Y	E	0,05
D	H	0.02
D	E	0.03
D	H	0,001

Kişinin kanser olma durumuna göre XRay taramasına girmiş olma olasılığı Çizelge 2.2.2' de gösterilmektedir [37].

Çizelge 2.2.2. Kanser durumuna göre XRay sonucu.

K	P(X=P I K)
E	0.90
H	0.20

Kişinin kanser durumuna göre nefes darlığı yaşıyor olma olasılığı Çizelge 2.2.3'te gösterilmektedir.

Çizelge 2.2.3. Kanser durumuna göre nefes darlığı sonucu.

K	P(N=E I K)
E	0.65
H	0.30

Tüm bu olasılıklara göre bileşke olasılık kütle fonksiyonu hesaplanabilir. Bileşke olasılık kütle fonksiyonunun sonucu 1' dir [37].

$P(H, S, K, X, N) = P(H).\ P(S).\ P(K \mid H, S).\ P(X \mid K).\ P(N \mid K)$

$= \sum P(H).\ P(S).\ P(K \mid H, S).\ P(X \mid K).\ P(N \mid K)$

$= \sum P(H).\ P(S).\ P(K \mid H, S).\ P(X \mid K).\ \sum_k P(N \mid K) \ --> \ \sum_k P(N \mid K) = 1$' dir.

$= \sum P(H).\ P(S).\ P(K \mid H, S).\ \sum_k P(X \mid K) \ --> \ \sum_k P(X \mid K) = 1$' dir.

$= \sum P(H).\ P(S).\ \sum_k P(K \mid H, S) \ --> \ \sum_k P(K \mid H, S) = 1$' dir.

$= \sum P(H).\ P(S)$ bulunur. H ve S birbirinden bağımsız olduğu için $P(H, S) = P(H).\ P(S)$ bileşke olasılık kütle fonksiyonuna göre toplam yazılır.

$= \sum_H P(H).\ \sum_S P(S) \ --> \ \sum_H P(H) = 1$ ve $\sum_S P(S) = 1$

=1 olarak bulunur.

Hava kirliği yüksek, sigara içen ve kanser olma riski % 5 olarak verilen bir hastanın nefes darlığı çekmeme olasılığını hesaplayabiliriz [37].

$P(N=H \mid K) = P(N= H \mid K= E)/\ P(K= E)$

Kanser olma veya olmama olasılığının(P(K)) tek başına bir sonucu henüz elimizde olmadığından H ve S parametrelerine bağlı olarak bulunacaktır. Bu olasılığa marjinal olasılık kütle fonksiyonu denilmektedir [37].

2.2.1. Bayes ağlarında bağımsızlık ve şartlı bağımsızlık

2.2.1.1. Bağımsızlık

α ve ß bir olay olmak üzere **α**, ß; P olasılığı altında birbirinden bağımsızsa yani P ⊨ **α** ⊥ ß ise ve eğer P(**α**, ß)= P(**α**). P(ß) veya P(**α** ∩ ß)= P(**α**). P(ß) şeklinde ise **α** ve ß olayları bağımsız olaylardır ve gösterimleri de bu şekildedir. Eğer bu şart geçerliyse P(**α** | ß)= P(**α**), P(ß | **α**)= P(ß) olacaktır [38].

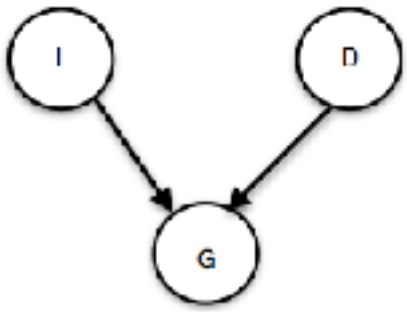
X,Y rastgele değişken olmak üzere X, Y; P olasılığı altında birbirinden bağımsızsa yani X,Y,P ⊨ X ⊥Y ise ve eğer P(X,Y)= P(X). P(Y) şeklinde ise bu durumda P(X|Y)= P(X), P(Y|X)= P(Y) olacaktır. Tüm bu bilgilerden yola çıkarak eğer Xve Y bağımsız değişkenler veya olaylar ise ⋁ X, Y, P(X,Y)= P(X). P(Y) şeklinde hesaplanır.

Bir Bayes ağına eğer G şeklinde gösterilecek bir ağ tanımlanacak olursa bu ağ üzerinde P olasılıklarını kullanarak faktörlere ayırma işleminde kullanılacaktır [38]. I, D ve G'nin olasılıkları Çizelge 2.2.1.1.1' de gösterilmektedir.

Çizelge 2.2.1.1.1. I, D ve G'nin olasılıkları [38].

I	D	G	Olasılıkları
i_0	d_0	g_1	0,126
i_0	d_0	g_2	0,168
i_0	d_0	g_3	0,126
i_0	d_1	g_1	0,009
i_0	d_1	g_2	0,041
i_0	d_1	g_3	0,126
i_1	d_0	g_1	0,252
i_1	d_0	g_2	0,0224
i_1	d_0	g_3	0,0056
i_1	d_1	g_1	0,06
i_1	d_1	g_2	0,036
i_1	d_1	g_3	0,024

Burada I ve D' nin G'ye bağlı gösterimi Şekil 2.2.1.1.1'deki gibi elde edilmektedir.



Şekil 2.2.1.1.1. G'nin I ve D'ye bağlı gösterimi [38].

I ve D' nin bileşke olasılıkları Çizelge 2.2.1.1.2' de gösterilmektedir.

I	D	Olasılıkları
i_0	d_0	0,42
i_0	d_1	0,18
i_1	d_0	0,28
i_1	d_1	0,12

Çizelge 2.2.1.1.2. I ve D'nin olasılıkları [38].

