All Sciences Proceedings
http://as-proceeding.com/

2nd International Conference on Recent Academic Studies
October 19-20, 2023 : Konya, Turkey

© 2023 Published by All Sciences Proceedings

ICRAS
https://as-proceeding.com/index.php/icras

Regresyon ve Sınıflandırmada Performans Değerlendirme Ölçütleri

Sıddık Keskin[1*]

[1]Biyoistatistik Anabilim Dalı, Temel Tıp Bilimleri, Bölümü / Van Yüzüncü Yıl Üniversitesi Tıp Fakültesi, Türkiye

skeskin@yyu.edu.tr

Özet – Bilimsel çalışmalarda değişkenler (özellikler) arasındaki ilişkiler, ilişkinin yapısına göre doğrusal ve doğrusal olmayan ilişkiler olmak üzere iki başlık altında incelenebileceği gibi aralarındaki ilişkilerin inceleneceği değişken sayısına göre de basit ilişkiler ve çoklu ilişkiler olarak ifade edilebilir. Değişkenler arasındaki basit veya çoklu ilişkilerin düzeyi (veya miktarı), yüzde (%) olarak belirlenebileceği gibi herhangi bir model (veya eşitlik) yardımıyla mutlak miktar olarak da belirlenebilir. Araştırmalarda bağımlı değişken (output değişkeni, çıktı değişkeni, cevap değişkeni veya yanıt değişkeni) olarak varsayılan değişkenle, bağımsız değişken veya değişkenler (açıklayıcı değişken veya tahminleyici değişken) arasındaki ilişkiyi belirlemek üzere kullanılan eşitlikler, genel olarak regresyon modelleri olarak ifade edilir. Regresyon modellerinde minimum açıklayıcı değişken ile en yüksek doğruluk derecesine sahip uygun bir modelin belirlenmesi amaçlanmaktadır. En uygun modeli belirlemek üzere karar verme sürecinde farklı ölçütlerden (kriterlerden) yararlanılmaktadır. Bu ölçütler genel anlamda, model performansını değerlendirme ölçütleri olarak bilinmektedir. Bağımlı değişkenin sürekli değişken olması durumundaki modeller için bu performans ölçütleri 'uyum iyiliği ölçütleri' (goodness of fit) olarak adlandırılırken, kategorik değişken olması durumunda, 'tahmin (kestirim) iyiliği (goodness of prediction) ölçütleri olarak ifade edilmektedir. Bu çalışmada; bağımlı değişkenin sürekli ve kategorik olması durumunda kullanılabilen model performans ölçütlerine değinilmiştir. Sonuç olarak, çalışmanın, model performans değerlendirme ölçütlerini kısmen de olsa derleyip birlikte sunan bir çalışma olması özelliğiyle literatüre katkı sağlayabileceği ve araştırmacılara yardımcı olabileceği vurgulanmıştır.

Anahtar Kelimeler – Belirleme Katsayısı, Karışıklık Matrisi, F1 Skoru, RMSE, MAPE, AIC

I. GİRİŞ

Bilimsel çalışmalara (araştırmalara) konu olan değişkenler (özellikler) arasındaki ilişkiler, ilişkinin yapısına göre doğrusal ve doğrusal olmayan ilişkiler olmak üzere iki başlık altında incelenebileceği gibi aralarındaki ilişkilerin inceleneceği değişken sayısına göre de basit ilişkiler ve çoklu ilişkiler olarak ifade edilebilir. Aralarındaki ilişki yapısı doğrusal olmayan durumlarda; ilişkilerin incelenmesi ve yorumlanmasının, doğrusal ilişkilere göre bir miktar daha zor ve karmaşık olduğu düşünülebilir. Bu nedenle uygulamalarda, doğrusal ilişkileri incelemek üzere geliştirilen istatistik yöntemlerin çoğunlukta olduğu ve yaygın kullanıldığı söylenebilir. Diğer yandan, her ne kadar, basit ilişkilerin hesaplanması ve yorumlanması kolay olsa da çoklu ilişkiler, doğadaki ilişki yapısını açıklama bakımından basit ilişkilerden daha üstündür. Bu nedenle, değişkenler arası ilişki yapısının çoklu ilişki düzeyinde incelenmesinin daha avantajlı olduğu söylenebilir.

