



ARAŞTIRMA MAKALESİ (Research Article)

YAPAY ZEKA YÖNTEMLERİYLE ERKEN EVRE DİYABET RİSK TAHMİNİ

Hüseyin YILMAZ¹, Abdulkadir BULDU², Yılmaz KAYA³, Fatma KUNCAN^{4*}

¹ Siirt Üniversitesi, Elektrik Elektronik Mühendisliği Bölümü, Siirt, ORCID: 0000-0002-7068-6429

² Siirt Üniversitesi, Elektrik Elektronik Mühendisliği Bölümü, Siirt, ORCID: 0000-0002-9161-4862

³ Batman Üniversitesi, Bilgisayar Mühendisliği Bölümü, Batman, ORCID: 0000-000-5167-1101

^{4*} Siirt Üniversitesi, Bilgisayar Mühendisliği Bölümü, Siirt, ORCID: 0000-0003-0712-6426

Geliş Tarihi(Received Date):16.12.2022

Kabul Tarihi(Accepted Date):14.02.2023

ÖZ

Makine öğrenmesi ve yapay zekâ günümüzde birçok alanda ve çalışmada yer almaktadır. Bu alanlardan biri de sağlık alanıdır. Biyomedikal işaretlerin son zamanlarda hızlıca elde edilmesi ve elde edilen birçok kaynağın teknoloji sayesinde süreç entegresindeki hız, sağlık alanında da birçok projenin başarı ile gelişmesine olanak sağlamıştır. Ülkemizde ve dünyamızda son zamanlarda oldukça fazla artış gösteren diyabet hastalığı, halk dilinde bilenen aksine yüksek miktarda şeker tüketiminden değil, vücutta bulunması gereken insülin hormonunun vücut tarafından üretilmemesi ya da insülin hormonunun emilmemesi sonucu olarak diyabet hastalıkları görülmektedir. Makine öğrenmesi ve sınıflandırma algoritmalarının çeşitli biyomedikal alanlarda hastalığın teşhisi ve erken evre tedavi yöntemlerinde başarılı sonuçlar elde etmesi, günümüzde hasta sayısının çok hızlı artış gösterdiği şeker hastalığının da teşhisinde, bizlere risk faktörleri artmadan öngörülen semptomlardan pozitif çıkarımlar elde ederek hastalığa karşı erken müdahale imkânı sunabilecektir. Yaptığımız çalışmada 2 farklı tipe sahip olan diyabet hastalığının teşhisi için 5 farklı makine öğrenmesini (KNN, Lojistik Regresyon, Karar Ağaçları, DVM, Random Forest), python programlama dili ile sınıflandırma sürecine dahil edilerek algoritmaların başarı oranlarının kıyaslanmasını yaptık. Kullandığımız algoritmalar, yapısal olarak sınıflandırma işlemlerinde oldukça başarılı sonuçlar vermektedir. KNN, Lojistik Regresyon, Karar Ağaçları, DVM, Random Forest algoritmaları Sınıflandırma süreçlerinde yüksek başarılar sağlamak ve bu başarılar hastalık tanı kitleri oluşturulmasına olanak sunmaktadır. Çalışmada en yüksek sonucu elde eden makine öğrenmesi algoritması olan Random Forest algoritmasının erken evre diyabet hastalığının teşhisinde %96 gibi yüksek bir oranda doğru sonuç elde edildi. Yapılan çalışmada Random Forest algoritmasının başarı oranının yanı sıra Kesinlik ve hassasiyet konusunda da %94'lük bir başarı yakaladı.

Anahtar Kelimeler: *Diyabet, Erken Evre, Sınıflandırma Algoritmaları, Makine Öğrenmesi*

EARLY STAGE DIABETES RISK PREDICTION WITH ARTIFICIAL INTELLIGENCE METHODS

ABSTRACT

Machine learning and artificial intelligence are involved in many fields and studies today. One of these fields is the field of health. The rapid acquisition of biomedical signs in recent times and the

speed of integration of many sources into the process thanks to technology have enabled the successful development of many projects in the field of health. Diabetes disease, which has increased considerably recently in our country and in the world, is not due to the consumption of high amounts of sugar, contrary to what is known in the public language, but as a result of the body not producing the insulin hormone that should be present in the body or not absorbing the insulin hormone. In our study, we compared the success rates of algorithms by including 5 different machine learning processes for the diagnosis of diabetes, which has 2 different types. The successful results of machine learning and classification algorithms in the diagnosis of the disease in various biomedical fields and in early stage treatment methods will provide us with the opportunity of early intervention against the disease by obtaining positive inferences from the predicted symptoms without increasing the risk factors, in the diagnosis of diabetes, where the number of patients is increasing very rapidly today. In our study, we compared the success rates of algorithms by including 5 different machine learning methods in the classification process with the python programming language for the diagnosis of diabetes, which has 2 different types. The algorithms we use give very successful results in structural classification processes. KNN, Logistic Regression, Decision Trees, DVM, Random Forest algorithms provide high success in classification processes and these successes allow the creation of disease diagnostic kits. In the study, the Random Forest algorithm, which is the machine learning algorithm that achieved the highest result, achieved an accurate result of 96% in the diagnosis of early stage diabetes. In the study, besides the success rate of the Random Forest algorithm, it also achieved a 94% success in precision and precision.

