

GÖĞÜS KANSERİ TEŞHİSİNDE MAKİNE ÖĞRENMESİ UYGULAMALARI

MACHINE LEARNING APPLICATIONS IN THE DIAGNOSIS OF BREAST CANCER

Abdullah ÖZTÜRK*
Bensu GÜNSEL*
Muzaffer TÖNGEL*
Elif YILDIRIM*
Berrin DENİZHAN**

Öz

Göğüs kanserinin erken tanı ve doğru teşhisinde hızlı ve doğru sonuçlar elde edilmesi insan hayatı için kritik öneme sahiptir. Yanlış ve geç tanı ve yanlış tedavinin sonuçları insan hayatının sonra ermesiyle sonuçlanabilmektedir. Bu sebeple bu çalışmada 569 veriye ait göğüs kanseri teşhis veri seti kullanılarak en kötü iç bükey noktalar, çevre en kötüsü, iç bükey noktalar ortalaması, en kötü yarıçap, çevre ortalaması, en kötü alan gibi farklı parametreler göz önünde bulundurularak ve makine öğrenmesi metotlarından Gradient boosting decision tree (GBDT), Multilayer Perceptron Classifier (MLPC) ve Catboost kullanılmıştır. Bu uygulama ile teşhis süresi azaltılmış ve aynı zamanda sonuçların güvenilirliği de artmıştır. Makine öğrenmesi sayesinde meme kanserini etkileyen unsurların birbiriyle olan ilişkileri incelenerek şimdiki ve gelecekteki teşhis ve tanılarda büyük kolaylık sağlanması amaçlanmaktadır. Bu çalışmada kullanılan algoritmalar ile tanı ve teşhis belirlenmesinde en yüksek doğruluk oranı Multilayer Perceptron Classifier (MLPC) algoritmasında %91,95 sonucu ile elde edilmiştir.

Anahtar Kelimeler: Gradient Boosting Decision Tree (GBDT), Multilayer Perceptron Classifier (MLPC), Catboost.

Abstract

Human life must obtain fast and accurate results in the early diagnosis and correct diagnosis of breast cancer. The results of wrong and late diagnosis and wrong treatment in breast cancer can result at the end of human life. For this reason, using the breast cancer diagnosis dataset with 569 data, considering different parameters such as the worst concave points, the worst circumference, the mean of the concave points, the worst radius, the mean of the perimeter, the worst area, and the Gradient boosting decision tree from machine learning methods. (GBDT), Multilayer Perceptron Classifier (MLPC) and Catboost

* Sakarya Üniversitesi Endüstri Mühendisliği Bölümü, Sakarya, Türkiye

** Dr. Öğr. Üyesi, Sakarya Üniversitesi Endüstri Mühendisliği Bölümü, denizhan@sakarya.edu.tr

are used. With this application, the diagnostic time has been reduced, and at the same time, the reliability of the results has increased. It is aimed to provide a great convenience in current and future diagnoses and diagnoses by examining the interrelationships of the factors affecting breast cancer through machine learning. With the algorithms used in this study, the highest accuracy rate in diagnosis was obtained with 91.95% in the Multilayer Perceptron Classifier (MLPC) algorithm.

Keywords: Gradient Boosting Decision Tree (GBDT), Multilayer Perceptron Classifier (MLPC), Catboost.

1. GİRİŞ

Teknolojik ilerlemelerden sonra bilgi ve bilgi yönetimi önemli bir konuma gelmiştir. Bilgiyi işleme yollarının geliştirilmesi sonucunda makine öğrenmesi gibi sistemlerin ortaya çıkmıştır. Günümüzde makine öğrenmesi algoritmaları sağlık sektöründe doktorlara ve çalışanlara hastalık tanı ve teşhisinde büyük ölçüde yardımcı olmaktadır. Bununla birlikte makine öğrenmesi algoritmaları kişilerin sağlık sorunuyla ilgili yapılan testlerin hastanın geçmiş verilerinden de yararlanarak hastalık teşhisine önemli bir katkı sağlamaktadır (Keleş ve Karslı, 2019)