Değişkenler arasındaki basit veya çoklu ilişkilerin düzeyi (veya miktarı), yüzde (%) olarak belirlenebileceği gibi herhangi bir model (veya eşitlik) yardımıyla mutlak miktar olarak da belirlenebilir.

Araştırmalarda bağımlı değişken (output değişkeni, çıktı değişkeni, cevap değişkeni veya yanıt değişkeni) olarak varsayılan değişkenle, bağımsız değişken (açıklayıcı değişken veya tahminleyici değişken) olarak ifade edilen değişken veya değişkenler arasındaki ilişkiyi belirlemek üzere kullanılan eşitlikler genel olarak model kavramıyla ifade edilir. Bu tür modeller de istatistik

analizlerde, regresyon modelleri olarak adlandırılır. Regresyon analizleri ise; bağımlı değişkenle bağımsız değişkenler arasındaki ilişkileri belirlemek ve bağımlı değişkene ilişkin tahminlerde (kestirim) bulunmak üzere en uygun modelin ve katsayıların bulunması işlemlerini içerir.

Regresyon modellerinde minimum açıklayıcı değişken ile en yüksek doğruluk derecesine sahip uygun bir modelin belirlenmesi amaçlanmaktadır. En uygun modeli belirlemek üzere karar verme sürecinde, farklı ölçütlerden (kriterlerden) yararlanılmaktadır. Bu ölçütler genel anlamda, model performansını değerlendirme ölçütleri olarak bilinmektedir.

Bağımlı değişkenin sürekli değişken olması durumundaki modeller için bu performans ölçütleri 'uyum iyiliği ölçütleri' (goodness of fit) olarak adlandırılırken, Kategorik değişken olması durumunda, 'tahmin (kestirim) iyiliği (goodness of prediction) ölçütleri olarak ifade edilebilir [1].

Bu çalışmada; bağımlı değişkenin sürekli ve kategorik olması durumunda kullanılabilen model performans ölçütlerine değinilmiştir.

II. MATERYAL VE YÖNTEM

Bağımlı değişkenin sürekli olması durumunda, tahmin (veya kestirim) yapmak üzere kullanılabilecek ideal bir modelde tahmin edilen değer ile gerçek değerin birbirine eşit olması beklenir. Uygulamalarda, bu durum hemen hemen hiç mümkün olmamakla birlikte, araştırmalarda bu farklılığın (sapmanın veya tahmin değerinin gerçek değerden olan uzaklığının) miktarı belirlenmek istenir.

Bağımlı değişkene ait ölçülen veya gerçek değer 'y' ile tahmin edilen veya kestirim değeri de '$\hat{y}$' ile gösterildiğinde; iki değer arasındaki fark, (uzaklık veya sapma) hata (error) olarak ifade edilir ve 'e' ile gösterilir. Bu durumda ideal bir modelde, bütün gözlem değerleri için 'e = y - $\hat{y}$' farklarının sıfır veya sıfıra yakın olması istenir.

A. *Bağımlı değişkenin sürekli olduğu durum için ölçütler*

Belirleme (Determinasyon, R^2) katsayısı

Model performansını değerlendirmede, en yaygın kullanılan ölçütlerden birisi, 'R^2' olarak gösterilen Belirleme (Determination) katsayısıdır. Belirleme katsayısı;

$$R^2 = (1 - \frac{HKT}{GKT}) \text{ veya}$$

$$R^2 = \frac{\sum_{1=1}^{n}(\hat{y}_i - \bar{y}_i)^2}{\sum_{1=1}^{n}(y_i - \bar{y}_i)^2} \text{ veya}$$

$$R^2 = 1 - \frac{\sum_{1=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{1=1}^{n}(y_i - \bar{y}_i)^2} \quad (1)$$

olarak ifade edilir [2], [3]. (1) no'lu eşitlik grubunda; HKT; Hata Kareler Toplamı, GKT; Genel Kareler Toplamıdır.

