

GÜL HASTALIĞI TEŞHİSİ İÇİN KARAR AĞACI ALGORİTMALARININ KARŞILAŞTIRMALI BİR ANALİZİ**

** Bu çalışmanın özeti 13-14 Mayıs 2022 tarihlerinde gerçekleştirilen Uluslararası Marmara Bilim Kongresi'nde sunulmuştur.

Burcu DURMUŞ^{1,a,*}, Öznur İŞÇİ GÜNERİ^{1,b}, Nevin GÜLER DİNCER^{1,c}

¹ Muğla Sıtkı Koçman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Muğla

^a burcudurmus@mu.edu.tr, ORCID: 0000-0002-0298-0802

^b oznur.isci@mu.edu.tr, ORCID: 0000-0003-3677-7121

^c nguler@mu.edu.tr, ORCID: 0000-0003-0361-1803

ÖZET

Veri madenciliği konuları arasında sıklıkla kullanılan sınıflandırma yöntemi, pek çok algoritmaya kıyasla pratik ve hızlı çözümler sunan alternatif bir yöntemdir. Sınıflandırma yönteminde veriler içerisindeki bilgiler, bağıntılar, desenler ve benzerliklerden yola çıkılarak çeşitli algoritmalar yardımıyla model oluşturulur. Bu model üzerinden yeni gözlemler için sınıf tahmini yapılır. Bu çalışmada, farklı karar ağacı algoritmaları ile gül hastalığı teşhisi için (hasta-hasta değil) sınıflandırma analizi yapılmıştır. Çalışmada amaç; gözlemin hasta-hasta değil şeklinde sınıflandırılmasından ziyade, bu ayırım yapılırken kullanılan değişkenlerin neler olduğunun belirlemesi ve literatür ile kıyaslanmasıdır. Eğitim ve test aşamasında veriler, çapraz doğrulama ile karşılaştırılmıştır. En başarılı yöntem tespit edilirken doğruluk, kesinlik, duyarlılık, F-ölçütü, MCC, ROC Area, PRC Area ve Kappa değerleri göz önüne alınmıştır. Yapılan analizler sonucunda en yüksek sınıflandırma başarısı, REP Tree algoritması ile %91.5 olarak elde edilmiştir. REP Tree algoritmasının sonucuna göre gül hastalığı için; kök düğümde yer alan güneş değişkeni güneş kremi kullanımının ve güneşten kaçınmanın önemini, iç düğümde yer alan stres değişkeni ise stresten uzak kalmanın önemini vurgulamaktadır.

Anahtar Kelimeler: Karar ağaçları, sınıflandırma, makine öğrenmesi, gül hastalığı.

*Sorumlu Yazar (Corresponding Author)

Atrif (Citation): Durmuş, B., Güneri, Ö. İ., Dincer, N. G., "Hastalık Teşhisi için Karar Ağacı Algoritmalarının Karşılaştırmalı Bir Analizi", UMÜFED Uluslararası Batı Karadeniz Mühendislik ve Fen Bilimleri Dergisi, 5(2): 13-35, 2023.

Geliş (Received): 11/10/2023

Kabul (Accepted): 28/12/2023

Yayın (Published): 31/12/2023

A COMPARATIVE ANALYSIS OF DECISION TREE ALGORITHMS FOR ROSE DISEASE DIAGNOSIS

ABSTRACT

The classification method, which is frequently used among data mining topics, is an alternative method that offers practical and fast solutions compared to many algorithms. In the classification method, a model is created with the help of various algorithms based on the information, correlations, patterns and similarities in the data. Class prediction is made for new observations using this model. In this study, classification analysis was performed for rosacea diagnosis (patient-non-patient) with different decision tree algorithms. The main purpose of the study is; rather than classifying the observation as patient or non-patient, the aim is to determine the variables used in making this distinction and compare them with the literature. During the training and testing phase, the data were compared with cross-validation. While determining the most successful method, accuracy, precision, sensitivity, F-measure, MCC, ROC Area, PRC Area and Kappa values were taken into consideration. As a result of the analysis, the highest classification success was achieved as 91.5% with the REP Tree algorithm. According to the results of the REP Tree algorithm, for rosacea; The sun variable located in the root node emphasizes the importance of using sunscreen and avoiding the sun, and the stress variable located in the internal node emphasizes the importance of staying away from stress.

Keywords: Decision trees, classification, machine learning, rosacea.

1. GİRİŞ

Veri madenciliği, veri yığınları içerisinde yer alan ortaya çıkarılmamış gizli ve faydalı bilgilerin belirli kurallar çerçevesinde ortaya çıkarılarak analiz edilmesi sürecidir (Ata ve Erbudak, 2022). Son yıllarda teknolojinin gelişmesi ve veriye erişimin kolaylaşmasıyla birlikte veri madenciliği yöntemlerinin uygulamalarında önemli bir artış gözlemlenmiştir.

Sağlık alanında veri madenciliği çalışmalarına ise ilk kez 1970'lerde temelleri atılan ve daha sonraki yıllarda geliştirilen uzman sistemler ile başlanmıştır. Ancak sağlık alanındaki verilerin hızla değişmesinden dolayı uzmanlar arasında görüş farklılıkları meydana gelmiş ve sağlık alanında veri madenciliği çalışmaları kesintiye uğramıştır. 1990'lı yıllara gelindiğinde sağlık durumu ve tedavi maliyet tahminleri için sinir ağları kullanılmış ve bu yıllardan sonra

sağlık alanında veri madenciliği yöntemlerinin kullanımı yaygınlaşmaya başlamıştır (Yıldırım vd., 2008). Veri madenciliği yöntemleri sağlık ile ilgili olarak test sonuçlarının tahmini, hastalık tanı ve teşhisi, hastalık düzeyinin belirlenmesi, ölüm ve hastalık varlığının tahmin edilmesi, klinik dokümanların analiz edilmesi, hastane, ilaç ve tedavi maliyet hesaplanması gibi pek çok konuda kullanılmaktadır.

Rozase (gül hastalığı), erişkinlerde sıkça rastlanan kronik dermatolojik hastalıklardan biridir (Webster, 2003). Halk arasında gül hastalığı olarak bilinmektedir. Gül hastalığına ait semptomlar her yaşta görülebilir; ancak 30-60 yaşları arasındaki bireylerde hastalık varlığı ile daha sık karşılaşılır (Zuber, 2000). Alevlenme ve iyilik halleri gibi durumlar görülen bu hastalık kadınlarda daha çok görülmesine rağmen, erkeklerde daha ciddi seyretmektedir. Gül hastalığı ırk ayrımı yapmamaktadır (Litt, 1997). İsveç'te yapılan bir çalışmada hastalığın toplumda görülme sıklığı %10 olarak raporlanmıştır (Berg ve Liden, 1989). Hastaların %57'si bu hastalığın sosyal yaşamlarını olumsuz etkilediğini ifade etmiştir (Zuber, 2000).

Son yıllarda gül hastalığı üzerine yapılan araştırmalarda hastalığın altta yatan sebeplerinin araştırılması üzerine yoğunlaşmıştır. Birçok gözlemsel çalışmada gül hastalığının bazı sistemik hastalıklarla ilişkisi olduğu ortaya konmuştur. Hastalığın oluşumunun ve gelişiminin nedenleri tam olarak bilinmemektedir; ancak sigara ve alkol tüketimi, Helicobacter pylori enfeksiyonu, depresyon, hipertansiyon, kardiyovasküler hastalıklar, anksiyete bozukluğu, dislipidemi, diyabet, migren, romatoid artrit, ülseratif kolit ve demans gibi birçok sistemik hastalığın gül hastalığına daha sık eşlik ettiği düşünülmektedir (Aldrich vd., 2015; Chosidow ve Cribier, 2011; Haber ve El Gemaye; 2018; Spöndlin vd., 2012; Yorulmaz ve Kulcu, 2015). Bu durum, gül hastalığının deriye sınırlı lokalize bir hastalık mı yoksa bazı sistemik bulgu ve sonuçları olabilecek bir hastalık mı olduğu sorusunu gündeme getirmektedir (Holmes vd., 2018).

