



Makine Öğrenmesi Enfeksiyonu Önleyebilir Mi? Sağlık Hizmetleri Meslek Yüksekokulu Öğrencilerinin Enfeksiyon Hastalıklarına Yatkınlık Değerlendirmesi*, **

Fehmi Mete ARISOY, **Ümit KARAKAŞ, ***Mustafa Serhat ŞAHİNOĞLU, *Sevil ALKAN, *****Hatice ÖNTÜRK AKYÜZ, *****Hamit Emre KIZIL**

Öz

Son yıllarda, makine öğrenmesi (MÖ) teknolojilerinden tıp ve halk sağlığı araştırmalarında yaygın olarak yararlanılmaktadır. MÖ algoritmaları, klinik veri tabanlarındaki çoklu, karmaşık değişkenler arasındaki etkileşimleri analiz etmede ve doğru tahminler yapmada oldukça etkilidir. Bu çalışmada, MÖ tabanlı tahmin modellerini oluşturmak için etik onayı alınmış, yaş, cinsiyet, önceki enfeksiyon bilgileri ve enfeksiyon algılarından oluşan veriler sağlık hizmetleri meslek yüksekokulu öğrencilerinden toplandı. Python yazılım dili kullanılarak makine öğrenmesi altında Logistic Regression, Random Forest, Decision Tree ve XGBoost modelleri kullanıldı. Verilerin modeller ve makine öğrenmesi analiz sonuçlarına göre en etkili makine öğrenme modeli %74.7'lik tahmin başarı oranı ile Logistic Regression modeli olduğu belirlendi. Bu model kullanılarak bireylerin enfeksiyona yakalanmaya yatkınlıkları yüksek bir doğrulukla hesaplanabildiği gösterildi. Dolayısıyla makine öğrenmesi kullanılarak profesyonel sağlık personeli olarak çalışacak ya da mesleki eğitim alacak genç bireylerin enfeksiyona yatkınlıkları ya da yeterli bilgi birikimine sahip olup olmadıkları hızlıca belirlenebilecektir. Böylelikle çalışma ortamlarında, enfeksiyona yakalanma ihtimallerinin en aza indirgenmesinin eğitim ile önüne geçilmeye çalışılacaktır. Modele veri girdisi devam ettikçe tahmin yüzde başarısının artacağı da göz önünde bulundurulmalıdır.

Anahtar kelimeler: *Enfeksiyon, makine öğrenmesi, lojistik regresyon, sağlık eğitimi*

Can Machine Learning Prevent Infection? Assessment of Health Services Vocational School Students' Susceptibility to Infectious Diseases

Abstract

Machine learning (ML) Technologies have been widely utilized in medical and public health research in recent years. ML algorithms effectively analyze interactions between multiple, complex variables in clinical data bases and make accurate predictions. In this study, ethical approval was obtained to build ML-based prediction models, and data on age, gender, previous infection knowledge, and perceptions of infection were collected from students at a vocational school of health services. Logistic Regression, Random Forest, Decision Tree, and XGBoost models were used in machine learning using Python software. According to the data models and machine learning analysis results, the most effective machine learning model was the Logistic Regression

model, with a prediction success rate of 74.7%. Using this model, it was shown that the susceptibility of individuals to infection can be calculated with high accuracy. Therefore, by using machine learning, it will be possible to quickly determine whether young individuals working as Professional healthcare personel or receiving vocational training are susceptible to infection or have sufficient knowledge. This way, it will be tried to minimize the possibility of contracting infections in their working environment through training. The prediction percentage success should also increase as data input to the model continues.

Keywords: Health education, infection, machine learning, logistic regression

Geliş Tarihi: 15.11.2023

Kabul Tarihi:19.12.2023

GİRİŞ

Avrupa Hastalık Önleme ve Kontrol Merkezi (ECDC) Raporu, birinci basamakta enfeksiyon önleme ve kontrol önlemlerini ele almaktadır. Bu rapora göre tüm personelin enfeksiyon hastalığı ile ilgili semptom kontrolünden geçirilmesi, hastalığın önemi yeni çalışanların eğitimine ve güncellenmesine önem verilerek tüm çalışanların güncel tutulması gerektiği vurgulanmaktadır. TC Aile Sağlığı Merkezleri (ASM), COVID-19 Pandemisi Sırasında Sağlık Tesislerinde Çalışma ve Enfeksiyon Kontrol Tedbirleri Rehberi Bölümü, eğitimleri tavsiye etmektedir. Ayrıca eğitimin kayıt altına alınması ve Sağlık Bakanlığı tarafından verilen materyallerle uyumlu olması gerektiğini ifade etmişlerdir (Özer & Özcan, 2020). Çünkü enfeksiyonun önlenmesi ve kontrol edilmesine yönelik uygulamaların sağlık hizmetlerinin sürdürülmesinde ve enfeksiyonun toplum üzerindeki etkilerini azaltmada kritik rolü olduğu bilinmektedir.