Düzeltilmiş (Adjusted) Belirleme (Determinasyon, R^2) katsayısı

Düzeltilmiş R^2;

$$R^2_{Adj} = 1 - \left[\left(\frac{n-1}{n-k-1}\right) x (1 - R^2)\right] \text{ veya}$$

$$R^2_{Adj} = 1 - \left[\frac{HKO/(n-p)}{GKT/(n-1)}\right] \text{ veya}$$

$$R^2_{Adj} = R^2 - \frac{k}{n-k-1}(1 - R^2) \quad (2)$$

olarak yazılır [2], [3]. (2) no'lu eşitlikte; n; gözlem sayısı, k' da bağımsız değişken sayısıdır.

Hata Kareler Ortalaması (Mean Squared Error, MSE)

Hata kareler ortalaması, regresyon analizlerinde model performansını değerlendirmede yaygın kullanılan ölçütlerden birisidir. Bu ölçüt;

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (3)$$

eşitliği ile hesaplanır ve en düşük değeri sıfır (0) olmak üzere, pozitif değerler alır [2], [3].

Hata Kareler Ortalamasının Karekökü (Root Mean Squared Error, RMSE)

Hata kareler ortalamasının karekökü, Hata kareler ortalamasına benzer şekilde; regresyon analizlerinde model performansını değerlendirmek üzere yaygın kullanılan ölçütlerden birisidir.

Bu ölçüt;

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

eşitliği ile hesaplanmaktadır [[2], [3] ,[4].

Ortalama Mutlak Hata (Mean Absolute Error, MAE)

Ortalama mutlak hata;

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (5)$$

eşitliği ile hesaplanır [2], [3] ,[4].

Ortalama Mutlak Yüzde Hata (Mean Absolute Percent Error, MAPE)

Ortalama Mutlak Yüzde Hata;

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i-\hat{y}_i|}{|y_i|} \text{x}100 \quad (6)$$

olarak hesaplanır [2], [3] ,[4].

B. Model Seçim Ölçütleri

Regresyon analizinde, bağımlı değişkeni tahmin etmek için en uygun modeli belirlemede, model seçim ölçütlerinden (model selection criteria) yararlanılmaktadır. Bu ölçütler, aynı zamanda modele dahil edilecek veya modelden çıkarılacak değişkenleri belirlemek üzere karar verme sürecine de yardımcı olur. Bu ölçütlerden yaygın olarak kullanılanları aşağıda kısaca verilmiştir.

Mallows' C_p kriteri (ölçütü)

Mallows' C_p kriteri (ölçütü)

$$C_p = \frac{1}{n}(HKT + 2\,d\,[\frac{HKT_p}{(n-p-1)}] \quad (7)$$

olarak hesaplanır [2]. (7) no'lu eşitlikte:

HKT; k adet değişkenin modele dahil edilmesi elde edilen Hata Kareler Toplamı, HKT_p; Tüm değişkenlerin modele dahil edilmesi ile elde edilen Hata Kareler Toplamıdır. Uygulamada C_p istatistiği küçük olan model seçilir.

Akaike bilgi kriteri (AIC, Akaike Information Criterion)

Akaike bilgi kriteri (AIC)

$$AIC = n\ln\left(\frac{HKO}{n}\right) + 2p \quad (8)$$

eşitliği ile hesaplanır [5].

(8) no'lu eşitlikte HKO; Hata Kareler ortalamasıdır. Uygulamada AIC değeri küçük olan model seçilir.

