

Karaciğer Yetmezliği Teşhisinde Özellik Seçimi Kullanarak Makine Öğrenmesi Yöntemlerinin Başarılarının Ölçülmesi

Emre SAYGIN^{1*}, Muhammet BAYKARA²

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

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

*¹ emresaygin@hotmail.com.tr, ² mbaykara@firat.edu.tr

(Geliş/Received: 27/11/2020;

Kabul/Accepted: 21/03/2021)

Öz: Karaciğer, vücudumuzda pek çok fizyolojik ve metabolik süreçte hayati görevler üstlenir. Karaciğer eksikliği ya da karaciğerin işlevsel bozuklukları sağlık sorunlarına ve nihayetinde hastalıklara yol açmaktadır. Karaciğer hastalıklarının erken dönemde teşhisi, gereksiz ilaç kullanımını azaltarak hem tedavi süresinin kısaltılmasında hem de karaciğer tahribatının engellenmesinde etkili olabilmektedir. Gelişen teknoloji ile beraber tıp dünyasında makine öğrenmesi yöntemleri hastalıkların teşhisinde büyük bir başarı sağlamaktadır. Bu çalışmada, karaciğer hastalıklarının erken teşhis edilmesinde kullanılması için sarmal (wrapper) özellik seçimi yöntemlerinden İleri Yönde Seçim (SFS) ile veri setinde bulunan en önemli öz nitelikler belirlenmiştir. Belirlenen önemli öz nitelikler üzerinde makine öğrenmesi yöntemleri uygulanarak karaciğer yetmezliği hastalığı teşhis başarıları ölçülmüştür. Makine öğrenmesi yöntemlerinde model iyileştirilmesi yapılarak hastalık teşhis başarıları artırılmıştır. Elde edilen sonuçlar literatürde bu veri seti ile yapılan diğer çalışmalar ile karşılaştırılmıştır. Indian Liver Patient Dataset (ILPD) üzerinde karaciğer yetmezliği hastalığının tahmini için; Destek Vektör Makinası (SVM), Çok Katmanlı Algılayıcı (MLP), Karar Ağaçları (DT), Lojistik Regresyon (LR) ve Hafif Gradyan Güçlendirme Makinesi Sınıflandırıcısı (LGBM) yöntemleri kullanılmıştır. Kullanılan yöntemlerde en yüksek teşhis başarıları sırasıyla; LGBM %82.12, MLP %81.13, DT %81.13, SVM %77.87 ve LR %77.80 olarak elde edilmiştir. Sonuçlar literatürdeki çalışmalarla kıyas edildiğinde tatmin edici olarak değerlendirilmiştir. Elde edilen modellerin karaciğer hastalığının erken teşhisinde hekimlere destek olabileceği düşünülmektedir.

Anahtar kelimeler: Makine öğrenmesi, karaciğer yetmezliği teşhisi, sınıflandırma, ILPD.

Measuring the Success of Machine Learning Methods Using Feature Selection in Diagnosis of Liver Failure

Abstract: The liver plays vital roles in many physiological and metabolic processes in our body. Liver deficiency or functional disorders of the liver lead to health problems and ultimately disease. Early diagnosis of liver diseases can be effective in shortening the treatment period and preventing liver damage by reducing unnecessary drug use. With the developing technology, machine learning methods in the medical world provide great success in diagnosing diseases. In this study, for detection of liver failure disease the most important attributes of the data set were determined with Forward Selection (SFS), one of the wrapper feature selection methods, to be used in the early diagnosis of liver diseases. The success of disease diagnosis has been increased by improving the model in machine learning methods. The results obtained were compared with other studies conducted with this data set in the literature. Diagnostic success of liver failure disease was measured by applying machine learning methods on the on Indian Liver Patient Dataset (ILPD); Support Vector Machine (SVM), Multilayer Perception (MLP), Decision Trees (DT), Logistic Regression (LR) and Light Gradient Reinforcement Machine Classifier (LGBM) methods were used. The highest diagnostic success in the methods used, respectively; LGBM was 82.12%, MLP 81.13%, DT 81.13%, SVM 77.87% and LR 77.80%. The results were considered satisfactory when compared with the studies in the literature. It is thought that the models obtained can support physicians in the early diagnosis of liver disease.

Key words: Machine learning, diagnosis of liver failure, classification, ILPD.

1. Giriş

Karaciğer diğer organlara göre daha fazla görev yapan, hacim olarak en büyük ve en önemli organımızdır. Karaciğer hastalıklarının teşhis edilmesinde karaciğer enzimleri büyük bir öneme sahiptir. Karaciğer denildiğinde akla ilk gelen enzimler Alanin Aminotransferaz (SGPT) (=ALT) ve Aspartat Aminotransferaz (SGOT)

* Sorumlu yazar: emresaygin@hotmail.com.tr. Yazarların ORCID Numarası: ¹ 0000-0001-7462-2011, ² 0000-0001-5223-1343

(=AST)'dır. Sigara kullanımı ve alkol tüketimi gibi alışkanlıklar karaciğer hastalıklarını tetikleyerek bu enzimlerin yükselmesinde rol oynamaktadır [1].

Diyabet, Hepatit B, Hepatit C gibi bazı hastalıklar karaciğerde hasar oluşturarak karaciğer yetmezliğine yol açar. Karaciğer hasarına yol açan hastalıklar tedavi edilmediği takdirde karaciğer yapısı zamanla bozulur ve karaciğer zamanla işlevini yerine getirememeye başlar. Aşırı derecede hasar alan ve doğal yapısı bozulan karaciğerin tedavisi mümkün olmadığından bu durumda organ nakline ihtiyaç duyulmakta veya ölümler görülebilmektedir [2].

Makineleşmenin ve teknolojinin gelişmesiyle makine öğrenmesi birçok alanda kullanılmaktadır. Makine öğrenmesi yöntemleri aracılığıyla, dijital ortamlarda depolanan anlamlı veya anlamsız veriler bir bütün olarak ele alındığında anlamlı sonuçlar üretilebilmektedir. Makine öğrenmesi yöntemleri, bu veriler üzerinde istatistiksel ve matematiksel yöntemler kullanarak bir sayısal öğrenme üzerinden model oluşturur. Makine öğrenmesi yöntemleri oluşturulan model üzerinden bir tahminde bulunur. Tahmin başarısı oluşturulan modelle doğrudan ilişkilidir.