Son yıllarda, makine öğrenmesi (MÖ) teknolojilerinden tıp ve halk sağlığı araştırmalarında yaygın olarak yararlanılmıştır. MÖ algoritmaları, klinik veri tabanlarındaki çoklu, karmaşık değişkenler arasındaki etkileşimleri analiz etmede ve doğru tahminler yapmada oldukça etkilidir (Chiu et al., 2022). Veri tabanı taramalarında sağlık alanında “makine öğrenmesi” içeriğine ait çok sayıda makale bulunmaktadır. Bu çalışmalarda özellikle enfeksiyon hastalıkları, ilaç kullanımı dozajları ve etkileri, kanser teşhisi gibi insan hayatında yer eden birçok sağlık durumunu tahmin eden modeller dikkat çekmektedir (Keshavamurthy, Dixon, Pazdernik, & Charles, 2022; Peiffer-Smadja et al., 2020).

Bu çalışmada, enfeksiyon hastalıklarının getireceği zorlukları koruyucu hekimlik açısından ele almayı amaçlamaktayız. Buna göre, enfeksiyondan önce ciddi hastalık riski altındaki sağlık profesyonelleri popülasyonu belirlemek için MÖ tabanlı tahmin modellerini oluşturmak için yaş, cinsiyet, önceki enfeksiyon bilgileri ve enfeksiyon algılarından oluşan verileri dahil etmeyi planlamaktayız. Elde edilecek MÖ modellerinin, sağlık politika yapımcılarının yüksek riskli popülasyonları belirlemesi ve ardından buna göre öncelikli bir eğitim stratejisi geliştirmesi açısından önemli değerdedir. Son COVID-19 pandemisinde, topluluk salgınları olan hemen hemen tüm ülkelerde, sağlık sistemlerinin çökmesi nedeniyle benzeri görülmemiş ölümler yaşanmıştır (Chiu et al., 2022). Gelecekteki benzer salgın senaryolarında, ön saflardaki sağlık profesyonelleri için tahmin modellerimiz eğitimde ve/veya aşılama öncelikli bireyleri tahmin etmede etkili olabilecektir.

Makine öğrenimi alanı hızla ilerlemektedir. Makine öğrenmesi, olaylar ve deneyimler yoluyla sürekli olarak öğrenmekte ve hem becerisini hem de karar verme yeteneğini geliştirmektedir (Mahmoud, Al Dhoayan, Bosaeed, Al Johani, & Arabi, 2021). Makine öğreniminin anket sorularına verilen cevapları analiz etme yoluyla sağlık hizmetleri öğrencilerinin ve dolayısıyla gelecekte birer sağlık profesyoneli olan bireylerin herhangi bir bulaşıcı hastalığa yakalanma ihtimalini tahmin etme doğruluğunu artıracaklarını varsaymaktayız. Çalışmamızın ilk amacı; bu anket verilerini kullanarak sağlık çalışanının enfeksiyona yakalanma riskini tahmin etmektir. Çalışmanın ardıl amacı ise enfeksiyonun önlenmesi ve kontrol edilmesine yönelik uygulamaların sağlık hizmetlerinin sürdürülmesine katkıda bulunmak için makine öğrenmesi modelini oluşturmaktır. Bu çalışma sonuçlarının yaygınlaştırılması ile başta sağlık çalışanlarının enfeksiyondan korunma ile ilgili bilgilerinin güncellenmesi konusunda farkındalık oluşması ve enfeksiyon zincirinin bulaşma yolu ve kaynak basamaklarında kırılması hedeflenmektedir.

YÖNTEM

Veri Toplama Araçları/Teknikleri

Örneklem büyüklüğü seçilmeden, Sağlık Hizmetleri Meslek Yüksekokulu öğrencilerinden 18 yaş ve üzeri gönüllüler çalışmaya dâhil edildi. Etik onay alındıktan sonra (Bitlis Eren Üniversitesi Etik İlkeleri ve Etik Kurulu 2022/10-3 sayılı ve E.2450 sayılı karar) bilgilendirilmiş olur formunu

onaylayan sağlık hizmetleri meslek yüksekokulu öğrencileri ile çevrim içi iletişime geçildi. Araştırma verileri 15 Temmuz 2022 ile 15 Ağustos 2022 tarihleri arasında sosyal medya (Whatsapp, Facebook, Twitter) çevrimiçi anketi aracılığıyla toplandı. Anket formu Google Formlar aracılığıyla oluşturuldu ve sosyal medya aracılığıyla öğrencilere gönderildi. Kişilerin isim ve kimlik bilgileri yer almıyordu. Farklı sağlık hizmetleri alanları ve akademik yıllar arasında temsili sağlamak için amaca yönelik bir örnekleme tekniği kullanıldı. Çalışma örnekleme 873 öğrencinin katılımı ile oluşturuldu.

