

Kalp Yetmezliği Hastalığının Erken Teşhisinde Makine Öğrenimi Algoritmalarının Performans Karşılaştırması

Performance Comparison of Machine Learning Algorithms for Early Diagnosis of Heart Failure

Gülay Gürgen^{1*} , **Soydan Serttaş²** 

¹ Kütahya Dumlupınar Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Kütahya, Türkiye

² Dr. Öğr. Üyesi, Kütahya Dumlupınar Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Kütahya, Türkiye

* Corresponding author: gurgenngulay@gmail.com

Geliş Tarihi / Received: 14.06.2023
Kabul Tarihi / Accepted: 21.07.2023

Araştırma Makalesi/Research Article
DOI: 10.5281/zenodo.8238065

ÖZET

Kalp yetmezliği, kalbin yeterli miktarda kan pompalayamadığı bir durumdur ve tedavi edilmediği takdirde ciddi sağlık sorunlarına yol açabilir. Erken teşhis, hastalığın ilerlemesini önleyebilir ve yaşam kalitesini artırabilir. Bu makalede farklı makine öğrenimi algoritmalarının kalp yetmezliği hastalığının erken teşhisindeki performansı değerlendirilmiştir. Kaggle veri tabanından alınan veri seti, kalp yetmezliği hastalığına sahip ve sağlıklı bireylerden oluşan kapsamlı bir hasta veri tabanından 11 bağımsız değişken içerecek şekilde oluşmaktadır. Çalışmada Classification and Regression Tree (CART), K-Nearest Neighbours (KNN), Logistic Regression, Random Forest (RF), AdaBoost, XGBoost, LightGBM ve CatBoost olmak üzere 8 adet algoritma kullanılmıştır. Makalenin sonuçları, kullanılan makine öğrenimi algoritmalarının kalp yetmezliği hastalığının erken teşhisinde etkili bir şekilde kullanılabileceğini göstermektedir. Algoritmalar, yüksek doğruluk oranları ve düşük hata değerleri sergilemiştir. Ayrıca, farklı algoritmalar arasında performans farklılıkları da belirlenmiştir. Random forest, F1 skoru (0.98(1)), ROC AUC (0.999) ve doğruluk (0.99) değerlerine göre çalışmanın en iyi tahminini yapan algoritması olmuştur. Bu bulgular, kalp yetmezliği hastalığının erken teşhisi için makine öğrenimi algoritmalarının potansiyelini vurgulamaktadır.

Anahtar Kelimeler: Kalp yetmezliği, Makine Öğrenmesi, Tahmin, Sınıflandırma

ABSTRACT

Heart failure is a condition where the heart is unable to pump an adequate amount of blood and can lead to serious health problems if left untreated. Early diagnosis can prevent the progression of the disease and improve the quality of life. This article evaluates the performance of different machine learning algorithms in early detection of heart failure disease. The data set from the Kaggle database consists of 11 independent variables from a comprehensive database of patients with heart failure and healthy individuals. Eight algorithms were used in the study, namely Classification and Regression Tree (CART), K-Nearest Neighborhoods (KNN), Logistic Regression, Random Forest (RF), AdaBoost, XGBoost, LightGBM and CatBoost. The results of the article demonstrate that the machine learning algorithms used can be effectively employed in the early diagnosis of heart failure disease. The algorithms exhibit high accuracy rates and low error values. Additionally, performance differences among different algorithms are identified. Random forest was the best estimator of the

study, with F1 score (0.98(1)), ROC AUC (0.999), and accuracy (0.99). These findings emphasize the potential of machine learning algorithms for the early diagnosis of heart failure disease.

Keywords: Heart Failure, Machine Learning, Prediction, Classification

1. GİRİŞ

Kalp yetmezliği, kalbin yeterli kan pompalama kapasitesine sahip olamadığı kronik bir hastalık olup hayat kalitesinde kötüleşmeye, hastaneye yatışlara ve en önemlisi de ölümlere neden olabilmektedir. Genellikle ileri safhada saptanır ve dünya çapında önemli bir sağlık sorunudur. Kalp yetmezliği tedavisi oldukça pahalı ve zor bir süreçtir. Bu nedenle erken teşhis ile doğru tedavi hayati öneme sahiptir ve böylece ölümlerin önüne geçilebilir.

