©2019 Published in 7th International Symposium on Innovative Technologies in Engineering and Science 22-24 November 2019 (ISITES2019 SanliUrfa - Turkey)
https://doi.org/10.33793/acperpro.02.03.47

Meme Kanserinin Teşhis Edilmesinde Karar Ağacı Ve KNN Algoritmalarının Karşılaştırmalı Başarım Analizi

[1]Muhammed Ali PALA, [1]Murat Erhan ÇİMEN, [1]Ömer Faruk BOYRAZ
[1]Mustafa Zahid YILDIZ, [1]Ali Fuat BOZ acperpro.02.03.29
[1]Sakarya Uygulamalı Bilimler Üniversitesi, Teknoloji Fakültesi, Elektrik Elektronik Mühendisliği Bölümü, Sakarya, TÜRKİYE

Özet

Makine öğrenmesi yöntemlerinin sınıflandırma ve tanımlama açısından başarısının artması, bilgisayar teknolojisine karar verme yeteneği kazandırmış ve analiz yapma metotlarını geliştirmiştir. Makine öğrenmesi yöntemlerinin bu avantajları, medikal alanda hastalıkların teşhis ve tedavi süreçlerinde uzmanlara yardımcı olabilecek gelişmiş karar destek sistemlerini ortaya çıkarmıştır. Kanser teşhisi süreçlerinde kullanımı oldukça yaygınlaşan makine öğrenmesi yöntemleri, sonuç çıkarımında önemli roller üstlenmektedir. Meme kanseri, kadınlar arasında en yüksek ölüm oranına sahip kanser türü olmakla birlikte dünyada görülen en yüksek ikinci kanser türüdür. Hasta sayısı dolayısı ile elde edilen verilerin büyüklüğü göz önünde tutulduğunda, bu verilerin hızlı bir şekilde analizinin yapılması, hastalığın erken teşhisi için önemli bir adımdır. Bu çalışmada, 699 hastadan toplanılmış görüntülerin sayısallaştırılması ile elde edilen 10 adet öznitelik içeren Wisconsin Üniversitesi meme kanseri veri seti, K en yakın komşu ve karar ağacı algoritmalarının çeşitli parametreleri değiştirilerek sınıflandırılması yapılmıştır. Eğitim ve test verileri çapraz doğrulama ile karşılaştırılmış ve en yüksek sınıflandırma başarısı %97,30 ile K en yakın komşu algoritması ile elde edilmiştir.

Key words: K-En Yakın Komşu, Karar Ağaçları, Sınıflandırma, Meme Kanseri, Makine Öğrenmesi

Comparative Performance Analysis of Decision Tree and KNN Algorithms in Diagnosing Breast Cancer

Abstract

Increasing the success of machine learning methods in terms of classification and identification has gained the ability to decide on computer technology and developed methods of analysis. These advantages of machine learning methods have developed advanced decision support systems that can help specialists in the diagnosis and treatment of diseases in the medical field. Machine learning methods, which are widely used in cancer diagnosis processes, play important roles in conclusion. Breast cancer is cancer with the highest mortality rate among women and it is also the second-highest cancer type in the world. Given the size of the data obtained due to the number of patients, the rapid analysis of these data is an important step for the early diagnosis of the disease. In this study, the University of Wisconsin breast cancer data set containing 10 features obtained by digitizing images collected from 699 patients were classified by modifying various parameters of K nearest neighbor and decision tree algorithms. Training and test data were compared with cross validation and the highest classification success was obtained %97,30 with the K nearest neighbor algorithm.