Makine öğrenmesi, son yıllarda birçok alanda başarı ile kullanılmaktadır. Sağlık sektöründe genellikle hastalık teşhis ve tedavi teknolojisinde kullanıldığı görülmektedir. Yapılan çalışmalarda makine öğrenmesi ve veri madenciliğinin önemli bir kolu olarak tanınan makine öğrenmesi Algoritmalarının gizli ilişki açığa çıkarma yeteneğini uygulamaktadır (Keleş ve Karslı, 2019). Bu sayede oluşturulan algoritmalar benzer semptomları olan fakat yüksek belirsizliği olan ve birbirinden ayırt edilemeyen hastalıkların teşhisinde karar verici olabilmektedir. Yapılan çalışmalarda başarı oranının yüksek olduğu görülmektedir (Yokuş ve Çakır, 2012), örneğin Şentürk ve Şentürk (2015) yapay sinir ağlarının (YSA) ve makine öğrenmesinin bu özelliklerini göğüs kanseri tanısında uygulamıştır.

Kanser hastalığı, dünya çapında ölüm istatistiklerine yüksek bir orana sahiptir (Global Burden of Disease Cancer Collaboration, 2017). Dünyada her yıl ortalama 14 milyon kanser teşhisi konulmaktadır. 2018 yılında kanser hastalığı nedeniyle hayatını kaybedenlerin sayısı 9,6 milyondur (HSAE, 2019) Göğüs kanseri göğüs hücresinde başlayan kanserin bir türüdür. Göğüs kanseri kadınlardaki göğüs dokusundaki süt kanalını oluşturan ve süt yapıcı hücrelerin kontrolsüz olarak çoğalmasıyla ortaya çıkmaktadır (Siyah, 2012). Göğüs kanserine sebep olan tümör iyi veya kötü huylu olarak adlandırılmaktadır (Tuncer, 2009). Kötü huylu tümörler Komşu organ ve dokulara yayılmaktadırlar bunun yanı sıra sınırları belirsizdir (Sezer ve ark., 2005)

Göğüs kanserinin tedavisini diğer kanser türlerinin tedavisinden ayıran en önemli detaylardan biri kanserin diğer organlara yayılmadan tedavi ve teşhisidir. Yani göğüs kanserinde doğru ve erken teşhis çok önemlidir. Bunun yanı sıra meme kanserinin temelini oluşturan genetik faktörlerin ortaya çıkarılması tedavi ve teşhiste önemli rol oynamaktadır (Segal ve ark., 2003). – (Yıldız ve ark., 2012) Dünyada meme kanseri kadınlar arasında en yaygın olarak görülen kanser türüdür. Gün geçtikçe sayısı artan bu kanser türüne Batı ülkelerindeki her 8-9 kadından birinde rastlanmaktadır. (Aydıntuğ, 2014) 2020'de en yaygın olan kanser türü 2,26 milyon vaka ile göğüs kanseridir ve göğüs kanseri sebebiyle 685 bin ölüm gerçekleşmiştir.

Günümüzde makine öğrenmesi kanser teşhislerinde önemli bir rol oynamaktadır. Makine öğrenmesi kullanan sistemler geçmiş hastalık verilerini kullanarak oluşturulabilmektedir. Oluşturulmuş sınıflandırma algoritmaları daha sonrasında gerekli testleri yaptıran kişilerin sonuçlarının analizinde doktorlara büyük kolaylık sağlamaktadır.

Bu çalışmada göğüs kanserine ait 569 veriye ait göğüs kanseri teşhis veri seti üzerinde Multi-layer Perceptron Classifier, Gradient Boosting Machines ve Catboost algoritmaları kullanılarak sınıflandırma yapılmıştır. Makine öğrenmesi sayesinde göğüs kanserini etkileyen unsurların birbiriyle olan ilişkileri incelenerek şimdiki ve gelecekteki teşhis ve tanılarda büyük kolaylık sağlanması amaçlanmaktadır.