Bayesian bilgi kriteri (BIC, Bayesian Information Criterion)

Bayesian bilgi kriteri

$$BIC = n\ln\left(\frac{HKO}{n}\right) + p\,x\,\ln(n) \quad (9)$$ eşitliği ile hesaplanır [5].

(9) no'lu eşitlikte HKO; Hata Kareler ortalamasıdır. Uygulamada BIC değeri küçük olan model seçilir.

C. Sınıflandırma (Bağımlı Değişkenin Kategorik olduğu durum) için ölçütler

Sınıflandırma (Bağımlı değişkenin kategorik olduğu durum) için genel olarak tahmin ve gerçek değerler arası uyum iyiliği ölçütleri kullanılır. Bu amaçla, Karışıklık Confusion) tablosu veya matrisinden yararlanılır.

Karışıklık tablosu (matrisi)

Makine öğrenmesi terminolojisinde, karışıklık matrisi (Confusion matrix) veya karışıklık tablosu (Confusion table) olarak adlandırılan bu tablo, istatistikte karar tablosu (Decision table) olarak da adlandırılır (Tablo 1). Tablo 1'de, birinci diyagonalde yer alan gözlemler, gerçek değere göre doğru olarak sınıflanan gözlemler iken, ikinci diyagonaldekiler yanlış sınıflanan gözlemlerdir. Bunlardan yanlış pozitif olan (YP) aynı zamanda Tip I tip hata, yanlış negatif olan (YN) Tip II hata olarak adlandırılır [6].

Tablo 1. Karışıklık tablosu

		Gerçek	
		Pozitif (+)	Negatif (-)
Tahmin	Pozitif (+)	Doğru Pozitif (DP, TP)	Yanlış Pozitif (YP, FP) [*Tip I hata*]
	Negatif (-)	Yanlış Negatif (YN, FN) [*Tip II hata*]	Doğru Negatif (DN, TN)

Doğruluk (Accuracy)

Kategorik değişkenler için model performansını değerlendirmek üzere kullanılan en yaygın ve basit ölçüt 'Doğruluk' tur. Doğruluk, gerçek değere göre doğru yapılan tahminlerin, toplam gözlem sayısına (n) bölümü olarak ifade edilir. Doğru tahminler, karışıklık matrisinde; gerçekte pozitif iken pozitif tahmin edilen ve gerçekte negatif iken negatif tahmin edilen birinci diyagonalde yer alan değerlerdir. Buna göre Doğruluk için eşitlik;

$$\text{Doğruluk} = \frac{DP + DN}{DP + DN + YP + YN} = \frac{TP + TN}{n} \text{ veya}$$

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

olarak yazılır [6].

Kesinlik (Precision)

Pozitif olarak tahmin edilen gözlemler veya bireyler (DP + YP) içerisinde, gerçekte pozitif olanların (DP) oranı olarak ifade edilir. Diğer bir ifade ile Kesinlik, pozitif olarak tahmin edilenlerin, yüzde ne kadarının gerçekte pozitif olduğunu gösteren ölçüttür. Buna göre Kesinlik için eşitlik;

$$\text{Kesinlik} = \frac{DP}{DP + YP} = \frac{TP}{TP + FP} \quad (11)$$

olarak yazılır [6].

Duyarlılık veya Sensitivite (Sensitivity, Recall)

Sensitivite (Sensitivity) olarak ifade edilen bu ölçüt, aynı zamanda 'doğru (gerçek) pozitif oran' (true positive rate) olarak da ifade edilir. Gerçekte pozitif olan gözlemler veya bireyler (DP + YN) içerisinde, pozitif (DP) olarak tahmin edilen bireylerin oranıdır. Diğer bir ifade ile Duyarlılık, gerçekte pozitif olanların, yüzde ne kadarının (yöntemle) pozitif olarak tahmin edildiğini gösteren ölçüttür. Buna göre Duyarlılık için eşitlik;

$$\text{Duyarlılık} = \frac{DP}{DP + YN} = \frac{TP}{TP + FN} \quad (12)$$

olarak yazılır [6].

