

Veri Seti - Sınıflandırma İlişkisinde Performansa Etki Eden Faktörlerin Değerlendirilmesi

Abdullah ALAN^{1*}, Murat KARABATAK²

¹ Yazılım Mühendisliği Anabilim Dalı, Fen Bilimleri Enstitüsü, Fırat Üniversitesi, Elazığ, Türkiye

² Yazılım Mühendisliği Bölümü, Teknoloji Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye

*¹ alanabdullah2001@hotmail.com, ² mkarabatak@firat.edu.tr

(Geliş/Received: 15/05/2020;

Kabul/Accepted: 11/08/2020)

Öz: Veri madenciliği yöntemlerinden biri olan sınıflandırma, en çok kullanılan veri madenciliği yöntemidir. Bir veri veya veri grubunun mevcut sınıflardan hangisine ait olduğunun belirlenmesi olarak tanımlanmaktadır. Sınıflandırma çalışmalarında en önemli kriter yüksek başarımlı bir sınıflandırıcı model oluşturabilmektir. Ancak başarımlı etkileyen birçok neden bulunmaktadır. Kullanılan test yöntemlerinin yanı sıra veri setine ait özellikler de başarımlı etkileyen unsurlardan biridir. Bu makalede, hem test tekniklerinin hem de veri setinin özelliklerinin başarımlı etkisi detaylı olarak incelenmektedir. Bu amaçla, literatürde yaygın olarak kullanılan 32 farklı veri seti kullanılarak deneyler gerçekleştirilmiştir. Ayrıca çalışmada 9 farklı sınıflandırma algoritması da kullanılarak bu algoritmaların veri setleri üzerindeki başarımları da kıyaslanmıştır. Elde edilen sonuçlar, algoritmaların etkinliğinin yanı sıra test yöntemlerinin başarımlı etkilerini de ortaya çıkarmıştır. Veri setine ait parametrelerin sınıflandırma başarımlına etkileri makalede detaylı olarak ortaya konulmuştur.

Anahtar kelimeler: Sınıflandırma, veri seti, başarımlı, test teknikleri.

Evaluation of the Factors Affecting Performance on the Data Set - Classification Relationship

Abstract: Classification, which is one of the data mining methods, is the most used data mining method. It is defined as determining the class of the data or data group. The most important criterion in classification studies is to obtain classifier model has high accuracy performance. However, there are many reasons that affect the classifier accuracy. In addition to the test methods, the properties of the data set are also one of the factors affecting the performance. In this paper, the effects of both test techniques and data set on accuracy are examined in detail. For this purpose, experiments were carried out using 32 different data sets, which are widely used in the literature. In addition, the performance of these algorithms on datasets was compared by using 9 different classification algorithms. The results obtained revealed the effectiveness of algorithms as well as the effects of test methods on success. The effects of the parameters of the data set on the classification performance are presented in detail in the paper.

Key words: Classification, Dataset, accuracy, test techniques.

1. Giriş

Son yıllarda farklı kaynaklar aracılığıyla oldukça büyük miktarlarda veri toplanmaktadır. Bu veriler günümüzde günlük yaklaşık olarak 2,5 exabyte boyutuna ulaşmaktadır. Elde edilen veriler 2,5 milyon adet 1 terabyte kapasiteli diske depolanabilecek kapasiteye karşılık gelmektedir. Toplanmakta olan bu veriler farklı çalışma alanlarında kullanılmaktadır. Bunlar; pazarlama stratejileri oluşturmak, halkla ilişkiler çalışmaları, bankacılık sektörü, güvenlik ve bilim insanlarının yaptıkları çalışmalarda kullanılmaktadır [1].

Günümüzde insanlar istem dışı olarak makine öğrenmesi çalışmalarına destek vermektedir. Günlük çekilen fotoğraf ve videolar, sosyal medya üzerinden yapılan yorumlar, video izleme platformlarından izlenen videolar, hastanelerde yapılan testler, çevrim içi alışveriş sitelerinden yapılan işlemler gibi çalışmalar daha sonra makine öğrenmesi modellerinde kullanılmak üzere veri tabanları oluşturmaktadır. Sürekli üretilmekte olan bu veriler çoğu zaman istatistiksel veri olarak kalmakta farklı bir amaç için kullanılmaktadır[2]. Veri bilimi o kadar önemli bir hale gelmiş bulunmaktadır ki NASA uzaya göndermek için yetiştirdiği astronotların yanı sıra veri bilimi üzerinde çalışmalar yapmak için R veya Python programlama dillerine hakim veri bilimcilerde yetiştirmektedir[3].

Makine öğrenmesi, matematiksel ve istatistiksel yöntemler kullanılarak öğrenme gerçekleştirebilen ve mevcut verilerle çıkarımlarda bulunan, bu çıkarımlar sayesinde bilinmeyen hakkında tahminlerde bulunan yöntemlerdir. Makine öğrenmesi yöntemleri ne kadar çok veri ile eğitilirse oluşturulan modelin tahmin etme olasılığı artmaktadır. Makinelerin yaptıkları çıkarımlara en güzel örneklerden biri olarak Google'ın geliştirdiği

* Sorumlu yazar: alanabdullah2001@hotmail.com. Yazarların ORCID Numarası: ¹ 0000-0002-1585-1698, ² 0000-0002-6719-7421

DeepMind'ın Go şampiyonunu yenmesi gösterilebilir [4]. Aslında DeepMind'ın yaptığı işlem insanların geçmişte yaptıkları hamleleri deneyimleyerek çıkarımlarda bulunmaktadır.

Makine öğrenmesi modeli oluşturulurken öğrenmenin ne oranda gerçekleştiği yani modelin ne kadar doğru tahminlerde bulunduğu büyük öneme sahiptir. Örneğin; kredi kartı sahtekârlığını tespit etmek için oluşturulan bir modelde, modelin yapılan sahte işlemleri doğru tahmin etmesi beklenmektedir. Makine öğrenmesi modelininin oluşturulduğu yöntem, verinin fazlalığı ve verinin dağılımı gibi etkenler ile doğru sonuç üretmeye çalışmaktadır. Burada karşılaşılan en büyük problemlerden biri karar sınıfına ait veri dağılımının dengeli olmamasıdır. Model oluşturulurken karar sınıfına ait azınlık veri gurubu eğitim veri seti içerisinde düşük miktarda ya da hiç yer almayacaktır. Bu da çoğu zaman modelin performansını ölçerken yanılgıya düşülmesine sebep olmaktadır.