2. YÖNTEM

2.1. Makine Öğrenmesi

Makine öğrenmesi çeşitli istatistiksel ve optimizasyonel teknikleri bulundurup, eldeki verilerden öğrenerek büyük ve karmaşık veri setleri içerisindeki ilişkileri hızlı bir şekilde tespit etmeye yarar (Belciug ve ark., 2010). Oluşturulan algoritmalar ile veri seti en iyi performans elde edilecek şekilde kurulmaktadır. Bu sebeple kullanılan makine öğrenmesi algoritmalarından bazıları: K en yakın komşu algoritması, destek vektör makineleri ve yapay sinir ağlarıdır. Bu algoritmaların tahmin, kümeleme veya sınıflandırma yapmaktadır (Selvi, 2019). Şekil 1'de görüldüğü gibi Makine öğrenmesi iki temel öğrenme metodu üzerine kurulmuştur ve bunlar gözetimli ve gözetimsiz öğrenmedir. Gözetimli öğrenmenin temel çalışma prensibi, sonuçları bilinen verilerin kullanılıp sonuçları bilinmeyen verilerin tahmin edilmesidir. Algoritmaya örnek olarak; en yakın komşuluk algoritması ve karar ağaçları verilebilir. Gözetimsiz öğrenmede ise verilerin detaylı özellikleri bilinmektedir sonrasında ise veriler gruplandırılıp etiketlenmektedir. En çok kullanılan gözetimsiz öğrenme algoritmaları kümeleme algoritması ve temel bileşen analizidir. Göğüs Kanseri Teşhisinde Farklı Makine Öğrenmesi Tekniklerinin Performans Karşılaştırılması çalışmasında; farklı makine öğrenmesi algoritmaları kullanılarak gerekli sınıflandırma çalışmaları yapılmış olup en yüksek doğruluk oranı %98,24 ile Lojistik regresyondan alınmıştır.



Şekil 1. Makine Öğrenmesi Türleri

Makine öğrenme algoritmaları kullanan meme kanseri tanılarında çeşitli yapay sinir ağları üzerinden farklı teknikler uygulanmıştır. Bunlardan birkaçı MLP, ileri besleme geri yayılma ve genetik algoritma teknikleridir. Bu çalışmalar arasında en iyi başarıyı veren yapay sinir ağı algoritması ise %99,28 ile MLP'dir (Paulin ve Santhakumaran, 2019). Makine öğrenmesi ile meme kanseri sınıflandırması çalışmasında çeşitli sınıflandırılma metotları kullanılmıştır ve bunlar Destek Vektör Makineleri, Fuzzy Cognitive Map, Bulanık-genetik hibrit algoritma, Lojistik regresyon ve Euclidian tabanlı centroidir (ETC). Bu çalışmalar sonucunda elde edilen en yüksek başarı oranı %99,04 ile ETC'dir (Kör, 2019). Al-Shargabi ve arkadaşları ile yapmış olduğu 'Göğüs Kanseri İçin Çok Katmanlı Algılayıcıyı Geliştirme ve Kanseri Tahmini' çalışmasında Wisconsin Diagnostic Breast Cancer (WDBC) veri kümesi kullanılarak Multi Layer algoritmasıyla deneysel sonuçlar için %97,70 doğruluk skoruna ulaşmıştır (Chakravarthy ve Rajaguru, 2019) Sevil'in 2019 yılında yapmış olduğu 'Göğüs Kanseri Teşhisinde Farklı Makine Öğrenmesi Tekniklerinin Performans Karşılaştırması' çalışmasında Destek Vektör Makineleri Naïve Bayes., rastgele orman, K-en yakın ve lojistik regresyon algoritmalarıyla yapmış olduğu çalışmada Lojistik Regresyon ile %98.24 doğruluk skoru elde etmiştir (Sevil, 2019) Derangula arkadaşları Profesör Edara ve Karri'nin yapmış olduğu 'Gradyan Artırma Kullanarak Meme Kanseri Verilerinin Özellik Seçimi ve Makine Öğrenimi Teknikleri'nin Uygulanması' çalışmasında Light Gradient Boosting Model (LGBM), Catboost ,Extremegradient boosting (XGB) ve Naïve Bayes algoritmalarıyla öznetelik seçiminden önce %94; öznetelik seçiminden sonra ise doğruluk skor %96 olmuştur (Derangula ve Edara, 2020). İbtahim ve arkadaşlarının 2015'te Uluslararası Hesaplama, Kontrol, Ağ Oluşturma, Elektronik ve Gömülü Sistem Mühendisliği Konferansı'nda yayınlanan çalışmalarında Göğüs Kanseri'nin Sınıflandırılmasında Çok Amaçlı Diferansiyel ile MLP kullanılarak %97,51 ortalama doğruluk skoruna ulaşılmıştır (İbrahim ve ark., 2020). Tabrizchi ve Tabrizchi'nin 2020 Mart ayında yayınlanmış olan Çok Aşamalı Optimize Edici Tabanlı Gradyan Kullanarak Meme Kanseri Teşhisi ve Karar Ağacını Güçlendirme çalışmasında Wisconsin Diagnostic Breast Cancer veri setinden yararlanılmıştır. Gradient boosting decision tree ve (GBDT) Multi-verse optimizer (MVO) algoritmalarını birlikte kullanarak %98.76 ortalama doğruluk skoruna ulaşmıştır (Tabrizchi ve Tabrizchi, 2020).