Veri setlerinde birden fazla öznelik bulunmaktadır. Özneliklerin uyumu, sınıflandırma algoritmalarının modelleme süreçlerinde çok önemli bir yere sahiptir. Bazı durumlarda gereksiz öznelikler modelin performansına bir katkı sağlamazken bazen de modelin performansını düşürmektedir. Öznelik seçme yöntemleri ile veri seti içinde en yararlı öznelikler seçilerek makine öğrenmesi sınıflandırma modelleri oluşturmada kullanılmaktadır.

Bu çalışmada ILPD veri seti üzerinde özellik seçimi için, sarmalayıcı (wrapper) yöntemlerinden olan ileri doğru seçim yöntemi kullanılmıştır. Özellik seçimi sonrasında elde edilen en önemli öznelikler kullanılarak, karaciğer yetmezliği hastalığı teşhisi için SVM, MLP, DT, LR ve LGBM makine öğrenmesi yöntemleri ile teşhis başarıları ölçülmüştür. Teşhis başarısı, makine öğrenmesi yöntemleri ile kurulan modellerde kullanılan hiper parametreler ile artırılmıştır.

Bu çalışma, kısaca dört aşamada karaciğer yetmezliği hastalığının tahmin doğruluğunu iyileştirmek için hibrid bir model ve karşılaştırılmalı bir analiz sunmaktadır. Birinci aşamada sınıflandırma algoritmaları veri setine uygulanır. İkinci aşamada özellik seçimi ile veri setinden önemli öz nitelikler seçilir ve bu önemli öz nitelikler kullanılarak veri setinin bir alt kümesi oluşturulur. Oluşturulan alt kümeye sınıflandırma algoritmaları uygulanır. Üçüncü aşamada, ikinci aşamada uygulanan sınıflandırma algoritmalarında model iyileştirilmesi yapılarak modeller yeniden kurulur, doğruluk performansları artırılır ve ilk üç aşamanın karaciğer hastalığı tahmin doğruluğu performansları karşılaştırılır. Dördüncü aşamada ise model iyileştirilmesi uygulanmış sınıflandırma algoritmalarının doğruluk performans sonuçları ile literatürde bu çalışmada kullanılan veri seti ile yapılan çalışmaların sonuçları karşılaştırılmaktadır.

Çalışmanın ikinci bölümünde bu alanda yapılan ilgili çalışmalar, üçüncü bölümünde veri seti ve makine öğrenmesi yöntemlerinin nasıl kullanıldığı ve model iyileştirilmesi hakkında detaylı bilgiler verilmiştir. Dördüncü bölümde ise bu çalışmada kullanılan veri seti kullanılarak yapılan çalışmaların sonuçları ile bu çalışmanın sonuçları karşılaştırılmıştır.

2. İlgili Çalışmalar

Karaciğer hastalıkları konusunda birçok veri seti üzerinde çalışmalar yapılmıştır. Bu kısımda konu ile ilgili literatürde öne çıkan bazı önemli çalışmalar özetlenmiştir.

Mamdouh ve Mabrouk [2] Diabetes, BUPA ve ILPD veri setleri üzerinde diyabet ve kronik karaciğer hastalığı teşhisinde SVM yöntemini, MATLAB çalışma ortamında modellemişlerdir. BUPA veri setinin ilk 4 özneliği ile yapılan teşhiste %63, ilk 6 öznelik ile %70, ilk 8 öznelik ile %70 oranında başarı göstermişlerdir. ILPD veri seti üzerinde ise veri setinin ilk 4 özneliği ile yapılan teşhiste %71, ilk 6 öznelik ile %73, ilk 8 öznelik ile %73.2 oranında başarı gösterdiğini belirtmişlerdir.

Gulia ve arkadaşları [3] ILPD veri seti üzerinde karaciğer hastalığı teşhisi için J48, MLP, SVM, Random Forest (RF) ve Bayes Net makine öğrenmesi yöntemlerini kullanmışlardır. Birinci aşamada veri seti üzerinde öznelik seçmeden, veri setinin orijinal haliyle teşhis başarı ölçümleri yapmışlardır. Birinci aşamada en yüksek teşhis başarısını %71.35 oran ile SVM sınıflandırma yönteminden elde etmişlerdir. Diğer algoritmaların sırasıyla RF %70.32, J48 %68.78, MLP %68.26 ve Bayes Net %67.23 başarı gösterdiğini belirtmişlerdir. İkinci aşamada, veri setinde Greedy Stepwise (Açgözlü algoritması) özellik seçimi kullanılması ile önemli özellikleri içeren öz nitelikleri belirlemişlerdir. Önemli öz niteliklerin *Total Bilirubin*, *Direct Bilirubin*, *Total Protiens*, *Albumin* ve *A/G oranı* olduğunu ifade etmişlerdir. İkinci aşamada RF algoritmasının doğruluğu %71,86 oran ile tespit edilmiştir. Diğer yöntemler sırasıyla SVM %71.35, MLP %70.84, J48 %70.66 ve Bayes Net %69.12 oranında başarı göstermişlerdir.

Alkuşak ve Gök [4] BUPA ve ILPD veri setleri üzerinde karaciğer yetmezliği teşhisi için YSA, RF, Functional Tree ve Radial Based Functional makine öğrenmesi yöntemlerini kullanmışlardır. BUPA verilerini %86.95'i eğitim (300 örnek), %13.05 'ini (45 örnek) test olarak kullanmışlardır. ILPD veri seti için %87.48'i (510 örnek) eğitim, %12.52'sini (73 örnek) test için kullanmışlardır. YSA'yı MATLAB ortamında, diğer yöntemleri WEKA aracında test etmişlerdir. Analizlerde WEKA'da kullandıkları yöntemler için 10-kat çapraz doğrulama test tekniği ve YSA için 10 gizli katman ve ileri beslemeli nöron ağı kullanmışlardır. Bu çalışma sonucunda en yüksek teşhis başarısı YSA ile BUPA veri seti için %76, ILPD veri seti için %78 olmuştur.

Priya ve arkadaşları [5] ILPD veri setini iki farklı aşamada değerlendirmişlerdir. Tahmin doğruluğunu iyileştirmek için bir özellik modeli ve karşılaştırmalı analiz uygulamışlardır. Birinci aşamada, orijinal veri setine minimum maksimum normalizasyon filtresi uygulamışlardır. İkinci aşamada, veri setinde PSO (Parçaçık Sürü Optimizasyonu) özellik seçimi kullanılması ile önemli özellikleri içeren öz nitelikleri belirlemişlerdir. Önemli öz niteliklerin *Total Bilirubin*, *Direct Bilirubin*, *Total Protiens*, *Albumin*, *A/G oranı*, *SGPT*, *SGOT* ve *Alkphos* olduğunu belirtmişlerdir. Birinci aşamada Bayes Net, %74.25 doğruluk oranı ile en başarılı yöntem olmuştur. Diğer yöntemlerde sırasıyla J48 %73.33, SVM %71.45, MLP %69.23 ve RF ile %68.44 oranında başarı elde etmişlerdir. İkinci aşamada J48 yöntemi %95.04 oran ile en başarılı algoritma olmuştur. İlgili çalışmada Bayes Net %90.33, RF %80.22, MLP %77.54 ve SVM %73.44 oranında başarı göstermiştir.