Makine öğrenmesi ile ilgili literatür taraması yapıldığında konu ile ilgili yapılan çalışmaların sayısında son zamanlarda bariz bir artış tespit edilmiştir. Aydın [5], çalışmasında yapay zekâ ve öğrenmesi teknikleri kullanarak bir tıp bilişimi uygulaması geliştirmeyi amaçlamış ve tek bir makine öğrenmesi modeli kullanarak tek bir metrik ile sonucu değerlendirmiştir. Hacıfendioğlu [6], glokom hastalığının göz sinirleri zarar görmeden önce teşhis edilebilmesi ve dünya çapında körlük nedenleri arasında ilk sırada bulunan bu hastalığın tahmin edilmesini amaçlamıştır. Çalışmasında üç farklı makine öğrenmesi modeli kullanmış ve destek vektör makineleri ile oluşturulan modelin en iyi performansı elde ettiği sonucuna ulaşmıştır. Kartal [7], çalışmasında kalp ameliyatı sırasında ya da kalp ameliyatı geçirdikten kısa bir süre sonra hastaya ait hayati riskin makine öğrenmesi yöntemleriyle belirlenmesini amaçlamıştır. Çalışmasında dört farklı makine öğrenmesi modeli kullanmış ve farklı metrikler kullanarak karar ağacı modelinin bu veri seti için en iyi model olduğunu tespit etmiştir. Şeker[8], çalışmasında iyi-kötü koku verilerine yönelik Elektro Ensefolo Grafi (EEG) işaretlerinin analizi ve sınıflandırmasını amaçlamıştır. Çalışmasında dört farklı makine öğrenmesi modeli kullanmış ve rastgele orman modelinin en iyi sonucunu sadece doğruluk metriğini kullanarak tespit etmiştir. Turgut[9], mikro dizi verileri kullanarak makine öğrenmesi yöntemleri ile sınıflandırma yapmıştır. Çalışmasında sekiz farklı makine öğrenmesi modeli kullanmış ve çalışmasının sonucunu değerlendirirken sadece doğruluk metriğinden faydalanmıştır. Pekel[10], çalışmasında sınıflandırma problemi üzerinde durmuş, dört makine öğrenmesi modeli kullanmış ve Naive bayes modelinin en iyi sonucu verdiğini doğruluk metriği kullanarak tespit etmiştir. Yukarıda bahsedilen makine öğrenmesi yöntemleri ile oluşturulan çalışmalara bakıldığında sadece tek veri seti ile inceleme yapıldığı ve başarımları kriteri olarak da genellikle doğruluk metriği kullanıldığı gözlemlenmiştir. Makine öğrenmesi yöntemlerinden sınıflandırma ile oluşturulan modellerin performanslarını ölçerken farklı metriklerin kullanımı ve farklı veri dağılımı olduğu durumlarda nasıl değiştiği hakkında yeterince çalışmaya rastlanmamıştır. Yapılan sınıflandırma çalışmalarında ise genelde sonuçlar sayısal olarak verilmekte, sonuçlar hakkında değerlendirme yeterince yapılmadığından dolayı sonuçların anlamı yetersiz kalmaktadır.

Bu araştırmada, diğer çalışmalardan farklı olarak farklı dağılımlara sahip 32 farklı veri seti incelenerek dokuz farklı sınıflandırma algoritması uygulanmıştır. Bu sınıflandırma algoritmaları k-en yakın komşu, karar ağaçları, rasgele orman, Naive bayes, lojistik regresyon sınıflandırma, destek vektör makineleri, gradyan artırma, adaboost ve yapay sinir ağlarıdır. Daha sonra ise karışıklık matrisinden faydalanılarak doğruluk, F_1 ölçütü, Matthews korelasyon katsayısı (MCC-Matthes Corelation Coefficients) ve alıcı işlem karakteristikleri altında kalan alan (AUC-Area under the ROC Curve) metrikleri kullanılmıştır. Ayrıca bu metriklerin ne ifade ettiği, temel düzeyde ifade edilmiştir. Böylelikle model seçimi yapılırken en iyi modelin seçimi için bir yol haritası oluşturulmuştur.

2. Materyal ve Yöntem

Bu çalışmada, literatürde yaygın olarak kullanılan makine öğrenmesi yöntemleri kullanılarak sınıflandırma çalışması yapılmıştır. Makine öğrenmesi uygulamalarında oluşturulan modelin doğruluğunun test edilmesi büyük önem arz etmektedir. Modelin performansını sadece doğruluk metriği ile ölçmek çoğu zaman yanılgıya yol açabilmektedir. Bundan dolayı oluşturulan modelin performansının değerlendirilmesinde sadece doğruluk metriği değil buna ek olarak kullanılacak metriklere ihtiyaç duyulmaktadır. Makine öğrenmesi yöntemlerinden olan sınıflandırmada bu değerlendirme için karışıklık matrisi olarak isimlendirilen bir tablodan faydalanılmaktadır. Verilen bilgiler ışığında bu çalışmada kullanılan yöntemlere ait bilgiler verilmiştir.

2.1. Çalışmada Kullanılan Veriler

Bu çalışmada, Kaggle[11] olarak bilinen açık kaynak veri seti sağlayıcısından elde edilen 32 farklı veri seti kullanılmıştır. Kullanılan veri setlerine ait nitelik sayısı, her bir veri setindeki örnek sayısı, veri setinde kaç sınıf olduğu ve sınıfların dağılımına ait denge düzeyi bilgileri Tablo 1'de verilmektedir. Burada yer alan denge düzeyi sütunu, veri setindeki sınıflarının dağılımının dengeli olup olmadığı hakkında bilgi vermektedir.

Tablo 1. Çalışmada kullanılan veri setleri.