Makine Öğrenmesi

MÖ modelleri kullanarak sağlık alanında enfeksiyon tahminleri yapan bir model planlandı. Anket formu aracılığıyla toplanan veriler Python yazılım dili kullanılarak makine öğrenmesi altında Logistic Regression, Random Forest, Decision Tree ve XGBoost modelleri kullanıldı. Bu modellerin tercih edilme sebebi, verilerin bu modellerin arkasındaki yatan formüllere daha uygun olması, geliştireceğimiz yapay zekânın bulunacağı tahminlerin daha yüksek skorlar elde edilebilme ihtimalidir. Analiz sonrası modellerden alınan skorlar birbirleriyle kıyaslandı ve en iyi skorun elde edildiği Logistic Regression modeline göre veriler modelin enfeksiyona yakalanma skor tahminini yapabilmesi için hazır hale getirildi (verilerin deploy edilmesi) (Keshavamurthy et al., 2022; Peiffer-Smadja et al., 2020).

BULGULAR

Çalışma anketini cevaplayan 873 kişinin 650'si kadın, 224'ü ise erkektir. Çalışmanın verilerinde genelleme yapabilmek adına çalışmaya katılanlar yüzdesel olarak ifade edildi. Çalışmaya katılanların %74.4'ünün kadın, %25.6'sının erkek olduğu ve %95.7'sinin de 18 ile 24 yaş aralığında olduğu görüldü (Tablo 1).

Tablo 1. Epidemiyolojik Bulgular

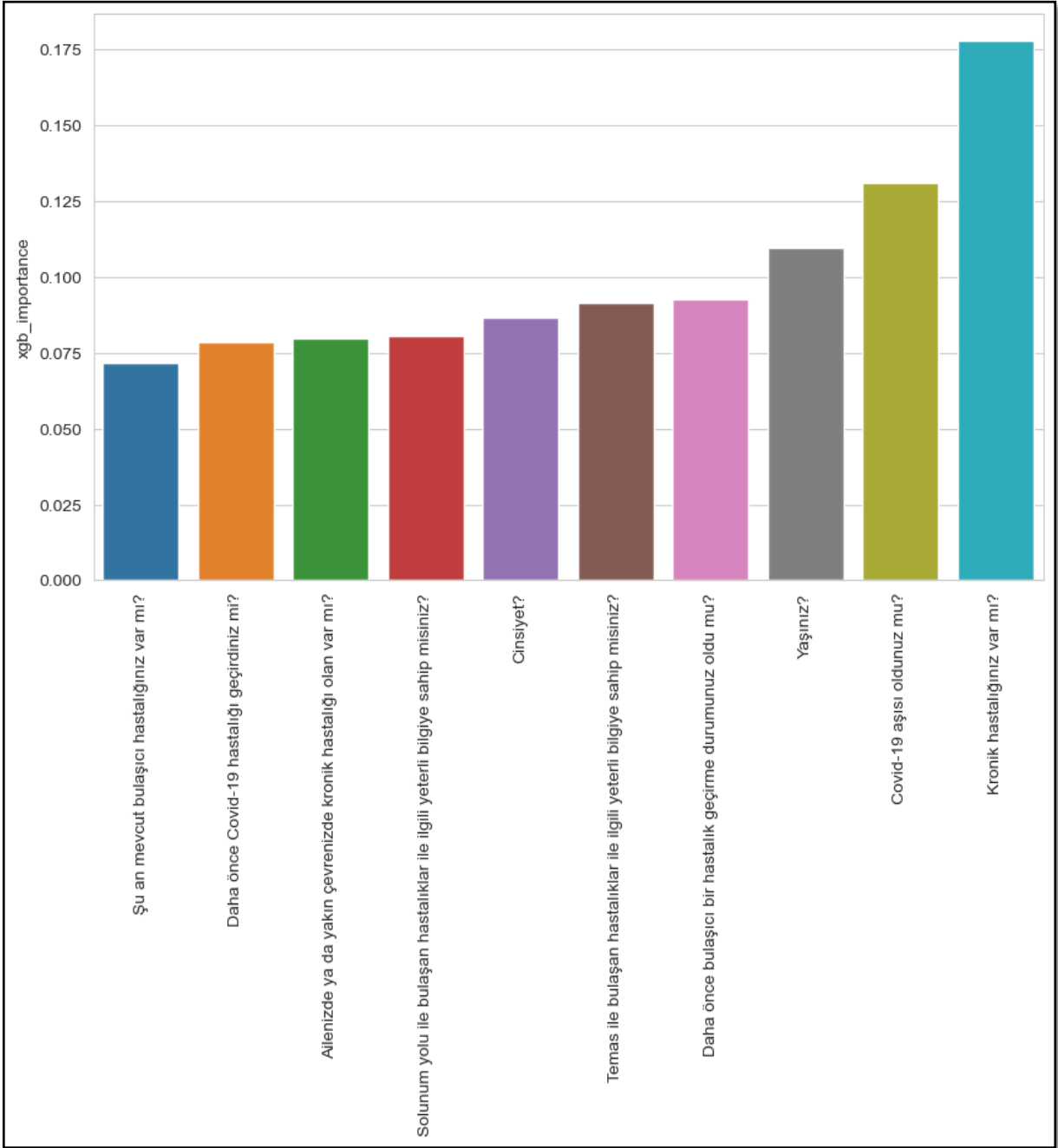
Cinsiyet	% 74.4 Kadın	%25.6 Erkek
Yaşınız*	18-24 yaş arası %95.7	>24 yaş % 4.3