Keywords: *Diabetes, Early Stage, Classification Algorithms, Machine Learning*

1. GİRİŞ

Diyabet, Dünya Sağlık Örgütü'nün (WHO) 2018 yılındaki raporuna göre dünya çapında 422 milyon kişiyi etkilemiş ve en hızlı büyüyen, yaşamı tehdit eden kronik hastalıklardan biridir [1]. Bu hastalık genellikle Tip2 ve Tip1 olmak üzere 2 gruba ayrılır. Kronik bir rahatsızlık olan diyabet hastalığı vücutta varlığını uzun bir süre gizli olarak da sürdürebilir. Aniden ortaya çıkan ya da zaman içerisinde ileri seviyelere ulaşan diyabet durumu, zamanla vücutta bulunan belli başlı organlara büyük zararlar verebilecek hale gelmektedir. En çok etkilenenler, kalp ve göz gibi insan yaşamında yeri bir hayli fazla olan organlardır [18].

Tip1 diyabet hastalığı, pankreasta bulunan hücrelerin görevini yerine getirememesinden kaynaklı insülin eksikliği sonucunda genelde erken çocukluk ve gençlik dönemlerinde görülebilen bir diyabet tipidir [2]. Tip 2 diyabet, sıklıkla 30 yaş ve üstünde görülen, insülin hormonuna karşı vücudun direnç göstermesi ya da insülin emiliminin yeterince gerçekleşmemesi durumunda ortaya çıkan bir diyabet türüdür. Dünya da ve Türkiye de en sık görülen tiplerdendir [3]. Diyabetin erken dönemde tespit edilmesi günümüzde hızla artan diyabet hastası sayısını düşürmek için oldukça önemlidir.

Diyabetin erken tespitinde, yaş, vücut kitle endeksi, aşırı sıvı kaybı vb. veriler diyabetin erken tespit edilmesi için doktorlara kolaylıklar sağlamaktadır. Son zamanlarda teknolojinin sağlık alanında da kullanılmaya başlaması, hastalıkların erken teşhisinde aktif olarak kullanılmasını ve makine öğrenme algoritmalarının da bu noktada uzman kişilere yardımcı olabilecek birer seçenek olduğu ortaya çıkmıştır.

Tıp alanındaki birçok teşhis, elektronik ve bilişim alanındaki gelişmeler sonucunda insan vücudundan alınan biyo-elektriksel sinyaller veya kan değerleri ile rahatlıkla ölçülmekte ve farklı alanlara disiplinler arası bir çalışma yapılmasına olanak tanımaktadır [4]. Makine öğrenmesi ya da yapay zekâ tam bu noktada birbiri ile ilişkileri zayıf ya da üst düzey olan noktalar arasında gerekli bağları kurabilir ve veri kümelerinden yeni sonuçların elde edilmesi noktasında bizlere yardımcı olabilir [5]. Bu sayede direkt göz ile görülemeyen ve aralarındaki bağlantı tespit edilemeyen durumlar kolaylıkla tespit edilebilir, örüntüler kurulur ve süreç hızlanır. Tüm bu gelişmeler beraberinde sağlık alanında birçok hastalık ve teşhisin yapay zekâ ve makine öğrenmesi yöntemleri ile daha kısa sürede ve yüksek bir doğruluk oranı ile çözülebilmeye yardımcı olmaktadır.

Bu çalışmada, Makine öğrenmesi tekniklerinden sınıflandırma algoritmalarından olan 5 farklı tekniğin (KNN, Lojistik Regresyon, Karar Ağaçları, DVM, Random Forest), Erken evre diyabet risk tahminindeki performanslarının karşılaştırılması yapılmıştır. Elde edilen tutarlı, yüksek doğruluk ve F1 skoruna sahip algoritmaların, hastalığın erken teşhisindeki başarı oranları incelenmiştir.

2. LİTERATÜR ÇALIŞMALARI

Early stage diabetes risk prediction veri seti, 2020 yılında Bangladeş de bulunan Sylhet Diabetes hastanesindeki hastaların kayıtlarından elde edilen verilerden yeni diyabetik hasta teşhisi için hazırlanmış bir veri setidir. Yakın zamanda hazırlanmış olan bu veri seti çıkış olarak 1 ve 0 verisinden oluştuğu için makine öğrenmesi sürecinde sınıflandırma algoritmalarında sıkça kullanılmaktadır. Literatürde bu veri seti kullanılarak yapılan bazı çalışmalar şunlardır;