No	Veri seti	Özellik Sayısı	Örnek Sayısı	Sınıf Sayısı	Denge Düzeyi(ρ)
1	Avustralya'da Yangın (AY)	24	142.193	2	3,4
2	Wisconsin Meme Kanseri (WMK)	31	569	2	1,7
3	Tic-Tac-Toe Oyunu (TTTO)	10	958	2	1,8
4	Sloan gökyüzü araştırması (SGA)	18	10.000	3	5,8
5	Pulsar yıldızı tahmini (PYT)	9	17.898	2	9,9
6	Ortopedik Hastaların Biyomekanik Özellikleri (OHBÖ)	7	310	2	2,1
7	Ortopedik Hastaların Biyomekanik Özellikleri (OHBÖ)	7	310	3	2,5
8	Kalp Hastalığı (KH)	14	303	2	1,1
9	Yapısal protein dizileri (YPD)	6	111.338	3	1,3
10	Farelerin protein ekspresyonu (FPE)	82	1080	8	2,0
11	Seyahat sigortası (SS)	11	63.326	2	63,9
12	Kas aktivitesini okuyarak jestlerin sınıflandırılması (KAOJS)	64	11.678	4	1,0
13	Parkinson hastalığı sınıflandırma (PHS)	755	756	2	2,9
14	Portekiz bankası pazarlama (PBP)	18	40.841	2	7,8
15	Genetik çeşitlilik sınıflandırma (GÇS)	46	65.188	2	2,9
16	Mobil cihaz fiyat sınıflandırması (MCFS)	21	2.000	4	1,0
17	Türkiye siyasi görüşleri (TSG)	16	885	6	13,1
18	Banka pazarlama (BP)	17	11.162	2	1,1
19	İris çiçeği (İÇ)	6	150	3	1,0
20	Şarap kalitesi (ŞK)	13	6.497	7	545,6
21	Hepatoselüler Karsinom (HCC)	50	165	2	1,6
22	Bireysel kredi sınıflandırma problemi (BKSP)	14	5.000	2	10,9
23	Sahte şirketleri denetlemek için denetim riski (SSDİDR)	25	776	2	1,5
24	İnternette alışverişte kullanıcı tercihleri (İAKT)	18	12.330	2	5,6
25	Şarap için müşteri segmentasyonu (ŞİMS)	14	178	3	1,4
26	Ph tanıma (PT)	4	653	15	1,1
27	Gelir sınıflandırması (GS)	15	32.561	2	3,1
28	Kriyoterapi analiz (KA)	7	90	2	1,1
29	Kredi kartı sahtekarlığı tespiti (KKST)	31	284.807	2	554,5
30	Deniz kulağı (DK)	9	4.177	3	1,0
31	Sesle cinsiyet tanıma (SCT)	21	3.168	2	1,0
32	Pima Kızılderilileri diyabet (PKD)	9	768	2	1,8

Tablo 1'de yer alan denge düzeyi bilgisi, karar sınıfındaki çoğunluk sınıfa ait veri sayısının, azınlık sınıfa ait veri sayısına bölünmesi ile elde edilmekte olup bu hesaplama Denklem 1'de gösterilmektedir. Veri setinin denge durumunun tespiti için kullanılmakta olan bu denklemden elde edilen değer, 1'in üzerinde olduğu durumlarda veri setinin dengesiz olduğunu ifade etmektedir. Bu denklemde yer alan C_i ifadesi, veri setinde yer alan karar sınıfını temsil etmektedir. $\text{Max}_i\{|C_i|\}$ ve $\text{Min}_i\{|C_i|\}$ ifadeleri karar sınıfının yer alan örneklerden en fazla ve en az etiketi içeren sınıfları temsil etmektedir [12].

$$\text{Denge düzeyi } \rho = \frac{\text{Max}_i\{|C_i|\}}{\text{Min}_i\{|C_i|\}} \quad (1)$$

2.2. Sınıflandırma

Denetimli öğrenme yöntemi olan sınıflandırma, makine öğrenmesinde en sık kullanılan yöntemlerden birisidir. Verinin sınıflandırılması işlemi, daha önce sınıf etiketleri belli olan veri setini eğitim ve test olmak üzere iki parçaya bölme işlemi ile başlamaktadır. Burada ki eğitim verisi ile oluşturulacak model tasarlanmakta ve test verisi ile de oluşturulan model test edilmektedir. Oluşturulan model yardımı ile yeni bir örnek ile karşılaşıldığında örneğin hangi sınıfa ait etiketle etiketleneceği belirlenmektedir. Sınıflandırmanın amacı, benzer özelliklerdeki verilerin önceden etiketlenmiş veri gruplarından hangisine ait olduğunu tahmin edilmesidir [13]. Literatürde sınıflandırma işlemi için çok fazla algoritma bulunmaktadır. Bunlardan bazıları; k-en yakın komşu,

karar ağaçları, naive bayes, destek vektör makineleri, gradyan artırma, yapay sinir ağları, adaBoost, lojistik regresyon sınıflandırıcı ve rasgele orman'dır.

2.2.1. k-en yakın komşu

K-en yakın komşu algoritması mesafeye dayalı algoritmaların birisidir. Temel çalışma mantığı, verilerin birbirleriyle olan uzaklıkları ve benzerliklerini kullanarak sınıflandırma işlemi gerçekleştirmektedir. Bu teknikte veri bir örüntü uzayında saklanmaktadır. Karar sınıfı bilinmeyen bir veri geldiğinde verinin hangi sınıfa ait olduğunun belirlenmesi için sınıfı bilinmeyen veriye en yakın k adet veri belirlenir. Daha sonra ise, veri kendisine yakın olan bu k adet veriden hangisine daha çok benziyorsa onun sınıfında etiketlenir[14].

2.2.2. Karar ağaçları

Sınıflandırma problemlerinde, oluşturulması, yorumlanması ve veri seti ile bütünleştirilmesinin kolaylığından dolayı en sık tercih edilen yöntemlerden birisidir. Karar ağaçları, veri yapılarındaki ağaç yapısı ile benzerlik göstermektedir. Karar ağaçları, düğüm ve dallardan oluşan anlaşılması kolay, her bir dalın bir olasılık durumunu temsil ettiği ve prensip olarak veriyi rekürsif olarak alt gruplara dallanma yaparak bölen bir algoritmadır. Bu ayırım aşamasında oluşan dalların her biri bir kuralı temsil etmektedir[15-16].