Gül hastalığının kesin bir tedavisi yoktur. Hastalara uygulanan tedavinin amacı ciltteki belirtileri hafifletmek, hastalığın kötüye gitmesini önlemek ve şikayetlerin azalmasını sağlamaktır. Bunun için çeşitli yöntemler kullanılmaktadır (Aybey, 2022; Bingöl, 2021; Anadolu Sağlık, 2021). Oral akne ilacı; kistli akneyi veya diğer hafif ilaçlara karşı inatçı olan akneyi tedavi etmek için kullanılan ilaçlardır. Oral antibiyotikler ise; iltihaplanmanın çoğalmasını engelleyerek hastalığın gerilemesine yardımcı olur. Ayrıca akne sayısının azalmasında da etkilidir. Oral antibiyotikler şiddetli aknesi olan ya da kremlerden fayda

göremeyen hastalarda tercih edilmektedir. Ağrı kesiciler; hastalığın ilerleyen aşamalarında hissedilen sancılar için etkili olmaktadır. Kızarıklık giderici kremler; cilt üzerindeki kırmızı görüntüyü azaltmaya yardımcı olmaktadır. Lazer ve diğer ışık bazlı cihazlar; ciltteki kırmızı görüntünün giderilmesi için kullanılan alternatif bir yöntemdir. Düzenli uygulanan cilt bakım rutinleri; hastalığın alevlenme sıklığı, kırmızı görünümü ve hastalığa bağlı oluşan diğer belirtilerin azaltılmasına yardımcı olmaktadır. Cilt bakım rutini esnasında yüze çok dokunulmamalı, yüz bölgesi keselenmemelidir. Temizlik için uygun dermatolojik ürünler tercih edilmeli ve cilt düzenli olarak nemlendirilmelidir. Tetikleyicilerin tespiti; hastalığın şiddetine doğrudan etki eden tetikleyici durumların, yiyeceklerin, içeceklerin belirlenmesi alınacak önlemler için gereklidir. Antioksidan özelliği olan ve yüksek miktarda B vitamini içeren yiyecekler (avokado, zencefil, lahana vb.) tüketilmelidir. Düzenli olarak güneş kremi kullanmak hastalığın belirtilerini azaltmaya yardımcı olan bir başka yöntem olarak değerlendirilebilir. Yaygın olmamakla birlikte cerrahi yöntemler hastalığın son aşamasında kullanılabilir. Ancak 13-cis-retinoic acid adı verilen bu tedavi, oldukça fazla yan etkiye sahiptir ve pek fazla önerilmemektedir.

Gül hastalığının teşhisi, hastalığın seyri ve uygulanacak tedavinin doğru ve etkili bir şekilde uygulanması açısından önem teşkil eder. Gül hastalığı, genel olarak klinik muayene ile tespit edilebilen bir hastalıktır. Ancak hastanın birden fazla cilt problemine aynı anda sahip olduğu bazı durumlarda hastalığın ayırt edilmesi zorlaşabilir. Bu gibi durumlarda, son yıllarda oldukça popüler olan ve üzerine sıklıkla çalışılan bir konu olan sınıflandırma yönteminden yararlanılabilir.

Heo ve arkadaşları tarafından yapılan çalışmada (2021) veri madenciliği yöntemi aracılığıyla akne, gül hastalığı, büllöz deri hastalığı ve piyodermagangrenozum gibi birçok cilt hastalığı için önerilen doksisisiklin adındaki antibiyotiğin olumsuz etkilerinin tespit edilmesi amaçlanmıştır. Çalışmada otuz yedi sinyal arasından gül hastalığı dahil on dokuzunun altı ülkenin hiçbir ilaç etiketinde yer almadığı ortaya konmuştur.

Cho ve arkadaşları (2020) klasik bir tıp metnini veri madenciliği yoluyla cilt bakımı için aday şifalı bitkilerin kapsamlı bir listesini hazırlamışlardır. Çalışmadan elde edilen sonuçlara göre inulae flos, gül hastalığı ile ilişkilendirilen tek bitki olmuştur.

Rajalingam ve arkadaşları 2023 yılındaki çalışmalarında 200.000'den fazla gönderiden oluşan denetimsiz bir makine öğrenimi analizi ile gül hastalığına sahip hastaların birincil

endişelerinin yanı sıra çevrimiçi hastalık arama kalıplarını daha iyi anlamayı araştırmışlardır. Gül hastalığı için büyük bir çevrimiçi forumda yapılan tüm gönderilerin analiz edilmesi, önemli endişe konularına ve tedavide belirtilen tercihlere ışık tutmuş ve buna ek olarak, olumlu veya olumsuz duyguların açığa vurulması ve destek mesajları da dahil olmak üzere daha önce keşfedilmemiş anlatılar keşfedilmiştir.

Kim ve Kim (2023) beş bitkiden oluşan bitkisel bir ilaç olan gyejibokryeong-hwan (GBH)'ın gül hastalığı üzerindeki potansiyel farmasötik mekanizmasını araştırmışlar ve ağ analizine dayalı gül hastalığı için 4 kılavuzda önerilen kimyasal ilaçlarla karşılaştırmalı analiz yoluyla GBH'ye özel terapötik noktaları araştırmışlardır. Yaptıkları çalışma GBH'nin gül hastalığı üzerindeki potansiyel altta yatan mekanizmasını aydınlatmıştır. GBH'nin, potansiyel olarak "IL-17 yolu" ve "nöroinflamatuvar yanıt" yoluyla kapsamlı bir farmakolojik şekilde gül hastalığı üzerinde etkili olabileceği ve gül hastalığı tedavisine yönelik mevcut kimyasal ilaçların hedef almadığı "damar yarası iyileşmesinde" rol oynayabileceği ortaya konmuştur.

Literatür incelendiğinde, gül hastalığının tespiti için karar ağaçlarına yönelik bir çalışmanın mevcut olmadığı görülmektedir. Oysa uygun karar algoritmaları ile veriler kolayca modellenebilir ve hastalık teşhisi yapılabilir. Bu çalışmada, gül hastalığının sınıflandırma yöntemlerinden karar ağaçları yapılarıyla modellenmesi üzerinde durulmuştur. İki sınıflı olan (hasta, hasta değil) gül hastalığı verileri için, farklı karar ağacı algoritmaları ile karar yapıları oluşturulmuş ve performans değerlendirilmesi yapılmıştır. Sınıflandırma sonuçları farklı performans kriterleri ile tartışılmıştır. Başarımı yüksek karar yapısına sahip algoritmalar tedavi sürecine yardımcı olmak amacıyla uzmanlara önerilmektedir.

2. MATERYAL VE YÖNTEM

Bu çalışmada Belli vd. (2016) tarafından Muğla Sıtkı Koçman Üniversitesi Eğitim ve Araştırma Hastanesi Dermatoloji Polikliniği'nde toplanan gül hastalığı verileri kullanılmıştır. Verilerin farklı karar ağacı algoritmaları yardımıyla sınıflandırması yapılmış ve algoritmaların performans analizleri karşılaştırılmıştır.

2.1. Veri Seti

Çalışmada kullanılan veri seti Belli vd. (2016) tarafından yapılan çalışmadan elde edilmiştir. Veri setinde yer alan kontrol grubu, polikliniğe gül hastalığı dışında çeşitli dermatolojik şikayetlerle başvuran hastalardan ardışık olarak seçilmiştir. 94 örnek içeren veri

seti teşhis sonucu olarak hasta ve hasta değil olarak işaretlenmiştir. Veri setinde sınıf kategorisi hariç 28 öznitelik bulunmaktadır. Tablo 1’de veri seti içerisinde bulunan 29 özniteliliğin tanımları, değer aralıkları, ortalamaları ve standart sapma değerleri verilmiştir.