Ö. N. Ergün, H. O. İlhan, Makine öğrenmesi algoritmalarının, hayatlarını doğrudan etkileyen, tedavisi olmayan ve ölümcül bir hastalık olan diyabetin kişilerin hayat standartlarını yükseltmeye faydalı olabileceğini göstermek için 8 farklı makine öğrenmesini kullanarak, hangi makine öğrenmesi algoritması tekniğinin hastalığın erken evre teşhisinde kullanılmasının daha hızlı ve yüksek başarı oranlı sonuç verdiğini test etmiştir. Yapılan bu çalışma diyabet hastalığı riskinin hastalığın daha erken evrelerinde tespit edilebilmesinde yapay zekâ ve makine öğrenmesi algoritmalarının başarı oranlarını göstermiştir. Yapılan çalışmada en iyi sonuç evrimsel sinir ağlarını tek boyutlu olarak kullanılarak elde edilmiştir. Yapılan çalışmada başarı oranı %99,04 olarak elde edilmiştir [6]

G. Bilgin, Türkiye’de son dönemlerde görülme sıklığı artan, insan vücudundaki birçok yapıya zarar veren bir hastalık olan diyabetin, makine öğrenmesi süreçleri ile tespitinin insan sağlığına zarar vermeden en az başarılı oranda tespit edilmesinin makine öğrenmesi süreçleri ile sağlanacağını gösterdiği çalışmada, Erken evre diyabet riskini tahmin edebilmek için makine öğrenmesi algoritmalarını kullanarak bir tanı kiti geliştirmiştir. Geliştirilen tanı kiti, 6 farklı algoritmanın karşılaştırılması sonucu elde edilmiştir. Benzer birçok çalışmada veri setinde var olan 16 tanımlayıcı 1 tane de sınıf belirten özneliğin 9 veya daha üstü seçilerek yapılırken, bu çalışmada tüm öznelikle sürece dahil edilmiş ve farklı algoritmalar kullanarak elde edilen sonuç skorları değerlendirilmiştir tanı kiti hazırlanırken. Çalışmada kNN sınıflandırma algoritması %99,81 doğruluk oranına ve %99,83 F1 skorunu elde etmiştir [7].

A. Yasar, Erken evre diyabet teşhisi için, Veri setine ön işleme ve öznelik çıkarma yöntemleri uygulayarak, farklı Wrapper metodlarının, makine öğrenmesi yöntemleri ile performanslarını analiz etmeye çalışmıştır. Çalışmada MATLAB programı kullanılmış, Öznelik seçim için PSO, CSA, TSA, SMA, ABC gibi Wrapper yöntemler ile 16 farklı öznelik bulunan veri setinden farklı sayıda ve

özelliğe yeni öznelikler elde edildi. Elde edilen her bir öznelik DT, RF, SVM, kNN, FFNN gibi makine öğrenmesi algoritmaları ile karşılaştırılmıştır. Elde edilen en başarılı sonuçlar, PSO+SVM=97,5 ile CSA+FFNN = 99,04 şeklindedir [8].

R. Ferdousi, M. E. Hossain, A. E. Saddik, Bulaşıcı olmayan hastalıkların erken evre risk tahminlemesi üzerine yaptıkları çalışmada, IOT teknolojisini veri seti ile entegre ederek dinamik olarak insanlardan alınacak veriler ile bu hastalıkların hangi evrede olduğuna dair dönüt alacakları giyilebilir bir IOT ve makine öğrenmesi çalışması yaptılar. Yapılan çalışmada algoritmaların tepki süreleri ve sonuçların başarı oranları ayrıca önemli idi. Kullanılan sensörlerden anlık olarak alınan veriler 0,1 saniye gibi kısa bir sürede işlenmektedir. Yapılan çalışma sonucunda RF, DT, MLP, RT ve KNN algoritmaları %91-94 arası başarılı sonuçlar elde edilmesini sağladı [9].

S. Pate, R. Pate, N. Ganatra, A. Pate, Makine öğrenmesi yaklaşımı ile veri seti üzerinde erken evre diyabet tespiti çalışmasında, 4 farklı makine öğrenmesi algoritması kullanarak erken evre tespit çalışması yaptı. Yapılan çalışmalarda elde edilen sonuçlara bakıldığında Random Forest algoritması %97,5 gibi bir oran ile en yüksek başarıyı gösteren algoritma oldu [10].

Sonar, P. ve Jaya Malini, K., 2019 yılında yayınladıkları çalışmada, diyabet risk olasılığını daha doğru şekilde tahmin edebilmek için, 4 farklı sınıflandırma (Decision Tree, ANN, Naive Bayes ve SVM) algoritmasından elde edilen sonuçları karşılaştırdı. Yapılan çalışma sonucunda elde edilen sonuçlardan Decision Tree algoritmasının %85 oranı ile en yüksek başarıyı elde ettiği görüldü [19].

Güneş Harman, 2021 yılında Python dilini kullanarak destek vektör makinesi ile Naive Bayes algoritmalarını karşılaştırmış ve diyabet hastalığını erken evre tespit etmeye çalışmıştır. Smote tekniği ile sınıflandırma işlemi üzerinde olumsuz etki bırakacak etmenleri en aza indirgeyerek başarılı sonuçlar elde etmeye çalışmıştır. Yapılan çalışma sonucunda Destek Vektör Makinaları algoritması diğer Makine öğrenmesi algoritmasına göre daha başarılı bir sonuç vermiştir. Destek vektör makinalarının başarı oranı %90'dır[20].