2.2.3. Rasgele orman

Rasgele sınıflandırma algoritması, ağaç tipindeki sınıflandırma algoritmalarının topluluğu olarak tanımlanabilmektedir. Rasgele orman, tüm öznelikler ile en iyi dalı kullanarak her bir düğümü dallara ayırmak yerine, her bir düğümden rastgele seçilen öznelikler arasından en iyisini kullanarak her bir düğümü dallara ayırır. Model oluşturulurken kullanılan her bir veri seti orijinal veri setinden yer değiştirilerek elde edilir. Sonra rasgele öznelik seçimi yapılarak ağaçlar geliştirilir. Geliştirilen ağaçlar budanmaz ve yapılan işlem rastgele orman algoritmasının doğruluğunu eşsiz yapmaktadır[17-19].

2.2.4. Naive bayes

Bayes teoreminden faydalanılarak oluşturulan bir algoritmadır. Literatürde en sık kullanılan algoritmaların biri olmasının sebebi kolay anlaşılabilir olması ve uygulanabilirliğidir. Bu yöntem kullanılarak bir verinin hedeflenen niteliğin sınıf değerine ait olma olasılığı bulunabilmektedir. Diğer bir ifade ile elde var olan, sınıflandırılmış veriler kullanılarak yeni gelen verinin mevcut bulunan sınıf etiketlerinden birisi olma ihtimalini hesaplayan bir yöntemdir[15].

2.2.5. Lojistik regresyon sınıflandırma

Regresyon analizi, iki veya daha fazla değişken arasındaki ilişkiyi ölçmek amacı ile kullanılmakta ve tanımlayıcı çıkarımsal istatistik sağlamaktadır. Hedeflenen, bağımlı ve bağımsız değişken arasındaki ilişkiyi en az sayıda değişken ile en iyi uyuma sahip olacak şekilde oluşturabilecek en iyi modeli kurmaktır[20].

2.2.6. Destek vektör makineleri

Destek vektör makineleri, son yıllarda regresyon ve sınıflandırma ile alakalı problemlerin çözümünde kullanılmaktadır. Destek vektör makineleri, sınıflandırma işlemi yaparken yüksek düzeyde başarımlı sağlamak için yüksek boyut özelliklerine sahip çekirdek fonksiyonları kullanmaktadır. Diğer sınıflandırma algoritmalarına göre daha başarılı sınıflandırma yaptığı gözlemlenmiştir[21-22].

2.2.7. AdaBoost

AdaBoost algoritması, zayıf sınıflandırıcıların bir araya gelmesiyle ortaya çıkan güçlü bir sınıflandırıcıyı temsil eden topluluk sınıflandırıcısı olarak isimlendirilmektedir. Modelin genel çalışma mantığı, her aşamada bir önceki aşamanın sonucunda yapılan yanlış tahminlerin ağırlığını artırarak sınıflandırıcının tekrar çalıştırılması ile başlamaktadır. Yapılan bu işlemlerle yanlış yapılan tahminlere odaklanıp, oluşturulan modelin sınıflandırmadaki doğruluk oranını yükseltmek amaçlanmaktadır[23].

2.2.8. Gradyan Artırma

Gradyan artırma algoritması, 2001 yılında Friedman tarafından ortaya atılmıştır. Yöntemin çalışma şekli; iteratif olarak hatayı tahmin edip, hatanın iyileştirilmesi hedeflenmektedir. Gradyan artırma algoritması, AdaBoost algoritmasında olduğu gibi örnekleri tekrar ağırlandırmaya çalışmamaktadır. Bir önceki aşamada bulunan kayıp fonksiyonunu negatif gradyan vektörüne uydurmaya çalışır. Hatayı azaltmak için, dereceli azaltma sonuçlarını fonksiyona ekler[24].

2.2.9. Yapay sinir ağları

İnsan beynini model olarak alarak geliştirilen bir teknolojidir. Yapay sinir ağları öğrenme yöntemi ile yeni bilgileri üretme, bu bilgileri keşfetme, düşünebilen sistemler tasarlama amacıyla geliştirilmiştir. Yapay sinir ağları kendi iç kurallarını üretir ve bu kurallarla elde edilen sonuçları karşılaştırma yaparak kendisini geliştirir. Yapay sinir ağları deneme yanılma yolu ile kendisini eğitir ve bir işin nasıl yapılması gerektiğini öğrenir[25].

2.3. Model performans değerlendirme metrikleri

Sınıflandırma algoritmaları ile oluşturulan modellerin değerlendirilmesi, hangi sınıflandırma modelinin daha doğru sonuçlar ürettiğinin belirlenmesinde bazı değerlendirme metrikleri kullanılmaktadır. Bunlar genel olarak karışıklık matrisi olarak isimlendirilen bir tabloya dayanmaktadır. Makine öğrenmesi ve istatistiksel sınıflandırma problemlerinde, bir sınıflandırıcının performansını görselleştirmek için geliştirilen bir tablo düzenidir[26-27]. Matrisin her satırı, gerçek değerleri ifade ederken, her sütun ise tahmin edilen değerleri temsil etmektedir. İki sınıflı bir karışıklık matrisi ve alacağı değerler Tablo 2’de verilmiştir.

Tablo 2. İki sınıflı karışıklık matrisi.

		Tahmin Edilen Sınıf		
		Sınıf	Pozitif	Negatif
Gerçek Sınıf	Pozitif		DP	YN
	Negatif		YP	DN

Karışıklık matrisinde geçen terimler ve anlamları aşağıda listelenmektedir;

- Doğru pozitif (DP); sınıflandırıcı tarafından pozitif sınıfa ait verilerden kaç tanesinin doğru şekilde sınıflandırıldığını temsil etmektedir.
- Doğru negatif (DN); sınıflandırıcı tarafından negatif sınıfa ait verilerden kaç tanesinin doğru şekilde sınıflandırıldığını temsil etmektedir.
- Yanlış Negatif (YN); gerçekte pozitif sınıfa ait olan bir verinin sınıflandırma sonucunda negatif sınıf olarak etiketlenmesidir.
- Yanlış pozitif (YP); gerçekte negatif sınıfa ait olan bir verinin sınıflandırma sonucunda pozitif sınıf olarak etiketlenmesidir.