Bu çizelgede 1.satır i_0 iken d_0 olduğu zamanki olasılık değerlerini, 2.satır i_0 iken d_1 olduğu zamanki olasılık değerlerini, 3.satır i_1 iken d_0 olduğu zamanki olasılık değerlerini, 4.satır i_1 iken d_1 olduğu zamanki olasılık değerlerini ifade etmektedir. Yine bu çizelgeye göre I'nın olasılık değerleri i_0 = 0,42 + 0,18 = 0,6 ve i_1 = 0,28 + 0,12 = 0,4 olarak bulunur. D'nin olasılık değerleri d_0 = 0,42 + 0,28 = 0,7 ve d_1 = 0,18 + 0,12 = 0,3 olarak bulunur.

Ayrıca bu çizelgedeki I ve D değerlerinin bağımsız olduğu olasılık hesabıyla ispatlanabilmektedir. P(I, D) = P(I). P(D) hesabıyla;

1.satır için i_0. d_0= 0,6. 0,7= 0,42

2.satır için i_0. d_1= 0,6. 0,3= 0,18

3.satır için i_1. d_0= 0,4. 0,7= 0,28

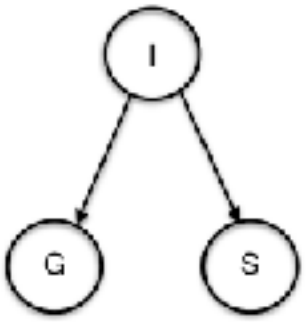
4.satır için i_1 . d_1 = 0,4. 0,3= 0,12 çarpımları sonucu I ve D değişkenlerinin bağımsız değişkenler olduğu bulunabilmektedir.

2.2.1.2. Şartlı bağımsızlık

X, Y ve Z birer rastgele değişken olmak üzere aşağıda verilen şartlar sağlandığında P olasılık ortamında yani $P \models (X \perp Y \mid Z)$ verildiği zaman; eğer P(X, Y | Z)= P(X | Z). P(Y | Z) ve aynı zamanda P(X | Y, Z)= P(X | Z) ve P(Y | X, Z)= P(Y, Z) şeklinde ise şartlı bağımsızlık söz konusu olur.

Bu üç değişkenin bileşke olasılık kütle fonksiyonu P(X, Y, Z) α ϕ_1(X, Z). ϕ_2(Y, Z) şeklinde iki tane ayrı olasılığa ayrıştırılmış olarak ifade edilmektedir.

Bir de bunun tersi durumunu düşünürsek G değişkenimiz I ve D değişkenlerimize bağlı olduğu için bu seferde I değişkenine bağlı iki değişken olduğunu düşünebiliriz. S ve G'nin I'ya bağlı gösterimi Şekil 2.2.1.2.1' de gösterilmektedir.



Şekil 2.2.1.2.1. S ve G'nin I'ya bağlı gösterimi [38].

I' ya bağlı olarak P(S, G/i_0) değerini elde etmek gerekmektedir. Çizelge 2.2.1.2.1'den S ve G'nin i_0' a bağlı olan olasılığını elde ettiğimizde Çizelge 2.2.1.2.2'yi elde edebilmekteyiz.

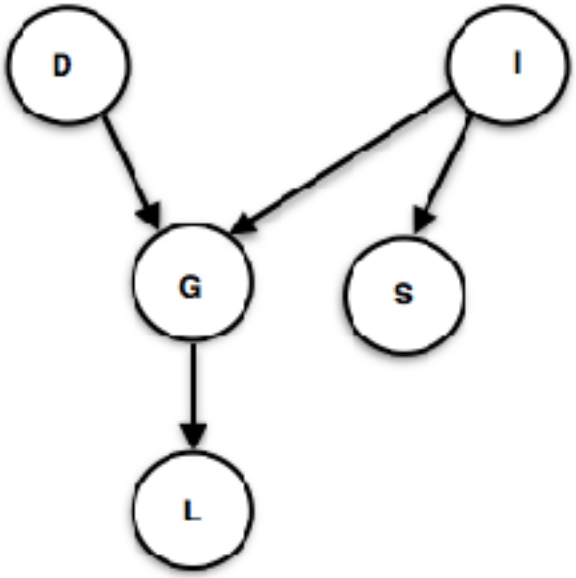
Çizelge 2.2.1.2.1. I, S ve G'nin olasılığı [38].

I	S	G	P
i_0	s_0	g_1	0,114
i_0	s_0	g_2	0,1938
i_0	s_0	g_3	0,2622
i_0	s_1	g_1	0,006
i_0	s_1	g_2	0,0102
i_0	s_1	g_3	0,0138
i_1	s_0	g_1	0,252
i_1	s_0	g_2	0,0224
i_1	s_0	g_3	0,0056
i_1	s_1	g_1	0,108
i_1	s_1	g_2	0,0096
i_1	s_1	g_3	0,0024

Çizelge 2.2.1.2.2. I' ya bağlı P(S, G/i_0) olasılığı [38].

P(S, G/ i_0)
s_0, g_1 için 0,19
s_0, g_2 için 0,323
s_0, g_3 için 0,437
s_1, g_1 için 0,01
s_1, g_2 için 0,017
s_1, g_3 için 0,023

Birden çok olay için Bayes Ağı Şekil 2.2.1.2.2' de gösterilmektedir.



Şekil 2.2.1.2.2. Birden çok olay için Bayes Ağı [38].