F1-Skoru (F1-Score)

F1 skoru, Kesinlik ve Duyarlılık ölçütlerinin bir kombinasyonu veya bu ölçütlerinin harmonik ortalaması olarak ifade edilebilir. Diğer bir ifade ile F1 skoru, bu iki ölçütü birleştirerek, tek bir değere indirger ve 0 ile 1 arasında değer alır. F1 skoru;

$$\text{F1 skoru} = \frac{2}{\frac{1}{Kesinlik} + \frac{1}{Duyarlılık}} \text{ veya}$$

$$\text{F1 skoru} = 2\frac{Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık} \quad (13)$$

olarak yazılır [6].

Özgüllük veya Spesifite (Specificity)

Tıp bilimlerinde, Duyarlılık (Sensitivite) ve Özgüllük (Spesifite) yaygın kullanılan ölçütlerdir. Özgüllük aynı zamanda 'doğru (gerçek) negatif oran' (true negative rate) olarak da ifade edilir. Gerçekte negatif olan gözlemler veya bireyler (TN + YP) içerisinde, negatif (DN) olarak tahmin edilen bireylerin oranıdır. Diğer bir ifade ile Özgüllük, gerçekte negatif olanların, yüzde ne kadarının (yöntemle) negatif olarak tahmin edildiğini gösteren ölçüttür. Özgüllük için eşitlik;

Özgüllük = DN/(DN+YP) = (TN)/(TN+FP) (14)
olarak yazılır [6].

III. TARTIŞMA

Daha önceden de belirtildiği üzere, regresyon modellerinde, model performansını belirlemek üzere en yaygın kullanılan ölçütlerden birisi belirleme katsayısıdır. Belirleme katsayısı iki şekilde yorumlanır. Birincisi, modeldeki bağımlı değişkeni tahmin etmedeki doğruluk derecesini yüzde (%) olarak ifade eder. Örneğin, Belirleme katsayısının %95 olarak bulunması, modelle yapılacak 100 tahminden 95'inin doğru tahmin olabileceğini ifade eder. İkincisi ise modeldeki bağımlı değişkende gözlenen varyasyonun (değişimin veya farklılığın), ne kadarının modelde yer alan bağımsız (açıklayıcı) değişkenlerle açıklanabildiğini ifade eder. Diğer bir ifade ile bağımlı değişkendeki varyasyonun, bağımsız değişkenlerle determine edilen (belirlenen) kısmını belirtir. Örneğin, Belirleme katsayısının %90 olarak bulunması, Bağımlı değişkendeki varyasyonun %90'ının bağımsız değişkenlerdeki varyasyon tarafından determine edildiğini veya bağımsız değişkenlerdeki varyasyondan kaynaklandığını ifade eder.

Belirleme katsayısı yaygın olarak kullanılsa da, küçük bir dezavantajından bahsedilir. Belirleme katsayısı, modele bağımsız değişkenlerin eklenmesi ile eklenen bağımsız değişkenin modele katkısı olmasa da daima artma eğilimi gösterir. Bu nedenle araştırıcılar, daha yüksek Belirleme katsayısı (R^2 değeri) elde etmek için modele açıklayıcı değişken ekleme eğiliminde olabilir. Bu durumda, katkı sağlamayan değişkenler de modele eklenmiş olabilir. Bunun çözümü için Düzeltilmiş belirleme katsayısı (Adjusted R^2)'nın kullanımı önerilmektedir.

Düzeltmiş belirleme katsayısı, Belirleme katsayısından ya daha düşük değer olarak ya da buna eşit değer olarak elde edilir. Böylece Düzeltilmiş belirleme katsayısı ile modele açıklayıcı değişken eklenmesinin, gerçekten de bu değişkenin modele katkı sağlayıp sağlamadığı belirlenmiş olur.