Key words: K-Nearest Neighbor, Decision Trees, Classification, Breast Cancer, Machine Learning

*Corresponding author: Muhammed Ali PALA Address: Faculty of Technology, Department of Electric Electronic Engineering Sakarya University Of Applied Sciences University, 54187, Sakarya TURKEY. E-mail address: pala@sakarya.edu.tr, Phone: +902646160309

1. Giriş

Dünya sağlık örgütü verilerine göre kanser vakaları 2018 yılında 9,6 milyon kişinin ölümüne yol açmıştır. Meme kanseri ise her yıl 2.089 milyon kadını etkilemekte ve 627 bin kadının hayatını kaybetmesine yol açmaktadır [1]. Orta yaşlı kadınlarda istatistiksel olarak daha sık görülen meme kanseri, meme dokusundaki süt bezleri ve süt kanalları arasında bulunan hücrelerin kontrolsüz olarak çoğalmasıdır. Kötü huylu (malignant) kanser türleri, iyi huylu (bening) kanser türlerinin aksine diğer dokulara yayılma özelliği göstermektedir. Meme kanseri vakalarında bu ayrımında tümörün morfolojik özellikleri önemli bir rol oynamaktadır. Kitlenin sınırlarının belli bir yapıda olması ve pürüzleşmenin düşük olması, kanserli dokunun iyi huylu olduğuna, sınırların belli ölçütlere sığmaması ve pürüzleşmenin yüksek olması ise kütü huylu kanser riski taşıdığının göstergesidir. Bu ayrım hastaların kanser türünün erken sürede teşhis edilmesini ve hastalığın tedavi süreçlerini iyileştirerek ve yaşam süresinin uzatılması açısından en çok tercih edilen metotlardan bir tanesidir [2]. Bu ayrımın hızlı ve efektif bir şekilde yapılabilmesinin makine öğrenmesi yöntemlerini medikal alanda da önemli roller üstlenmesine yol açmıştır [3].

Meme kanserinin türünün teşhisi, hastalığın seyri ve uygulanacak tedavinin efektif olarak uygulanması açısından oldukça önemlidir. Bu amaç doğrultusunda çeşitli yöntemler kullanılmaktadır. Mamografi, iğne uçlu aspirasyon yöntemi ve invaziv biyopsi bu yöntemlerden başlıca olanlarıdır. Mamografi ile tespiti yapılan kanser vakalarında son yıllarda yüksek doğruluğa görüntü işleme algoritmaları ve görüntüyü elde etme metotlarının gelişmesiyle ulaşılmıştır. Hastaya uygulanan iyonlaştırıcı radyasyon etkisi düşük olması ile birlikte yüksek doğruluğa ulaşılması amaçlandığı için ilk teşhis aşamalarında daha çok tercih edilmektedir [4]. Tümörü ve türünü en doğru şekilde tespit etme yöntemlerinin bir diğeri ise invaziv biyopsi yöntemidir. Bu yöntemin uygulanması diğer yöntemlere karşın daha zor olmakla beraber diğer yöntemlere göre daha başarılı sonuçlar üretmektedir. Ayrıca hasta üzerine bıraktığı fizyolojik ve psikolojik etki ise bu yöntemi son teşhis amacı ile kullanılmasına neden olmaktadır. İğne uçlu aspirasyon yöntemi ise son yıllarda gösterilen gelişmeler neticesinde en çok tercih edilen yöntem haline gelmiştir. Yüksek doğruluk oranı ve uygulamadaki hız iğne uçlu aspirasyon yönteminin en büyük avantajlarıdır.

Tahmin temeline dayalı sistemlerde, mevcut değişkenler göz önünde bulundurularak gelecekteki bilinmeyen değerler regresyon yöntemleriyle veya sınıflandırma yöntemleriyle elde edilir. Makine öğrenmesi, gelecek ile ilgili tahmin yapılmasını sağlayacak kural ve ilişkileri ortaya çıkaran ve veriyi çeşitli eşitliklerle tanımlayan yöntemlerdir [5]. Bu amaç doğrultusunda k-en yakın komşu, karar ağaçları veya çeşitli hibrit yöntemler literatürde sıklıkla kullanılmaktadır [6]. Tanımlayıcı yöntemlerde ise veri içerisindeki oluşumları ortaya çıkarır ve verinin daha kolay yorumlanmasına olanak sağlar. Kümeleme analizleri, birliktelik kuralları ve ardışık örüntü kuralları sıklıkla kullanılan tanımlayıcı yöntemlerdir [7].