Bu bilgilere göre şartlı olasılık birbirini etkilememe durumu ve faktörlere ayırma durumlarından oluşmaktadır. Bayes ağı üzerinde birbirinden bağımsız iki faktörü elde edebilmekteyiz.

$$P(D, I, G, S, L) = P(D).\ P(I).\ P(G/\ D, I).\ P(S/\ I).\ P(L/G)$$

$$P(D, S) = \sum_{G,L,I} P(D).\ P(I).P(S/\ I).\ P(G/\ D, I).\ P(L/G)$$

$$= \sum_{G,L,I} P(D).\ P(I).P(S/\ I).\ P(G/\ D, I).\ P(L/G) \text{ --> } \sum_{L} P(L/G) = 1$$

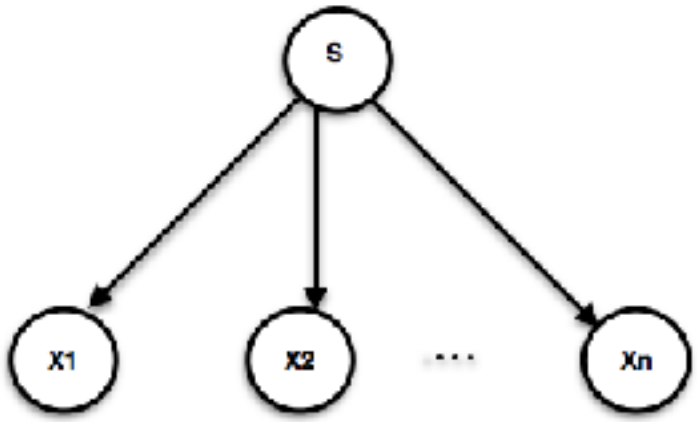
$$= \sum_{G,L,I} P(D).\ P(I).P(S/\ I).\ \sum_{G} P(G/\ D, I) \text{ --> } \sum_{G} P(G/\ D, I) = 1$$

$$= P(D).\sum_{I} P(I).P(S/\ I) \text{ elde edilir.}$$

Bu şekilde Bayes ağı üzerinde birbirinden bağımsız olan iki faktör yani D ve S değişkenleri $P(D) = \phi_1$, $P(I).P(S/\ I) = \phi_2$ şeklinde elde edilmiştir [38].

2.3. Naive Bayes Sınıflandırıcısı

Bayesian ağları kullanarak sınıflandırma problemleri üzerine çalışmalar yapılabilmektedir. Bunlardan en çok bilineni Naive Bayes modelidir. S gibi bir sınıflandırma modelimiz olduğunu varsayıp X verileri gelmiş olsun. S burada bir sınıf olarak kabul edilmiş ise X_1, X_2, …X_n bu sınıfa bağlı olan parametreler olarak alınır. Başka bir sınıf için ise bu parametrelerin değerleri değişecektir. Ağı ve olasılığı Şekil 2.3.1' deki gibi ifade edilmektedir [39].



Şekil 2.3.1. S Sınıfına bağlı parametreler.

Tek sınıf varsa olasılık:

$P(S, X_1,\ldots, X_n) = P(S).\prod_i P(X_i|\ S)$ şeklinde ifade edilebilmektedir.

İki sınıf varsa olasılık:

$$\frac{P(S = S_1 \mid X_1, \ldots, X_n)}{P(S = S_2 \mid X_1, \ldots, X_n} = \frac{P(S = S_1)}{P(S = S_2)} \cdot \prod_{i=1}^{n} \frac{P(X_i \mid S = S_1)}{P(X_i \mid S = S_2)}$$

şeklinde ifade edilebilmektedir [39].

Meyve tür ve miktarları Çizelge 2.3.1' deki gibi bir veri tablosu verilsin:

Çizelge 2.3.1. Sarı, tatlı, uzun meyveler için veri miktarları [39].

Meyve	Sarı	Tatlı	Uzun	Toplam
Mango	350	450	0	650
Muz	400	300	350	400
Diğer	50	100	50	150
Toplam	800	850	400	1200

Sarı, tatlı ve uzun olacak bir meyvenin hangi sınıfta olduğunu bulmak için Naive Bayes sınıflandırıcı kullanılabilmektedir. Meyve= {Sarı, Tatlı, Uzun}—> X şeklinde gösterebiliriz. X'i bulmak için olasılık hesabı:

P(X|Mango)= P(Mango|X). P(X) / P(Mango) formülü ile yapılmaktadır. Bu formül tüm niteliklere uygulanıp sonuç elde edilecektir.

1.Durum:

$$P(\text{Sarı}|\text{Mango}) = P(\text{Mango}|\text{Sarı}).\ P(\text{Sarı}) / P(\text{Mango}) = \frac{350}{800}.\ \frac{800}{1200} / \frac{650}{1200} = 0.53$$

P(Tatlı|Mango)= P(Mango|Tatlı). P(Tatlı) / P(Mango)= 0.69

P(Uzun|Mango)= P(Mango|Uzun). P(Uzun) / P(Mango)= 0

2.Durum:

P(Sarı|Muz)= P(Muz|Sarı). P(Sarı) / P(Muz)= 1

P(Tatlı|Muz)= P(Muz|Tatlı). P(Tatlı) / P(Muz)= 0.75

P(Uzun|Muz)= P(Muz|Uzun). P(Uzun) / P(Muz)= 0.875

3.Durum:

P(Sarı|Diğer)= P(Diğer|Sarı). P(Sarı) / P(Diğer)= 0.33

P(Tatlı|Diğer)= P(Diğer|Tatlı). P(Tatlı) / P(Diğer)= 0.66

P(Uzun|Diğer)= P(Diğer|Uzun). P(Uzun) / P(Diğer)= 0.33

P(X|Mango), P(X|Muz), P(X|Diğer) olasılıklarını aradığımız için meyvelerin sarı, tatlı ve uzun gelme olasılık sonuçlarını çarpacağız. Çıkan sonuçlardan en büyük olan, aranan meyveyi vermektedir.

P(X|Mango)= P(Sarı|Mango). P(Tatlı|Mango). P(Uzun|Mango)= 0

P(X|Muz)= P(Sarı|Muz). P(Tatlı|Muz). P(Uzun|Muz)= 0.65

P(X|Diğer)= P(Sarı|Diğer). P(Tatlı|Diğer). P(Uzun|Diğer)= 0.072 elde edilmektedir.

P(X|Muz) > P(X|Diğer) > P(X|Mango) olduğu için X meyvesi muz sınıfındadır[39].