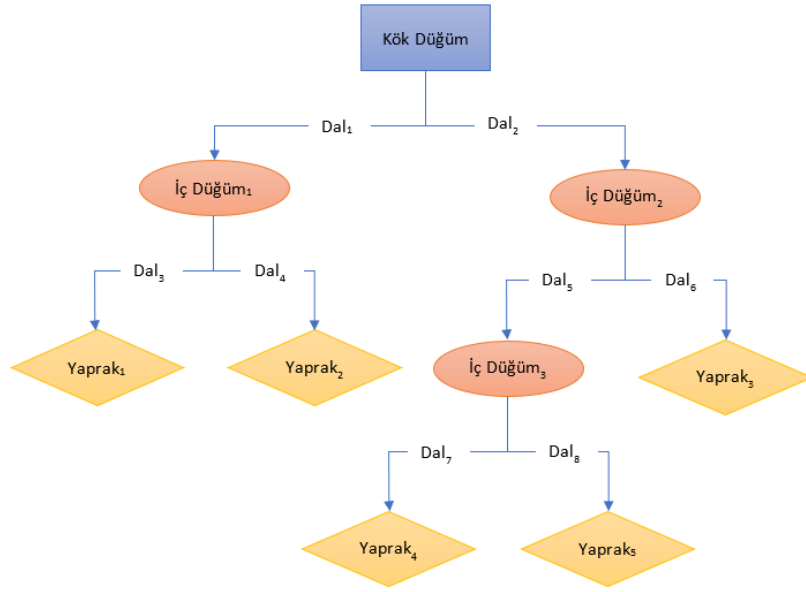
Tablo 1. Verilere ilişkin öznitelik açıklamaları ve değerleri

Numara	Öznitelik Tanımlaması	Değer Aralığı	Ortalama	Standart Sapma
1	cinsiyet	1-2	-	-
2	yaş	35-78	50.404	9.065
3	güneş	0-1	-	-
4	baharat	0-1	-	-
5	stres	0-1	-	-
6	sıcak	0-1	-	-
7	soğuk	0-1	-	-
8	sigara tüketimi	0-1	-	-
9	alkol tüketimi	0-1	-	-
10	egzersiz yapma	0-1	-	-
11	kilo	50-97	71.479	10.192
12	boy	150-182	162.521	7.028
13	bel çevresi	62-120	87.745	11.102
14	ldl	57-210	126.045	30.109
15	trigliserid	36-324	116.979	62.568
16	kolesterol	132-300	208.851	36.315
17	açlık kan şekeri	69-135	94.404	10.663
18	insülin	3.24-26.37	9.463	4.876
19	crp	0.06-89.51	3.693	9.755
20	hdl	31-104	58.138	15.122
21	sistolik kan basıncı	90-170	122.766	17.049
22	diyastolik kan basıncı	60-100	78.191	9.27
23	ailede hastalık öyküsü	0-1	-	-
24	bmi	20.9-36.5	27.074	3.779
25	insülin direnci	0.68-6.38	2.234	1.263
26	metabolik sendrom	0-1	-	-
27	hgba1c	0.06-6.96	4.655	2.202
28	antiht	0-1	-	-
29	sınıf	1-2	-	-

2.2. Karar Ağaçları

Sınıflandırma analizinde en yaygın kullanılan yöntemlerin başında gelen karar ağaçları, istatistiksel olarak birbirleriyle anlamlı olan grupları tespit ederek sonuçları kolay okunabilir ağaç diyagramlarına dönüştüren kurallar bütünüdür.

Karar ağacının genel yapısı dal, düğüm (kök düğüm ve iç düğüm) ve yapraklardan oluşmaktadır (Şekil 1). Karar ağacının yapısındaki düğümler nitelikleri temsil eder. Dallar ve yapraklar ise ağacın birer elemanlarıdır. Kök düğüm, ağaç yapısının en üst seviyesinde yer alır ve bir ağaç yapısında yalnızca bir kök düğüm vardır. Ağacın en altındaki elemanlar yapraklardır. Kök düğüm ve yapraklar arasında kalan elemanlar ise dal olarak isimlendirilir (Alpaydın, 2000).



Şekil 1. Karar ağacı yapısı

Karar ağaçlarında her bir dalın gerçekleşme olasılığı vardır. Dolayısıyla düğümlerden dallara veya dallardan düğümlere ulaşma olasılığı hesaplanabilir. Karar ağaçlarında kurallar (IF-THEN rules) kökten yaprağa doğru ilerleyerek oluşturulur.

Karar ağacı yönteminde sınıflandırma işlemi için veri seti, eğitim ve test olarak ikiye ayrılır. Eğitim verileri öğrenme işlemi için kullanılır. Bu aşamada kurallar oluşturulur. Test verileri bu kurallar çerçevesinde sınıflandırılır ve kuralların başarısı test edilir. Başarılı sonuçlar elde edilmesi durumunda kurallar, yeni verilerin sınıflandırılması amacıyla kullanılabilir (Sarmanova ve Albayrak, 2013).

Karar ağaçları ile yapılan sınıflandırmada eğitim verileri tek sınıf içeren alt kümelere bölünür. Bu işlemin sonucunda karmaşık ve büyük bir ağaç yapısı elde edilir. Bu durumda ağaç yapısının gereksiz karmaşıklığını azaltmak için 'budama' yapılır. Budama, oluşturulan

ağaç yapısındaki alt kümelere yaprak atanması ile ağaç yapısının daha küçük bir yapıya indirgenmesidir (Cheng vd., 1988; Friedl ve Brodley, 1997).

Sınıflandırma analizi için geliştirilmiş pek çok karar ağacı algoritması bulunmaktadır. Bölümün devamında çalışma boyunca kullanılan karar ağacı algoritmaları açıklanmıştır.

2.2.1. ID3, C4.5, J48 Algoritmaları

Karar ağaçları için geliştirilen ve yaygın bir şekilde kullanılan algoritmaların başında ID3, C4.5, J48 algoritmaları gelmektedir.

ID3 algoritması, eğitim verilerindeki gözlem ve nitelik sayısının fazla olması durumunda yoğun hesaplama yapmaya ihtiyaç duymadan bir karar yapısı oluşturur. Kategorik veriler için işlem yapan algoritma, Quinlan tarafından sayısal verilerin de kullanılmasını sağlayacak şekilde güncellenmiştir (Quinlan, 1993; Şatır, 2016).

C4.5 algoritması ise, ID3 algoritmasına dayanarak ID3 algoritmasındaki bazı eksikliklerin giderilmesi ile geliştirilmiştir (Bahety, 2014). Bu algoritmada bilgi kazancı oranı, öznitelik seçim ölçütü olarak kullanılır. Böylece en yüksek bilgi kazancı oranına sahip öznitelik seçilir (Bulut, 2016; Şatır vd., 2016).

J48 algoritması, ID3 ve C4.5 algoritmalarının bir türevidir ve bu algoritmaların Weka'ya uyarlanmış halidir. J48 algoritmasında, en yüksek bilgi kazancısına sahip nitelikler kullanılarak oluşturulan ağaçtan sonuç kararına ulaşılır (Quinlan, 1993). Bu algoritmada amaç, karar ağacını optimize etmektir (Coşkun ve Baykal, 2011).

ID3, C4.5 ve J48 algoritmaları entropi ve bilgi kavramlarına dayanmaktadır. Entropide amaç, veri setini sınıflara ayırmak için gerekli bilgi miktarını hesaplamaktır (Çetinkaya, 1986). Entropi, Eşitlik 1 ile hesaplanır.

$$Entropi(S) = \sum_{i=1}^n -p_i \log_2(p_i) \quad (1)$$

Burada p_1, p_2, \dots, p_n hedef değişkenin sınıflarına ait olasılıkları; S ise niteliğin sahip olabileceği farklı değerler miktarını göstermektedir.