Bu çalışma içerisinde k-en yakın komşu algoritmasının uzaklık ölçütlerinin ve komşuluk sayısının değişimi ile karar ağacı algoritmalarının bölünme sayısı parametresinin sınıflandırma başarısı üzerine etkileri kıyaslanmış ve meme kanseri veri seti üzerinde gösterilmiştir. Sınıflandırma başarısı en yüksek olan durumlarında performans değerlendirilmeleri yapılmıştır.

2. Materyal ve Metot

Bu çalışma içerisinde Dr. William H. tarafından Wisconsin Üniversitesi hastanesinde toplanılan meme kanseri verileri kullanılmıştır. Verilerin k-en yakın komşuluk ve karar ağacı algoritmalarının çeşitli parametreleri değiştirilerek sınıflandırması yapılmış ve algoritmaların performans analizleri kıyaslanmıştır.

2.1. Veri Setine Genel Bakış

Çalışmada kullanılan veri seti meme kanseri görüntülerinin sayısallaştırılması ile elde edilmişmiş 10 adet öznitelikten meydana gelmektedir. 699 örnek içeren veri seti teşhis sonucu olarak iyi huylu sınıf 2, kötü huylu sınıf 4 olarak işaretlenmiştir. Veri setinde 241 (%34,5) kötü huylu, 458 (%65,5) iyi huylu örnek bulunmaktadır [8]. Tablo 1'de veri seti içerisinde bulunan 10 adet özniteliğin tanımları, değer aralıkları, ortalamaları ve standart sapma değerleri verilmiştir.

Tablo 1. Öznitelik açıklamaları ve değerleri

Numara	Öznitelik Tanımlaması	Değer	Ortalama	Standart Sapma
1	Kapanma Kalınlığı	1-10	4.442	2.820
2	Boyut Eş biçimlilik	1-10	3.150	3.065
3	Şekil Eş biçimlilik	1-10	3.215	2.988
4	Yapışma	1-10	2.840	2.864
5	Epitelyal Boyut	1-10	3.234	2.223
6	Çıplak Çekirdek	1-10	3.544	3.643
7	Yumuşak Kromatin	1-10	3.445	3.449
8	Normal Nükleoli	1-10	2.869	3.050
9	Mitoz	1-10	1.603	1.732
10	Sınıf	2-4		

2.2. K En Yakın Komşu

K en yakın komşu algoritması, sınıflandırma problemlerinin çözümünde kullanılan örnek tabanlı algoritmalar sınıfındadır. Öğrenme işlemini veri içerisinde tutulan eğitim seti ile gerçekleştirir. Eğitim işlemi en yakın varsayılan k adet veriyi, belirli uzaklık ölçütü çerçevesinde benzerliklerinin hesaplanması ile yapmaktadır [9]. Bu uzaklık ölçütleri Minkowski, Öklid, Chebyshev ve kosinüs eşitlikleri kullanarak belirlenebilmektedir. Literatürde ise sıklıkla Öklid uzaklığı tercih edilmektedir. P ve Q iki noktalar kümesi olmak üzere, $P = x_{1,}x_{2,} \dots, x_n$ ve $Q = y_{1,}y_2, \dots, y_n$ arasındaki mesafe eşitlik 1'de gösterildiği gibi hesaplanır.

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

Yeni bir veri sınıflandırma amacıyla algoritmaya geldiğinde, eğitilmiş veri seti içerisinde bulunan k adet en yakın merkez komşunun sınıf etiketlerine bakılır. Daha sona sınıf etiketlerinin çoğunluğa göre yeni gelen veri o kümeye dahil edilir [10].

K-en yakın komşu algoritmasının gerçekleştirilmesinin kolay olması, lineer olmayan eğitim süreçlerini içermemesi ve gürültülü eğitim verilerine karşı başarılı sonuçlar vermesi en büyük

avantajlarıdır. Fakat k-en yakın komşu algoritması çeşitli dezavantajlarda sahiptir [11]. Veri setinin büyüklüğüne bağlı olarak yüksek düzeyde bellek kullanması ve işlem yükünün veri attıkça artması k-en yakın komşu algoritmasının fazla öznitelik içeren verilerde çalışmasını zorlaştırmaktadır. Ayrıca performans ölçümünün dışardan girilen komşu sayısına duyarlılığı ve belirlenen uzaklık ölçütüne karşı hassasiyeti en temel eksikliklerindendir [12].