3. BULGULAR

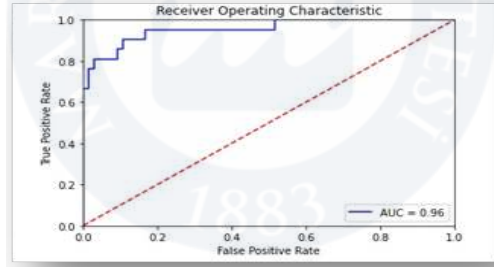
Çalışmaya ait bulgular MLPClassifier, catboost ve gradient boosting machine çözümleri ile sunulacaktır.

3.1. MLPClassifier

Mlpclassifier kütüphaneden eklendikten sonra daha önceden ayrılan eğitim verileri ile varsayılan parametreleri ile eğitilir. Bu eğitim sonucunda elde edilen doğruluk skoru %93'tür. Yeni model oluşturulur ve hiperparametreler belirlenir. Gizli katmanlarda modelde kullanılacak nöron sayılarını ve katmanları göstermektedir. Örneğin ilk gizli katmanda 3 katmanlı 100 nöron olduğu görülmektedir. Aktivasyon fonksiyonu gizli katmanlardaki nöronların aktivasyonunu göstermektedir. Hiperparametreler modele eklendikten sonra 10 katlı çapraz doğrulama ile oluşturulan modelde 240

iterasyon sonrasında eğitim verileri ile oluşturulan model %96,8 doğruluk skoruna ulaşmıştır. Test verileri üzerinde çalışma yapmadan önce, hiperparametreler ile çalışan modelin en iyi sonucu veren en iyi parametreleri fonksiyon yardımıyla çağrılır ve tekrardan bu parametrelerle eğitim sağlanır. Bu çalışma sonrasında modelin doğruluk skoru %93,1 seviyesindedir. Karmaşıklık matrisi oluşturulur tahmin ve gerçek değerler arasındaki sonuç gözlemlenir. Çıkan sonuçların analiz edilmesi için sınıflandırma raporu oluşturulduğunda sonuçlar daha net görülmektedir. F1 score sonuçları; ağırlıklı ortalamayla bulunmuş olup sırasıyla bağımlı değişkenler için 0,96 ve 0,84 olarak bulunmuştur. Recall (geri çağırma): Burada pozitif olarak tahmin etmemiz gereken değerlerin sırasıyla 0,98 ve 0,76 olarak doğru sonuçlandığı görülmektedir. Precision (keskinlik): Bu sütunda pozitif olarak tahminleildiğinde değerlerin gerçekte kaç adetinin pozitif olduğu görülmek üzere sırasıyla; 0,93, 0,94'tür.

Macro average ise; precision, recall, F1 score değerlerinin ortalamasıyla bulunmuş olup sırasıyla 0,93, 0,93 ve 0,93'tür. Weighted average ise; precision, recall, F1 score değerlerinin ağırlıklandırılmasıyla bulunmuş olup sırasıyla 0,93, 0,93 ve 0,93'tür.

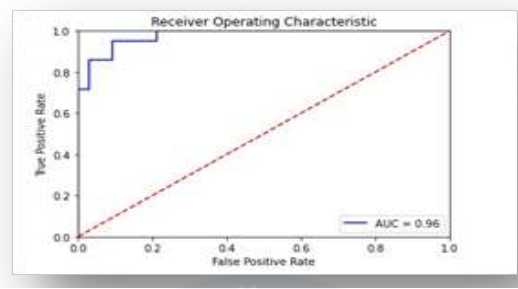


Şekil 1. MLPClasfıer Gerçek Pozitif – Yanlış Pozitif Grafiği

Şekil 2'ye bakıldığında 0.8'in üzerine hızlıca çıkan iki sınıfın (class = 0, class = 1) doğruluk oranlarının yüksek ve birbirlerine paralel olduğu söylenebilir. Aynı zamanda sınıfların geçtiği eğrinin altında yer alan bölüm AUC bölgesi yüksek yer kapladığından model başarısının yüksek olduğunu görülebilir. Yani yapılan bu çalışmada AUC Değerine bakarak modelimizdeki sınıfların birbirinden ne kadar iyi ayrıldığını görülmektedir.