3. YÖNTEM VE DENEYSEL SONUÇLAR

Çalışmada hesaplamalar için RStudio ve R dili kullanılmıştır. RStudio ücretsiz bir geliştirme ortamıdır. İstatistiki hesaplamalar yapabilmek ve grafik oluşturabilmek için R dili kullanılmaktadır[41].

Bu tez çalışmasında, ilk olarak veri seti siteden indirilip excel formatında incelendi, sonra RStudio' da kullanabilmek için csv formatına dönüştürüldü. Veri seti kullanılabilir hale geldikten sonra BnLearn adlı kütüphane eklenmiştir. 6 sütun ve 1728 satır için istenilen koşullu olasılık sonuçları bulunmuş, ilişkiler çizge olarak gösterilmiştir. Kullandığımız BnLearn kütüphanesi, Bayes ağlarının parametrelerini tahmin etmek, grafik yapısını öğrenmek ve faydalı çıkarımlar yapabilmek amacıyla kullanılmakta olan bir R paketidir [42]. R dili ilk kez 1993 yılında piyasaya duyurulmuştur [43].

Konsola gireceğimiz install.packages ("bnlearn") satırı ile BnLearn kütüphanesini yükleyebilmekteyiz. Ardından kullandığımız Tepe Tırmanma Algoritması fonksiyon olarak çağırılır ve Bayes Ağ yapısı oluşturulur.

Veri setleri anket oluşturularak hazırlanabileceği gibi hazır da alınabilmektedir. Bu tez çalışmasında, UCI makine öğrenme deposunda bulunan Araba Değerlendirme Veri Seti kullanılmıştır.

Kullanılan veri setinde toplam örnek sayısı 1728'dir. Veri seti 6 özellikten meydana gelmektedir. Bulunan özellikler Buying, Maint, Doors, Persons, Lug_boot ve Safety olarak görülmektedir.

Buying özelliği vhigh, high, med, low olmak üzere 4 değişken içerir ve alış fiyatını ifade etmektedir. Maint özelliği vhigh, high, med, low olmak üzere 4 değişken içerir ve bakım fiyatını ifade etmektedir. Doors özelliği 2, 3, 4, 5more olmak üzere 4 değişken içerir ve kapı sayısını ifade etmektedir. Persons özelliği 2, 4, more olmak üzere 3 değişken içerir ve taşınabilecek kişi sayısını ifade etmektedir. Lug_boot özelliği small, med, big olmak üzere 3 değişken içerir ve bagaj

büyüklüğünü ifade etmektedir. Safety özelliği low, med, high olmak üzere 3 değişken içerir ve arabanın olası güvenliğini ifade etmektedir.

Veri setinde arabanın verilen özelliklere göre bulunduğu sınıf gösterilmektedir. Class olarak adlandırılan bu sınıflar unacc, acc, good ve vgood olmak üzere 4 sınıftan oluşmaktadır[40]. Sınıf dağılımları Çizelge 3.1' de ve örnek veri seti Çizelge 3.2' de gösterilmektedir.

Çizelge 3.1. Sınıf dağılımları.

Sınıf	N	N(%)
unacc	1210	(70.023 %)
acc	384	(22.222 %)
good	69	(3.993 %)
vgood	65	(3.762 %)

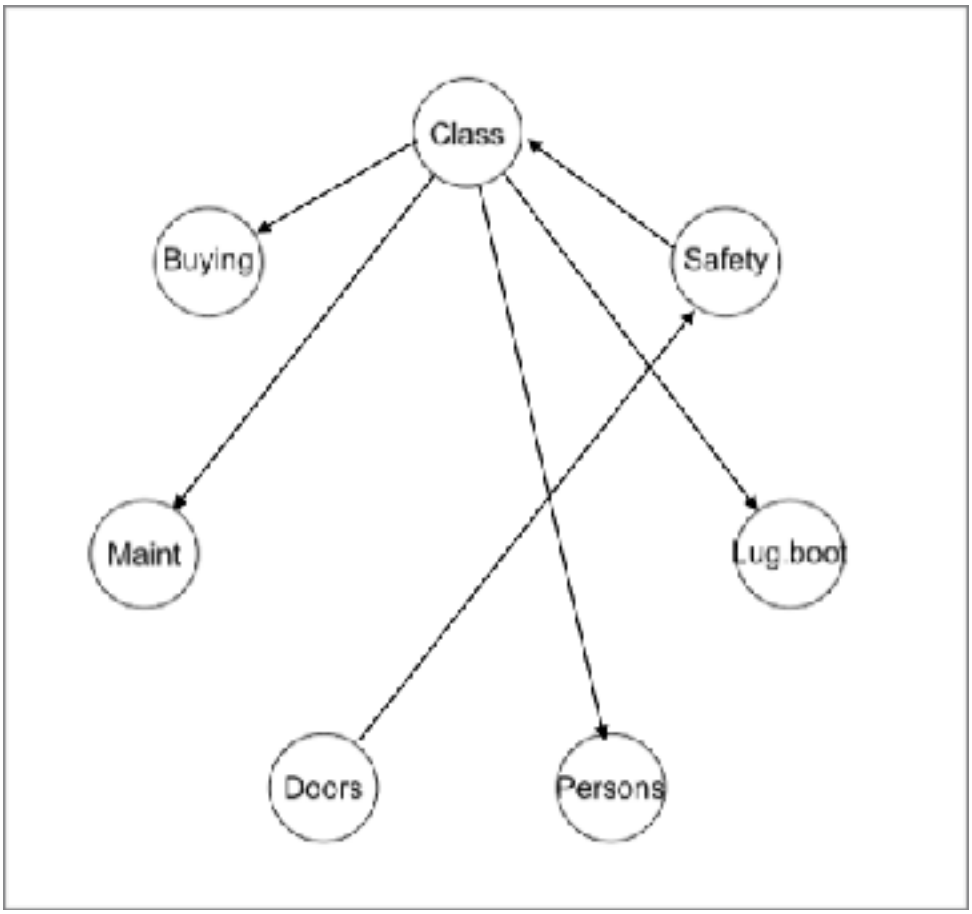
Çizelge 3.2. Örnek veri seti.