*Verilerin makine öğrenmesi modelleriyle doğru analizi amacıyla yaşlar, 18-24 yaş arası ve >24 olacak şekilde kategorilendirilmiştir.

Çalışmada enfeksiyon ve COVID-19 ile ilgili sorulardan MÖ'ye uygun olan ve modellemede kullanılacak olan sorular seçildi. Sorulara verilen cevaplar "1=hayır" ve "0 =evet" olarak kullanılarak makine öğrenmesi analiz edildi (Tablo 2). MÖ tahmininde özellik önemi karşılaştırması (xgbimportance) değerlerine bakıldığında enfeksiyona yatkınlığın modeller açısından en önemli değişken verisinin "Kronik hastalığınız var mı?" sorusuna verilen cevap olduğu görüldü (Şekil 1).

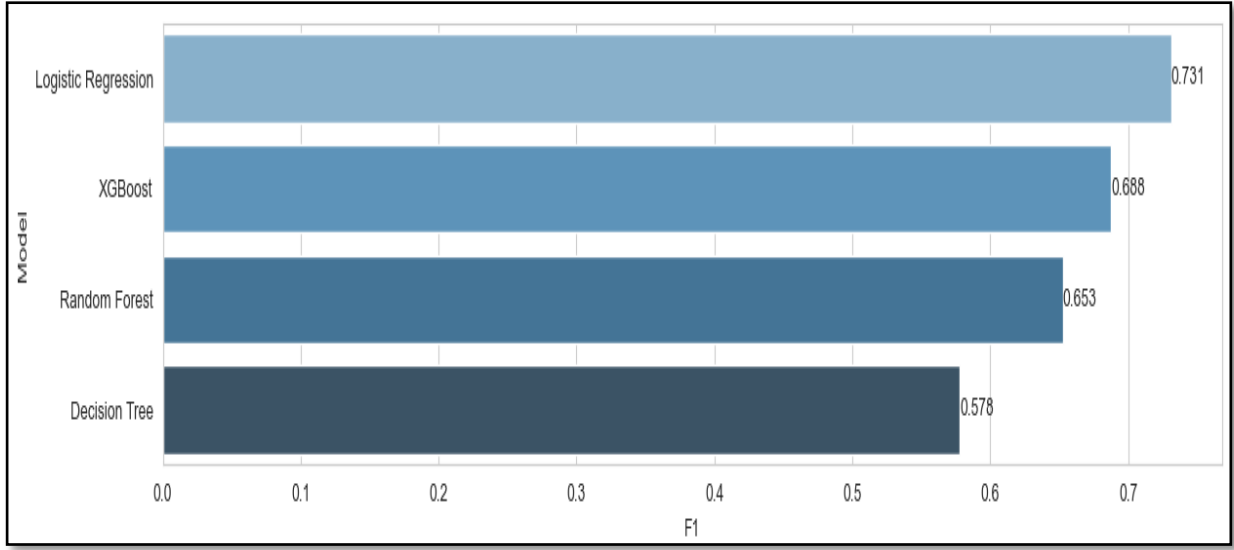
Tablo 2. Anket Sorularına Verilen Cevapların Evet, Hayır Cinsinden Yüzdesel Değerleri