3. VERİ SETİ

Kullandığımız veri kümesi, 520 kişiye ait diyabetle ilgili 17 farklı semptomun bilgilerini içermektedir. Veri seti'nin hazırlanması sırasında, diyabet teşhisine yönelik semptomlar erken evrede diyabet teşhisi konulan hastalardan doğrudan elde edilmiştir. Data seti, Bangladeş Sylhet Diyabet Hastanesinde hastalara uygulanan bir anket yöntemi kullanılarak toplanmıştır [1, 17].

Veri kümesinin içeriği Çizelge 1'de verilmiştir. Veri setinde toplamda 520 hasta verisi bulunmakta ve bu hastaların yaş aralıkları ekstra bir gruptandırmaya alınmamıştır, diğer veriler ise sadece 1 ve 0 verileri olarak gruplara ayrılmıştır. 1 verisi veri seti içerisinde evet cevabı ile özdeşleştirilmişken 0 verisi hayır cevabına karşılık olarak seçilmiştir.

Çizelge 1. Veri Seti İçeriği.

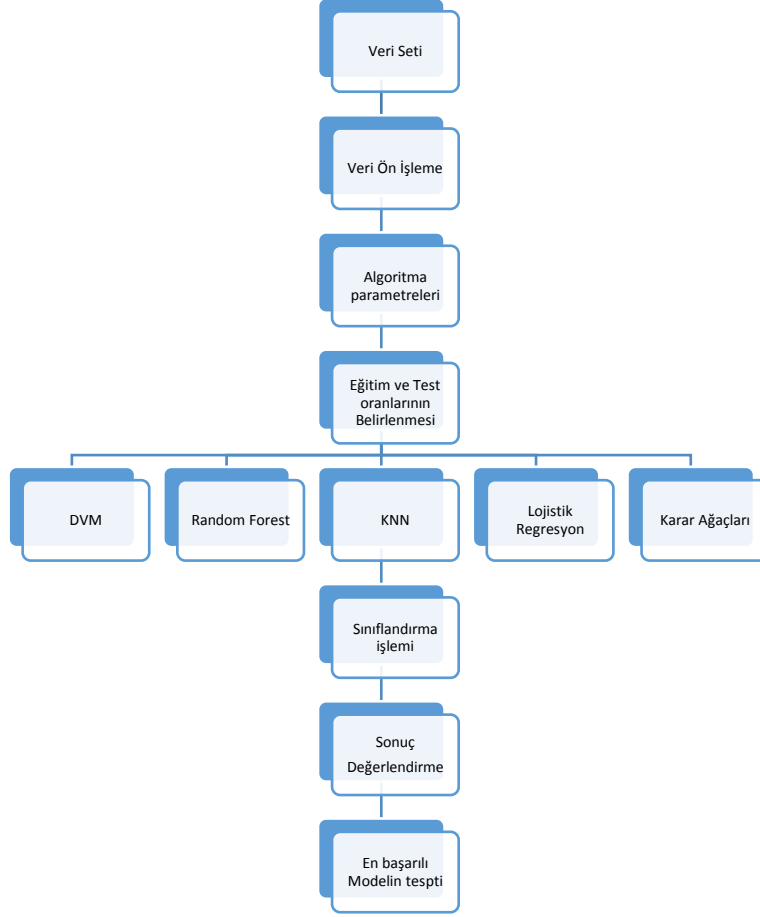
	Veri		Değerler
1	Yaş	20-65	
2	Cinsiyet	1.	Erkek- 0. Kadın
3	Poliüri	1.	Evet – 0. Hayır
4	Polidipsi	1.	Evet – 0. Hayır

5	Ani Kilo Kayıpları	1.	Evet – 0. Hayır
6	Halsizlik	1.	Evet – 0. Hayır
7	Polifaji	1.	Evet – 0. Hayır
8	Genital Bölgede Pamukçuk	1.	Evet – 0. Hayır
9	Gözlerde Bulanık görme	1.	Evet – 0. Hayır
10	Kaşınma	1.	Evet – 0. Hayır
11	Sinirlilik	1.	Evet – 0. Hayır
12	Geç iyileşme	1.	Evet – 0. Hayır
13	Kısmi parezi	1.	Evet – 0. Hayır
14	Kas sertleşmesi	1.	Evet – 0. Hayır
15	Alopesi	1.	Evet – 0. Hayır
16	Obezite	1.	Evet – 0. Hayır
17	Teşhis	1.	Pozitif – 0. Negatif