Buying	Maint	Doors	Persons	Lug_boot	Safety	Class
Vhigh	Vhigh	2	4	Small	High	Unacc
Vhigh	Vhigh	2	4	Med	Low	Unacc
Vhigh	Vhigh	2	4	Med	Med	Unacc
High	Low	2	2	Med	Med	Unacc
Med	High	2	More	Big	Med	Acc
Med	High	2	More	Big	High	Acc
Med	High	3	2	Small	Low	Unacc
Low	Med	4	4	Med	Med	Good
low	Med	4	4	Med	High	Vgood

Çizelge 3.3. Araba değerlendirme veri setinden bir kesit [40].

Buying	Maint	Doors	Persons	Lug.boot	Safety	Class
vhigh	vhigh	2	2	small	low	unacc
vhigh	vhigh	3	2	med	med	unacc
high	high	?	4	small	high	unacc
high	high	2	more	small	high	unacc
med	high	2	more	med	med	unacc
med	med	3	4	big	high	vgood
med	med	4	4	big	high	vgood
med	med	4	4	big	low	unacc
med	low	3	more	big	high	vgood
low	vhigh	3	more	med	high	acc
low	vhigh	5more	2	big	high	unacc
low	high	4	4	small	high	acc
low	med	4	4	small	high	good

Tez çalışmasında kullanılan özellikler için bulgular bu bölümde açıklanmıştır. Fakat öncelikle küçük bir veri seti için bazı aşamalar gösterildi. Çizelge 3.3' te verilen veri seti için uygulamamızı çalıştırdığımızda elde edilen Bayes Ağı Şekil 3.1' deki gibi bulunur. Bunlara ek olarak koşullu olasılıkların sonuçları hesaplandı. Buradaki amaç verilerden bir Bayes ağ elde edebilmektir.



Şekil 3.1. Veri setinden elde edilen Bayes Ağı.

Değişkenler arasındaki koşullu olasılıkların hesabı aşağıdaki satırlar ile elde edilmektedir. Bu sonuç değişkenin bağlı olduğu değişkenle arasındaki ilişkinin belirlenmesinde önemlidir.

- cpquery(fittedbn, event = (Buying=="high"), evidence = (Class=="acc"))=

0.1354167

- cpquery(fittedbn, event = (Maint=="vhigh"),evidence = (Class=="good"))=

0.2924791

- cpquery(fittedbn, event = (Safety=="med"),evidence = (Doors=="4"))=

0.1503856

- cpquery(fittedbn, event = (Safety=="med"),evidence = (Class=="acc"))=

0.1565657

- cpquery(fittedbn, event = (Lug.boot=="med"),evidence = (Class=="acc"))=

0.2177955

- cpquery(fittedbn, event = (Persons=="med"),evidence = (Class=="unacc"))=

0

Tüm veri setinde ise 5 özellik için Tepe Tırmanma fonksiyonu kullanılarak koşullu olasılıklar hesaplanmıştır. Bu da bir özelliğin diğer özelliğe bağlı olarak gerçekleşme ihtimalinin hesaplanması demektir ve bunların sonuçları Çizelge 3.4, Çizelge 3.5 Çizelge 3.6, Çizelge 3.7, Çizelge 3.8' de gösterilmektedir.

Çizelge 3.4. Buying özelliği için koşullu olasılık değerleri.

Buying		acc	good	unacc	vgood
	1	0	0	0	0
High	0	0.2812500	0.0	0.2677686	0.0
Low	0	0.2317708	0.6666667	0.2132231	0.0
Med	0	0.2994792	0.3333333	0.2214876	0.0
vhigh	0	0.1875000	0.0	0.2975207	0.0

Çizelge 3.5. Maint özelliği için koşullu olasılık değerleri.

Maint		acc	good	unacc	vgood
	1	0	0	0	0
High	0	0.2734375	0.0	0.2595041	0.20
Low	0	0.2395833	0.6666667	0.2214876	0.40
Med	0	0.2994792	0.3333333	0.2214876	0.40
vhigh	0	0.1875000	0.0	0.2975207	0.0

Çizelge 3.6. Persons özelliği için koşullu olasılık değerleri.

Persons		acc	good	unacc	vgood
	1	0	0	0	0
2	0	0.0	0.0	0.4760331	0.0
4	0	0.5156250	0.5217391	0.2578512	0.4615385
more	0	0.4843750	0.4782609	0.2661157	0.5384615