Alice [6] ILPD veri seti üzerinde karaciğer yetmezliği teşhisi için DT, Naive Bayes, RF, SVM ve YSA makine öğrenmesi yöntemlerini uygulamıştır. Çalışmada R programlama dili kullanılmıştır. En yüksek teşhis başarısı %81 ile DT algoritmasından elde edilmiştir. Diğer yöntemlerden RF %77, SVM %77, ANN %71 ve Naive Bayes %37 oranında başarı göstermiştir.

Venkata ve arkadaşları [7] tıbbi veri setlerinde makine öğrenmesi yöntemlerinin başarı oranlarını incelemişlerdir. Seçilen tıbbi veri kümeleri Göğüs Kanseri Verileri, Kronik Böbrek Hastalığı, Kriyoterapi, Hepatit, İmmünoterapi ve ILPD'dir. Veri kümelerini Naive Bayes, J48, MLP, JRip, IBk ve Bagging makine öğrenmesi yöntemleri ile analiz etmişlerdir. Çalışmalarında 10 kat çapraz doğrulama kullanmışlardır. ILPD veri setinde en yüksek teşhis başarısını %69.30 oran ile Bagging yönteminden elde etmişlerdir. Diğer yöntemlerde sırası ile MLP %68.95, J48 %68.78, Jrip %66.38, IBk %64.49 ve Native Bayes'de %55.75 oranında başarı elde ettiklerini belirtmişlerdir.

Jagdeep ve arkadaşları [8] sınıflandırma ve özellik seçme tekniğini kullanarak bir yazılım mühendisliği yaklaşımına dayalı olarak karaciğer hastalığını tahmin etmeye odaklanmışlardır. Akıllı karaciğer hastalığı tahmin yazılımının (ILDPS) geliştirilmesi, yazılım mühendisliği modeline dayalı özellik seçimi ve sınıflandırma tahmin teknikleri kullanılarak yapılmıştır. ILPD veri seti üzerinde doğruluğu bulmak için LR, SMO, RF, Naive Bayes, J48 ve IBk makine öğrenmesi yöntemlerini uygulamışlardır. Veri setini iki aşamada değerlendirmişlerdir. Birinci aşamada veri seti üzerinde öznitelik seçmeden, veri setinin orijinal haliyle teşhis başarı ölçümleri yapmışlardır. İkinci aşamada, veri setinde Greedy Stepwise özellik seçimi algoritması ile önemli özellikleri içeren öz nitelikleri belirlemişlerdir. Önemli öz niteliklerin *Total Bilirubin*, *Direct Bilirubin*, *Alkphos*, *SGPT* ve *SGOT* olduğunu belirtmişlerdir. Doğruluk oranlarını 10 kat çapraz doğrulama ile test etmişlerdir. Birinci aşamada en yüksek teşhis başarısını %72.53 oran ile RF makine öğrenmesi yönteminden elde etmişlerdir. Diğer yöntemlerden LR %72.50, SMO %71.35, J48 %68.78, IBk %64.15 ve Naive Bayes'de %55.74 oranında başarı gösterdiğini belirtmişlerdir. İkinci aşamada en yüksek teşhis başarısı %73.36 oran ile LR makine öğrenmesi yönteminden elde etmişlerdir. Diğer yöntemlerde RF %71.87, SMO %71.36, J48 %70.67, IBk %67.41 ve Naive Bayes'de %55.90 oranında başarı elde etmişlerdir. Önerilen ILDPS'ye çevrimiçi olarak <https://gndec.ac.in/~jagdeepmalhi/ildps/> adresinden ücretsiz olarak erişilebilir.

Joel ve arkadaşları [9] ILPD veri setleri üzerinde karaciğer yetmezliği teşhisi için LR, SVM, K En Yakın Komşu (K-NN) ve YSA makine öğrenmesi yöntemlerini kullanmışlardır. En yüksek doğruluğa sahip modeli, Python'da Tkinter paketi kullanılarak bir Grafik Kullanıcı Arayüzü (GUI) olarak sunmuşlardır. En yüksek teşhis başarısı %92.80 oranı ile ANN makine öğrenmesi yönteminden elde edilmiştir. Diğer yöntemlerden SVM %75.04, LR %73.23 ve K-NN %72.05 oranında başarı göstermiştir. YSA'da değişken olarak giriş sayısı 10, gizli katman sayısı 2, birinci gizli katmandaki nöron sayısı 400, ikinci gizli katmandaki nöron sayısı 400 olarak ayarlanmıştır.

Aytürk Keleş ve arkadaşları [10] ILPD veri seti üzerinde karaciğer yetmezliği teşhisi için J48, Decision Stump, Lojistik Model Ağacı (LMT), REP Tree, Hoeffding Tree, IBk, RF ve RT makine öğrenmesi yöntemlerini kullanmışlardır. En yüksek teşhis başarısı %81.9 oranı ile RF makine öğrenmesi yönteminden elde edilmiştir.

Diğer yöntemlerden IBk %81.60, J48 %74.40, RT %73.80, LMT %73.80, REP Tree %71.30, Hoeffding Tree %69.10 ve Decision Stump %67.1 oranında başarı göstermiştir.