2.3. Karar Ağaçları

Karar ağaçları ilk aşamada veriyi sınıflandırmak amacıyla ağaç yapısınu üstten başlayarak oluşturan algoritmalardır. Ağaç yapısı en üsten başlamak koşulu ile kök, dal ve yapraklar olarak isimlendirilir. Her dal üsteki köke bağlı olmak koşuluyla dallar düğümlere bağlanır. Veride bulunan her bir öznitelik sınıflandırma sonrasında ağaçta bir düğüm noktasını temsil etmektedir. Tüm bu ağaç yapısı arasında kalan düğüm noktaları birer sınıflandırma kuralıdır ve her bir yaprak bir sınıf kabul edilir [13].

Literatürde birçok karar ağacı algoritması bulunmaktadır. Kullanılan ağaç oluşturma esasları göz önüne alındığın veri çeşitliliğine bağlı olarak farklı sınıflandırma başarıları göstermektedirler. ID3, J48, C4.5 ve C5 algoritmaları en bilinen karar ağacı algoritmalarıdır. C4.5 algoritması sınıflandırma işlemi esnasında en ayırt edici özelliğin seçiminde entropi denklemini kullanmaktadır. Entropi denklemi ile veri içerisindeki belirsiz durumlar ve verideki rasgelelik oranı ölçülebilmektedir. $p_{1,}p_{2,\dots,}p_3$ olasılık durumlarını temsil etmek üzere, tüm bu durumların toplamı 1 olmalıdır. Bu durum göz önünde bulundurulduğunda entropi eşitlik 2'de verildiği haliyle hesaplanır [14].

$$I(P) = -\sum_{i=1}^{k} p_i * log\ (p_i) \tag{2}$$

Veri tabanında bulunan tüm özniteliklerin entropileri hesaplanmasının ardından, her bir özniteliğin $\text{Bilgi}(\text{D}, \text{S})$ değeri eşitlik 3'te verildiği gibi hesaplanır.

$$Bilgi(D,S) = \sum_{i}^{n} \frac{T_i}{T} I(T_i) \tag{3}$$

Bilgi değerleri hesaplanan elemanların kazanç değerleri eşitlik 4'de verildiği üzere hesaplanır ve kazanç değeri maksimum olan eleman en üstteki kök düğümüne yerleştirilir.

$$Kazanç(D,S) = Bilgi(S) - Bilgi(D,S) \tag{4}$$

Diğer karar ağacı algoritmaları da çeşitli veriler üzerinde yüksek sınıflandırma başarısı göstermektedir. J48 algoritması ağaç yapısını oluştururken sınıflandırma başarısını arttırmak için zayıf dallardan kurtulma temeline dayalıdır [15]. Ayrıca karar ağaçları ile uluşturulacak hibrit yöntemlerin sınıflandırma başarısını arttırdığı çalışmalar literatürde mevcuttur [16]. Karar ağacı algoritmaları yorumlaması kolay, birden çok çıktısı olan problemlerin çözümünde etkili bir yöntemdir. Fakat algoritmalarda karmaşık dallanmalar üretebileceği gibi, ezbere ağaç oluşturma gibi sorunlar meydana gelebilmektedir [17].

2.4. Performans Değerlendirme

Ver madenciliği yöntemleri ile yapılan sınıflandırma işlemlerinde, sınıflandırmanın başarısı çeşitli yöntemler ile test edilmesi gerekmektedir. Sistemi eğitmek için kullanılan verilerin bir kısmı test verisi olarak kullanılabileceği gibi, veri seti ile aynı öznitelikleri barındıran verilerde test için kullanılabilmektedir. Bu amaç için verinin belirli bir yüzdelik dilimi test verisi olarak ayrılabileceği gibi çapraz doğrulama metotları da kullanılabilmektedir [18].