Makine öğrenmesi, son yıllarda tıp alanında büyük bir ilgi odağı haline gelmiştir. Makine öğrenmesi algoritmaları, büyük veri kümelerindeki kalıpları tanıyarak hastalıkların teşhis edilmesinde ve tedavi planlarının belirlenmesinde kullanılabilir. Bu makale, kalp yetmezliği teşhisi için kullanılan makine öğrenmesi yöntemlerinin karşılaştırılmasına odaklanmıştır. Bu yöntemler, veri seti boyutu, model eğitimi için gereken zaman, model doğruluğu ve genelleme yeteneği açısından karşılaştırılacaktır. Bu çalışmanın amacı, kalp yetmezliği teşhisi için en iyi makine öğrenmesi yöntemini belirlemek ve hastaların erken teşhis ve tedavi edilmesine yardımcı olmak için bir katkıda bulunmaktır.

1.1. LİTERATÜR TARAMASI

Kalp yetmezliği, kalbin yeterli miktarda kan pompalayamaması sonucu ortaya çıkan ve dünya genelinde önemli bir halk sağlığı sorunu olan bir durumdur. Bu hastalık, yüksek ölüm oranı ve yüksek maliyetli tedavi yöntemleri nedeniyle ciddi bir endişe kaynağıdır. Son yıllarda, kalp yetmezliği hastalarının teşhisi ve tedavisinde yapay zekâ teknikleri, özellikle makine öğrenmesi algoritmaları kullanılmaktadır. Bu algoritmalar, kalp yetmezliği teşhisini ve prognozunu belirlemede yüksek doğruluk oranları elde etmiştir. Ancak, literatürde birçok makine öğrenmesi algoritması mevcut olup, bu algoritmaların hangisinin en iyi performansı sergilediği konusunda tartışmalar devam etmektedir. Bu çalışma kalp yetmezliği tahmin problemi için kullanılan popüler makine öğrenmesi algoritmalarının performanslarını karşılaştırmayı amaçlamaktadır. Bu amaç doğrultusunda, literatürdeki çalışmalar taranmıştır.

Davide Chicco ve Giuseppe Jurman tarafından gerçekleştirilen bir çalışmada, kalp hastalığı risk tahmini için sınıflandırma algoritmalarının performans değerlendirilmesi yapılmıştır. Kullanılan makine öğrenmesi algoritmaları sırasıyla Lojistik Regresyon, Karar Ağacı, Random Forest, Gradient Boosting Machine, Naïve bayes gibi algoritmalar kullanılmıştır. Çalışma sonucunda, Random Forest en yüksek MCC'yi (+0,384), en yüksek doğruluğu (%74) ve en yüksek ROC AUC'yi (0,800) elde ederek diğer tüm yöntemlerden daha iyi performans gösterdiği görülmüştür. Ayrıca, yalnızca serum kreatinin ve ejeksiyon fraksiyonunun tıbbi kayıtlardan kalp yetmezliği hastalarının sağkalımını tahmin etmek için yeterli olduğunu ve hastaların hayatta kalma durumunu tahmin etmek için serum kreatinin ve ejeksiyon fraksiyonu odaklanması önerilmektedir (Chicco, Jurman, 2020).

Ezgi Aktaş Potur ve Nihal Erginel çalışmasında kalp yetmezliği hastalarının sağ kalımlarının sınıflandırma algoritmaları ile tahmin edilmesi hedeflenmiştir. Bu amaçla, karar ağacı (decision tree), çoklu katmanlı algılayıcı (multilayer perceptron) ve destek vektör makineleri (support vector machines) gibi farklı sınıflandırma algoritmaları kullanılmıştır. Çalışma sonucunda, en yüksek

doğruluk oranının %90 ile çoklu katmanlı algılayıcıların elde ettiği belirlenmiştir (Aktaş Potur, Erginel, 2021).