3. Materyal ve Yöntem

Bu çalışmada karaciğer yetmezliği hastalığının teşhisi için UCI makine öğrenmesi havuzunda araştırmalar için paylaşılan ILPD veri seti kullanılmıştır. ILPD veri seti içerisinde 583 örnek bulunmaktadır. Veriler %80 eğitim seti (466 örnek) ve %20 test seti (117 örnek) olacak şekilde ayrılmıştır. Her veri seti ayrılma işleminde aynı eğitim ve test seti kullanılması için random state değeri 42 kullanılmıştır. Veri setinde cinsiyet öz niteliği dikkate alınmamıştır. Bu çalışma dört aşamada karaciğer yetmezliği hastalarının tahmin doğruluğunu iyileştirmek için hibrid model ve karşılaştırılmalı bir analiz sunmaktadır. Birinci aşamada “Yaş, Toplam Bilirubin, Doğrudan Bilirubin, Alkalen Fosfataz, Alanin Aminotransferaz, Toplam Protein, Albümin, Albümin/Globulin Oranı” öz niteliklerine sahip orijinal veri setine SVM, MLP, DT, LR ve LGBM makine öğrenmesi yöntemleri uygulanarak teşhis tahmin başarıları ölçülmüştür. İkinci aşamada veri seti üzerinde öz nitelik seçimi için, Sarmalayıcı (Wrapper) yöntemlerinden olan ileri doğru seçim kullanılmıştır. Özellik seçimi sonrasında “Doğrudan Bilirubin, Yaş, Alanin Aminotransferaz, Alkalen Fosfataz” öz nitelikleri kullanılarak SVM, MLP, DT, LR ve LGBM makine öğrenmesi yöntemleri ile karaciğer yetmezliği hastalığı teşhis tahmin başarıları ölçülmüştür. Üçüncü aşamada, ikinci aşamada uygulanan SVM, MLP, DT, LR ve LGBM makine öğrenmesi yöntemlerine model iyileştirilmesi uygulanarak modeller yeniden kurulmuş olup, doğruluk performansları artırılmıştır. İlk üç aşamanın tahmin doğruluğu performansları karşılaştırılmıştır. Dördüncü aşamada ise model iyileştirilmesi uygulanmış sınıflandırma yöntemlerinin doğruluk performans sonuçları, literatürde bu çalışmada kullanılan veri seti ile yapılan çalışmaların sonuçları karşılaştırılmıştır. Makine öğrenmesi yöntemlerinin modelleme işlemleri Python programlama dili kullanılarak JupyterLab ortamında gerçekleştirilmiştir. Makine öğrenmesi yöntemlerinin başarıları ölçülürken çapraz doğrulama yöntemi (cv) 10 kat olarak kullanılmıştır.

3.1. Veri Seti

Çalışmada kullanılan veri seti, UCI makine öğrenmesi havuzunda bulunan ücretsiz olarak kullanıma açık Hint Karaciğer Hastalarının verilerini barındıran ILPD’dir. Veri seti Hindistan’da bulunan Andhra Pradesh eyaletinin kuzey doğusundaki hasta insanlardan ölçüm yapılarak toplanmıştır [11].

Veri setinde 583 adet örnek hasta kaydı vardır. Bu kayıtların 441 adedi erkek, 142 adedi ise kadın hastalardan oluşmaktadır. Hasta kayıtlarının 416’sı karaciğer hastası iken 167 tanesi karaciğer hastası değildir. Karaciğer hastası olan erkek sayısı 324, karaciğer hastası olan kadın sayısı ise 92’dir.

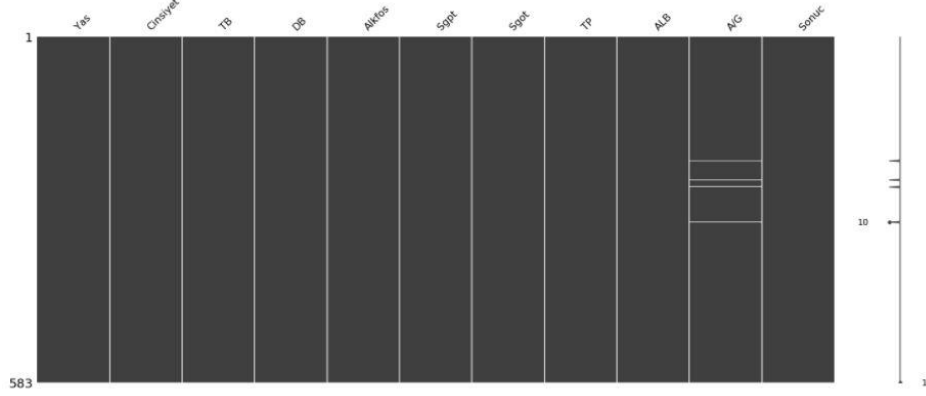
Veri setinde 11 adet özellik bulunmaktadır. 10 tanesi öznitelik olarak kullanılabilirken on birinci özellik ise hastalığın olup olmama durumunun gösterildiği alandır. Özniteliklerde bulunan SGPT ve SGOT günümüzde farklı isimlerde kullanılmaktadır. SGOT yeni terminolojide ALT, SGPT ise AST olarak adlandırılmaktadır. Veri seti özellikleri hakkında detaylı bilgi Tablo 1’de sunulmuştur.

Tablo 1. Veri seti özellikleri.

No	Özellikler	Açıklama	Veri Tipi	Özellik Aralık Değeri
1	Age	Yaş	Numeric	[4 – 90]
2	Gender	Cinsiyeti	Nominal	“male” or “female”
3	TB	Toplam Bilirubin	Numeric	[0.4 - 75]
4	DB	Doğrudan Bilirubin	Numeric	[0.1 - 19.7]
5	AlkPhos	Alkalen Fosfataz	Numeric	[63 - 2110]
6	Sgpt	Alanin Aminotransferaz	Numeric	[10 - 2000]
7	Sgot	Aspartat Aminotransferaz	Numeric	[10 - 4929]
8	TP	Toplam Protein	Numeric	[2.7 - 9.6]
9	ALB	Albümin	Numeric	[0.9 - 5.5]
10	A/G Ratio	Albümin/Globulin Oranı	Numeric	[0.3 - 2.8]
11	Selector Field	Secici Alan Bilgisi	Numeric (1,2)	1-Hasta, 2-Hasta Değil

3.1.1 Eksik Verilerin Giderilmesi

Veri seti incelendiğinde 4 hastanın Albümin/Globülin oranı alanının boş olduğu görülmüştür. Daha doğru bir sonuç için tüm hasta kayıtlarının Albümin/Globülin oranının ortalaması ile eksik veriler doldurulmuştur. Veri setinin eksik alanları Şekil 1’de sunulmuştur.



Şekil 1. Eksik veriler.

3.1.2 Normalizasyon Filtresi

Veriler arasındaki farklılıkların fazla olması bazı sınıflandırma yöntemlerinin öğrenme doğruluğunu etkilemektedir. Normalizasyon uygulanmasının amacı, matematiksel işlemler ile veriler arasındaki farklılıkları ortadan kaldırarak verilerin karşılaştırılmasını kolaylaştırmaktır. Bu çalışmada MLP ve SVM algoritmalarında Python programlama dili sklearn.preprocessing kütüphanesinde bulunan *Standart Scaler* kullanılarak veriler normalize edilmiştir.