Hata kareler ortalaması (MSE), gözlenen değerler ile tahmin edilen değerler arasındaki farkların karesini aldığı için sapmalar büyük değerler olarak bulunmaktadır. Bununla birlikte Plevris ve ark. [7], MSE'nin en önemli dezavantajının uç (outlier) değerlere karşı duyarlı olduğunu, diğer bir ifade ile sağlam (robust) bir ölçüt olmadığını vurgulamışlardır. Bu durumda, Hata kareler ortalamasının karekökü alınarak, RMSE değeri elde edilir. RMSE ile değerler küçültülmüş olmakla birlikte, hatanın ölçü biriminin, bağımlı değişkenin ölçü birimi ile aynı olması sağlanır.

Ortalama mutlak hata (MAE), gözlenen değerler ile tahmin edilen değerler arasındaki sapmaların (farkların) mutlak değerce toplamı alınarak, gözlem sayısına bölünmesiyle elde edilen değerdir. Diğer bir ifade ile tahmin değerleri ile gerçek değerler arasındaki mutlak değer olarak alınan farkların

ortalamasıdır. Negatif değerlerle pozitif değerlerin birbirini götürme ve böylece MAE için daha küçük değer elde etme riskini ortadan kaldırmak için mutlak değerce toplamlar alınmaktadır. MAE ile; RMSE'ye benzer şekilde, bağımlı değişkenin ölçü birimi ile aynı birimde değer elde edilmektedir.

Farklı çalışmalarda, bağımlı değişken farklı ölçü biriminde ölçülmüş olabilir. Örneğin bir çalışmada 0 ile 100 arasında değerler alırken, başka bir çalışmada standardize edilmiş ve 0 ile 1 arasında değer almış olabilir. Bu durumda MAE hesaplandığında, standardize edilmiş değerlerdeki MAE, diğerine göre daha düşük olacaktır. Ölçü biriminden bağımsız hale getirmek için yüzde olarak ifade edilen, Ortalama Mutlak Yüzde Hata (MAPE) ölçütü hesaplanmaktadır.

Kategorik değişkenler için Karışıklık tablosundaki Doğruluk (Accuracy) ölçütü, her ne kadar en basit ve en yaygın kullanılan ölçüt olsa da, Binary (ikili) değişkenler için dengesiz durumlarda (örneğin; $p = 0.9$; $q = 0.1$) Doğruluk ölçütü yüksek değer olarak bulunabilir. Bu gibi durumlarda, yanlış pozitif (YP) ve yanlış negatif (YN) tahminlerin de göz önüne alınması gerekmektedir. Bu nedenle; Kesinlik (Precision) olarak ifade edilen ölçüt, yanlış pozitifleri (YP) dikkate alırken, Duyarlılık (Recall) olarak ifade edilen ölçüt, yanlış negatifleri (YN) dikkate almaktadır.

Kesinlik aynı zamanda, pozitif tahmin değeri ya da pozitif kestirim değeri (positive predictive value) olarak da ifade edilebilir. Yanlış pozitiflerin (YP) düşük olması durumunda, 'Kesinlik' yüksek değer olarak bulunacaktır. Böylece Tip I hata yapılması durumunda, ortaya çıkacak riskin, maliyetin veya cezanın yüksek olacağı durumlarda, 'Kesinlik' yararlı bir önlem olabilir. Örneğin kanser teşhisinde kullanılan herhangi bir 'genetik test' için; bu testin kullanımı sonucunda, Pozitif sonucun (TP) yüksek olması, kanser gelişme riskini artırıyor olsun. Genetik testin yanlış pozitif sonuç vermesi, bireyin ya da kişinin gereksiz yere fiziksel ve zihinsel olarak kansere karşı önlem alma gayreti içine girmiş olmasına neden olacaktır. Dolayısıyla Yanlış negatif (YN) durumdan daha ziyade bu durumdan kaçınılmak istenilebilir. Zira ikinci durum; aslında kansere yakalanma riski yüksek olan bir kişi için genetik testin negatif sonuç verdiği anlamına gelmektedir. Ancak riskin yüksek olması veya artması, kişinin kesinlikle kansere yakalanacağı anlamına gelmeyeceği gibi durum böyle olsa bile zamanında teşhis edilebilmesi mümkün olabilecektir.