Serdar Gündoğdu tarafından gerçekleştirilen bir çalışmada, kalp hastalık riskinin tahmin edilmesi için farklı sınıflandırıcı algoritmalarının performansları değerlendirilmiştir. Bu amaçla, karar ağacı, k-en yakın komşu (k-nearest neighbours), doğrusal diskriminant analizi (linear discriminant analysis) ve destek vektör makineleri gibi algoritmalar kullanılmıştır. Çalışma sonucunda, en yüksek doğruluk oranı %90.2 ile random forest algoritması ile elde edildiği görülmüştür (Gündoğdu, 2021).

Kalp yetmezliği olan hastaların ölüm ya da sağ kalma durumlarının tahmini için yapılan bir diğer çalışmada, 299 örneklem ve 13 öznitelik ile oluşturulan bir veri kümesi kullanılmıştır. Çalışmada Yapay Sinir Ağları (YSA), Aşırı Gradyan Arttırma (AGA – Extreme Gradient Boosting) ve Rastgele Orman (RO) yöntemleri incelenmiştir. AGA algoritmasının kalp yetersizliği olan durumları tespit etmede daha başarılı olduğu; YSA ve RO algoritmalarının ise kalp yetersizliği olmayan durumları tespit etmede daha başarılı olduğu görülmüştür (Bozkurt Keser, Keskin, 2023).

Sağlıklı kişilerin ileride yakalanması ihtimal kalp hastalığını belirlemek amacıyla yapılan çalışmada, veri setinin kategorik yapısından dolayı yapay sinir ağları, karar ağaçları ve rasgele orman algoritmaları test edilmiş ve Rastgele Orman algoritmasının 90.22% ile en iyi sonucu verdiği belirtilmiştir (Karanfil, 2017).

Kalp ameliyatı sırasında veya kalp ameliyatı sonrası kısa bir süre içinde hastanın hayati riskinin belirlenmesi için yapılan bir çalışma da mevcuttur. Sınıflandırmaya dayalı makine öğrenimi tekniklerinden olan Naive Bayes Sınıflandırıcısı, k-En Yakın Komşu Algoritması, Lojistik Regresyon Analizi, ID3 ve C4.5 Karar Ağacı Algoritmaları gibi farklı modeller oluşturulan bu çalışmada kullanılan modellerin performansları karşılaştırılarak en iyi performansı C4.5 karar ağacı modelinin verdiği belirtilmiştir (Kartal, 2015).

Kalp rahatsızlığı tespiti için örnek bir veri seti üzerinde makine öğrenmesi algoritmalarını uygulayarak sonuçları kıyaslamayı amaçlayan diğer bir çalışmada ilk olarak veri seti incelenmiş ve kalp rahatsızlığına dair işaretler belirlenmiştir. Lojistik Regresyon, Random Forest ve KNN algoritmaları kullanılarak örnek bir model oluşturulmuş ve kalp rahatsızlığı olan bireyler tespit edilmiştir. Elde edilen sonuçlar kıyaslandığında, Random Forest algoritmasının %88 doğruluk değeriyle daha başarılı olduğu gözlemlenmiştir. Bu değeri sırasıyla %85 doğruluk değeriyle Lojistik Regresyon ve %70 doğruluk değeriyle KNN algoritması takip etmiştir (Coşar, Deniz, 2021).

2. MATERYAL ve METOT

2.1. VERİ KÜMESİ

Bu çalışmada Kaggle veri tabanında bulunan kalp yetmezliği veri kümesi kullanılmıştır. Veri kümesine Kaggle platformu üzerinden ulaşırken, veriler Faisalabad Institute of Cardiology ve Faisalabad da bulunan Allied Hastanesi'nden Nisan-Aralık 2015 tarihleri arasında toplanan tıbbi kayıtlardan gelmektedir. Veri seti toplam 299 kayıttan oluşmaktadır. Kullanılan 13 özellik yaş, cinsiyet, anemi, CPK enzim düzeyi, bireyin diyabet olup olmadığı durumu, ejeksiyon fraksiyon, hipertansiyon, trombosit miktarı, serum kreatinin, serum sodyum, sigara kullanımı, takip süreci ve hayatta kalım durumudur. Kullanılan Heart Failure Prediction veri kümesinin 13 özelliğinin almış olduğu değerler ve veri tipleri ile ilgili özet bilgiler Çizelge 1'de sunulmuştur. Özelliklerle ilgili tanımlayıcı aralıkların belirlenmesi için kullanılan kaynaklar belirtilmiştir