4. YÖNTEMLER

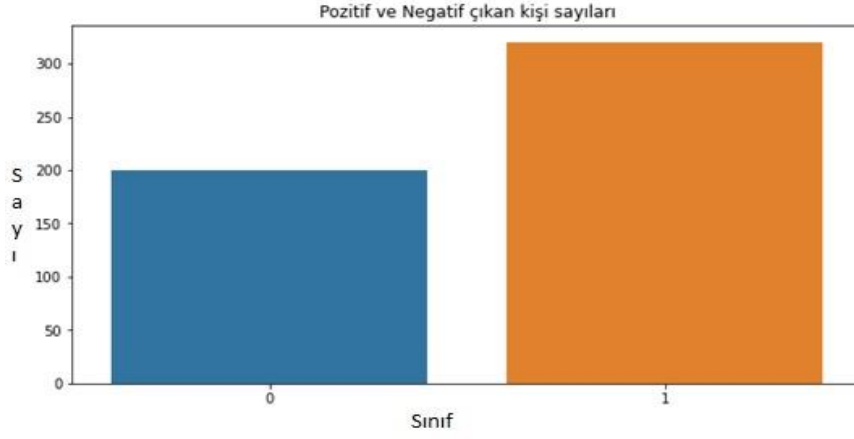
4.1. Veri Önışleme

Veri setine önışleme yapılarak eksik, dađınık ya da yanlış tanımlanmış verilerin tespiti denendi fakat herhangi bir eksik veriye rastlanmadı. Çizelge 1’de verilen bilgiler paralelinde veri seti üstünde gerekli dönüşümler yapılarak verinin işlenebilir şekilde olması sağlandı.

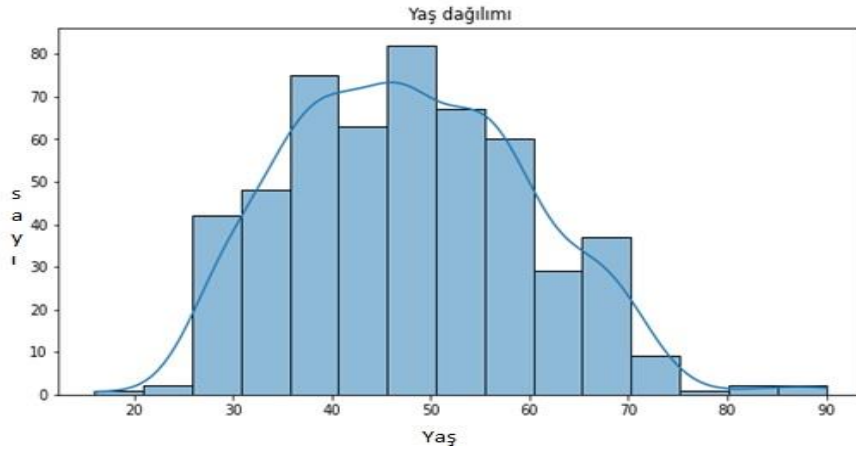


Şekil 1. Veri setine uygulanan işlem basamakları.

Bu çalışmada veri setinde bulunan bazı parametreler istatistik açısından daha çok dikkate alındı. Bunlar, Cinsiyet, yaş dağılımı ve hastalığın pozitif ve negatif olma durumları olarak grafiğe döküldü. **Şekil 1**'de veri setinde var olan kadın erkek sayısının hastalık oranları verilmiştir.

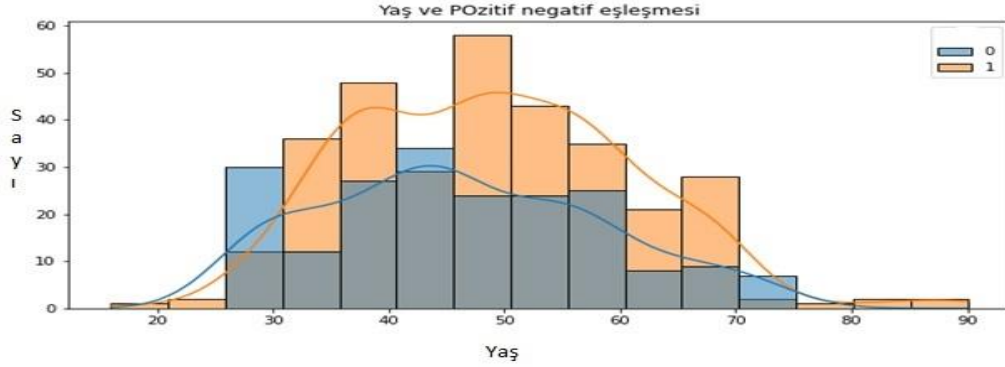


Şekil 2. Diyabet sonuçları pozitif ve negatif çıkan kişi sayıları.



Şekil 3. Diyabet veri setindeki kişilere ait yaş dağılımı.

Şekil 2’de veri setinde var olan 520 kişiye ait yaş dağılım grafiği figsize = (10,5) belirlenerek elde edildi ve yine aynı şekilde Şekil 3’te yaşa göre kadın ve erkek pozitif ve negatif eşleşme oranları verilmiştir.



Şekil 4. Yaş ve pozitif-negatif korelasyonu.

4.2. Sınıflandırma Yöntemleri

Veri setinin üstünde sınıflandırma algoritmalarının bazılarını sırası ile test ederek aralarından en yüksek Doğruluk, Hassasiyet, Kesinlik ve F skor oranı elde edilenlerin kendi aralarında sıralanması amaçlanmıştır. Sıralama işlemini yapmadan önce veri setine uygulanan sınıflandırma algoritmaları için standart olarak hepsinde eğitim ve test veri setlerinin oranlarını 0,3 olarak ayarlanmıştır. Bu şekilde verilerden %30'u test iken kalan %70'i eğitim veri seti olarak kullanılmıştır. Veri setine sırası ile KNN, Lojistik regresyon, Karar Ağaçları, Destek Vektör Makineleri (DVM), Random Forest algoritmaları sırası ile uygulanmış elde edilen skorlar kayıt altına alınmış ve Çizelge 2'de gösterilmiştir.