3.2. Catboost

Catboost kütüphaneden eklendikten sonra eğitim verileri ile herhangi bir hiperparametre ayarı yapılmadan varsayılan parametreleri ile model çalıştırılır. Varsayılan parametrelerle oluşturulan modelin doğruluk skoru %91,95 olarak çıkmıştır. Modelin öğrenme hızı ve modelde oluşturulacak düğüm sayıları belirlenmiştir. 10 katmanlı çapraz doğrulama ile oluşturulan eğitim modelinin doğruluk skoru %93,92 olarak çıkmıştır. Bu çalışma sonrasında modelin doğruluk skoru %90,8 seviyesindedir. Karmaşıklık matrisi oluşturulur şekil30'da tahmin ve gerçek değerler arasındaki sonuç gözlemlenir. F1 score sonuçları; ağırlıklı ortalamayla bulunmuş olup sırasıyla bağımlı değişkenler için 0.94 ve 0.79 olarak bulunmuştur.

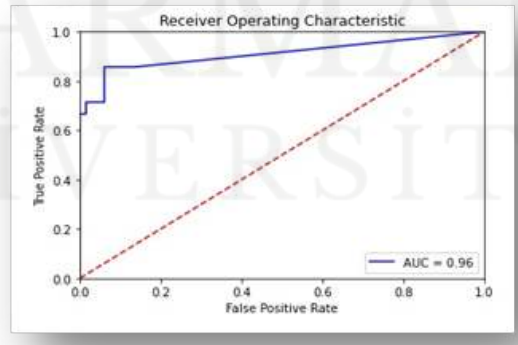


Şekil 2. Catboost Gerçek Pozitif – Yanlış Pozitif Grafiği

Şekil 3'teki grafiğe bakıldığında 0.8'in üzerine hızlıca çıkan iki sınıfın (class = 0, class = 1) doğruluk oranlarının yüksek ve birbirlerine paralel olduğu söylenebilir. Aynı zamanda sınıfların geçtiği eğrinin altında yer alan bölüm AUC bölgesi yüksek yer kapladığından model başarısının yüksek olduğunu görülebilir. Yani yapılan bu çalışmada AUC Değerine bakarak modelimizdeki sınıfların birbirinden ne kadar iyi ayrıldığını görülmektedir.

3.3. Gradient Boosting Machine (Gbm) Çözümü

Gradient boosting machine çözümünde varsayılan parametrelerle oluşturulan modelin doğruluk skoru %90,8 olarak bulunmuştur. 10 katmanlı çapraz doğrulama ile oluşturulan eğitim modelinin doğruluk skoru %92,78 olarak çıkmıştır. Bu çalışma sonrasında modelin doğruluk skoru %90,8 seviyesindedir. Çıkan sonuçların analiz edilmesi için Şekil 36'daki sınıflandırma raporu oluşturulduğunda sonuçlar daha net görülmektedir. F1 score sonuçları; ağırlıklı ortalamayla bulunmuş olup sırasıyla bağımlı değişkenler için 0.94 ve 0.79 olarak bulunmuştur.



Şekil 3. Gbm Gerçek Pozitif – Yanlış Pozitif Grafiği

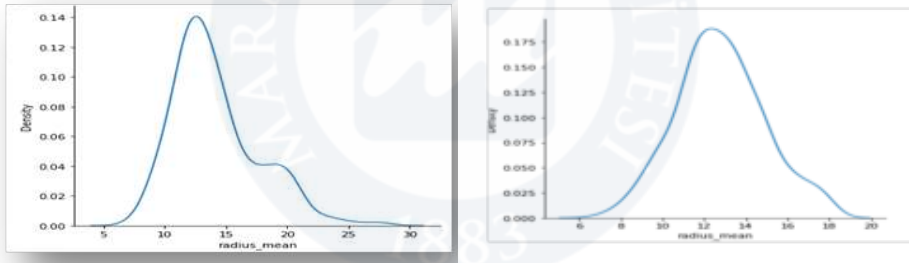
Şekil 4'dteki grafiğe bakıldığında 0.8'in üzerine diğer modeller kadar hızlı çıkılamadığı görülmüştür (class = 0, class = 1) doğruluk oranlarının yüksek ve birbirlerine paralel olduğunu söylenebilir. Aynı

zamanda sınıfların geçtiği eğrinin altında yer alan bölüm AUC bölgesi diğerlerinden az olsa da yüksek yer kapladığından model başarısının yüksek olduğunu görülebilir. Yani yapılan bu çalışmada AUC Değerine bakarak modelimizdeki sınıfların zamanla birbirinden ne kadar iyi ayrıldığını görülmektedir.