Çizelge 1. Veri kümesi özellikleri

Özellikler	Değerler	Değişken
Yaş (yıl)	Nümerik	Bağımsız
Anemi: Kırmızı kan hücrelerinin veya hemoglobinin azalması (0 = Hayır, 1 = Evet)	Kategorik	Bağımsız
Kreatin Kinaz: Kandaki CPK enzim düzeyi (mcq/L)	Nümerik	Bağımsız
Diyabet (0 = Hayır, 1 = Evet)	Kategorik	Bağımsız
Ejeksiyon Fraksiyonu: Her kasılmada kalpten çıkan kanın yüzdesi	Nümerik	Bağımsız
Hipertansiyon: Hipertansiyon (0 = Hayır, 1 = Evet)	Kategorik	Bağımsız
Trombosit: Trombositler vücudunuzun kanamayı durdurmak için pıhtı oluşumuna yardımcı olan küçük renksiz kan hücreleridir. (kiloplatelet/mL)	Nümerik	Bağımsız
Serum kreatinin: Kandaki serum kreatinin düzeyi (mg/dL)	Nümerik	Bağımsız
Serum Sodyum: Kandaki serum sodyum seviyesi (mEq/L)	Nümerik	Bağımsız
Cinsiyet (0 = Kadın, 1 = Erkek)	Kategorik	Bağımsız
Sigara Kullanımı (0 = Hayır, 1 = Evet)	Kategorik	Bağımsız
Zaman: Takip süreci	Nümerik	Bağımlı
Ölüm Olayı (0 = Hayır, 1 = Evet)	Kategorik	Bağımlı

2.2. YÖNTEMLER

Çalışmada kalp yetmezliği tespitinde kullanılan makine öğrenmesi yöntemlerinin performansını karşılaştırmak için Classification and Regression Tree (CART), K-Nearest Neighbours (KNN), Logistic Regression, Random Forest (RF), AdaBoost, XGBoost, LightGBM, CatBoost ve RandomizedSearchCV yöntemleri kullanılmıştır. Veriler öncelikle bir veri ön işleme aşamasından geçirilmiştir. Bu aşamada, eksik verilerin doldurulması, aykırı verilerin temizlenmesi ve veri ölçeklendirmesi yapılmıştır. Ardından veriler eğitim, doğrulama ve test olarak üçe ayrılmıştır. Eğitim seti modelin eğitimi için kullanılırken doğrulama seti modellerin hiperparametre optimizasyonunda kullanılmıştır. Test seti ise model performanslarının değerlendirilmesinde kullanılmıştır. CART, KNN, Logistic Regression, RF, AdaBoost, XGBoost, LightGBM ve CatBoost yöntemleri, Scikit-learn kütüphanesi kullanılarak uygulanmıştır. Hiperparametre optimizasyonu için RandomizedSearchCV yöntemi kullanılmıştır. Bu yöntem, tüm modeller için uygulanmış olup her bir model için en iyi hiperparametre seti belirlenmiştir.

2.2.1. Classification and Regression Tree (CART)

CART, karar ağaçları kullanarak verileri sınıflandırmak veya tahmin etmek için kullanılan bir yöntemdir (Breiman, Friedman, Stone, Olshen, 1984). Karar ağacı, verileri sınıflandırmak veya tahmin etmek için bir dizi karar noktasından oluşur.

2.2.2. K-Nearest Neighbours (KNN)

KNN, bir veri noktasının sınıfını, ona en yakın komşu veri noktalarının sınıflarına göre tahmin eder (Cover, Hart, 1967).

2.2.3. Logistic Regression

Logistic regresyon, bir veri kümesindeki değişkenlerin bir veya daha fazla bağımsız değişkenle ilişkisini modeller (Hosmer, Lemeshow, 1989). Bu yöntem, verilerin sınıflandırılması veya tahmin edilmesi için kullanılır.

2.2.4. Random Forest (RF)

RF, birden fazla karar ağacını bir araya getirerek daha doğru sonuçlar elde etmek için kullanılan bir yöntemdir (Breiman, 2001). Her karar ağacı, verileri rastgele alt kümeler halinde böler ve her alt kümeyle ait bir karar ağacı oluşturur.