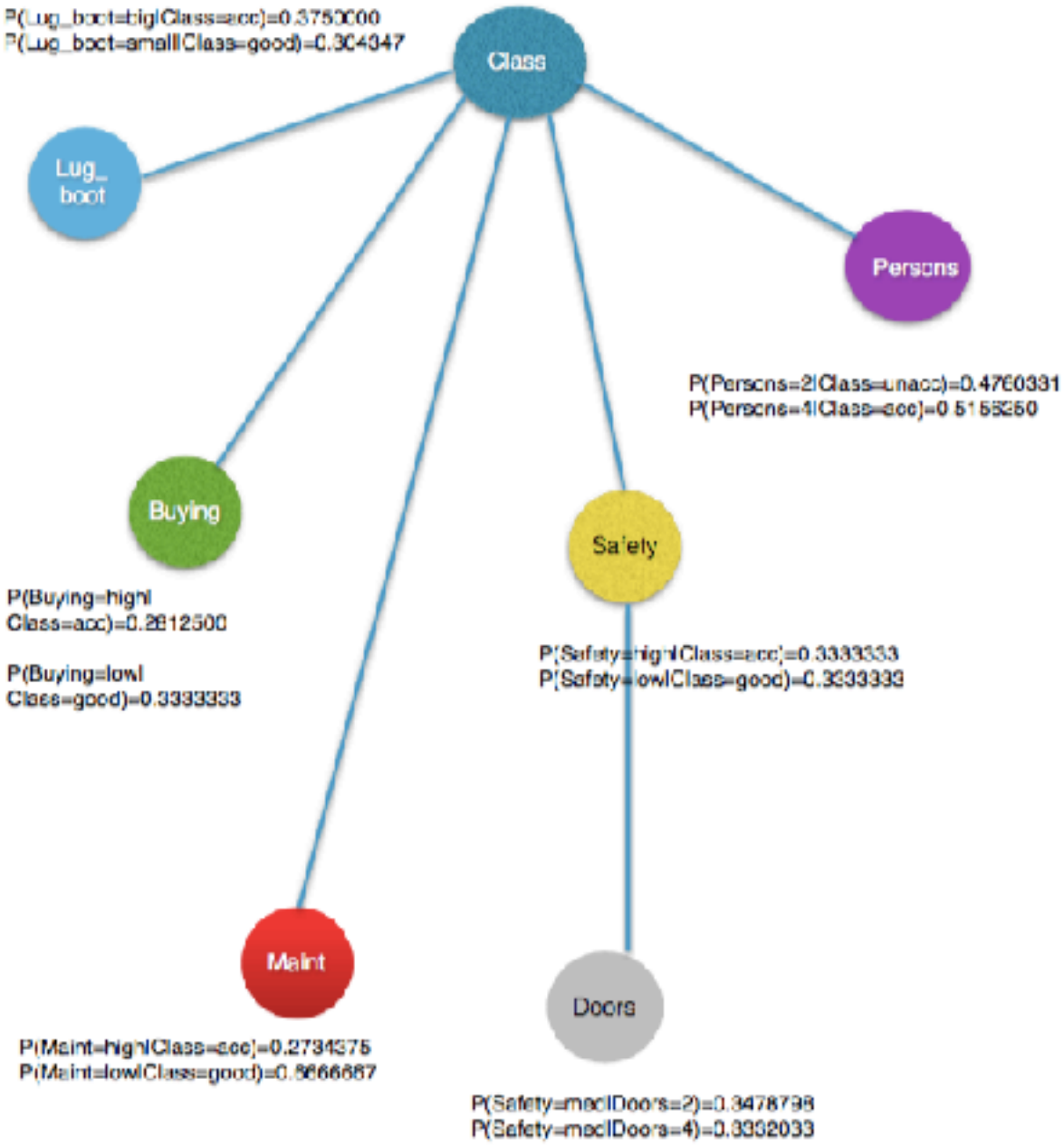
Çizelge 3.7. Lug_boot özelliği için koşullu olasılık değerleri.

Lug_boot		acc	good	unacc	vgood
	1	0	0	0	0
big	0	0.3750000	0.3478261	0.3041322	0.6153846
med	0	0.3515625	0.3478261	0.3239669	0.3846154
small	0	0.2734375	0.3043478	0.3719008	0.0000000

Çizelge 3.8. Safety özelliği için koşullu olasılık değerleri.

Safety		acc	good	unacc	vgood
	1	0	0	0	0
high	0	0.3333333	0.3333333	0.3333333	0.3333333
low	0	0.3333333	0.3333333	0.3333333	0.3333333
med	0	0.3333333	0.3333333	0.3333333	0.3333333

Ayrıca bu çalışmada özelliklerin birbirleri arasında bulunan ilişkileri gösterilmektedir. Verilen bilgilere göre ilişkilerin arasındaki bağı gösteren Bayes Ağı çizdirilmiştir ve Şekil 3.2' de gösterilmektedir.



Şekil 3.2. Bayes Ağı graf gösterimi.

Ek olarak, eğer özellik verilerek sınıf hakkında olasılık hesabı yapılırsa “cpquery” kod satırı kullanılabilmektedir. Buna göre çalıştırılan uygulamanın sonuçları aşağıda gösterilmiştir.

- cpquery(fittedbn, event = (Buying=="high"), evidence = (Class=="acc"))= 0.2727273

Bu satırdaki kodda Buying özelliği high değerine eşit olduğunda sınıfın değerinin acc olarak gelme ihtimali hesaplanmaktadır.

- cpquery(fittedbn, event = (Maint=="vhigh"),evidence = (Class=“good"))= 0

Bu satırdaki kodda Maint özelliği vhigh değerine eşit olduğunda sınıfın değerinin good olarak gelme ihtimali hesaplanmaktadır.

- cpquery(fittedbn, event = (Maint=="med"), evidence = (Doors=="2"))= 0.2504119

Bu satırdaki kodda Maint özelliği med değerine eşit olduğunda kapı değerinin 2 olarak gelme ihtimali hesaplanmaktadır.

- cpquery(fittedbn, event = (Maint=="med"), evidence = (Persons=="2"))= 0.2145034

Bu satırdaki kodda Maint özelliği med değerine eşit olduğunda kişiler değerinin 2 olarak gelme ihtimali hesaplanmaktadır.

- cpquery(fittedbn, event = (Persons=="more"), evidence = (Class=="vgood"))= 0.5107527

Bu satırdaki kodda Persons özelliği more değerine eşit olduğunda sınıfın değerinin vgood olarak gelme ihtimali hesaplanmaktadır.

- cpquery(fittedbn, event = (Safety=="med"), evidence = (Doors=="2"))= 0.3478798

Bu satırdaki kodda Safety özelliği med değerine eşit olduğunda kapı değerinin 2 olarak gelme ihtimali hesaplanmaktadır [44].

4. SONUÇ VE ÖNERİLER

Depolanan veri miktarı hızla arttığı için verileri analiz etmek, değerlendirmek ve verilerden sonuç çıkarmak gerekli bir hale gelmiştir. Çünkü insan hayatında bir konuda çok çeşitli unsurlar göz önünde olunca kimi zaman karar verici sistemlere ihtiyaç duyulmaktadır. Bu noktada bilişim alanındaki yazılımlar sayesinde veriler işlenip kullanıcılara önerilerde bulunabilmekte, kullanıcıların karar vermesine yardımcı olabilmektedir.