Karar ağaçları yöntemlerinde baskın değişkenin tespiti için her bir niteliğe ait bilgi kazancı değeri hesaplanır. M özelliğinin S örneği için bilgi kazancı Eşitlik 2 ile hesaplanır (Bahety, 2014). Bilgi kazancı en yüksek olan özellik, kök düğüm olarak ağaç yapısında yerini alır.

$$Bilgi Kazancı(M, S) = Entropi(S) - \sum_{j=1}^n \frac{S_j}{S} Entropi(S_j) \quad (2)$$

ID3, C4.5, J48 algoritmaları bazı avantajlara sahiptir (Akbal vd., 2017):

- Hem kategorik hem de sayısal veriler için kullanılabilir.
- Eğitim verilerinde eksik değer bulunması durumunda da uygulanabilir.
- Eğitim verilerinin istenmeyen değerlerini eleyebilir (Kaynar vd.; 2016Ture vd., 2009).
- Aşırı uygunluk problemini ortadan kaldırılır.

2.2.2. LMT Algoritması

LMT algoritması, lojistik regresyon ve karar ağacını birleştiren bir sınıflandırma modelidir. Sıradan karar ağaçları, yaprakları parçalı bir sabit model oluşturur. LMT algoritmasında ise ağacın yaprakları parçalı doğrusal regresyon modeline sahip bir karar ağacıdır (Landwehr vd., 2003; Landwehr vd., 2005). Bu algorithmada LogitBoost algoritması, ağacın her düğümünde lojistik regresyon modeli oluşturur. Ardından düğüm C4.5 algoritması kullanılarak ayrıştırılır. Her LogitBoost işleminde kendi sonuçlarından üst düğüm üzerinden yeniden başlatılır. Son adımda ağaç budanır (Sumner vd., 2005).

2.2.3. Random Tree Algoritması

Random Tree algoritması, her düğümde belirli sayıda rastgele seçilmiş özellikleri alır ve ağacı oluşturur (Breiman, 2001). Bu algorithmada budama işlemi yapılmaz. Random Tree algoritmasının veri setine dayalı sınıf olasılıklarının tahmin edilmesine izin veren bir opsiyonu vardır.

2.2.4. Random Forest Algoritması

Random Forest algoritmasında birden fazla karar ağacı oluşturulur. Elde edilen ağaçlar birleştirilerek karar ormanları elde edilir. Random Forest algoritmasında karar ağacı, her bir düğümde rastgele alınan niteliklerin en iyisi seçilerek ve tüm düğümler dallara ayrılarak oluşturulur. Ağaç yapısındaki her bir veri seti, gerçek veri setinden yer değiştirmeli olarak türetilir. Böylece Random Forest algoritması, diğer algoritmalara göre daha hızlı sonuçlar vermektedir. Rastgele özellik seçimi ile ağaçlar geliştirilir. Bu yöntemde budama işlemi uygulanmaz (Breiman, 2001).

2.2.5. LAD Tree Algoritması

LAD Tree algoritması, veri setindeki pozitif ve negatif örnekleri ayırt edebilen mantıksal bir ifadeye dayanır (Holmes vd., 2002). Model oluşturmak için ilk adım, büyük grup desenleri oluşturmaktır. Ardından büyük desen içerisinde modeldeki yaygınlık ve homojenlik açısından belirli gereksinimleri sağlayan gruplar belirlenir. Son aşamada bu gruplardan alt gruplar seçilir (Wisaeng, 2013).

2.2.6. REP Tree Algoritması

REP Tree algoritması, bölme kriteri olarak bilgi kazan ölçütünü kullanır ve azaltılmış hata yöntemine dayalı olarak budama yapar. Son adımda da karar ya da regresyon ağacını oluşturur (Witten vd., 2011). Bu yöntem ile sadece nümerik verilerin modellenmesi yapılır. Kayıp veriler için ise C4.5 algoritması kullanır (Zhao ve Zhang, 2008).

2.2.7. Decision Stump Algoritması

Decision Stump algoritması, tek seviyeli karar ağacı oluşturan bir yöntemdir. Bu yöntemde karar ağacının kök düğümü, direk olarak yaprak düğümlere bağlanır. Decision Stump, sınıflandırma işlemini direk olarak tek bir öznelik değerine bağlı gerçekleştirir. Algoritma genelde tek başına sınıflandırma yapmak için değil, boosting yöntemleri ile hibrit yöntemler oluşturmak için kullanılır (Witten ve Frank, 2005).

2.2.8. Hoeffding Tree Algoritması

Hoeffding Tree algoritması, her gözlemi en fazla bir kez okur ve işler. Büyük veri setlerinde etkin bir şekilde çalışır. Ayrıca Hoeffding Tree algoritması, C4.5 gibi temel (basic) karar ağacı algoritmalarının depolama sorununu çözer. Hoeffding Tree algoritması, oldukça karmaşık karar yapılarında bile kabul edilebilir bir hesaplama maliyeti ile çalışır. Algoritma, karar ağacının her bir düğümünün nasıl parçalanacağına Hoeffding sınırı adı verilen istatistiksel bir değer ile karar verir. Hoeffding Tree algoritmasının önemli bir özelliği, karar ağacının her bir düğümünün test edilmesinin tüm verileri kullanan sınıflandırıcılar ile hemen hemen aynı olmasıdır (Domingos ve Hulten, 2000).

2.2.9. Optimized Forest Algoritması

Optimized Forest algoritması, Random Forest veya Bagging algoritmaları ile optimal bir alt orman belirleyerek karar yapısı oluşturma temellerine dayanan bir yöntemdir (Adnan ve İslam, 2016).

2.2.10. AD Tree Algoritması

AD algoritması, karar ve tahmin düğümlerinden oluşur. Karar düğümleri bir eylemin sonucunu gösterir. Tahmin düğümleri tek bir sayı içerir. AD ağaçları daima hem karar hem de tahmin düğümlerine sahiptir. Test verisi (ya da herhangi bir gözlem) için sınıflandırma, her tahmin düğümünün ve bütün karar düğümlerinin doğru olduğu yollar takip edilerek yapılır (Freund ve Mason, 1999).

2.2.11. BF Tree Algoritması

BF algoritması, böl ve yönet mantığına dayanır. İlk adım olarak, bir nitelik kök olarak atanır. Bu nitelik üzerinden kriterlere göre gruplar ayrılır. Ardından eğitim verileri, kök düğüm esas alınarak alt gruplara bölünür. Her döngüde, genişlemeye en uygun olan alt grup seçilir (Shi, 2007). Bu süreç, belirli bir genişleme katsayısına göre tüm düğümler netleşinceye kadar sürdürülür.

2.2.12. Sys For Algoritması

Sys For algoritması hem ikili hem de çok sınıflı veri kümeleri için çok sayıda karar ağacı ile oluşturulmuş bir yöntemdir (Islam ve Giggins, 2011). Bir Sys For algoritması için aşağıdaki adımlar izlenir:

- Bölünme noktaları bulunur.
- Ağaç sayısı oluşturulur.
- Oluşturulan ağacın 1. seviye düğümü alınır ve rastgele bir nitelik seçilir.
- Önceki adımlarda oluşturulan tüm ağaçlar, Sys For olarak döndürülür.

Bu yöntemde model, karar ağaçlarına ve karar ormanlarına benzer şekilde çalışır (Al-Saggaf ve Nielsen, 2014).

2.3. Performans Göstergeleri

Performans göstergeleri, iki sınıflı bir problem için Tablo 2 ile verilen karmaşıklık (hata) matrisi üzerinden incelenebilir.