2.2.5. AdaBoost

AdaBoost, zayıf sınıflandırıcıları bir araya getirerek daha güçlü bir sınıflandırıcı oluşturmak için kullanılan bir yöntemdir (Freund, Schapire, 1996). Bu yöntem, her sınıflandırıcının yanlış sınıflandırdığı verileri vurgulayarak bir sonraki sınıflandırıcının bu hataları düzeltmesine yardımcı olur.

2.2.6. XGBoost

XGBoost, bir veri kümesindeki özellikleri modellemek ve tahmin etmek için kullanılan bir yöntemdir (Chen, Guestrin, 2016). Bu yöntem, bir özelliğin etkisini ölçmek ve en önemli özellikleri belirlemek için kullanılır.

2.2.7. LightGBM

LightGBM, büyük veri kümelerinde hızlı bir şekilde çalışabilen bir sınıflandırma ve regresyon modeli oluşturmak için kullanılan bir yöntemdir. Bu yöntem, veri kümesinin özelliklerini sıkıştırarak model oluşturma sürecini hızlandırır.

2.2.8. CatBoost

CatBoost, kategorik verileri işlemek için özel olarak tasarlanmış bir sınıflandırma ve regresyon yöntemidir. Bu yöntem, diğer yöntemlerdeki aşırı öğrenme sorunu çözmek için geliştirilmiştir.

2.2.9. RandomizedSearchCV

RandomizedSearchCV, bir hiperparametre ayarlama yöntemidir (Bergstra, Bengio, 2012). Bu yöntem, bir model için en iyi hiperparametreleri bulmak için rastgele bir örnekleme stratejisi kullanır.

2.3. DEĞERLENDİRME KRİTERLERİ

Modellerin başarılarının incelenmesi için ikili sınıflandırmaya dayalı performans değerlendirme ölçütleri kullanılmıştır. Bu ölçütler arasında doğruluk (accuracy), duyarlılık (recall), kesinlik (precision) ve F1 skoru yer almaktadır. Ayrıca, ROC eğrileri çizilerek, AUC (Area Under the Curve) metriği ile performans karşılaştırmaları yapılmıştır. Sonuç olarak, tüm modellerin performansları karşılaştırılmış ve en iyi performansı gösteren model saptanmıştır.

Doğruluk, modelin doğru sınıflandırılan örneklerin oranını ifade eder. Duyarlılık, gerçek pozitiflerin doğru şekilde tespit edilme oranını ifade ederken, kesinlik pozitif olarak tahmin edilen örneklerin gerçek pozitifler arasındaki oranını ifade etmektedir. F1 skoru ise duyarlılık ve kesinliği birleştirerek modelin performansını daha kapsamlı bir şekilde değerlendirir. ROC eğrisi altındaki alan (AUC) ise modelin pozitif ve negatif örnekleri ayırt edebilme başarısını gösterir.

2.3.1. KARMAŞIKLIK Matrisi

Karmaşıklık matrisi, sınıflandırma algoritmalarının performansını ölçmek için kullanılan bir metriktir. Bu matris, gerçek sınıf ve tahmin edilen sınıflar arasındaki ilişkiyi göstermektedir.

True Positive (TP): Gerçek pozitif sınıfların sayısı (Gerçek sınıf "1", Tahmin edilen sınıf "1")

False Positive (FP): Yanlış pozitif sınıfların sayısı (Gerçek sınıf "0", Tahmin edilen sınıf "1")

False Negative (FN): Yanlış negatif sınıfların sayısı (Gerçek sınıf "1", Tahmin edilen sınıf "0")

True Negative (TN): Gerçek negatif sınıfların sayısı (Gerçek sınıf "0", Tahmin edilen sınıf "0")

Karmaşıklık matrisi, bu değerleri kullanarak sınıflandırma algoritmasının doğruluğunu, kesinliğini, duyarlılığını ve F1 skorunu hesaplamak için kullanılabilir. Karmaşıklık matrisi Çizelge 2'de gösterilmiştir.