Yanlış negatiflerin (YN) düşük olması durumunda, 'Duyarlılık' yüksek olacaktır. Böylece Tip II hata yapılması durumunda, ortaya çıkacak riskin, maliyetin veya cezanın yüksek olacağı durumlarda, 'Duyarlılık' yararlı bir önlem olabilir. Örneğin, bankacılıkta herhangi bir transfer işleminin hileli olup olmadığını tahmin etmek üzere bir test veya algoritma olduğu varsayılsın. Yanlış negatif sonuç, gerçekte işlem hileli veya sahte iken algoritmanın, sahte veya hileli bu işlemi doğru bir işlem olarak kabul ettiği anlamına gelir. Böyle bir karar, geri dönüşü olmayan bir para kaybına neden olabileceği gibi banka açısından istenmeyen sonuçlar verir. Diğer yandan yanlış pozitif sonuç ise doğru veya meşru bir işlemin sahte, hileli veya yolsuzluk yapılmış bir işlem gibi görülmesine ve engellenmesine neden olur. Bu durumda, her ne kadar, doğru bir işlem gecikmiş olsa da gerekli kontrol yapıldıktan sonra işlem düzeltilebilir.

F1 skorunun, Kesinlik ve Duyarlılık değerlerinin optimum noktası için uygun bir ölçü olduğu söylenebilir. Bu değerlerden herhangi birisinin düşük olması durumunda, F1 skoru da düşük olma eğilimi gösterirken, her ikisinin de yüksek olması durumunda ise F1 skoru yüksek olma eğilimi gösterir. Kesinlik ve Duyarlılık ölçütlerinden herhangi birisinin sıfır olması durumunda, F1 skoru sıfır olur. F1 skoru, birden fazla modele ait kesinlik ve doğruluk değerleri arasından en iyi modeli seçmek için kullanılır.

IV. SONUÇLAR

Çalışmada, regresyon ve sınıflandırmada, diğer bir ifade ile bağımlı değişkenin sürekli veya kategorik olması durumunda, model performansını değerlendirmek üzere, yaygın olarak kullanılabilen performans ölçütlerine değinilmiş ve bu ölçütlerin hesaplanma şekli ile birlikte genel özelliklerinden bahsedilmiştir. Bu çalışmanın, model performans değerlendirme ölçütlerini kısmen de olsa derleyip birlikte sunan bir çalışma olması özelliği ile literatüre katkı sağlayacağı ve araştırmacılara yardımcı olabileceği ümit edilmektedir.

KAYNAKLAR

[1] Breiman, L. "Statistical modeling: The two cultures (with comments and a rejoinder by the author)". Statistical science, 16(3), 199-231. 2001.

[2] N. Yashwanth, “Evaluation metrics & Model Selection in Linear Regression”. 2021.
[3] A. Chugh, “MAE, MSE, RMSE, coefficient of determination, adjusted R squared—which metric is better?”. Retrieved December, 8, 2020..
[4] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation”. PeerJ Computer Science, 7, e623. 2021.
[5] C. Lindsey and S. “Sheather Variable selection in linear regression”. The Stata Journal 10, (4), 650–669.2010
[6] A. Bhandari, “Everything you Should Know about Confusion Matrix for Machine Learning”, Analytics Vidhya. [Online]. Available, 3, 10. 2020.
[7] V Plevris., G. Solorzano, N. P. Bakas and M.A.B. Seghier. (2022). “Investigation of performance metrics in regressıon analysıs and machıne learning-based prediction models”. The 8th European Congress on Computational Methods in Applied Sciences and Engineering ECCOMAS Congress 2022 5 – 9 June 2022, Oslo, Norway