İstatistiksel Analiz

Bu çalışmada göğüs kanseri teşhisinde makine öğrenmesinin etkileri araştırılmıştır. Çalışmada 569 adet veri kullanılmış ve sonuç olarak yapılan analizler sonucunda öz niteliklerin kendi içlerinde aykırılıklar tespit edilmiştir.

Şekil 4'e bakıldığında verilerin sola basık bir şekilde dağıldığı görülmektedir. Bu öz niteliğin standart sapmasına bakıldığında ise standart sapma 3,52 ortalaması 14,12 çıkmıştır.



Şekil 4 – 5. Radius ortalama aykırı veri analizi öncesi ve sonrası

Aykırı veri analizi için bu öz niteliğe özel Q1 ve Q3 değerleri hesaplanmış ve çeyrekler açıklığı yardımıyla yeni alt ve üst sınırlar belirlenmiş olup bu aralık dışındaki değerler aykırı veri (outlier) olarak tanımlanmış ve veri setinden çıkartılmıştır. Standart sapma 3σ seviyesinden Şekil 5'te 2σ seviyesine indirgenmiştir bu durum dağılımın sivrilüğünden de anlaşılmaktadır. Aykırı veri analizi sonrası standart sapma 2,12 ve ortalama 12,78 değerine çekilmiştir.

Tablo 1. Algoritmaların R^2 ve MSE Değerlerinin karşılaştırılması

MODEL	R^2	Ortalama Kare Hatası (MSE)
mlpc	0,6233	0,0919
mlpc_tuned	0,4978	0,0689
catboost	0,5606	0,0804
catboost_tuned	0,4978	0,0919
gbm	0,4978	0,0919
gbm_tuned	0,4978	0,0919

Tablo 1'de modellerin hiperparemetre ve hiperparametre olmadan R^2 ve Ortalama Kare Hatası (MSE) değerleri listelenmiştir. Çalışma, insanlar üzerinde yapıldığı için R^2 değeri düşük çıkmıştır.

Genel olarak insan davranışları ve insan doğası gibi fiziksel olmayan süreçlerde R^2 değeri yaklaşık 0,5 civarlarında olmaktadır.

Veri setinden aykırı verileri çıkarıldığında için mutlak hata (MAE) kullanma zorunluluğumuz ortadan kalmıştır. Ortalama Kare Hatasına bakıldığında MLPC algoritmasından en az olduğu görülmektedir. Ortalama Kare Hatası da 0'a yakın olmak gerektiğini 0'a ne kadar yakınsa modelin o kadar iyi olduğunu söyler. Model kendi normal parametreleri ile çalıştığında Ortalama Kare Hatası daha yüksek çıkmıştır. Catboost algoritmasında ise kendi parametreleri ile oluşturulan modelin Ortalama Kare Hatası hiper parametrelili modelden daha az çıkmıştır.

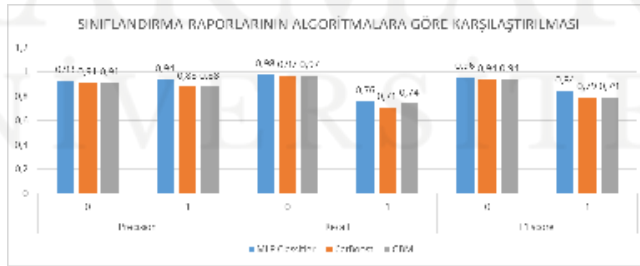
Tablo 2. Doğruluk Skoru

	Doğruluk Skorları
MLP Classifier	%93,1
CatBoost	%90,8
GBM	%90,8

Tablo 2'de ise yapılan çalışmada sınıfların ne kadar iyi ayrıldığını gösterilmektedir. Bu değerlerimiz yüksek çıkmakla beraber iki sınıf arasındaki ayrımın çok kuvvetli yapıldığını göstermektedir. Eğer değer 0,5'e yakın olsaydı kurulan modelin rastgele tahminleme yaptığını ve performansının kötü olduğu anlaşılırdı fakat MLPC modelinde skorun 0,87 olduğu ve sınıflandırmayı kuvvetli doğrulukla yaptığı anlaşılmaktadır.