3.1.3 Öznitelik Seçimi

Bu çalışmada veri setinde öznitelik seçmek için Sarmalayıcı (Wrapper) yöntemlerinden İleri Doğru Seçim yöntemi (Step Forward Selection (SFS)) kullanılmıştır. SFS yönteminde, kurulan modelin performansına en çok katkıda bulunan yani bağımlı değişken ile en yüksek korelasyona sahip öznitelik ile döngü süreci başlatılır. Belirlenen önem düzeyine (α) göre diğer öznitelikler sırası ile kontrol edilir. Önem düzeyine göre şartı sağlayan öznitelikler modele eklenir. Başta belirlenen önem düzeyi şartına uymayan öznitelikler modele eklenmez. Tüm değişkenler kontrol edildikten sonra döngü sonlanır [16]. Bu çalışmada önem düzeyi %95 olacak şekilde öznitelik seçilmiştir. SFS yöntemi ile belirlenen öznitelikler Tablo 2’de gösterilmiştir.

Tablo 2. SFS yöntemi sonrasında elde edilen önemli öznitelikler.

Öznitelik Adı	Kısaltması
Doğrudan Bilirubin	DB
Yaş	Yaş
Alanin Aminotransferaz	Sgpt
Alkalin Fosfataz	AlkFos

3.2. Destek Vektör Makinesi

Vapnik ve ekibi tarafından 1990’lı yıllarda Destek Vektör Makinesi (Support Vector Machine) algoritması geliştirilmiştir. Gözetimli makine öğrenmesi algoritmalarından olan Destek Vektör Makinesi, ikili temel sınıfa ait veriler üzerinde istatistiksel işlem uygulanması ile birbirinden ayırmak için kullanılır [12].

Veri setimizde iki sınıfa ait bir sınıflandırma kullanılacağından dolayı doğrusal destek vektör sınıflandırması tercih edilmiştir. Model iyileştirilmesi yapılırken model kurma aşamasında maksimum yinleme sayısının değerinin düşürülmesi doğruluk değerini artırmıştır. Bu nedenle doğruluk oranında en iyi performansı max_iter

hiper parametresi 23 olarak şekilde elde edilmiştir. İyileştirilmiş model oluşturulurken random state değeri 42 ve max iter sayısı 23 olarak kullanılmıştır. Bu çalışmada Python programlama dili sklearn.svm kütüphanesinde bulunan LinearSVC adında Linear SVM makine öğrenmesi yöntemi kullanılmıştır.

3.3. Çok Katmanlı Algılayıcılar

Çok Katmanlı Algılayıcılar (Multilayer Perceptron) girdi, ara ve çıktı katmanlarından oluşur. Tek katmanlı algılayıcıdan farklı olarak ara katman, giriş katmanı ile çıktı katmanı arasında bir köprü görevi görür. Ara katman girdi katmanından gelen girdileri çıktı katmanına göndermeden önce probleme göre değerlendirir. Değerlendirme sonucunda probleme göre daha iyi bir karar verilmesi sağlanır. Ara katman sayısı problemin durumuna göre artırılabilir [13].

MLP’de iyileştirilmiş model çalışmalarında beş katmanlı, her bir gizli katman 10 nörondan oluşan model kullanılmıştır. Gizli katmanlardaki aktivasyon işlevi için düzeltilmiş doğrusal birim fonksiyonu tercih edilmiştir. L2 ceza parametresi olarak 0.1 değeri doğruluğu arttırmıştır. Maksimum yineleme sayısının 215 olarak eklenmesiyle doğruluk oranında en iyi performans elde edilmiştir. İyileştirilmiş model oluştururken değiştirilen hiper parametreler Tablo 3’te gösterilmiştir. Bu çalışmada Python programlama dili sklearn.neural_network kütüphanesinde bulunan MLPClassifier sınıflandırma algoritması kullanılmıştır.

Tablo 3. MLP’de kullanılan hiper parametreler.

activation	relu
hidden_layer_sizes	(10,10,10,10,10)
alpha	0.1
max_iter	215
random_state	42

3.4. Lojistik Regresyon

Lojistik Regresyon (Logistic Regression) gözetimli algoritmalar grubunda bulunan, ikili sınıflandırma problemlerini çözen istatistiksel bir yöntemdir. Veri kümesinin regresyon analiz edilmesiyle iki olasılık (0 veya 1) sonucu belirtir [13]. Bu çalışmada kullanılan sonuç bağımlı değişkeni sayısal (1-karaciğer hastası 2- karaciğer hastası değil) olarak tutulmaktadır.

LR’da iyileştirilmiş model çalışmalarında, veri setimizin boyutu küçük olduğu için optimizasyon problemlerinde tercih edilen algoritma olarak *liblinear* kullanılmasının başarıyı arttırdığı gözlemlenmiştir. Sonuç değişkenimiz 2 sınıfa ait olduğu için (1-karaciğer hastası 2- karaciğer hastası değil) multi class özelliği ovr olarak kullanılmıştır. İyileştirilmiş model oluştururken değiştirilen hiper parametreler Tablo 4’te gösterilmiştir.

Bu çalışmada Python programlama dili sklearn.linear_model kütüphanesinde bulunan LogisticRegression sınıflandırma yöntemi kullanılmıştır.

Tablo 4. Lojistik regresyonda kullanılan hiper parametreler.

solver	liblinear
multi_class	ovr
random_state	42

3.5. Karar Ağaçları

Karar Ağaçları (Decision Tree) gözetimli öğrenmede en çok kullanılan algoritmalarındandır. Temel amaç veri seti içerisindeki karmaşık yapıları basit yapılara dönüştürmektir. Karar ağaçları ile hem sayısal hem de kategorik veriler işlenebilir. Test ve eğitimin hızlı olması ve sonuçların görsel olarak kolay yorumlanabilmesinden sınıflandırmada sıklıkla kullanılan yöntemlerdendir. Kök, düğüm ve dallanmalardan oluşmaktadır. Karar ağaçlarında işlemler iki adımda sonuçlanır. Birinci adımda ağaç oluşturulur. İkinci adımda ise oluşturulan ağaçtan sınıflandırma kuralları oluşturulur. Karar ağacını oluştururken kullanılan algoritmalara göre karar ağacının yapısı değişiklik gösterebilir [14].

Bu çalışmada Python programlama dili sklearn.tree kütüphanesinde bulunan DecisionTreeClassifier makine öğrenmesi yöntemi kullanılmıştır. Karar ağacı *Classification and Regression Tree (CART)* algoritması ile oluşturulmuştur. Dallanmalar için *entropy* algoritması kullanılmıştır. Ağacın maksimum derinliğinin 5 ile sınırlandırılması doğruluk oranını arttırmıştır. En iyi bölünme için $\text{max_features}=\sqrt{\text{n_features}}$ karşılığına gelen max_features parametresi auto kullanılmıştır. İyileştirilmiş model oluştururken değiştirilen hiper parametreler Tablo 5’te gösterilmiştir.