Karışıklık matrisi yardımı ile sınıflandırma performansını belirlemek için kullanılan metrikler ise açıklamaları ile beraber aşağıda verilmektedir;

- Doğruluk; eğitim kümesi kullanılarak oluşturulan modelin test kümesindeki verileri doğru sınıflandırma oranıdır [27]. Denklem 2’de gösterildiği şekilde hesaplanmaktadır.

$$\text{Doğruluk} = \frac{DP + DN}{DP + DN + YP + YN} \quad (2)$$

- Duyarlılık ya da doğru pozitif oran (DPO); sınıflandırıcının gerçekte pozitif sınıfa ait olan verileri doğru olarak tahmin etme oranını vermektedir[27]. Denklem 3’de gösterildiği şekilde hesaplanmaktadır.

$$\text{Duyarluluk} = \text{Anma} = \text{DPO} = \frac{DP}{DP + YN} \quad (3)$$

- Özgüllük ya da doğru negatif oran (DNO); sınıflandırıcının gerçekte negatif sınıfa ait olan verileri doğru olarak tahmin etme oranını vermektedir[27]. Denklem 4’de gösterildiği şekilde hesaplanmaktadır.

$$\text{Özgüllük} = \text{DNO} = \frac{DN}{DN + YP} \quad (4)$$

- Kesinlik ya da Pozitif tahmin değeri (PTD); sınıflandırma sonucunda pozitif olarak tahmin edilenlerin ne oranda doğru olarak tahmin edildiğini vermektedir[27]. Denklem 5’de gösterildiği şekilde hesaplanmaktadır.

$$\text{Kesinlik} = \text{Pozitif Tahmin Değeri} = \frac{DP}{DP + YP} \quad (5)$$

- Negatif tahmin değeri (NTD); sınıflandırma sonucunda negatif olarak tahmin edilenlerin gerçekte ne kadarının negatif sınıfta olduğunu göstermektedir[27]. Denklem 6’da gösterildiği şekilde hesaplanmaktadır.

$$\text{NTD} = \frac{DN}{DN + YN} \quad (6)$$

- F ölçütü; kesinlik ve anma ölçütleri tek başlarına anlamlı bir karşılaştırma sonucu çıkarılmasında yeterli olmamaktadır. İki ölçütü bir arada değerlendiren F ölçütü sayesinde daha anlamlı sonuçlar üretebilmektedir. F ölçütü kesinlik ve anma değerlerinin harmonik ortalaması alınarak elde edilmektedir. F-ölçütü genel olarak F_1 , $F_{0.5}$ ve F_2 olarak kullanılmaktadır. F-ölçütü, 0 ile 1 arasında değerler almaktadır. Doğru tahminlerde bulunan bir sınıflandırıcının f-ölçütü değerinin 1’e yakın olması beklenmektedir[27]. Denklem 7’de gösterildiği şekilde hesaplanmaktadır.

$$f_\beta - \text{ölçütü} = \frac{(1 + \beta^2)x(\text{kesinlik} * \text{anma})}{\beta^2 + \text{kesinlik} * \text{anma}} \quad (7)$$

- Matthews korelasyon katsayısı(MCC); iki ve çok sınıflı sınıflandırma çalışmalarında performansı ölçmek için MCC olarak adlandırılan bir yöntem önerilmektedir. Bu ölçüm diğer metriklerin aksine karışıklık matrisinde yer alan bütün değerleri kullanmaktadır. MCC, -1 ile +1 arasında değerler almaktadır. Sınıflandırıcı +1’e yakın değerler almışsa doğru tahminlerde bulunduğunu, 0 değerini almışsa rastgele tahminlerde bulunduğunu ve -1’e yakın değerler almışsa yanlış tahminlerde bulunduğunu göstermektedir[28]. Denklem 8’de gösterildiği şekilde hesaplanmaktadır.

$$\text{MCC} = \frac{DP * DN - YP * YN}{\sqrt{(DP + YP) * (DP + YN) * (DN + YP) * (DN + YN)}} \quad (8)$$

Alıcı işlem karakteristikleri eğrileri altında kalan alan(AUC); sınıflandırıcıların performansını test etmek için genel olarak kullanılmaktadır. AUC, 0 ile 1 arasında değerler üretmektedir. AUC değerinin bire yaklaşması sınıflandırıcının doğru tahminlerde bulunduğunu, 0.5 değerini alması sınıflandırıcının rastgele tahminlerde bulunduğunu ve bu değer altında aldığı değerlerde ise sınıflandırıcının doğru çalışmadığını göstermektedir [29].

3. Deneysel Sonuçlar

Bu çalışma da; ilk olarak Bölüm 2.1’de yer alan veri setlerinin her biri için Bölüm 2’de yer alan sınıflandırma algoritmaları uygulanmıştır. Her bir veri seti için iki farklı değerlendirme yöntemi kullanılarak sınıflandırıcı modeller oluşturulmuştur. Kullanılan yöntemler Hold-out ve 10 katlı çapraz doğrulama (10-CV) yöntemidir. Hold-

out yönteminde, veri setini modeli oluşturmak için eğitim ve test parçalarına bölünme işlemi yapılmaktadır. Eğitim verisi ile model oluşturulurken, test verisi ile de bu modelin doğruluğunun testi yapılmaktadır. Bu çalışmada Hold-out yönteminde eğitim ve test verisi bölme işlemi %70 eğitim ve %30 test verisi olacak şekilde yapılmıştır. Çapraz doğrulama yönteminde ise veri seti belirtilen sayıda parçaya bölünmektedir. Bu parçalardan 1 tanesi test diğerleri ise eğitim amaçlı kullanılır. Modelin performansını belirlemek için bu işlem her parça için ayrı ayrı belirlenen sayıda yapılarak her aşamada bir başarımlı değeri elde edilmektedir. Bu elde edilen değerlerin ortalaması alınarak oluşturulacak modelin gerçek performans değeri elde edilmiş olur. Tablo 3'te en yüksek doğruluk değerlerinin hangi sınıflandırma algoritmaları ile elde edildiği ve hangi test yöntemi ile edildiği verilmektedir.

Tablo 3. Çalışmada kullanılan veri setleri için sınıflandırma sonuçları.