4.2.1. K-nearest neighbors algoritması

K-Nearest Neighbors algoritması, bağımsız değişkenler tarafından oluşturulan vektörlerin komşularının, hangi sınıfta daha fazla yoğunluk gösterdiği üzerine tahmine dayalı çalışır. K en yakınlık değeri genelde tek sayı seçilir [11]. Çalışmada K komşuluk değeri ($k_{neighbors}=3$) 3 olarak seçilmiştir. Komşuluk ve uzaklık değerleri algoritmanın çalışma performansını etkileyen 2 farklı ana değerdir. Bulunan K değerinin diğer değerlere olan uzaklıkları farklı şekillerde ölçülebilir. Öklid uzaklık ölçümü de bu değerlerden biridir [12]. Denklem 1'de KNN algoritmasının mesafe hesaplama işleminin ifadesi gösterilmiştir.

$$Mesafe = \sqrt{(X_i - X_{Yeni})^2 + (Y_i - y_{Yeni})^2} \quad (1)$$

Denklem 1'den de anlaşılacağı üzere uygun seçilen K değeri, öklit uzaklığı olarak yeni gelen bir değer için hangi sınıfa gireceğini göstermektedir.

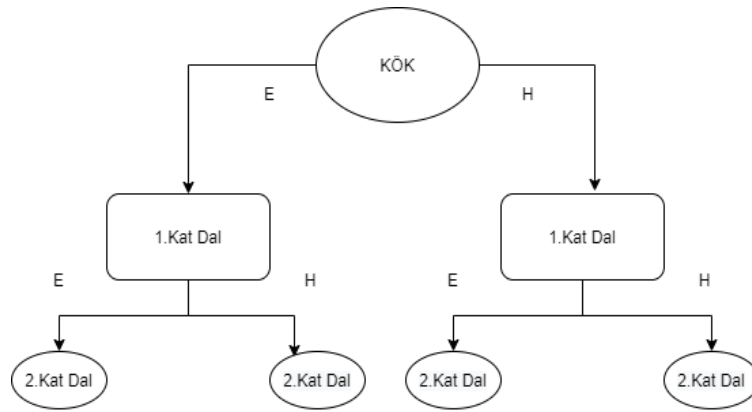
4.2.2. Lojistik regresyon algoritması

Birden fazla bağımsız değişkenin sonucu belirlediği veri kümelerini analiz için kullanılan istatistiksel bir yöntemdir. Doğru-Yanlış, Evet-Hayır gibi direkt sonuç bildiren kategorik çıktıya sahip verilerde işlem yapan bir Regresyon örneğidir. Lojistik regresyon Maksimum olasılık ile lineer regresyondan ayrılır [13]. Sigmonoid fonksiyonu, Lojistik regresyonun 0 ile 1 arasındaki gösterimi için kullanılır. Denklem 2'de bir Sigmonoid fonksiyonu görülmektedir.

$$G(z) = \frac{1}{1+e^{-z}} \quad (2)$$

4.2.3. Karar ağaçları algoritması

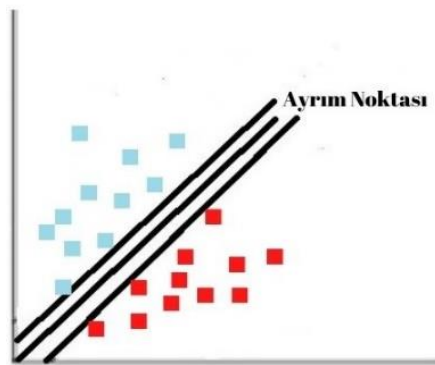
Parametrik olmayan makine öğrenmesi algoritmasıdır. Ters duran bir ağacı andıran karar algoritmaları, Kökten alt dallara ve alt dallardan sonra çoğalan yeni adımlara uzanan bir yapıdadır. Her oluşan yeni dal gövdede bağlı olduğu ana dalın özelliklerini taşır Şekil 5'te gösterilmiştir. Veri setinden seçilen bir kolondan yola çıkılır. Seçilen bu kolondan elde edilen veriler tüm ağaca ya da veri setine uygulanır [14].



Şekil 5. 2 dallı karar ağacı algoritması gösterimi.

4.2.4. Destek vektör makineleri algoritması

Destek vektör makineleri algoritması, uzay vektör alanında bulunan 2 farklı sınıfa birbirinden ayırmak için bir karar sınırı oluşturan makine öğrenme algoritmasıdır [15]. Algoritmadaki amaç 2 farklı sınıfa ait olan verileri birbirinden ayırıştırın margin noktası, arttıkça veriler daha homojen bir dağılım gösterir ve ayrışma sağlar. Şekil 6'da bir DVM algoritmasının sonuç değişkenini margin ile ayrımı gösterilmektedir.



Şekil 6. DVM sonucu oluşan margin.