Sınıflandırma yöntemi sonucu olarak üretilen tahmin sınıfları ile veride bulunan gerçek sınıflar ayrı ayrı kümelerde ifade edilir. Bu kümeler ile sınıflandırma sonucunun çeşitli parametreleri hakkında bilgi sahibi olunabilmektedir. Tablo 2'de bu kümelerin isimlendirme kriteri verilmiştir. Tahmin edilen pozitif değer ile gerçek pozitif değerin kesişimi doğru pozitif (TP), tahmin edilen pozitif değer ile gerçek negatif değerin kesişimi yanlış pozitif (FP), tahmin edilen negatif değer ile gerçek pozitif değerin kesişim kümesi yanlış negatif (FN) ve tahmin edilen negatif değer ile gerçek negatif değerin kesişim kümesi doğru negatif (TN) olarak isimlendirilir. Bu isimlendirmelerin hata matrisine yerleştirilmesi Tablo 2 'de gösterildiği gibidir.

Tablo 2. Hata matrisi yapısı

		Gerçek Değer	
		Pozitif	Negatif
Tahmin Edilen Değer	Pozitif	TP	FP
	Negatif	FN	TN

Veri madenciliği yöntemleri ile elde edilen sınıflandırma başarısını ölçmek için kullanılan bazı performans ölçütleri Tablo 3'te verildiği gibidir. Doğruluk değeri, sınıflandırma işlemi sonucu elde edilen değerlerin gerçek değerleri ne kadar ifade ettiğini belirtmektedir. Duyarlılık denklemi, doğru pozitif değerlerin tespit edilmesinde sınıflandırma işleminin yeteneğini ortaya koyar. Benzer şekilde belirleyicilik değeri ise sınıflandırma sonucu sınıflama işleminin negatif değere karşı olan yeteneğini ortaya koymaktadır. Kesinlik denklemiyle sınıflama işleminin yanlış pozitif değerleri elimine edilmesi kabiliyetini ortaya koymaktadır. F skor değeri ise kesinlik ve hassasiyet değerlerinin hesaplanması sonucu ortaya çıkabilecek olağan dışı durumları elimine etmek için kullanılan bir hesaplama yöntemidir.

Tablo 3. Veri madenciliğinde kullanılan bazı performans ölçütleri

Ölçüt	Denklem
Doğruluk	$\frac{TP + TN}{TP + TN + FP + FN}$
Duyarlılık	$\frac{TP}{TP + FN}$
Belirleyicilik	$\frac{TN}{FP + TN}$
Kesinlik	$\frac{TP}{TP + FP}$
F Skor	$2.\frac{Kesinlik.Duyarlılık}{Kesinlik + Duyarlılık}$

3. Deneysel Sonuçlar

Deneyler açık kaynak kodlu Python 3.6 kullanılarak gerçekleştirilmiştir. K en yakın komşu ve karar ağacı algoritmalarının çeşitli parametrelerinin değiştirilmesiyle kullanılan veri setinin sınıflandırma başarısının hangi ölçüde etkilediği incelenmiştir. Test verisi ise çapraz doğrulama kullanılmıştır. K-en yakın komşu algoritmasında ağırlık değeri atamasında kullanılan mesafe ölçütünün ve komşu sayısının, karar ağacı algoritması için ise maksimum ayrılma parametrelerinin sınıflandırma başarısına etkisi kıyaslanmıştır. K-en yakın komşu algoritmasında uzaklık ölçütü olarak Öklid, Minkowski, Chebyshev ve kosinüs uzaklığının sınıflandırma başarına etkisi ve ayrıca her bir uzaklık ölçüt için 1 ile 10 arasında değişen küme sayıların sınıflandırma başarısına etkisi gösterilmiştir. Karar ağaçlarında ayırma kriteri olarak Gini endeksi kullanılmış olup maksimum ayrılma sayısı 1 ile 10 arasındaki değiştirilerek sınıflandırma başarısı ölçülmüştür. K-en yakın komşu ve karar ağacı algoritması ile elde edilen en yüksek sınıflandırma başarılarının hata matrisleri elde edilmiş ve bu matrisler sonucunda en başarılı sonuçların performans değerlendirilmeleri yapılmıştır.