Çizelge 2. Karmaşıklık matrisi tablosu

		Tahmin Edilen Sınıf	
		Sınıf = 1	Sınıf = 0
Gerçek Sınıf	Sınıf = 1	True Positive (TP)	False Negative (FN)
	Sınıf = 0	False Positive (FP)	True Negative (TN)

2.3.2. HESAPLAMALAR

Doğruluk (Accuracy) doğru sınıflandırma oranıdır.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN}$$

Kesinlik (Precision) pozitif sınıf (1) tahminlerinin başarı oranıdır.

$$\text{Kesinlik} = \frac{TP}{TP + FP}$$

Duyarlılık (Recall) pozitif sınıfın (1) doğru tahmin edilme oranıdır.

$$\text{Duyarlılık} = \frac{TP}{TP + FN}$$

F1 Skor Kesinlik ve Duyarlılık değerlerinin harmonik ortalamasını göstermektedir.

$$F1 \text{ Score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

2.3.3 Receiver Operating Characteristic Curve (ROC Eğrisi)

ROC eğrisi, sınıflandırma modelinin performansını gösteren bir grafik yöntemidir. Bu eğri, sınıflandırma modelinin duyarlılık ve kesinlik performansını farklı eşik değerleriyle gösterir. ROC eğrisi, yatay ekseninde False Positive Rate (FPR) ve dikey ekseninde True Positive Rate (TPR) gösterir. Bir ROC eğrisi, sol üst köşeye yakın olduğunda, modelin daha iyi performans gösterdiği anlamına gelir. Ayrıca, ROC eğrisi altında kalan alan (AUC), sınıflandırma modelinin performansını ölçen

bir değerdir. AUC değeri 0 ile 1 arasında değişir ve 1'e ne kadar yakınsa, modelin performansı o kadar iyidir.

Her bir eşik değeri için, sınıflandırma sonuçlarına dayanarak aşağıdaki dört ölçüm hesaplanır:

$$\text{True Positive Rate (TPR)} = \frac{TP}{TP + TN}$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN}$$

$$\text{True Negative Rate (TNR)} = \frac{TN}{TN + FP}$$

$$\text{False Negative Rate (FNR)} = \frac{FN}{TP + FN}$$

TPR ve FPR ölçümlerinin farklı eşik değerleri için hesaplanması sonucu bir çift (FPR, TPR) oluşur. Bu çiftler, ROC eğrisinin noktalarını oluşturur.

Son olarak, ROC eğrisi çizilir. Bu eğri, FPR eksenindeki değerlere karşı TPR eksenindeki değerlerin karşılaştırmasını gösterir. Optimal sınıflandırma performansı, ROC eğrisinin sol üst köşesine (FPR = 0, TPR = 1) yakın bir noktada bulunur (Fawcett, 2006-Hand, 2001).

3. BULGULAR ve TARTIŞMA

Yapılan tüm işlemleri gerçekleştirdiğimiz bilgisayar donanımının özellikleri aşağıda belirtilmiştir:

- 11th Gen Intel(R) Core (TM) i5-11400H @ 2.70GHz işlemci,
- 16 GB DDR4 RAM
- Windows 10 Professional 64 bit işletim sistemi.

Kalp yetmezliği tahmin problemi, hastaların klinik özellikleri (yaş, cinsiyet, tıbbi geçmiş vb.) temel alınarak kalp yetmezliği riskinin tahmin edilmesini amaçlamıştır. Çalışmada kullanılan sınıflandırıcı yöntemlerinin karşılaştırılması Çizelge 3'te gösterilmiştir.

Çizelge 3. Algoritmaların başarı değerlendirmelerinin kıyaslanması

Algoritma	Doğruluk	Kesinlik (0)	Kesinlik (1)	Duyarlılık (0)	Duyarlılık (1)	F1 Skor (0)	F1 Skor (1)	ROC AUC	Süre (Saniye)
CART	0.85	0.85	0.83	0.94	0.66	0.89	0.73	0.796	4.64 s
KNN	0.78	0.77	0.81	0.94	0.66	0.89	0.73	0.881	5.22 s
Logistic Regression	0.84	0.86	0.79	0.91	0.70	0.89	0.74	0.899	0.41 s
Random Forest	0.99	0.98	1.00	1.00	0.96	0.99	0.98	0.999	71.92 s
AdaBoost	0.85	0.85	0.83	0.94	0.66	0.89	0.73	0.799	12.89 s
XGBoost	0.95	0.94	0.98	0.99	0.88	0.97	0.92	0.981	13.33 s