Tablo 3. Algoritmaların Sınıflandırma Raporu

	Precision		Recall		F1 score		Support	
	0	1	0	1	0	1	0	1
MLP Classifier	0.93	0.94	0.98	0.76	0.96	0.84	66	21
CatBoost	0.91	0.88	0.97	0.71	0.94	0.79	66	21
GBM	0.91	0.88	0.97	0.74	0.94	0.79	66	21



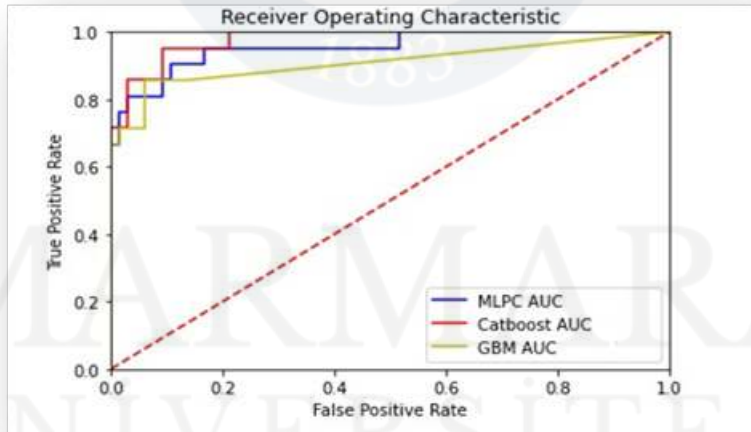
Şekil 6. Sınıflandırma raporlarının algoritmalara göre karşılaştırılması

Tablo 3 ve Şekil 6'daki karşılaştırmada ise modellerin sınıflara (0-1) göre duyarlılık f1 skoru ve hassasiyet analizleri bulunmaktadır. Test veri setinde 0 sınıfı için 66, 1 sınıfı için 21 veri bulunmaktadır. Burada bize hesaplanma yönteminden dolayı en iyi sonucu verecek olan değer F1 skorudur. Burada

MLPC algoritmasının diğerlerine göre daha iyi sınıflandırma yaptığı görülmektedir. Dikkat edilmesi gereken bir diğer nokta ise bulunan sınıfların doğruluk oranıdır, bütün algoritmalar 0 sınıfı için yapılan çalışmalarda yüksek doğruluk oranları tahmin etse de 1 sınıfı için olan sınıflamada GBM ve Catboost düşük doğruluk oranlarına sahiptir. Yani dışarıdan bir kişiye ait test sonuçları bu algoritmalara işlendiğinde, tahminlemesini sağlandığında ve eğer bu kişi 1. Sınıfa dahilse bu iki modelin bunu doğru tahminlemesi zorlaşmaktadır. Fakat MLPC algoritmasında bu sınıf için de doğruluk skoru yine diğerlerine göre oldukça yüksektir.

4. TARTIŞMA ve SONUÇ

Karşılaştırma grafiğinde yapılan makine öğrenmesi çalışmalarının sonucu şekil 7'de bulunmaktadır. En iyi doğruluk ve hassasiyete sahip olan algoritma bu verilerle yapılan çalışma için MLPC sınıflandırma algoritması olmuştur. Bu çalışmanın temel amacı göğüs kanseri teşhisinde sürelerin azaltılması ve doğruluğun artırılmasıdır. MLPC algoritması ile yapılan uygulama sonuçları mevcut bir vakanın doğru sınıflandırma oranını %93,1 çıkarmıştır. Şekil 7'deki grafikte en az AUC alanına sahip olan algoritma GBM olmuştur yani bu modeldeki sınıflar arasındaki ayrım diğerlerine göre yeterince keskin değildir.