No	Veri seti	Sınıf Sayısı	Sınıflandırıcı	Yöntem	Doğruluk	F ₁	MCC	AUC
1	AY	2	KA, GA	10-CV	83,54	0,711	0,412	0,646
2	WMK	2	KNN	Hold-out	98,25	0,980	0,960	0,973
3	TTTO	2	GA	10-CV	100	1,000	1,000	1,000
4	SGA	3	GA	10-CV	98,89	0,976	0,980	1,000
5	PYT	2	RO, AB	Hold-out	98,01	0,938	0,876	0,922
6	OHBÖ	2	RO	Hold-out	86,02	0,832	0,664	0,835
7	OHBÖ	3	RO	Hold-out	84,95	0,792	0,753	0,98
8	KH	2	LRS	10-CV	84,16	0,840	0,680	0,837
9	YPD	3	RO	Hold-out	92,69	0,927	0,889	0,940
10	FPE	8	LRS, YSA	Hold-out	100	1,000	1,000	1,000
11	SS	2	DVM, YSA	10-CV	98,54	0,494	0,000	0,500
12	KAOJS	4	KA, NB, DVM, GA, YSA	10-CV	100	1,000	1,000	1,000
13	PHS	2	RO	Hold-out	86,22	0,780	0,559	0,753
14	PBP	2	DVM	10-CV	88,75	0,610	0,168	0,533
15	GÇS	2	LRS, YSA, DVM	Hold-out	75,32	0,430	0,000	0,500
16	MCFS	4	YSA	10-CV	95,85	0,958	0,942	0,984
17	TSG	6	DVM	Hold-out	47,58	0,300	0,267	0,480
18	BP	2	GA	Hold-out	85,85	0,859	0,717	0,859
19	İÇ	3	KNN, GA, AB, YSA	Hold-out	97,78	0,978	0,967	0,980
20	ŞK	7	RO	Hold-out	69,03	0,390	0,523	0,500
21	HCC	2	KA, RO, NB, LRS, DVM, GA, AB, YSA	10-CV	100	1,000	1,000	1,000
22	BKSP	2	GA	10-CV	98,76	0,963	0,926	0,948
23	SŞDİDR	2	KA, GA, AB	10-CV	100	1,000	1,000	1,000
24	İAKT	2	RO	Hold-out	90,51	0,796	0,592	0,765
25	ŞİMS	3	LRS, YSA	Hold-out	100	1,000	1,000	1,000
26	PT	15	GA	Hold-out	78,57	0,480	0,770	0,960
27	GS	2	GA	Hold-out	87,63	0,816	0,632	0,800
28	KA	2	RO	10-CV	94,44	0,945	0,890	0,946
29	KKST	2	RO	Hold-out	99,95	0,917	0,833	0,881
30	DK	3	RO	Hold-out	64,59	0,645	0,468	0,840
31	SCT	2	KNN	Hold-out	98	0,980	0,960	0,980
32	PKD	2	YSA	Hold-out	80,95	0,790	0,581	0,778

Yapılan deneysel çalışmalar sonucunda, 32 veri seti arasından toplam 20 veri seti için Hold-out yöntemi ile oluşturulan modellerin daha iyi performans sergilediği görülürken diğer 12 veri seti için ise 10 katlı çapraz doğrulama yöntemi ile oluşturulan modellerin daha iyi performans verdiği görülmüştür. Çapraz doğrulama yöntemi makine öğrenmesi modellerinde genellikle tercih edilen bir yöntem olmaktadır. Bunun en önemli nedeni modeli eğitirken ve test ederken verinin tamamını hem eğitim hem de test amaçlı kullanmasıdır. Hold-out yöntemi ise verinin sadece ilk aşamada yani model eğitme ve test aşamasında bir kez bölünmesi ile gerçekleştiğinden model oluşturulurken test bölümünde kalan bir verinin modelin yanlış öğrenmesine sebep olma durumu olabilmektedir. Hold-out yöntemi verinin çok fazla olduğu durumlarda zaman kaybını gidermek için tercih edilebilmektedir. Tablo 4'de yer alan a ve b karışıklık matrislerinde bu durum incelenmektedir.

Tablo 4. PHS veri seti için oluşturulan karışıklık matrisleri.

		Tahmin Edilen Sınıf				Tahmin Edilen Sınıf			
		Sınıf	Parkinson	Sağlıklı			Sınıf	Parkinson	Sağlıklı
Gerçek Sınıf	Parkinson		27	21	Gerçek Sınıf	Parkinson	89	103	
	Sağlıklı		10	167		Sağlıklı	37	527	
a) Rasgele Orman ve Hold-out					b) Rasgele Orman ve 10-CV				

Tablo 4 (a)'da yer alan karışıklık matrisi, rasgele orman sınıflandırıcısı ve Hold-out yöntemi ile oluşturulan modeli temsil etmekte ve Tablo 4 (b)'deki karışıklık matrisi de yine rasgele orman ve 10 katlı çapraz doğrulama yöntemi ile oluşturulan modeli temsil etmektedir. Oluşturulan karışıklık matrislerinden de görüldüğü gibi (a) modeli toplam veri setinin %30'unu temsil ettiğinden daha az veri ile modelin testi gerçekleştirilmektedir. Çapraz doğrulama yöntemi ile oluşturulan model ise tüm veri setinin kullanımı ile oluşturulmuştur. Oluşturulan (a) ve (b) modellerinin doğruluk değeri sırasıyla %86,22 ve %81,48'dir. Hold-out yöntemi ile oluşturulan modelin başarımları değeri daha yüksek gibi görünse de hem daha az veri ile test gerçekleştirilmesi hem de eğitim ve test verilerinin dağılımının modelin sonucunu etkilemesinden dolayı çapraz doğrulama ile oluşturulan modelin tercih edilmesi daha doğru bir yaklaşım olacaktır.

Oluşturulan modeller incelendiğinde dengesiz veri setlerinde sadece doğruluk metriğinin yeterli olmadığı tespit edilmiştir. Bu metriğe ek olarak MCC, F ölçütü ve AUC değerlerinin de verilmesi modelin seçiminde etkili olmaktadır. Tablo 5 (a), destek vektör makineleri ile oluşturulan modeli ve Tablo 5 (b), Naive Bayes ile oluşturulan model sonuçlarını göstermekte ve bu durumu en iyi şekilde ifade etmektedir.

Tablo 5. KKST veri seti için oluşturulan karışıklık matrisleri.