Tablo 4'te k-en yakın komşu algoritması ile yapılan deneylerle elde edilen sonuçlar verilmiştir. Veri setinde bulunan tüm özniteliklerin kullanıldığı sınıflandırmada kosinüs uzaklığı ile yapılan sonuçların ortalaması diğer uzaklık ölçütleri ile elde edilen sonuçların ortalamasından daha yüksektir. En başarılı sınıflandırma %97,30 ile 4 komşuluğun kosinüs uzaklığı ile yapılan deneylerde elde edilmiştir.

Tablo 4. Uzaklık ve komşu sayısı parametrelerine göre sınıflandırma başarısı

Uzaklık Ölçütü	Sınıflandırma Doğruluk Performansı (%)									
	1-NN	2-NN	3-NN	4-NN	5-NN	6-NN	7-NN	8-NN	9-NN	10-NN
Öklid	95,40	95,40	96,70	96,40	96,40	96,70	96,90	96,60	96,60	96,70
Minkowski	94,60	94,60	96,70	96,60	96,60	96,40	96,60	96,40	96,40	96,30
Chebyshev	95,00	94,70	96,30	96,60	96,10	96,30	96,40	96,40	96,90	96,70
Kosinüs	96,00	94,70	97,30	97,30	96,70	97,10	97,00	97,10	97,10	97,10

Tablo 5'te k-en yakın komşu algoritmasında %97,30 ile en başarılı olan deneye ait hata matrisi verilmiştir.

Tablo 5. KNN ile elde edilen en yüksek sınıflandırmanın hata matrisi

		Gerçek Değerler	
		2	4
Tahmin Edilen	2	446	7
Değerler	4	12	234

Tablo 6'da Gini endeksi kullanılarak oluşturulan karar ağacında bölünme sayısının sınıflandırma başarısına etkisi verilmiştir.

Tablo 6. Bölünme sayısının sınıflandırma başarısına etkisi

Bölünme Sayısı	Sınıflandırma Başarısı (%)
1	92,00
2	92,10
3	92,10
4	94,10
5	94,10
6	94,30
7	94,70
8	94,40
9	94,60
10	94,40

Tablo 7'de %94,70 başarı oranı ile sınıflandırılmış 7 bölümlemeli sonucun hata matrisi verilmiştir.

Tablo 7. Karar Ağaçları ile elde edilen en yüksek sınıflandırmanın hata matrisi

		Gerçek Değerler	
		2	4
Tahmin Edilen	2	433	12
Değerler	4	25	229

K en yakın komşu ve karar ağacı algoritmaları ile elde edilen sınıflandırma performansının karşılaştırılması Tablo 8'de verilmiştir.

Tablo 8. Elde edilen en başarılı sonuçların performans değerleri

Ölçüt	KNN	Karar Ağacı
Doğruluk	0,973	0,947
Duyarlılık	0,974	0,945
Belirleyicilik	0,970	0,971
Kesinlik	0,984	0,984
F Skor	0,979	0,964

4. Tartışma Ve Sonuç

Meme kanserinin teşhisinde kullanılan yöntemlerden elde edilen verilerde bulunan anlamlı ilişkileri çıkarabilmek hastalığın teşhis ve tedavi süreçlerini olumlu yönde etkilemektedir. Bu çalışmada meme kanseri veri setinin k-en yakın komşu ve karar ağacı algoritmaları yardımıyla bu anlamlı birliktelikleri yüksek başarı oranında yapılacağı gösterilmiştir. Her iki algoritma ile elde edilen %97,30 ve %94,70 oranında doğruluk oranı, çalışmasının hastalığın teşhis süreçlerinin kısalmasına yardımcı olacağını göstermektedir. Literatürde meme kanseri veri seti kullanılarak çeşitli çalışmalarda mevcuttur [19]. Kullanılan yöntemler ile çalışmamızda kullanılan yöntemler kıyaslandığında, yöntemimizin eğitilme ve sonuç çıkarma hızı diğer yöntemlere göre oldukça iyi performans sergilemiştir. Bu yönüyle yüksek işlem gücü gerektirmemesinden dolayı gömülü

bilgisayarlar ile oldukça hızlı çalışabilecek potansiyele sahiptir.