Tablo 2. Karmaşıklık matrisi

		gözlemlenen sınıf	
		sınıf=1	sınıf=2
gerçek sınıf	sınıf=1	TP	FN
	sınıf=2	FP	TN

TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative

2.3.1. Doğruluk (Accuracy, ACC):

Model performansının ölçülmesinde en yaygın kullanılan yöntem doğruluk oranıdır. Bu değer Eşitlik 3'deki gibi hesaplanır.

$$ACC = \frac{TP+TN}{TP+TN+FN+FP} \quad (3)$$

Hata oranı, doğruluk değerini 1'e tamlayan değerdir. Hesabı için gerekli denklem Eşitlik 4 ile verilmiştir.

$$ER = \frac{FP+FN}{TP+TN+FP+FN} \quad (4)$$

2.3.2. Kesinlik (Precision, PRE):

Kesinlik kavramı, sınıfı 1 olarak bulunan pozitif örnek sayısının sınıfı 1 olan örnek sayısına oranı olarak tanımlanabilir. Eşitlik 5'deki gibi hesaplanır.

$$PRE = \frac{TP}{TP+FP} \quad (5)$$

2.3.3. Duyarlılık (Recall, REC):

Doğru sınıflandırılmış pozitif gözlem sayısının toplam pozitif gözlem sayısına oranıdır. Eşitlik 6 ile verilen değer duyarlılık oranını hesaplar.

$$REC = \frac{TP}{TP+FN} \quad (6)$$

2.3.4. F-Ölçütü (F-Measure):

Kesinlik ve duyarlılık ölçütleri tek başına anlamlı bir karşılaştırma sonucu çıkarmakta her zaman yeterli olmayabilir. Bu nedenle F-ölçütü (F) tanımlanmıştır. F-ölçütü, kesinlik (PRE) ve duyarlılığın (REC) harmonik ortalamasıdır (Eşitlik 7).

$$F = \frac{2*PRE*REC}{PRE+REC} \quad (7)$$

2.3.5. Kappa İstatistiği (Kappa Statistic, K):

Kappa katsayısı (K), -1 ile +1 arasında değişen ve uyumu ölçen bir değerdir. Eşitlik 8 ile hesaplanır (Bilgin, 2017).

$$K = \frac{P_0 - P_c}{1 - P_c} \quad (8)$$

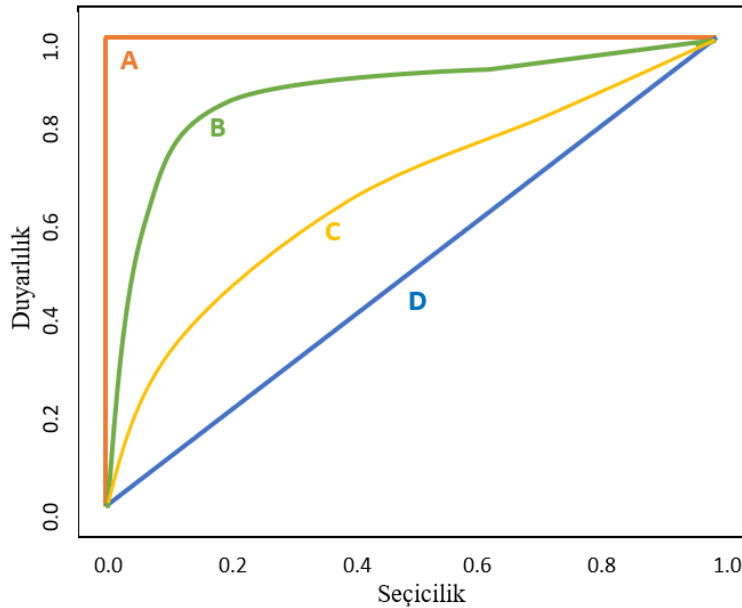
2.3.6. Matthews Korelasyon Katsayısı (MCC):

Sınıflandırma problemlerinde, modelin ya da fonksiyonun performans derecesini değerlendirmek için kullanılan bir ölçüttür (Cohen, 1960). Eşitlik 9, MCC kriterinin nasıl hesaplanacağını göstermektedir.

$$MCC = \frac{TP * TN - FP - FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (9)$$

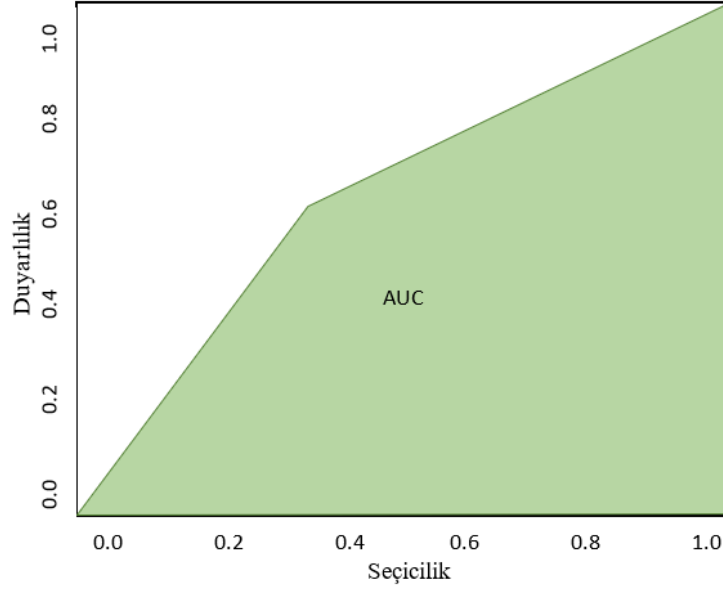
2.3.7. ROC Eğrisi ve ROC Eğrisi Altında Kalan Alan (AUC):

ROC eğrisi; x-ekseninde yanlış pozitif oran değeri (seçicilik) ve y-ekseninde doğru pozitif oran değeri (duyarlılık) alınarak çizilen eğridir. Şekil 2, farklı eşik değerlerini içeren bir ROC eğrisini göstermektedir. Grafiğe göre; A eğrisinde seçicilik değeri 0 olduğundan testin seçiciliği %100'dür. Duyarlılık değeri 1'dir. Yani testin duyarlılığı %100'dür. D eğrisi ise, en başarısız testtir. Çünkü burada seçicilik değeri 1'dir. B ve C eğrileri ise, orta derecede sonuçlar veren testlere ait eğrilerdir.



Şekil 2. ROC eğrisi

Testlerin performansını değerlendirmede kullanılan ölçütlerden bir diğeri, ROC eğrisi altında kalan alan AUC'dur. AUC değeri, ROC eğrisi altında kalan alanın bir ifadesidir (Şekil 3). AUC değerinin alabileceği en küçük değer 0.50 ve en büyük değer 1'dir.



Şekil 3. ROC eğrisi altında kalan alan (AUC)

3. BULGULAR VE TARTIŞMA

Bu çalışmada, karar ağaçları yardımıyla gül hastalığına sebep olabilecek değişkenleri belirleyerek hastalığın belirtilerini azaltmaya yönelik uygulanacak yöntemlerin seçilmesinde uzmanlara yardımcı olmak amaçlanmıştır. Bu doğrultuda, gül hastalığı üzerine bir veri seti ele alınmıştır. Veri setindeki bağımlı değişken, hasta ve hasta olmayan şeklinde sınıflandırılmıştır. Veriler, Weka 3.8.5 programı üzerinde farklı karar ağaçları yardımıyla modellenmiştir. Sınıflandırma analizi için J48, LMT, LAD Tree, REP Tree, AD Tree, BF Tree, Hoeffding Tree, Decision Stump, Random Tree, Random Forest, Optimized Forest, Sys For algoritmaları kullanılmıştır. Sonuçları değerlendirmek amacıyla doğruluk başta olmak üzere; duyarlılık, kesinlik, F-ölçütü, Kappa istatistik değeri, MCC sayısı, ROC eğrisi ve ROC eğrisi altında kalan alan değeri (AUC) ölçütleri göz önünde bulundurulmuştur.