4.2.5. Random forest algoritması

Karar ağaçları algoritmasının anlaşılması, Random Forest algoritmasını daha iyi anlamamıza yardımcı olur. Birden fazla karar ağacı kullanarak daha uyumlu modeller üretilmesine fırsat tanır, bu şekilde elimizdeki veri setinin daha isabetli bir sınıflandırma yapmasını bu model sayesinde sağlayabiliriz [16].

5. PERFORMANS DEĞERLENDİRME

Toplam 520 kişiye ait 17 farklı değer bulduğu makine öğrenmesi algoritmasından elde edilen sonuçlar Çizelge 2’de verilmiştir. Her algoritmadan elde edilen karışıklık matrisi değerlerine göre, yöntemlerin kesinlik, hassasiyet ve F1 skorları elde edilerek başarı oranları karşılaştırması yapılmıştır.

Çizelge 2. Veri setine uygulanan sınıflandırma algoritmalarından elde edilen sonuçlar.

Yöntemler	Accuracy (Doğruluk)	Precision (Kesinlik)	Recall (Duyarlılık)	F1 Score (F1 Skoru)
KNN	0,91	0,85	0,92	0,88
Lojistik regresyon	0,89	0,84	0,86	0,8498
Karar Ağaçları	0,96	0,84	0,94	0,94
DVM	0,95	0,97	0,89	0,92
Random Forest	0,96	0,94	0,94	0,94

6. TARTIŞMA

Makine öğrenmesi birçok alanda bizlere karar verme ve uygulama noktasında elimizi güçlendirecek argümanlar sunmaktadır. Tıbbi alanlar başta olmak üzere borsa ve kripto para gibi birden fazla değişkenin sürece dahil olduğu durumlarda bu tarz makine öğrenmeleri süreç içerisinde insanların zarar görmelerini ve yakın zamanda meydana gelebilecek durumları kestirmeleri açısından önem arz etmektedir.

Erken evre diyabet hastalığı da günümüzde tedavisi olmayan erken evrede teşhis sonucu yaşam kalitesini artıran bir hastalıktır. Kalıtsal boyutunun yanı sıra özellikle genç çocukluk dönemlerinde ortaya çıkan Tip-1 diyabet kişi sağlığını etkilemekten öte zamanla uzuv kaybına da neden olmaktadır. Makine Öğrenmesi, sağlık alanında çalışan personele, gizli şekilde varlığını sürdüren, teşhisi zaman alan diyabet hastalığının en kısa sürede ve en yüksek başarı oranları ile teşhis konması için imkân sunmaktadır bizlere. Uyguladığımız algoritmalar incelendiğinde %87 ile %96 arasında yüksek bir oranda başarı elde edilmiştir. Elde edilen sonuçlar %80 - 20 test eğitim oranları ile elde edilmiştir. Makine öğrenmesi sürecinde eğitim test oranlarını farklı seviyelerde denemiş ve en yüksek oranın 80/20 oranında olduğu görülmüştür. Çalışmanın devamında çapraz doğrulama sayısını artırarak elde edilecek başarı oranlarında artış gözlemlenebilir. Yapılan çalışma bizlere makine öğrenmesi algoritmalarının diyabet alanında teşhis konusunda elde ettiği başarıların, gündelik hayatta da aktif şekilde kullanılabileceğini, yapay zekâ ile teşhis tanı kitleri oluşturmasıyla kullanımının artacağı göstermiştir.

7. SONUÇLAR

Bu çalışma kapsamında veri setine uygulanan 5 farklı sınıflandırma yöntemi yardımıyla elde edilen doğruluk oranları kıyaslanmış ve elde edilen bulgulara göre en yüksek sınıflandırma başarısı %96 oranla DVM ve Random Forest yöntemlerinde elde edilmiştir. Ayrıca Lojistik regresyon kullanılarak elde edilen %89'luk doğruluk oranı da çalışma kapsamında elde edilen en düşük orandır. Yani 5 farklı sınıflandırma metodu %89-%96 arasında bir doğruluk oranı skalası oluşturmuştur. Bu çalışmada özellikle başarı oranları yüksek çıkacak algoritmalar yerine, topluluk öğrenmesine yatkın, uzay vektör bağlamları yüksek algoritmaları tercih edilmiştir. Kullanılan algoritmaların sınıflandırma performanslarının yüksek oluşu ve başarı oranlarındaki keskinlik ve F skorları bizlere bu algoritmaların en ideal algoritmalar olduğunu göstermiştir. Literatürde yapılan diğer karşılaştırmalarda genelde tek algoritma üstünden süreçler incelenmiş ve algoritmanın başarı oranı tek bir açıdan irdelenmiştir. Çalışmada 5 farklı algoritmanın hepsinde yüksek doğruluk elde edilecek derecede eğitim-test ve komşuluk derecesi gibi ayarlar yapılmıştır.