Şekil 7. Karşılaştırma grafiği

Yapılan çalışmada sağlık sektörü ele alındığında yapılan çalışmanın sürdürülebilir olması ön plandadır. Bu anlamda sürekli yapılacak düzenlemeler ve iyileştirmeler sistemin doğruluk ve güvenilirlik oranını yüksek tutacaktır. Bu çalışmanın sağlık alanında özellikle göğüs kanseri tanı ve teşhisinde önemli bir rol alacağı ön görülmüştür. Göğüs kanserinde hızlı ve doğru teşhisin ne derecede önemli olduğu bilinmektedir. Veri analizi ve makine öğrenmesi ile tanı ve teşhisinde ölçülen parametrelerle ilgili analizler yapılmış olup hangi değerlerin daha önemli olduğu belirlenmiştir.

KAYNAKLAR

- Aydıntuğ S., “Meme Kanseri Erken Tanı”, Sürekli Tıp Eğitimi Dergisi, Cilt. 13, Sayı. 6, ss 228, 2014.
- Belciug, S., Salem, A. B., Gorunescu, F., Gorunescu, M., Clustering-based approach for detecting breast cancer recurrence, In Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on, 533-538, 2010.
- Chakravarthy S. and Rajaguru H, “Ensemble Flower-Pollination Algorithm for Automatic Breast Cancer Classification using Haar Features, Department of Electronics and Communication Engineering, 2021.
- Derangula A. and Edara S. “Feature Selection of Breast Cancer Data Using Gradient Boosting Techniques of Machine Learning”, Department of Computer Science and Engineering, 2020.
- E. Segal, H. Wang ve D. Koller, “Discovering molecular pathways from protein interaction and gene expression data”, Bioinformatics, Cilt 19, 264-272, 2003.
- Halk Sağlığı Anabilim Dalı Ekibi (HSAE),” Dünya kanser günü”, Şubat 2019, URL Adresi: <https://ankara.baskenthastaneleri.com/uploads/files/halk-sagligi-dunya-kanser-gunu.pdf> (Erişim Tarihi: 10.05.2021)
- Keleş A., Karşlı Ö.B. “Makine öğrenme yöntemleri ile karaciğer hastalığının teşhisi”, Ağrı, 2019Yokuş B, Çakır D.Ü. “Kanser Biyokimyası”, Dicle Üniv Vet Fak Derg: 1(2): 7-18, 2012.
- Kör H., Classification of Breast Cancer by Machine Learning Methods”, Department of Computer Technology, Sungurlu Vocational School, University of Hitit, Corum,2019.
- Paulin A. Santhakumaran, Classification of Breast cancer by comparing Back propagation training algorithms”, International Journal on Computer Science and Engineering, India,2011.
- Sevli, Onur, “Göğüs Kanseri Teşhisinde Farklı Makine Öğrenmesi Tekniklerinin Performans Karşılaştırması”, Avrupa Bilim ve Teknoloji Dergisi, (16), 176-185, 2019.
- Şentürk Z.K. ve Şentürk A. “Yapay Sinir Ağları İle Göğüs Kanseri Tahmini”, Mühendislik Fakültesi, Bilgisayar Müh. Bölümü, Düzce Üniversitesi, 2015.
- Sevli O.,” Göğüs Kanseri Teşhisinde Farklı Makine Öğrenmesi Tekniklerinin Performans Karşılaştırması” Mehmet Akif Ersoy Üniversitesi, Mühendislik Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, Burdur, Mayıs 2019.
- Sezer, O.G., Erçil, A., Keskinöz, M. “Destek Vektör Makinesi Kullanarak Bağımsız Bileşen Tabanlı 3B Nesne Tanıma (in Turkish)”, Mühendislik ve Doğa Bilimleri Fakültesi, Sabancı Üniversitesi, SIU 2005.
- Siyah B. “Destek Vektör Makineleri ve Çok Katmanlı Algılayıcılar ile Göğüs Kanseri Teşhisi”, MKÜ Bilgisayar Mühendisliği Bölümü, 2012.
- Tabrizchi H. and Tabrizchi M., “Breast cancer diagnosis using a multi-verse optimizer-based gradient boosting decision tree”, 2020.
- Tuncer M. “Türkiye’de Kanser Kontrolü”. Koza Matbaacılık. Ankara 2009.
- Yıldız, O., Tez, M., Bilge, H. Ş., Akcayol, M. A., & Güler, İ., “Meme Kanseri Sınıflandırması İçin Veri Füzyonu Ve Genetik Algoritma Tabanlı Gen Seçimi”, Journal of the Faculty of Engineering & Architecture of Gazi University, 27(3), 2012.