		Tahmin Edilen Sınıf				Tahmin Edilen Sınıf			
		Sınıf	Hileli işlem var	Hileli işlem yok			Sınıf	Hileli işlem var	Hileli işlem yok
Gerçek Sınıf	Hileli işlem var		284.315	0	Gerçek Sınıf	Hileli işlem var	282.397	1.918	
	Hileli işlem yok		492	0		Hileli işlem yok	176	316	
a) Destek Vektör Makineleri ve 10-CV					b) Naive Bayes ve 10-CV				

Oluşturulan iki model incelendiğinde (a) modeli %99,83 oranında doğruluk değeri verirken, (b) modeli %99,26 doğruluk değeri vermektedir. İlk model (a) incelendiğinde 'Hileli işlem yok' sınıfına ait hiçbir verinin doğru tahmin edilmediği oluşturulan karışıklık matrisinde görülmektedir. Bu sınıflandırıcılara ait olan MCC, F ölçütü ve AUC değerleri incelendiğinde, doğruluk değeri yüksek olan sınıflandırıcının aksine daha düşük doğruluk değerinin elde edildiği sınıflandırıcının aslında daha iyi sonuç verdiği gözlemlenmiştir. DVM ile elde edilen MCC, F ölçütü ve AUC değerleri sırasıyla 0, 0,5 ve 0,49 olduğu ve bu değerlerin sınıflandırıcının rasgele tahmin yaptığını göstermektedir. Aynı değerler, Naive Bayes sınıflandırıcı (b) ile hesaplandığında ise sırasıyla 0,29, 0,8 ve 0,67 olduğu görülmektedir. Bu da model seçimi yaparken sadece doğruluk metriğini kullanılmaması gerektiği, ek olarak farklı metriklerden faydalanılması gerektiğini göstermektedir.

Yapılan testlerin bazılarında %90 ve üzeri, bazılarında ise daha düşük başarımları değerleri elde edildiği Tablo 3'de görülmektedir. Bu çalışmalardan Tablo 4'de yer alan Kredi kartı sahtekarlığı tespiti (KKST) veri seti için oluşturulan model ve Ph Tanıma (PT) veri seti ile oluşturulan modeller incelendiğinde birinci veri setinde %99,83 oranında bir doğruluk değeri elde edilmişken diğer veri setinde %78,57 oranında bir doğruluk değerinin elde edildiği görülmektedir. Tablo 6'da PT veri seti ile oluşturulan modele ait karışıklık matrisi görülmektedir.

Tablo 6. PT veri seti için oluşturulan karışıklık matrisi.

Gerçek Sınıf	Sınıf	Tahmin Edilen Sınıf													
		0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	32	5	1	0	0	0	0	0	0	0	0	0	0	0
1	1	5	36	2	1	0	0	0	0	0	0	0	0	0	0
2	2	1	5	34	2	1	1	0	0	0	0	0	0	0	0
3	3	1	3	6	29	4	0	1	0	0	0	0	0	0	0
4	4	0	2	2	5	33	2	0	0	0	0	0	0	0	0
5	5	0	0	2	1	5	28	4	2	0	2	0	0	0	0
6	6	0	0	0	3	1	6	31	3	0	0	0	0	0	0
7	7	0	0	0	0	1	1	6	33	2	1	0	0	0	0
8	8	0	0	0	0	2	0	1	1	36	2	1	1	0	0
9	9	0	0	0	0	0	2	0	2	1	33	3	3	0	0
10	10	0	0	0	0	0	0	0	3	2	0	32	3	3	0
11	11	1	0	0	0	0	0	0	0	2	1	3	30	5	1
12	12	0	0	0	0	0	0	0	0	0	1	3	0	34	3
13	13	0	0	0	0	0	0	0	0	0	0	0	1	3	33
14	14	0	0	0	0	0	0	0	0	0	0	0	0	2	5
															37

Burada Tablo 3'te yer alan diğer metriklere bakıldığında ise KKST veri seti ile oluşturulan modelin aslında rasgele tahminler de bulunduğu ikinci modelin daha doğru başarımlar elde ettiği görülmektedir. Burada ikinci önemli nokta ise birinci veri setinin üç sınıflı, diğerinin ise 15 sınıflı bir veri seti olmasıdır. Karar sınıfına ait dağılımın artması model başarımlarındaki yüzdeyi düşürdüğü fakat diğer metriklerle desteklendiğinde aslında daha doğru bir model oluşturulduğu görülmektedir.

4. Sonuçlar ve Tartışma

Bu makalede, sınıflandırma çalışmalarında karşılaşılan özel durumlar ve model seçimi yapılırken referans alınan metrikler dikkate alınarak sınıflandırma başarımının doğru bir şekilde değerlendirilebilmesi yönünde deneysel çalışmalar yapılmıştır. Doğruluk metriğinin yanı sıra MCC, F ölçütü ve AUC gibi metriklerinde kullanılması gerektiğinin önemine dikkat çekilmiştir. Çalışmada ayrıca sınıf dengesizliği ve sınıf sayısının fazla olduğu durumlarda modelin başarımının nasıl değerlendirileceği hakkında incelemeler yapılmıştır. Genel olarak incelendiğinde üç farklı durum dikkat çekmiştir.

İlk olarak, sınıflandırma modeli oluşturulurken kullanılan test tekniğinin önemi yapılan çalışmada ön plana çıkmaktadır. Sınıflandırıcının performansının doğru olarak değerlendirilebilmesi için mutlaka tüm verilerin test ve eğitim aşamasında kullanılması gerçeği ortaya çıkmaktadır. Bu nedenle mümkün olduğu sürece Çapraz doğrulama testlerinin uygulanması gerektiği tespit edilmiştir. Hold-Out gibi yöntemler kullanıldığı durumlarda daha yüksek doğruluk değerleri elde edilebilse de, eğer çapraz doğrulama testi veri setine uygulanabiliyor ise buradan elde edilen doğruluk değerinin gerçek sonucu yansıttığı ve bu değer kullanılması gerektiği sonucuna varılmıştır. Sadece çapraz doğrulama testinin uygulanmadığı durumlarda Hold-Out yönteminin kullanılması önerilmektedir.

İkinci olarak, veri setlerindeki sınıf sayısının da sonuçları doğrudan etkilediği görülmektedir. Sınıf sayısının fazla olmasına paralel olarak doğruluk değerinin düşüğü gözlemlenebilmektedir. Ancak, diğer metrikler birlikte değerlendirildiğinde bu düşük başarımın daha az sınıflı veri setlerinden elde edilen yüksek başarımlardan daha olumlu sonuçlar verdiği görülmüştür. Örneğin 15 sınıflı bir veri setinden %70 gibi bir doğruluk değerinin elde edilmesi, 2 sınıflı bir veri setinde %80 gibi bir doğruluk değerinden daha anlamlı ve başarılı olduğu sonucunu ortaya koyabilmektedir.