Karar ağaçları ve k-en yakın komşu algoritmalarının başarı oranlarının kıyaslandığı bu çalışmada; meme kanseri teşhisinde k-en yakın komşu algoritması ile yapılan performans testlerinin karar ağacı algoritmasına göre daha iyi sonuçlar verdiği fakat karar ağaçlarının daha hızlı çalıştığı gözlemlenmiştir. Bu yönleriyle makine öğrenmesi tekniklerinin medikal alanda bulunan problemlere uygulaması kolay çözümler getirileceği öngörülmektedir.

Uzmanların karar verme süreçlerine destek olabilecek karar-destek sistemlerinin tasarlanmasında kullanılacak makine öğrenmesi teknikleriyle hastalıkların teşhisinde yüksek başarıya ulaşmak mümkündür. Makine öğrenmesi yöntemlerinde hedef %100 doğrulukla sınıflandırma yapmaktır. İleriki çalışmalarda hedefimiz etki değeri yüksek özniteliklerin belirlenmesi ve öğreticili/öğreticisiz derin öğrenme algoritmalarını aynı veri setine veya farklı problemleri içeren veri setlerine uygulayarak sınıflama doğruluğunu %100'e çıkarmaktır.

5. Referanslar

[1] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 68(6), 394-424.

[2] Chang, R. F., Wu, W. J., Moon, W. K., & Chen, D. R. (2005). Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. Breast cancer research and treatment, 89(2), 179.

[3] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 13, 8-17.

[4] Kuhl, C. K., Schrading, S., Leutner, C. C., Morakkabati-Spitz, N., Wardelmann, E., Fimmers, R., ... & Schild, H. H. (2005). Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. Journal of clinical oncology, 23(33), 8469-8476.

[5] Wawre, S. V., & Deshmukh, S. N. (2016). Sentiment classification using machine learning techniques. International Journal of Science and Research (IJSR), 5(4), 819-821.

[6] Geppert, H., Vogt, M., & Bajorath, J. (2010). Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. Journal of chemical information and modeling, 50(2), 205-216.

[7] Al-Maolegi, M., & Arkok, B. (2014). An improved Apriori algorithm for association rules. arXiv preprint arXiv:1403.3948.

[8] Wolberg, W. H., & Mangasarian, O. L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. Proceedings of the national academy of sciences, 87(23), 9193-9196. Veri Seti URL: https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/ ,Erişim Tarihi: 22.10.2019

[9] Dudani, S. A. (1976). The distance-weighted k-nearest-neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics, (4), 325-327.

[10] Muja, M., & Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. VISAPP (1), 2(331-340), 2.

[11] Bhatia, N. (2010). Survey of nearest neighbor techniques. arXiv preprint arXiv:1007.0085.

[12] Liu, H., & Zhang, S. (2012). Noisy data elimination using mutual k-nearest neighbor for classification mining. Journal of Systems and Software, 85(5), 1067-1074.

[13] Kotsiantis, S. B. (2013). Decision trees: a recent overview. Artificial Intelligence Review, 39(4), 261-283.

[14] Danacı, M., Çelik, M., & Akkaya, A. E. (2010). Veri madenciliği yöntemleri kullanılarak meme kanseri hücrelerinin tahmin ve teşhisi. Akıllı Sistemlerde Yenilikler ve Uygulama Sempozyumu, 21-24.

[15] Bhargava, N., Sharma, G., Bhargava, R., & Mathuria, M. (2013). Decision tree analysis on j48 algorithm for data mining. Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, 3(6).

[16] Polat, K., & Güneş, S. (2007). Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. Applied Mathematics and Computation, 187(2), 1017-1026.

[17] Rokach, L., & Maimon, O. Z. (2008). Data mining with decision trees: theory and applications (Vol. 69). World scientific.

[18] Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics surveys, 4, 40-79.

[19] Abdel-Zaher, A. M., & Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. Expert Systems with Applications, 46, 139-144.