Tablo 5. Karar ağacı için kullanılan hiper parametreler.

criterion	entropy
max_depth	5
max_features	auto
random_state	42

3.6. Hafif Gradyan Güçlendirme Makinesi Sınıflandırıcısı

Hafif Gradyan Güçlendirme Makinesi Sınıflandırıcısı (Light Gradient Boosting Machine Classifier) 2017 yılında Microsoft tarafından geliştirilen bir kütüphanedir. LGBM, XGBoost’un eğitim süresi performansının arttırmaya yönelik geliştirilen Gradient Boosting Machine (GBM) türlerindedir. XGBoost ilk aramayı enlemesine yaparken LGBM ilk aramayı derinlemesine yapar [15]. LGBM’de iyileştirilmiş model çalışmalarında default parametrelerden öğrenme oranının artırılması ve ağaç sayısının 150 ile sınırlandırılması tahmin başarısını arttırmıştır. İyileştirilmiş model oluştururken değiştirilen hiper parametreler Tablo 6’da gösterilmiştir. Bu çalışmada Python programlama dili lightgbm kütüphanesinde bulunan LGBMClassifier sınıflandırma algoritması kullanılmıştır.

Tablo 6. LGBM’de kullanılan hiper parametreler.

learning_rate	0.2
n_estimators	150
random_state	42

4. Sonuçlar ve Değerlendirme

Kullanılan makine öğrenmesi yöntemlerinin veri seti üzerindeki sınıflandırma başarılarını değerlendirmek için bazı ölçütler vardır. Bu ölçütlerin hesaplanmasında Tablo 7’de belirtilen karışıklık matrisi (confusion matrix) kullanılır. Karışıklık matrisi oluşturulan modeldeki tahmin doğruluk başarısını 2x2 matris halinde belirtir. Gerçek değerler ile yapılan tahminlerin karşılaştırılmasını sağlar [10].

Tablo 7. Karışıklık matrisi (confusion matrix).

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

Tablo 7’de belirtilen terimler şöyle açıklanabilir;

- **TP (True Positive):** Doğru tahmin edilen karaciğer hastası sınıf değerini belirtir.
- **FP (False Positive):** Yanlış tahmin edilen karaciğer hastası sınıf değerini belirtir.
- **FN (False Negative):** Yanlış tahmin edilen karaciğer hastası olmayan sınıf değerini belirtir.
- **TN (True Negative):** Doğru tahmin edilen karaciğer hastası olmayan sınıf değerini belirtir.

Literatürde sınıflandırma başarılarını değerlendirmek için kullanılan birçok metrik bulunmaktadır. Bu çalışma kapsamında da kullanılan bu metrikler ve formülleri bu kısımda açıklanmaktadır.

- **Doğruluk (Accuracy):** Doğru tahmin edilmiş karaciğer hastalarının tüm değerlere olan oranıdır.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- **Duyarlılık (Sensitivity):** Doğru tahmin edilmiş karaciğer hastalarının, doğru tahmin edilmiş karaciğer hastaları ve yanlış tahmin edilmiş karaciğer hastası olmayanların toplamına oranıdır. Doğru tahmin edilmiş karaciğer hastalarını tespit etme yeteneğini gösterir.

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (2)$$

- **Özgüllük (Specificity):** Doğru tahmin edilmiş karaciğer hastası olmayan sınıf değerinin yanlış veya doğru tahmin edilmiş karaciğer hastası olmayanların toplamına oranıdır. Karaciğer hastası olmayan sınıf değerlerinin ne ölçüde karaciğer hastası olmadığını tahmininde bulunduğu oranını gösterir.

$$\text{Özgüllük} = \frac{TN}{TN + FP} \quad (3)$$

- **Kesinlik (Precision):** Doğru tahmin edilmiş karaciğer hastası değerlerinin, yanlış tahmin edilmiş karaciğer hastaları değerlerinin ve doğru tahmin edilmiş karaciğer hastaları değerlerinin toplamına oranıdır. Karaciğer hasta sınıf değerlerinin ne ölçüde karaciğer hastası olarak tahminde bulunduğu oranını gösterir.

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (4)$$

- **F-Ölçütü (F-Measure):** Harmonik ortalaması alınan Kesinlik ve Duyarlılık değerlerin bize verdiği değerdir. Modelin kesinliği ve sağlamlığının bir ölçüsüdür.

$$\text{F-Ölçütü} = 2 \times \frac{\text{Duyarlılık} \times \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (5)$$

- **Ortalama Mutlak Hata (Mean Absolute Error (MAE)):** İki sürekli değişken arasındaki farkın ölçülmesidir. Sıfıra yakın değerler daha iyi performans göstermektedir.

Veri setinden orijinal hali ve öz nitelik çıkarımı sonrası uygulanan makine öğrenmesi yöntemleri sonuç değerleri Tablo 8’de verilmiştir.

Tablo 8. Öznitelik çıkarımı öncesi ve sonrası doğruluk değerleri karşılaştırılması.

Sınıflandırma Algoritmaları	Doğruluk Oranları (%)	
	Özellik Seçimi Öncesi	Özellik Seçimi Sonrası
MLP	74.31	78.63
SVM	75.22	76.13
LR	78.09	77.80
DT-CART	75.90	77.04
LGBM	76.21	77.95

Birçok veri seti kümesinde ilgili veya ilgisiz birçok veri bulunabilmektedir. Veri setinde bulunan 11 öznitelik özellik seçimi işlemi sonrasında 4 özneliğe indirgenmiştir. Tablo 8 incelendiğinde öznitelik çıkarımı sonrasında MLP, SVM, DT, LGBM makine öğrenmesi algoritmalarının doğruluk oranları arttığı görülmektedir. Özellik seçimi sonrasında model iyileştirmesi uygulandıktan sonraki doğruluk değerleri karşılaştırması Tablo 9’da verilmiştir.

Tablo 9. Özellik seçimi öncesi ve sonrası doğruluk değerleri karşılaştırılması.