Yapılan analizler sonucunda en yüksek sınıflandırma başarısı, REP Tree algoritması (0.915) ile elde edilmiştir. Diğer algoritmalara ilişkin sonuçlar sırasıyla LMT 0.915; Sys For ve BF Tree 0.894; J48, LAD Tree, Random Forest ve Optimized Forest 0.883; AD Tree

0.862; Decision Stump 0.798; Hoeffding Tree 0.787; Random Tree 0.713 şeklinde bulunmuştur. Tablo 3'te karar ağaçları ile yapılan analizlerden elde edilen sonuçlar verilmiştir. Tablo 4'te algoritmalarından elde edilen doğru ve yanlış sınıflandırılmış gözlem sayıları verilmiştir. Tablo 3 ve 4'ten görüleceği üzere, diğer performans değerlendirme kriterleri, doğruluk oranı ile paralel sonuçlar elde edildiğini göstermektedir. Sonuç olarak en iyi sonucu REP Tree algoritması vermiş olmasına rağmen, diğer algoritmaların sonuçları da önemsenmeyecek kadar kötü değildir.

Tablo 3. Karar ağacı algoritmaları için sınıflandırma sonuçları

Algoritma	Doğruluk	Duyarlılık	Kesinlik	F-Ölçütü	MCC	ROC Eğrisi	PRC Bölgesi	Kappa İstatistiği
J48	0.883	0.884	0.883	0.766	0.892	0.863	0.863	0.764
LMT	0.915	0.927	0.915	0.914	0.84	0.958	0.958	0.827
LADTree	0.883	0.884	0.883	0.883	0.766	0.895	0.882	0.764
REPTree	0.926	0.935	0.926	0.925	0.859	0.874	0.863	0.849
ADTree	0.862	0.844	0.862	0.862	0.723	0.875	0.859	0.723
BFTree	0.894	0.896	0.894	0.893	0.788	0.894	0.883	0.785
HoeffdingTree	0.787	0.787	0.787	0.787	0.572	0.862	0.84	0.572
DecisionStump	0.798	0.854	0.798	0.786	0.642	0.712	0.714	0.583
RandomTree	0.713	0.724	0.713	0.712	0.438	0.718	0.658	0.43
RandomForest	0.883	0.884	0.883	0.883	0.766	0.932	0.928	0.764
OptimizedForest	0.893	0.896	0.894	0.893	0.788	0.93	0.928	0.785
SysFor	0.894	0.911	0.894	0.892	0.802	0.897	0.876	0.783

Tablo 4. Sınıflandırma algoritmaları ile elde edilmiş doğru ve yanlış sınıflandırılmış gözlem sayısı

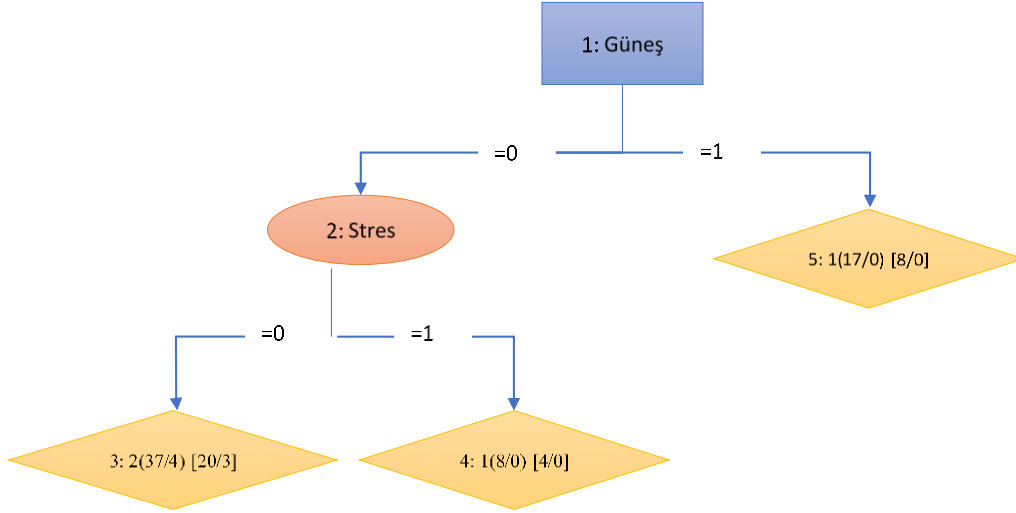
Algoritma	Doğru sınıf. gözlem say.	Yanlış sınıf. gözlem say.
J48	83	11
LMT	86	8
LADTree	83	11
REPTree	87	7
ADTree	81	13
BFTree	84	10
HoeffdingTree	74	20
DecisionStump	75	19
RandomTree	67	27
RandomForest	83	11
OptimizedForest	84	10
SysFor	84	10

Karar ağaçlarından elde edilen modeller, kök düğüm ve iç düğümlerden, sonuç çıktısı olarak da yapraklardan oluşmaktadır. Kök düğüm, modeldeki en yüksek bilgi kazancı değerine sahip değişkendir. İç düğümler ise, modeli kurmada etkili olan diğer değişkenleri içermektedir. Tablo 5, gül hastalığı verileri ile sınıflandırma analizi yapıldıktan sonra elde edilen modellere ilişkin kök ve iç düğüm değişkenlerini göstermektedir. İç düğüm değişkenleri, modeldeki bilgi kazancı değeri sırasına göre verilmiştir. Bu sonuçlar, hastalar için uygun tedavinin ne olabileceği ya da hastaya verilebilecek tavsiyelerin neler olabileceği yönünde fikirler vermektedir.

J48 algoritması ile elde edilen karar ağacında kök düğümü güneş değişkeni ve iç düğümü stres ve baharat değişkenleri oluşturmaktadır. Bu sonuçlar; güneş kreminin önemini, güneşten kaçınmayı ve stresten ve baharatlı yiyeceklerden uzak kalmanın hastalık için önemini göstermektedir.

LAD Tree algoritmasında kök düğümde yer alan güneş değişkeni, güneş kreminin önemini ve güneşten kaçınmayı göstermektedir. İç düğümlerde yer alan stres ve baharat değişkenleri stresten ve baharatlı yiyeceklerden uzak durmayı vurgulamaktadır. Hastanın kilosu, sistolik kan basıncı, metabolik sendrom varlığı ve egzersiz yapma durumları birbirleriyle yakından ilişkilidir. Bu değişkenler birlikte değerlendirildiğinde, hastalarda yağlı ve sağlıksız beslenmenin hastalık için tetikleyici olabileceği düşünülebilir. Ailede hastalık öyküsü değişkeni ise, hastalığın genetik yatkınlıktan kaynaklanabileceğini ortaya koymaktadır.

En iyi doğruluk değerini vermiş olan REP Tree algoritmasının sonucuna göre; kök düğümde yer alan güneş değişkeni güneş kremi kullanımının ve güneşten kaçınmanın önemini, iç düğümde yer alan stres değişkeni ise stresten uzak kalmanın önemini vurgulamaktadır. Şekil 4.'te REP Tree algoritmasına ilişkin karar ağacı verilmiştir.



Şekil 4. REP Tree algoritması karar ağacı

AD Tree algoritmasında karar için üç model oluşturulmuştur. Birinci modelde güneş kök düğümü oluşturmakta ve güneş kremi kullanımının ve güneşten kaçınmanın önemli olduğunu göstermektedir. İç düğümlerde yer alan crp değişkeni vücutta enfeksiyon başladığının bir habercisi olabileceken, sistolik kan basıncı, metabolik sendrom ve hdl değişkenleri kalp ile ilgili bir hastalığın habercisi olabilir. İnsülin değişkeni ise direkt olarak vücuttaki glikoz ile ilgilidir ve sağlıksız bir beslenme düzenine işaret etmektedir. Bu değişkenler başka ciddi bir hastalığı olmayan hastalarının tedavisi için ele alındığında, hastalara sağlıklı beslenme tavsiye edilmektedir. İkinci model için kök düğüm strestir ve stresten kaçınmanın önemini göstermektedir. İç düğüm güneş ve baharattır. Güneş kremi kullanımının ve güneşten ve baharatlı yiyeceklerden kaçınmanın önemini göstermektedir. Üçüncü model sadece açlık kan şekeri kök düğümünden oluşmaktadır. Açlık kan şekeri, insülin ile doğrudan ilişkili olduğundan benzer şekilde değerlendirilebilir.