LightGBM	0.91	0.92	0.91	0.96	0.81	0.94	0.86	0.974	18.85 s
CatBoost	0.88	0.88	0.88	0.96	0.72	0.92	0.79	0.954	18.56 s

Karmaşıklık matrisini oluşturan parametrelerden elde edilen performans kriterlerine göre, Random Forest algoritmasının diğer algoritmalara kıyasla en yüksek doğruluk (accuracy) değerine sahip olduğu görülmüştür. Ayrıca Random Forest algoritması, diğer algoritmalara göre daha yüksek precision, recall, F1-score ve ROC AUC değerleri elde etmiştir. Bu sonuçlar, modelin çok iyi performans sergilediğini göstermektedir. Diğer taraftan, KNN algoritması diğer algoritmalara göre daha düşük performans sergilemiştir.

XGBoost, LightGBM, CatBoost, CART, AdaBoost ve Logistic Regression algoritmaları da yüksek doğruluk, precision, recall, F1-score ve ROC AUC değerleri elde etmiştir. Bu sonuçlar, bu algoritmaların da iyi performans sergilediğini göstermektedir.

KNN algoritmasının sınıflandırma performansını gösteren kriterlerden ROC AUC, F1 skoru ve doğruluk değerleri sırasıyla 0.881, 0.89(0), 0.73(1) ve 0.78 olarak bulunmuştur. Bu değerlere göre, KNN, çalışmanın en kötü tahminini yapan algoritmasıdır.

Random Forest algoritmasının sınıflandırma performansına ait ROC AUC, F1 skoru ve doğruluk değerleri sırasıyla 0.999, 0.99(0), 0.98(1) ve 0.99'dur. Bu sonuçlara göre, Random Forest çalışmanın en iyi tahminini yapan algoritması olmuştur.

İşlem için sınıflandırıcılardan en az süre harcayan 0.41 sn ile Logistic Regression iken; en başarılı tahmin de bulunan Random Forest ise sınıflandırma işlemini en fazla zaman harcayarak 71.92 sn'de bitirmiştir.

Sonuçlar, modelin kalp yetmezliği riskini %99 doğruluk oranıyla tahmin edebildiğini göstermiştir. Modelin en önemli özelliklerinin serum kreatinin ve ejection fraksiyonu olduğu görülürken, sırasıyla önem sıralaması şu şekildedir; trombosit, kreatin kinaz, yaş, serum sodyum, hipertansiyon, diyabet, sigara kullanımı, anemi, cinsiyet. Ayrıca erkeklerin kadınlara göre her yaş grubu için geçerli olmak üzere kalp yetmezliği riskinin daha fazla olduğu gözlemlenmiştir.

4. SONUÇLAR

Bu çalışmada kalp yetmezliği hastalarının sağ kalımlarının yüksek bir doğrulukla tahmin edilmesi için 299 kayıt ve 13 değişkenden oluşan kalp yetmezliği veri seti kullanılarak 8 makine öğrenmesi algoritması performansı değerlendirilmiştir. Kullanılan algoritmalar Classification and Regression Tree (CART), K-Nearest Neighbours (KNN), Logistic Regression, Random Forest (RF), AdaBoost, XGBoost, LightGBM, CatBoost iken en yüksek doğruluk oranına ulaşan %99 ile Random Forest olmuştur. En düşük performans gösteren algoritma ise %78 doğruluk ile KNN'dir. Ayrıca değişkenlerin önem düzeylerine bakıldığında en yüksek öneme sahip özelliklerin serum creatinine ve ejection fraction olduğu görülmüştür. Erkekler'in kadınlara göre kalp yetmezliğine yakalanma riskleri de daha yüksek çıkmıştır. Bu çalışma, kalp yetmezliği hastalığının erken teşhisine destek olması amacıyla yapılmış olup bu konuda literatüre katkı sunacaktır.