Sınıflandırma Algoritmaları	Doğruluk (%)		
	Özellik Seçimi Öncesi	Özellik Seçimi Sonrası	Model İyileştirme Sonrası
MLP	74.31	78.63	81.13
SVM	75.22	76.13	77.87
LR	78.09	77.80	77.80
DT-CART	75.90	77.04	81.13
LGBM	76.21	77.95	82.12

Tablo 9 incelendiğinde özellik seçimi sonrasında tüm makine öğrenmesi yöntemlerinde doğruluk oranlarının arttığı gözükmemektedir. Kullanılan makine öğrenmesi yöntemlerinin öz nitelik çıkarımı sonrasında model iyileştirilmesi uygulaması aşamasından sonraki performans değerleri Tablo 10'da verilmiştir.

Tablo 10. Makine öğrenmesi yöntemleri performans değerleri.

Algoritma	Doğruluk	Duyarlılık	Özgüllük	Kesinlik	F-Ölçütü	Ortalama Mutlak Hata
SVM	%77.87	%85.05	%56.66	%85.05	%85.05	0.22
MLP	%81.13	%90.80	%46.66	%83.15	%86.81	0.20
DT-CART	%81.13	%76.14	%50	%95.40	%84.69	0.25
LR	%77.80	%95.40	%23.33	%78.30	%86.01	0.23
LGBM	%82.12	%81.39	%45.16	%80.45	%80.92	0.28

Tablo 10 incelendiğinde en yüksek teşhis başarısı %82.12 oran ile LGBM makine öğrenmesi yönteminden elde edilmiştir. LGBM algoritmasına yakın olarak %81.13 başarı oranı ile DT-CART ve MLP ikinci yüksek başarı gösteren yöntemdir. Duyarlılık ölçütüne göre en yüksek oran %95.40 ile LR yönteminden elde edilmiştir. Sonuçlar kesinlik ölçüsüne göre değerlendirildiğinde %95.40 oran ile en başarılı yöntem DT-CART olmuştur. F-ölçütüne göre %86.81 oran ile en iyi sonuç veren yöntem MLP'dir. Bu alanda yapılan ilgili çalışmalar incelendiğinde ILPD veri seti üzerinde daha önceden LGBM makine öğrenmesi yöntemi uygulanmadığı görülmüştür. Bu çalışma kapsamında ele alınan MLP, SVM, LR, DT makine öğrenmesi yöntemleri uygulanan çalışmalar ele alınmış olup, bu çalışma ile Doğruluk, Duyarlılık, Özgüllük Kesinlik, F-Ölçütü, Ortalama Mutlak Hata değerlerinin karşılaştırmaları Tablo 11'de gösterilmiştir.

Tablo 11 incelendiğinde daha önceki çalışmalarda kullanılan ortak makine öğrenmesi yöntemlerinde doğruluk metriğine göre en yüksek başarı sağlandığı görülmektedir. MAE'ye göre değerlendirilen çalışmalar ele alındığında ise en düşük MAE hata oranı elde edilmiştir. Makine öğrenmesi sağlık alanında da sıklıkla kullanılmaktadır. Teknolojinin gelişmesiyle beraber hastalık teşhislerinde geleneksel teşhis yöntemlerin yanında akıllı uygulamalar kullanılarak hastalıkların erken teşhis edilmesi amaçlanmaktadır. Akıllı uygulamaların geliştirilmesinde genellikle makine öğrenmesi algoritmalarından faydalanılır. Bu çalışmada son yıllarda sıklıkla görülen ve erken dönemlerde teşhis edilmediği takdirde hayat kayıplarına dahi yol açan karaciğer yetmezliği hastalığının, yüksek doğruluk ile erken teşhis edilmesi üzerine odaklanılmıştır. Özellik seçim yöntemiyle 11 öz nitelikten 4 öz niteliğe indirgenmiş olup daha yüksek tahmin başarıları elde edilmiştir. Kullanılan beş farklı makine öğrenmesi yöntemi üzerinde modeller oluşturulmuştur. Bu modellere model iyileştirme yapıp modellerin hastalık tahmin başarıları artırılmıştır. Model seçiminde en yüksek doğruluk, duyarlılık, kesinlik ve F Ölçütü oranları dikkate alınmıştır. En başarılı sonuç %82.12 doğruluk oranıyla LGBM yönteminden elde edilmiştir. Bu çalışmanın sonucu, ILPD veri seti ile yapılan benzer çalışmalar ile karşılaştırılmıştır. Karşılaştırılan çalışmaların başarı oranları Tablo 12'de verilmiştir.

Tablo 11. Yapılan çalışmanın literatürde aynı makine öğrenmesi yöntemleri kullanan çalışmalar ile karşılaştırılması.

Çalışmalar/ Algoritmalar	Başarı Ölçme Metrikleri	[2]	[3]	[5]	[6]	[8]	[9]	Önerilen Yöntem
MLP	Doğruluk (Accuracy)	-	%70.84	%77.54	-	-	-	%81.13
	Duyarlılık (Sensitivity)	-	-	-	-	-	-	%90.80
	Özgüllük (Specificity)	-	-	-	-	-	-	%46.66
	Kesinlik (Precision)	-	-	-	-	-	-	%83.15
	F-Ölçütü (F-Measure)	-	-	-	-	-	-	%86.81
Mean Absolute Error	-	0.34	0.70	-	-	-	0.20	
SVM	Doğruluk (Accuracy)	%73.2	%71.35	%73.44	%77	-	%75.04	%77.87
	Duyarlılık (Sensitivity)	%90	-	-	%74	-	%79	%85.05
	Özgüllük (Specificity)	%30	-	-	%85	-	%71.11	%56.66
	Kesinlik (Precision)	-	-	-	-	-	%77.09	%85.05
	F-Ölçütü (F-Measure)	-	-	-	-	-	-	%85.05
Mean Absolute Error	-	0.28	0.71	-	-	-	0.22	
LR	Doğruluk (Accuracy)	-	-	-	-	%74.36	%73.23	%77.80
	Duyarlılık (Sensitivity)	-	-	-	-	-	%88	%95.40
	Özgüllük (Specificity)	-	-	-	-	-	%30.62	%23.33
	Kesinlik (Precision)	-	-	-	-	-	%78.57	%78.30
	F-Ölçütü (F-Measure)	-	-	-	-	-	-	%86.01
Mean Absolute Error	-	-	-	-	0.40	-	0.23	
DT	Doğruluk (Accuracy)	-	-	-	%81	-	-	%81.13
	Duyarlılık (Sensitivity)	-	-	-	%37	-	-	%76.14
	Özgüllük (Specificity)	-	-	-	%95	-	-	%50
	Kesinlik (Precision)	-	-	-	-	-	-	%95.40
	F-Ölçütü (F-Measure)	-	-	-	-	-	-	%84.69
Mean Absolute Error	-	-	-	-	-	-	0.25	

Tablo 12. Literatür karşılaştırması.