Random Tree algoritması, kanda bulunan bir lipid (yağ) türü olan trigliserid kök düğümünden oluşmaktadır. Bu değişken kilo, egzersiz, glikoz gibi değişkenlerde olduğu gibi sağlıksız beslenme şeklinde yorumlanabilir. İç düğümlerde yer alan egzersiz, açlık kan şekeri, hdl, bmi, kilo, kolesterol, crp değişkenleri de yine aynı şekilde beslenme ile yorumlanabilir. Stres değişkeni, stresten uzak kalma; güneş değişkeni, güneş kremi kullanımı ve güneşten kaçınma; baharat değişkeni, baharatlı yiyeceklerden kaçınma olarak değerlendirilebilir. Yaş ve boy değişkenleri, boyun yaşa bağlı olarak değişmesinden dolayı birlikte değerlendirilebilir.

Gül hastalığına ilerleyen yaş gruplarında daha sık rastlandığı düşünüldüğünde, modelde yerini alması beklenen bir durum olarak değerlendirilebilir. Benzer şekilde kadınlarda erkeklere göre daha sık görülmesi de model için beklenen bir durumdur. Sigara tüketimi değişkeninin modelde yer alması, hastanın sigaradan da kaçınması gerektiğini göstermektedir.

BF Tree, Decision Stump, Optimized Forest ve Sys For algoritmalarından elde edilen modellerde; yukarıda bahsedilen değişkenler yer almaktadır ve benzer şekilde yorumlanabilirler. Random Forest ve Hoeffding Tree algoritmaları ile yüksek başarımlar elde edilmesine rağmen, modelde yer alan değişkenlerin neler olduğu saptanamamıştır.

Tablo 5. Karar ağacı ile elde edilmiş modellere ait kök düğüm ve iç düğüm değişkenleri

Algoritma	Kök Düğüm	İç Düğümler
J48	güneş	stres, baharat
LAD Tree	güneş	stres, baharat, sistolik kan basıncı, egzersiz, metabolik sendrom, kilo, ailede hastalık öyküsü
REP Tree	güneş	stres
AD Tree	1.güneş, 2.stres, 3.açlık kan şekeri	1.crp, sistolik kan basıncı, hdl, metabolik sendrom, insülin 2.güneş, baharat -
BF Tree	güneş	stres, baharat, sistolik kan basıncı
Hoeffding Tree	-	-
Decision Stump	güneş	-
RandomTree	trigliserid	egzersiz, açlık kan şekeri, hdl, yaş, stres, crp, güneş, cinsiyet, baharat, bmi, kilo, boy, sigara tüketimi, kolesterol
Random Forest	-	-
	1.egzersiz	1.diyostalik kan basıncı, stres, güneş, hdl, kolesterol, boy, insülin
	2.stres	2.crp, güneş, cinsiyet, trigliserid, egzersiz, kolesterol, hdl
Optimized Forest	.	.
	.	.
	.	.
	85.trigliserid	85.boy, kolesterol, insülin direnci, güneş, yaş, stres, açlık kan şekeri, hdl
	1.güneş	1.stres
	2.stres	2.güneş
Sys For	.	.
	.	.
	.	.
	13.açlık kan şekeri	13.stres

4. SONUÇ VE ÖNERİLER

Gül hastalığı, deri üzerinde kırmızı kaşıntılı döküntüler ile seyreden bir rahatsızlıktır. Gül hastalığı, görüntüsünün kötü olması ve günlük yaşamı olumsuz etkilemesinden dolayı hastaları hem fiziksel hem de psikolojik olarak etkilemektedir. Hastalığının kesin bir tedavisi yoktur, ancak bazı yöntemler ile hastalığın belirtileri ve etkileri azaltılabilmektedir. Tedavi için belirli bir yöntem olmadığından dolayı uzmanlar, tedavi seçiminde hastalığın fizyolojik özellikleri, bulunduğu yer, yayılım miktarı gibi belirtileri dikkate almaktadırlar.

Bu çalışmanın temel amacı, pek çok sınıflandırma analizine dayalı çalışmada olduğu gibi kişinin gül hastası olup olmadığını belirlemekten ziyade; hangi değişkenlerin hasta üzerinde etki bıraktığını tahmin etmektir. Bu amaçla, farklı karar ağacı algoritmaları ile çalışılmıştır. Öncelikle verilerin farklı algoritmalar ile nasıl performans sergilediği araştırılmıştır. İnceleme sonucunda, tüm karar ağaçlarının %70'in üzerinde başarılı performans sergilediği görülmüştür. İkinci adımda, karar ağaçları ile elde edilmiş modellerdeki kök ve iç düğümlerde yer alan değişkenler incelenmiş ve sonuç olarak hastalığı etkileyen ya da etkileyebilecek olan sebepler belirlenmiştir. Elde edilen bu sonuçlar, gül hastalığının tedavisinde izlenilecek yol hakkında ipuçları vermektedir.

Farklı algoritmalar ile elde edilen ve Tablo 5 ile sunulan sonuçlar, literatürde gül hastalığı ile ilgili yapılan dermatolojik araştırma sonuçları ile benzerlik göstermektedir. Bu yönüyle çalışmadan elde edilen sonuçlar hem yapılan önceki çalışmalarını desteklemekte hem de yeni yapılacak çalışmalar için bir yol gösterici olmaktadır. Ancak literatür incelendiğinde, gül hastalığının tespiti için makine öğrenimine ilişkin bir çalışmanın mevcut olmadığı görülmektedir. Bu çalışma, uygun karar algoritmaları ile verilerin kolayca modellendiğini ve hastalık teşhisi yapılabildiğini ortaya koymaktadır. Böylece başarıyı yüksek karar yapısına sahip algoritmalar tedavi sürecine yardımcı olmak amacıyla uzmanlara önerilmektedir. İlerleyen çalışmalarda, farklı makine öğrenimi yöntemleri ile gül hastalığı verileri yeniden modellenebilir ve elde edilen sonuçların karşılaştırılması yapılabilir.

KAYNAKÇA

- [1] Adnan, N. ve Islam, Z. (2016), Optimizing the number of trees in a decision forest to discover a subforest with high ensemble accuracy using a genetic algorithm, Knowledge-Based Systems, 110, 86-97.

- [2] Aldrich, N., Gerstenblith, M., Fu, P., Tuttle, M. S., Varma, P., Gotow, E., Cooper, K. D., Mann, M. ve Popkin, D. L. (2015), Genetic vs environmental factors that correlate with rozasea: a cohort-based survey of twins, *Jama Dermatol*, 151, 1213-1219.
- [3] Alpaydın, E. (2000), Zeki veri madenciliği: ham veriden altın bilgiye ulaşma yöntemleri, *Bilişim 2000 Veri Madenciliği Eğitim Semineri*.
- [4] Al-Saggaf, Y., ve Nielsen, S. (2014), Self-disclosure on Facebook among female users and its relationship to feelings of loneliness, *Computers in Human Behavior*, 36, 460-468.
- [5] Anadolu Sağlık (2021), Rozase Hastalığı: Nedir, Nedenleri, Belirtileri ve Tedavisi, *Anadolu Sağlık Merkezi, Deri Hastalıkları Bölümü*.
- [6] Ata, O. ve Erbudak, A. E. (2022), Veri madenciliği ve makine öğrenimi ile döviz kuru tahmini uygulaması, *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 34(2), 553-563.
- Freund, Y. ve Mason, L. (1999), The alternating decision tree learning algorithm, *Paper Presented at the Proceedings of the Sixteenth International Conference on Machine Learning*.
- [7] Aybey, B. (2022), Gül Hastalığı (Rozasea) Nedir? Belirtileri ve Tedavisi, *Florence Nightingale Hastaneleri, İstanbul*.
- [8] Bahety, A. (2014), Extension and evaluation of ID3-Decision Tree Algorithm, *Entropy*, 2(1), 1-8.
- [9] Belli, A. A., Gök, S. Ö., Akbaba, G., Etgu, F. ve Doğan, G. (2016), The relationship between rosacea and insulin resistance and metabolic syndrome, *European Journal of Dermatology*, 26, 260-264.
- [10] Berg, M. ve Liden, S. (1989), An epidemiological study of rosacea, *Acta Dermato Venereologica*, 69, 419-423.
- [11] Bilgin, M. (2017), Gerçek veri setlerinde klasik makine öğrenmesi yöntemlerinin performans analizi, *Breast*, 2(9), 683-688.
- [12] Bingöl, M. (2021), Rozasea'da Hastalık Şiddetiyle Sosyal Fobi, Anksiyete, Benlik Saygısı, Uyku Kalitesi ve Dermatolojik Yaşam Kalite İndeksi Arasındaki İlişki, *Uzmanlık Tezi, Atatürk Üniversitesi Tıp Fakültesi, Erzurum*.
- [13] Breiman, L. (2001), Random Forests, *Machine Learning*, 45(1), 5-32.