Yapılan bu çalışma bizlere gösteriyor ki, sağlık alanındaki birçok durumda artık teşhis ve yorumlama süreçleri yerine makine öğrenmesi ve yapay zekâ yöntemleri yer alacaktır. Bunlar kanser gibi, hızlı yaş tahmini, damar sorunları vb. birçok farklı sağlık alanında günden güne daha sık kullanılarak hastalıkların erken teşhisinde önemli bir yol kat edecek hale gelecektir. Hali hazırda günümüzde biyomedikal alanda yapay zekâ ve makine öğrenmesi algoritmaları kullanarak birçok proje yapılmakta ve yapılmaya devam etmektedir. Bu çalışmalar ilerleyen zamanlarda da sağlık alanında birçok alanın makine öğrenmesi algoritmaları ile tespit edilmesini sağlamıştır.

TEŞEKKÜR

Bu makale çalışması Siirt Üniversitesi Mühendislik Fakültesi İnsan Bilgisayar Etkileşimi Laboratuvarında gerçekleştirilmiştir. İnsan Bilgisayar Etkileşimi Laboratuvarı personeline destekleri için ayrıca teşekkür ederim.

KAYNAKÇA

- [1] Islam, M. M., Ferdousi, R., Rahman, S., and Bushra, H. Y. (2020). Likelihood prediction of diabetes at early stage using data mining techniques. In Computer vision and machine intelligence in medical image analysis (pp. 113-125). Springer, Singapore.
- [2] Gregg, E. W., Cheng, Y. J., Srinivasan, M., Lin, J., Geiss, L. S., Albright, A. L., and Imperatore, G. (2018). Trends in cause-specific mortality among adults with and without diagnosed diabetes in the USA: an epidemiological analysis of linked national survey and vital statistics data, *The Lancet*, 391(10138), 2430-2440.
- [3] Kılınc, E. (2022). TİP 2 Diyabetli yetişkinlerde bilgi, motivasyon ve davranış becerileri modeli temelli diyabet eğitimi ve motivasyonel görüşmenin bakım sonuçlarına etkisi: Randomize kontrollü çalışma.
- [4] Demir, N., Kuncan, M., Kaya, Y., and Kuncan, F. (2022). Multi-Layer Co-Occurrence Matrices for Person Identification from ECG Signals, *Traitement du Signal*, 39(2).

- [5] Karakoyun, M., and Hacibeyoğlu, M. (2014). Biyomedikal Veri Kümeleri ile Makine Öğrenmesi Sınıflandırma Algoritmalarının İstatistiksel Olarak Karşılaştırılması, Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 16(48), 30-42.
- [6] Ergün, Ö. N., and İlhan, H. O. (2021). Early Stage Diabetes Prediction Using Machine Learning Methods. Avrupa Bilim ve Teknoloji Dergisi, (29), 52-57.
- [7] Bilgin, G. (2021). Makine öğrenmesi algoritmaları kullanarak erken dönemde diyabet hastalığı riskinin araştırılması, Journal of Intelligent Systems, Theory and Applications, 4(1), 55-64.
- [8] Yasar, A. (2021). Data Classification of Early-Stage Diabetes Risk Prediction Datasets and Analysis of Algorithm Performance Using Feature Extraction Methods and Machine Learning Techniques, International Journal of Intelligent Systems and Applications in Engineering, 9(4), 273-281.
- [9] Ferdousi, R., Hossain, M. A., and El Saddik, A. (2021). Early-stage risk prediction of non-communicable disease using machine learning in health CPS, IEEE Access, 9, 96823-96837.
- [10] Patel, S. (2021). Predicting a risk of diabetes at early stage using machine learning approach, Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(10), 5277-5284.
- [11] Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M., and Özçift, A. (2016). KNN algoritması ve r dili ile metin madenciliği kullanılarak bilimsel makale tasnifi, Marmara Fen Bilimleri Dergisi, 28(3), 89-94.
- [12] Guo, G., Wang, H., Bell, D., Bi, Y., and Greer, K. (2003). KNN model-based approach in classification. In On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings (pp. 986-996). Springer Berlin Heidelberg.
- [13] Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., and Klein, M. (2002), Logistic regression, New York: Springer-Verlag (p. 536).
- [14] Quinlan, J. R. (1996). Learning decision tree classifiers. ACM Computing Surveys (CSUR), 28(1), 71-72.
- [15] Pisner, D. A., and Schnyer, D. M. (2020). Support vector machine. In Machine learning (pp. 101-121). Academic Press.
- [16] Hegde, C., Wallace, S., and Gray, K. (2015). Using trees, bagging, and random forests to predict rate of penetration during drilling. In SPE Middle East Intelligent Oil and Gas Conference and Exhibition. OnePetro.
- [17] UCI Machine Learning Repository. (2020, July 12). Early stage diabetes risk prediction dataset. <https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset>.

- [18] Diyabet nedir, Eriřim Adresi: <https://www.medicalpark.com.tr/seker-hastaligi-diyabet-nedir/hg-1703> (Ziyaret Tarihi:04.11.2022)
- [19] Sonar, P., and JayaMalini, K. (2019, March). Diabetes prediction using different machine learning approaches. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 367-371). IEEE.
- [20] Harman, G. (2021). Destek Vektör Makineleri ve Naive Bayes Sınıflandırma Algoritmalarını Kullanarak Diabetes Mellitus Tahmini. *Avrupa Bilim ve Teknoloji Dergisi*, (32), 7-13.