Birçok çalışmada verilerin işlenmesi üzerine araştırmalar yapılmıştır. Literatürde çok sayıda yöntem mevcuttur. Tepe tırmanma algoritması hızlı olduğu için tercih edilmiş ve çalışmaya fonksiyon olarak eklenmiştir. Literatürde veriler arasındaki ilişkileri belirlemek için bu şekilde bir kaynak bulunmamaktadır. Bu nedenle bu çalışma ileriki çalışmalar için önemli bir adımdır.

Böylece araba veri setinin değişkenlere göre koşullu olasılık değerleri hesaplanmış, tepe tırmanma algoritmasına tabi tutulmuş ve Bayes Ağı çizilerek verilerin birbirleri arasındaki ilişkileri belirlenmiştir.

Bu çalışmada elde edilen bulgular kullanılan veri seti için otomobil sektöründe değerlendirilebilir. Yani verilen özelliklere göre bir arabanın alınabilir veya alınamaz olup olmadığı bu karar mekanizması ile belirlenebilmektedir. Uygulama sonucunda doğruluk oranı %52 bulunmuştur. Görüldüğü gibi sonuç iyileştirilmeye açıktır. Olasılık fonksiyonu iyileştirilirse bu oran artar ve bu da Denetimli Tavlama gibi bir algoritma ile sağlanabilir. İleriki çalışmalarda bu algoritmanın kullanılması önerilebilir.

Özetlemek gerekirse, bu tez çalışmasında gerekli tanımlar yapıldıktan sonra verilerin depolanması, ardından depolanmış verinin ne şekilde değerlendirilebileceği, günlük hayatın neresinde olduğu gösterilmektedir. Tüm bunlar için gereken işlem, adım ve hesaplamalar açıklanmıştır.

Hesaplamalar sonucu UCI veri seti depolama sitesinden seçilen araba değerlendirme veri setinde 1728 veri ile 6 özellik arasındaki ilişki RStudio ortamı

kullanılarak Tepe Tırmanma Algoritması fonksiyonu ile Bayes Ağı oluşturulup gösterilmiştir. Tüm özellikler için koşullu olasılık hesabı yapılmış ve ağdan sonuç çıkarılmıştır [44].

KAYNAKLAR

[1] Bharati M. Ramageri, Data Mining Techniques and Applications, Indian Journal of Computer Science and Engineering, Vol. 1 No. 4 301-305, 2010.

[2] T.Pala, Tıbbi Karar Destek Sisteminin Veri Madenciliği Yöntemleriyle Gerçekleştirilmesi, (Yüksek Lisans Tezi, Marmara Üniversitesi ,Elektronik Bilgisayar Eğitimi Anabilim Dalı Bilgisayar - Kontrol Programı), İstanbul, 2013.

[3] L.Fan, K.Poh, P.Zhou, Partition-Conditional ICA for Bayesian Classification of Microarray Data, Expert Systems with Applications, 8188-8192, 2010.

[4] N.B.Sebik, H.İ.Bülbül, Veri Madenciliği Modellerinin Akciğer Kanseri Veri Seti Üzerinde Başarılarının İncelenmesi, Türk Bilim Araştırma Vakfı, 1-7, 2018.

[5]. B.Kır Savaş, S.İlkin, S.Hangişi, S.Şahin, Gölge Tespitinde Kullanılan Bayes Sınıflandırma, Otsu Bölütleme ve Histogram Dağılımı Yöntemlerinin Karşılaştırılması, Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 345-355, 2017.

[6] M.O.Olgun, G.Özdemir, İstatistiksel Özellik Temelli Bayes Sınıflandırıcı Kullanarak Kontrol Grafiklerinde Örüntü Tanıma, Journal of the Faculty of Engineering and Architecture of Gazi University, 303-311, 2012.

[7] R.Solmaz, M.Günay and A.Alkan, Fonksiyonel Tiroit Hastalığı Tanısında Naive Bayes Sınıflandırıcının Kullanılması, Akademik Bilişim'14 - XVI. Akademik Bilişim Konferansı Bildirileri, 891-896, 2014.

[8] S.Çelik, M.Şişeci Çeşmeli, İ.Pençe, A.Kalkan, Siğil Tedavisinde Kullanılan Immunotherapy Yönteminin Uygunluğunun Bayes Yöntemi ile Tespiti, 5th International Management Information Systems Conference, 103-106, 2018.

[9]. S.Tufféry . Data mining and statistics for decision making. Chichester: John Wiley & Sons,Ltd.,Publication; 2011, 301-553.

[10] F.Gorunescu, Data Mining: Concepts, models and techniques. Berlin: Springer Science & Business Media, 2011, 186-191.

[11] E.Korkem, *Mikroarray Gen Ekspresyon Veri Setlerinde Random Forest ve Naive Bayes Sınıflama Yöntemleri Yaklaşımı*" Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara, 2013.

[12] D.Lowd and P.Domingos, Naive bayes models for probability estimation, Proceedings of the 22nd International Conference on Machine Learning, 2005.

[13] M.Gitmez, M.Karabatak, A.Varol, Optik Çoğuşma Anahtarlama Ağ Verisi Üzerinden Sınıflandırma Algoritmalarının Karşılaştırılması, 2019, IEEE.

[14] Al-Hudairy, Hazem H. M. Abd Al-Rahman. Data mining and decision making support in the governmental sector" Master Thesis, Louisville University, Kentucky, 2004.

[15] P.Domingos and M.Pazzani, On the Optimality of the Simple Bayesian Classifier under Zero-One Loss, Machine Learning, 29 (1997) 103-130.

[14] Al-Hudairy, Hazem H. M. Abd Al-Rahman. Data mining and decision making support in the governmental sector" Master Thesis, Louisville University, Kentucky, 2004.