Son olarak, veri setindeki sınıfların dengesiz olarak dağılımı da doğruluk değerinin değerlendirilmesinde sorunlar ortaya çıkarabildiği sonucuna varılmıştır. Bu durumda, sadece doğruluk metriğinin kullanılması yanlış model seçimine sebep olabilmektedir. Bu çalışmada yer alan KKST ve SS veri setleri ile oluşturulan modellerde sadece doğruluk metriğinden faydalanılması rasgele tahminlerde bulunan modellerin daha iyi sonuç verdiği sonucunu yansıtabilmektedir. Bu durumun önüne geçmek için MCC, F ölçütü ve AUC metriklerinden yararlanılması faydalı olacaktır. Böylece doğru sınıflandırıcı seçilmesi daha kolay hale gelebilecektir.

Sonuç olarak, sınıflandırma çalışmalarında doğru sınıflandırıcının seçilmesinin çok önem arz ettiği ve bu hususta dikkat edilmesi gereken birçok parametrenin olduğu ortaya konulmuştur. Özellikle, veri setinin boyutu, sınıfların dağılımları, sınıf sayısı ve kullanılacak olan test yöntemi ile beraber değerlendirme metriklerinin doğru bir şekilde kullanılmasının, iyi bir sınıflandırma yapılmasında çok önemli parametreler olduğu sonucuna varılmıştır.

Teşekkür

Bu makale, Abdullah ALAN'ın "Makine Öğrenmesi Sınıflandırma Yöntemlerinde Performans Metrikleri ile Test Tekniklerinin Farklı Veri Setleri Üzerinde Değerlendirilmesi" başlıklı yüksek lisans tezinden üretilmiştir.

Kaynaklar

- [1] Gürsakal N. Makine Öğrenmesi ve Derin Öğrenme. Bursa, 2017.
- [2] Göbekçin T. (Ed.) Master algoritma. İstanbul, 2017.
- [3] <https://open.nasa.gov/blog/datanaut-fall-2017-class-announcement/> (Erişim Tarihi: 21.01.2020)
- [4] <https://www.theguardian.com/media-network/2015/mar/05/digital-oligarchy-algorithms-personal-data> (Erişim Tarihi: 1.01.2020)
- [5] Aydın F. Kalp ritim bozukluğu olan hastaların tedavi süreçlerini desteklemek amaçlı makine öğrenmesine dayalı bir sistemin geliştirilmesi, Yüksek Lisans Tezi, Trakya Üniversitesi, Fen Bilimleri Üniversitesi, Edirne, 2011.
- [6] Hacıfendioğlu Ş. Makine öğrenmesi yöntemleri ile glokom hastalığının teşhisi, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya, 2012.
- [7] Kartal E. Sınıflandırmaya dayalı makine öğrenmesi teknikleri ve kardiyolojik risk değerlendirmesine ilişkin bir uygulama, Doktora Tezi, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2015.
- [8] Şeker M. İyi-kötü kokular ile ilişkili EMOTIV-EPOC tabanlı EEG kayıtlarının makine öğrenmesi yöntemleri ile sınıflandırılması, Yüksek Lisans Tezi, Dicle Üniversitesi, Fen Bilimleri Enstitüsü, Diyarbakır, 2017.
- [9] Turgut S. Makine öğrenmesi yöntemleri kullanarak kanser teşhisi, Yüksek Lisans Tezi, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2017.
- [10] Pekel E. Farklı makine öğrenmesi algoritmalarının karşılaştırılması, Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Samsun, 2018.
- [11] <https://kaggle.com> (Erişim Tarihi: 05.05.2020)
- [12] Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *Journal of Big Data* 2019; 6(1): 27.
- [13] Fawcett T. An introduction to ROC analysis, *Pattern recognition letter* 2006; 27 (8): 861-874.
- [14] Ayık YZ, Özdemir A, Yavuz U. Lise türü ve lise mezuniyet başarısının, kazanılan fakülte ile ilişkisinin veri madenciliği tekniği ile analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* 2007; 10(2): 441-454.
- [15] Bozkır AS, Sezer E, Gök B. Öğrenci seçme sınavında (öss) öğrenci başarımını etkileyen faktörlerin veri madenciliği yöntemleriyle tespiti. 5. Uluslararası İleri Teknolojiler Sempozyumu, 2009, 13-15 Mayıs, Karabük, s:1-7.
- [16] Albayrak AS, Koltan Yılmaz Ş. Veri madenciliği: Karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi 2009; 14(1): 31-52.
- [17] Breiman L. Random forests, *machine learning*, 2001 Kluwer Academic Publishers 2001; 45(1): 5-32.
- [18] Archer KJ. Empirical characterization of random forest variable importance measure. *Computational Statistics & Data Analysis* 2008; 52(4): 2249-2260.
- [19] http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (Erişim Tarihi: 24 Nisan 2019)
- [20] Balaban ME, Kartal E. Veri madenciliği ve makine öğrenmesi temel algoritmalar ve R dili ile uygulamaları. İstanbul, 2015.
- [21] Boser B, Guyon I, Vapnik V. A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory: Pittsburgh, Pennsylvania, USA 1992*; 144-152.
- [22] Melgani F, Bruzzone L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 2004; 42(8): 1778-1790.
- [23] Kégl B. Introduction to AdaBoost, 2009; 11-14.
- [24] Friedman J. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 2001; 29(5): 1189-1232.
- [25] Özdemir ME, Yıldırım E, Yıldırım S. Classification of emotional valence dimension using artificial neural networks. In *Signal Processing and Communications Applications Conference (SIU) 2015*; 23: 2549-2552.
- [26] Han J, Kamber M, Pei J. *Data mining: Concepts and techniques*. Morgan Kaufmann Publishers 2012.
- [27] Japkowicz N. *Performance evaluation for learning algorithms*, Cambridge University Press, Cambridge 2011.
- [28] Akosa J. Predictive accuracy: a misleading performance measure for highly imbalanced data. In *Proceedings of the SAS Global Forum 2017*; 12.
- [29] Kılıç S. Klinik karar vermede ROC analizi. *Journal of Mood Disorders* 2013; 3 (3): 135-40.