Çalışmalar/ Algoritmalar	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	Önerilen Yöntem
MLP	-	%70.84	-	%77.54	-	-	-	-	-	%81.13
DT	-	-	-	-	%81	-	-	-	-	%81.13
LR	-	-	-	-	-	-	%74.36	%73.23	-	%77.80
SVM	%73.2	%71.35	-	%73.44	%77	-	-	%75.04	-	%77.87
LGBM	-	-	-	-	-	-	-	-	-	%82.12
Native Bayes	-	-	-	-	%37	%55.75	%55.90	-	-	-
RF	-	%71.86	%72.84	%80.22	%77	-	%71.87	-	%81.9	-
YSA	-	-	%78	-	%71	-	-	%92.80	-	-
J48	-	%70.66	-	%95.04	-	%68.78	%70.67	-	%74.4	-
KNN	-	-	-	-	-	-	-	%72.05	-	-
Bayes Net	-	%69.12	-	%90.33	-	-	-	-	-	-
Bagging	-	-	-	-	-	%69.30	-	-	-	-
IBk	-	-	-	-	-	%64.49	%67.41	-	%81.6	-
JRIP	-	-	-	-	-	%66.38	-	-	-	-
MP	-	-	-	-	-	%68.95	-	-	-	-
SMO	-	-	-	-	-	-	%71.36	-	-	-
Radyal Tabanlı İşlevsel (RTI)	-	-	%72.32	-	-	-	-	-	-	-
Lojistik Model Ağacı (LMT)	-	-	-	-	-	-	-	-	%73.8	-
Decision Stump	-	-	-	-	-	-	-	-	%67.1	-
Hoeffding Tree	-	-	-	-	-	-	-	-	%69.1	-
REP Tree	-	-	-	-	-	-	-	-	%71.3	-
Random Tree	-	-	-	-	-	-	-	-	%73.8	-

Tablo 12 incelendiğinde bu çalışmada kullanılan makine öğrenmesi yöntemlerinin sonuçları, ILPD veri seti kullanılarak aynı makine öğrenmesi yöntemleri ile yapılan çalışmaların sonuçlarına göre daha başarılı olduğu görülmektedir. İncelenen çalışmalar arasında en yüksek teşhis başarısı %95.04 oranı ile J48 makine öğrenmesi yönteminden elde edilmiştir. İleriki çalışmalarda ILPD veri seti üzerinde özellik seçimi ile denenmemiş makine

öğrenmesi yöntemlerinin kullanılması planlanmaktadır. Ayrıca derin öğrenme yöntemlerinin de ilgili veri seti üzerinde denenmesi hedeflenmektedir.

Kaynaklar

- [1] Ersoy O.” Karaciğer Enzim Yüksekliğinin Değerlendirilmesi”. Ankara Medical Journal 2012; 12 (3): 129-135.
- [2] Mamdouh E. ve Mabrouk M. “A Study of Support Vector Machine Algorithm For Liver Disease Diagnosis”, American Journal of Intelligent Systems; 2014; 4 (1): 9-14.
- [3] Gulia A. Vohra R. Rani P. “Liver Patient Classification Using Intelligent Techniques”, International Journal of Computer Science and Information Technologies; 2014; 5 (4): 5110-5115.
- [4] Alkuşak E. ve Gök M., 2015. “Karaciğer Yetmezliğinin Teşhisinde Makine Öğrenmesi Algoritmalarının Kullanımı”, ISITES 2014 Sempozyumu (ISITES’2014); 18-20 Haziran 2014; Karabük. 703-707.
- [5] Priya M.B. Juliet P. L. Tamilselvi P.R.“ Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms”, International Research Journal of Engineering and Technology (IRJET); Ocak 2018; 206-211.
- [6] Auxilia L.A. “Accuracy Prediction using Machine Learning Techniques for Indian Patient Liver Disease”, 2nd International Conference on Trends in Electronics and Informatics (ICOEI); 11-12 Mayıs 2018; Tirunelveli, India. 45-50.
- [7] Bendi R.V. Boddu R.S.K. “ Performance Comparison of Classification Algorithms on Medical Datasets”, 9th Annual Computing and Communication Workshop and Conference (CCWC); 7-9 Ocak 2019; USA. No:2322.
- [8] Singh J. Bagga S. Kaur R. “Software-based Prediction of Liver Disease with Feature Selection and Classification Techniques”. International Conference on Computational Intelligence and Data Science (ICCIDIS 2019); 6 -7 Eylül 2019; Gurugram. 1970-1980.
- [9] Jacob J. Mathew J. C. Mathew J. Issac E.“ Diagnosis of Liver Disease Using Machine Learning Techniques”, International Research Journal of Engineering and Technology (IRJET); 4 Nisan 2018; 5(4): 4011-4014.
- [10] Keleş A. Kaslı Ö.B. Keleş A. “Makine Öğrenme Algoritmaları ile Karaciğer Hastalığının Teşhisi”, Turkish Studies-Information Technologies and Applied Sciences; 2020; 75-83.
- [11] [https://archive.ics.uci.edu/ml/datasets/ILPD+\(Indian+Liver+Patient+Dataset\)](https://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+Dataset)).
- [12] Selvi O. ”Göğüs Kanseri Teşhisinde Farklı Makine Öğrenmesi Tekniklerinin Performans Karşılaştırması”, Avrupa Bilim ve Teknoloji Dergisi; 2019; (16): 176-185.
- [13] Özdemir E. Ballı S. “Türkiye Erkekler Basketbol Ligi Maç Sonuçlarının Makine Öğrenmesi Yöntemleri ile Tahmini”, Mühendislik Bilimleri ve Tasarım Dergisi, 2020; 8 (3): 740 - 752.
- [14] Kaya Ç. Yıldız O. “Makine Öğrenmesi Teknikleriyle Saldırı Tespiti: Karşılaştırmalı Analizi”, Marmara Fen Bilimleri Dergisi, 2014; (3): 89-104.
- [15] Omar A. B. K. “XGBoost and LGBM for Porto Seguro’s Kaggle challenge: A comparison”, Distributed Computing Group Computer Engineering and Networks Laboratory ETH Zurich , Semester Project. 2018
- [16] Kayaalp T.G. Güney Ç. M. Cebeci Z. “Çoklu Doğrusal Regresyon Modelinde Değişken Seçiminin Zootekniye Uygulanışı”, Çukurova Üniversitesi Ziraat Fakültesi Dergisi. 2015; 30 (1): 1-8