[16] Burhan Yamuk. *Elektronik postaların ayrıştırılmasında naive bayesian ve bulanık mantık yöntemlerinin karşılaştırılması*" Yüksek Lisans Tezi, Gazi Üniversitesi Türkiye, 2011.

[17] C.İnal, S.Günay, Olasılık ve matematiksel istatistik. Ankara: Hacettepe Üniversitesi; 2010.

[18] Sema Güzel. Veri Madenciliğinde Sınıflandırma Algoritmaları Kullanılarak Hepatit Hastalığının Tespiti" Yüksek Lisans Tezi, Kahramanmaraş Sütçü İmam Üniversitesi Türkiye, 2018.

[19] B.Lantz, Machine learning with R. Birmingham: Packt Publishing Ltd, 2015:89-124.

[20] R.Zafarani, M.A.Abbasi, H.Liu, Social media mining: an introduction.Cambridge University Press, 2014, 135-171.

[21] J.Han, M.Kamber, J.Pei, Data Mining: Concepts and Techniques, San Francisco: Morgan Kaufmann Publishers, 2006, 285-378.

[22] M.Fatima, M.Pasha, Survey of Machine Learning Algorithms for disease diagnostic, *Journal of Intelligent Learning Systems and Applications*, Vol.9, No.1 (2017) 1-16.

[23] E.Dünder, *Bayesci Ağlarda Öğrenme Algoritmalarının Karşılaştırılması"* Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi, Samsun, 2013.

[24] M.Karabıyık, B.Yet, Bayes ağları ile futbol analitiği: FutBA modeli, Pamukkale Univ. Muh. Bilim Derg., 121-131, 2019.

[25] M.Atalay, H.Yorulmaz, O.Önay and E.N.Çinicioğlu, Trafik Kazaları Analizi İçin Bayes Ağları Modeli, Yöneylem Araştırması ve Endüstri Mühendisliği 31. Ulusal Kongresi, Sakarya, 2011.

[26] S.C.Yücebaş, Hipokrat-I: Bayes ağı tabanlı tıbbi teşhis destek sistemi,(Başkent Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Yüksek Lisans Programı), Ankara, 2006.

[27] Z.D.Akşehir, E.Pekel, S.Akleylek, E.Kılıç and Y.Oruç, İş Sağlığı ve Güvenliği Sektöründe Bayes Ağları Uygulaması, Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 47-59, 2019.

[28] L. Euler, Solutio Problematis and Geometriam Situs Perinents, 1736.

[29] K. R. Saoub, A Tour Through Graph Theory, London: Chapman and Hall CRC, 2017.

[30] F. Harary, Graphical Enumeration, New York & London: Academic Press, 1973.

[31] M.Guan, Graphic Programming Using Odd or Even Points, Chinese Mathematics 1, 1962, 273-277.

[32] Anonymous. (2018). https://www.matematikrehberim.com/oku.php?sef=kopruden-gecmek-icin-graf-modeli&makaleid=122 (on-line access on 07 June, 2020).

[33] Anonymous. (2020). https://www.matematikrehberim.com/oku.php?sef=kopruden-gecmek-icin-graf-modeli&makaleid=122 (on-line access on 09 Dec, 2020).

[34] V.Yiğit, O.Türkbey, Tesis Yerleşim Problemlerine Sezgisel Metotlarla Yaklaşım, Gazi Üniv. Müh. Mim. Fak. Der. Cilt 18, No 4, 45-56, 2003.

[35] B.E. Türkay, F. Küçüktezcan ve A. Bulut, Elektrik Enerjisinin Bölgeler Arası Alışverişinin Optimizasyonu, EMO Bilimsel Dergi, 1:1(2011) 31-38. Makale

[36] Anonymous. (2009).http://bilgisayarkavramlari.com/2009/12/02/tepe-tirmanma-algoritmasi-hill-climbing-algorithm/ (on-line access on 13 Şubat 2020).

[37] Anonymous. (2020). https://www.youtube.com/watch?v=SxjesZtT-tc (on-line access on 25 March, 2020).

[38] Anonymous. (2020). https://www.youtube.com/watch?v=EHcEnuG5QXo&t=1081s (on-line access on 25 March, 2020).

[39] Anonymous. (2020).https://www.youtube.com/watch?v=p_VQhdkKw8U&t=727s (on-line access on 27 March, 2020).

[40] Anonymous. (2020). http://archive.ics.uci.edu/ml/datasets/Car+Evaluation (on-line access on 09 Dec, 2020).

[41] Anonymous. (2020). https://oyademirkesen.wordpress.com/2016/11/27/ilk-blog-gonderisi/ (on-line access on 09 Dec, 2020).

[42] Scutari, M. (2011). Measures of variability for graphical models.

[43] R programlama dili nedir? (2020). https://medium.com/datarunner/r-nedir-4375f53ba1d4(on-line access on 09 Dec, 2020).

[44] Oymak, E., Karcı, A. (2019).Örnek Verilerle İnanç Ağlarının İnşa Edilmesi (pp: 17-24). 2. Uluslararası Mühendislik ve Teknoloji Yönetimi Kongresi, Ekim 24-25, Mardin.

ÖZGEÇMİŞ

Ad Soyad **:** Elif Aslı Oymak

ÖĞRENİM DURUMU:

- **Lisans:** 2015, İnönü Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü

MESLEKİ DENEYİM:

- İnönü Üniversitesi Fen Bilimleri Enstitüsü, Yazılım Mühendisliği Ana Bilim Dalı, Araştırma Görevlisi, 2020- devam ediyor .

YÜKSEK LİSANS TEZİNDEN TÜRETİLEN ÇALIŞMALAR

- **Oymak, E., Karcı, A. (2019).** Örnek Verilerle İnanç Ağlarının İnşa Edilmesi (pp: 17-24). 2. Uluslararası Mühendislik ve Teknoloji Yönetimi Kongresi, Ekim 24-25, Mardin (Bildiri).