KAYNAKLAR

Aktaş Potur, E., Erginel, N. (2021). Kalp Yetmezliği Hastalarının Sağ Kalımlarının Sınıflandırma Algoritmaları ile Tahmin Edilmesi. *Avrupa Bilim ve Teknoloji Dergisi*. (24), 112-118.

- Bergstra, J. ve Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*. 13(Feb), 281-305.
- Breiman, L. (2001). Random forests. *Machine learning*. 45(1), 5-32.
- Breiman, L., Friedman, J., Stone, C. J., ve Olshen, R. A. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC, New York, USA.
- Chen, T. ve Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 785-794.
- Chen, W., Liu, T., Ying, X., Wei, D. Q., ve Li, Y. (2016). Deep learning in bioinformatics: introduction, application, and perspective in big data era. *Methods*. 93, 3-12.
- Chicco, D. ve Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*. 20(1), 16.
- Cover, T. ve Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*. 13(1), 21-27.
- Davison, A. (2020). Heart Failure Clinical Data. Kaggle. <https://www.kaggle.com/andrewmvd/heart-failure-clinical-data> (Erişim tarihi: 21.10.2022).
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*. 27(8), 861-874.
- Freund, Y. ve Schapire, R. E. (1996). Experiments with a new boosting algorithm. *Proceedings of the Thirteenth International Conference on Machine Learning*. 148-156.
- Galen Lab. CPK Testi. <https://www.galenlab.com.tr/cpk-testi/> (Erişim tarihi: 12.01.2023).
- Goodfellow, I., Bengio, Y. ve Courville, A. (2016). *Deep Learning*. MIT press, Massachusetts, USA.
- Gündoğdu S. (2021). Kalp hastalık risk tahmini için Python aracılığıyla sınıflandırıcı algoritmalarının performans değerlendirmesi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*. 23(69), 1005-1013.
- Hand, D. J., ve Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine learning*, 45(2), 171-186.
- Hosmer Jr, D. W. ve Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley & Sons, New Jersey, USA.
- Huang, G., Liu, Z., Van Der Maaten, L., ve Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700-4708.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... ve Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*. 3146-3154.

- Karanfil, S. (2017). Yapay Öğrenme ile Hastalık Riski Tahmini. (Yüksek Lisans Tezi), İstanbul Kemerburgaz Üniversitesi, Fen Bilimleri Enstitüsü.
- Kartal, E. (2015). Sınıflandırmaya Dayalı Makine Öğrenmesi Teknikleri ve Kardiyolojik Risk Değerlendirmesine İlişkin Bir Uygulama. (Doktora Tezi), İstanbul Üniversitesi, Fen Bilimleri Enstitüsü.
- Medical Park (2022). <https://www.medicalpark.com.tr> (Erişim tarihi: 16.01.2023).
- Memorial (2022). <https://www.memorial.com.tr/> (Erişim tarihi: 03.02.2023).
- Coşar, M. ve Deniz, E. (2021). Makine Öğrenimi Algoritmaları Kullanarak Kalp Hastalıklarının Tespit Edilmesi, Avrupa Bilim ve Teknoloji Dergisi. (28), 1112-1116.
- Özdemir, C. ve Erdil, E. (2021). Kalp Yetersizliği Hastalarının Sağ Kalım Tahmini. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi. 21(1), 91-99.
- Provost, F. ve Fawcett, T. (2001). Robust classification for imprecise environments. Machine learning. 42(3), 203-231.
- Karanfil, S. (2017). Yapay Öğrenme ile Hastalık Riski Tahmini. (Yüksek Lisans Tezi), İstanbul Kemerburgaz Üniversitesi, Fen Bilimleri Enstitüsü.
- Shaikhina, T. ve Khovanova, N. A. (2017). Handling limited datasets with neural networks in medical applications: A small-data approach. Artificial Intelligence in Medicine. 75, 51-63.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. Science. 240(4857), 1285-1293.
- University of California, Irvine. (1999). Heart Disease Data Set. <https://archive.ics.uci.edu/ml/datasets/heart+disease> (Erişim tarihi: 21.10.2022).
- Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., ... ve Liu, Z. (2017). Targeting NEK2 attenuates glioblastoma cell proliferation and radioresistance via inhibiting the ROS/NF-κB signaling pathway. Scientific Reports. 7(1), 1-13.