- [14] Bulut, F. (2016), Çok katmanlı algılayıcılar ile doğru meslek tercihi, *Anadolu University Journal of Science and Technology-An Applied Sciences and Engineering*, 17(1), 97-109.
- [15] Cem Sönmez, Sinyalize Kavşaklarda Trafik Akımının Modellenmesi, Yayınlanmamış Yüksek Lisans Tezi, İTÜ Fen Bilimleri Enstitüsü, İstanbul, 2005.
- [16] Cheng, J., Fayyad, U. M., Irani, K. B. ve Qian, Z. (1988), Improved decision trees: a generalized version of ID3, *Proceedings of the Fifth International Conference on Machine Learning*, 100-106.
- [17] Cho, G., Park, H. M., Jung, W. M., Cha, W. S., Lee, D. ve Chae, Y. (2020) Identification of candidate medicinal herbs for skincare via data mining of the classic Donguibogam text on Korean medicine, *Integrative Medicine Research*, 9(4), 1-9.
- [18] Chosidow, O., Cribier, B. (2011), Epidemiology of rosacea: updated data, *Annales de Dermatologie et de Venereologie*, 138, 179-183.
- [19] Cohen, J. (1960), A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20(1), 37-46.
- [20] Coşkun, C. ve Baykal, A. (2011), Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması, XIII. Akademik Bilişim Konferansı Bildirileri, 51-58.
- [21] Çetinkaya, O. (2011), Belirsizliğin ölçülmesi ve entropi, *İstanbul Üniversitesi İktisat Fakültesi Mecmuası*, 44, 1-4.
- [22] Domingos, P. ve Hulten, P. G. (2000) Mining high-speed data streams, *Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining*, 71-80.
- [23] Friedl, M. A. ve Brodley, C. E. (1997), Decision tree classification of land cover from remotely sensed data, *Remote Sensing Of Environment*, 61(3), 399-409.
- [24] Haber, R. ve El-Gemayel, M. (2018) Comorbidities in rosacea: A systematic review and update, *Journal of American Academy of Dermatology*, 78, 786-792.
- [25] Heo, J. Y., Cho, M. K. ve Kim, S. (2022), Data mining for detecting signals of adverse drug reaction of doxycycline using the Korea adverse event reporting system database, *Journal of Dermatological Treatment*, 33(4), 2192-2197.
- [26] Holmes, A.D., Spoenclin, J., Chien, A.L., Baldwin, H. ve Chang, A. L. S. (2018) Evidence-based update on rosacea comorbidities and their common physiologic pathways, *Journal of American Academy of Dermatology*, 78(1), 156-166.

- [27] Holmes, G., Pfahringer, B., Kirkby, R., Frank, E. ve Hall, M. (2002), Multiclass Alternating Decision Trees, Paper Presented at the Proceedings of the 13th European Conference on Machine Learning.
- [28] Islam, Z. ve Giggins, H. (2011), Knowledge discovery through SysFor: A systematically developed forest of multiple decision trees, In Proceedings of the Ninth Australasian Data Mining Conference, 121, 195-204.
- [29] Kaynar, O., Görmez, Y., Yıldız, M. ve Albayrak, A. (2016), Makine öğrenmesi yöntemleri ile duygu analizi, International Artificial Intelligence and Data Processing Symposium, 234-241.
- [30] Kim, J. ve Kim, K. (2023), Elucidating the potential pharmaceutical mechanism of Gyejibokryeong-hwan on rosacea using network analysis, *Medicine (Baltimore)*, 102(9), 1-11.
- [31] Landwehr, N., Hall, M. ve Frank, E. (2003), *Logistic model trees*, Springer, Berlin.
- [32] Landwehr, N., Hall, M. ve Frank, E. (2005). *Logistic model trees*. *Machine Learning*, 59, 161-205.
- [33] Litt, J. Z. (1997), Rosacea: how to recognize and treat an age-related skin disease, *Geriatrics*, 52, 39-47.
- [34] Quinlan, J. R. (1993), *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers Inc, Massachusetts.
- [35] Rajalingam, K., Levin, N., Marques, O., Grichnik, J., Lin, A. ve Chen, W. S. (2023), Treatment options and emotional well-being in patients with rosacea: An unsupervised machine learning analysis of over 200,000 posts, *Journal of The American Academy of Dermatology*, 13, 172-178,
- [36] Sarmanova, A. ve Albayrak, S. (2013), Alleviating class imbalance problem in data mining, In *Signal Processing and Communications Applications Conference*, 1-4.
- [37] Shi, H. (2007). *Best-First Decision Tree Learning*, The University of Waikato, Hamilton, NewZealand.
- [38] Spöndlin, J., Voegel, J. J., Jick, S. S., Meier, C. R. (2012), A study on the epidemiology of rozasea in the U.K., *British Journal of Dermatology*, 167, 598-605.
- [39] Sumner, M., Frank, E. ve Hall, M. (2005), *Speeding Up Logistic Model Tree Induction*, Springer, Berlin.

- [40] Şatır, E., Azboy, F., Aydın, A., Arslan, H. ve Hacıfendioğlu, Ş. (2016). Veri indirgeme ve sınıflandırma teknikleri ile glokom hastalığı teşhisi, *El-Cezeri Journal of Science and Engineering*, 3(3), 485-497.
- [41] Ture, M., Tokatli, F. ve Kurt, I. (2009), Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4. 5 and ID3) in determining recurrence-free survival of breast cancer patients, *Expert Systems with Applications*, 36(2), 2017-2026.
- [42] Webster, G. F. (2003), *Rosacea and Related Disorders*, Mosby, USA.
- [43] Wisaeng, K. (2013), A Comparison of different classification techniques for bank direct marketing, *International Journal of Soft Computing and Engineering*, 3(4), 116-119.
- [44] Witten, I. H. ve Frankk, E. (2005), *Data Mining: Practical Machine Learning Tools and Techniques*, San Elsevier, Francisco.
- [45] Witten, I. H., Frank, E. ve Hall, M. A. (2011), *Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques*, Elsevier Science.
- [46] Yıldırım, P., Uludağ, M. ve Görür A. (2008), Hastane bilgi sistemlerinde veri madenciliği, *Akademik Bilişim, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale*, 429-434.
- [47] Yorulmaz, A., Kulcu, S. C. (2015), *Helicobacter pylori and inflammatory skin diseases*, *World Journal of Dermatology*, 4, 120-128.
- [48] Zhao, Y., ve Zhang, Y. (2008), Comparison of decision tree methods for finding active objects, *Advances in Space Research*, 41(12), 1955-1959.
- [49] Zuber, T. J. (2000), *Rosacea*, *Dermatology*, 27, 309-318.