	Evet (%)	Hayır (%)
Kronik hastalığınız var mı?	7.9	92.1
Ailenizde ya da yakın çevrenizde kronik hastalığı olan var mı?	45.0	55.0
Daha önce bulaşıcı bir hastalık geçirme durumunuz oldu mu?	24.9	75.1
Şu an mevcut bulaşıcı hastalığınız var mı?	1.0	99.0
Daha önce Covid-19 hastalığı geçirdiniz mi?	35.2	64.8
Covid-19 aşısı oldunuz mu?	89.2	10.8
Temas ile bulaşan hastalıklar ile ilgili yeterli bilgiye sahip misiniz?	75.7	24.3
Solunum yolu ile bulaşan hastalıklar ile ilgili yeterli bilgiye sahip misiniz?	67.5	32.5
Sıklıkla enfeksiyon rahatsızlığı geçiririm.	50.1	49.9

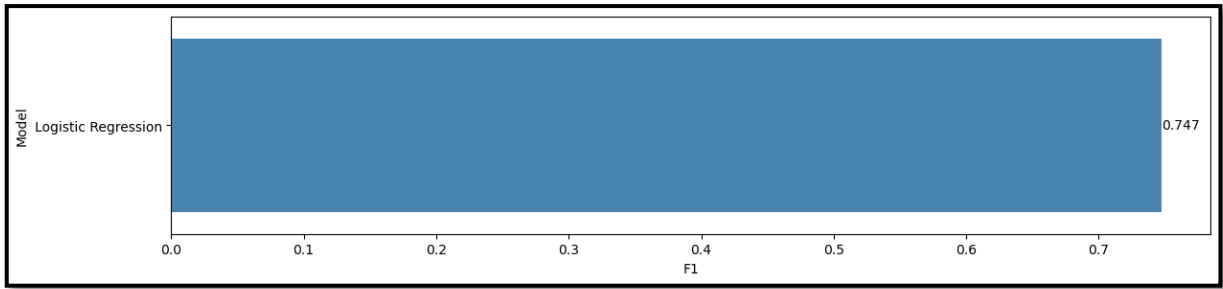


Şekil 1. Makine Öğrenmesi Tahmininde Özellik Önemi Karşılaştırması (xgbimportance: -XGBoost Feature Importance- Genel olarak önem, model içindeki güçlendirilmiş karar ağaçlarının oluşturulmasında her bir özelliğin ne kadar yararlı veya değerli olduğunu gösteren bir puan sağlar)(Bastidas, Garcia-Zapirain, Totoricagüena, Zahia, &Carpio, 2021)

Çalışmada *Logistic Regression*, *XGBoost*, *Random Forest* ve *Decision Tree* olmak üzere dört farklı MÖ modeli test edilmiştir. Bu öğrenme modellerinde *Logistic Regression* tahmin başarısı 0.731, *XGBoost* tahmin başarısı 0.688, *Random Forest* ve *Decision Tree* tahmin başarısı da sırasıyla 0.653 ve 0.578 olarak hesaplandı (Şekil 2). Yapılan bu testler sonucunda en yüksek skor *Logistic Regression* modeline ait olduğundan çalışmada temel olarak *Logistic Regression* modeli kullanıldı. Veri düzeltmeleri, eksik verilerin temizlenmesi ve veriye uygun yapay veri üretimi sonrasında *Logistic Regression* modelinin tahmin skorunun 0.747'ye yükseldiği görüldü (Şekil 3).



Şekil 2. Tahmin doğruluğu bakımından kullanılan modellerin karşılaştırılması



Şekil 3. Model kararı verildikten sonra doğru tahmin oranını arttırmak amacıyla suni veriler ile elde edilen skor

SONUÇ, TARTIŞMA VE ÖNERİLER

Makine öğrenimini (MÖ) kullanan yapay zeka, verilerden kalıpları otomatik olarak öğrenen ve verinin yapısına ilişkin hiçbir varsayım gerektirmeyen bir teknikler bütünüdür. Bu tekniklerin güçlü yanı, tahmin ediciler arasındaki etkileşimin yanı sıra verilerdeki doğrusal olmayan ilişkileri de yakalamalarıdır. Birçok çalışma bunların hastalık tahmini için umut verici performanslarını ortaya koymuştur (Nusinovici et al., 2020). Bulaşıcı hastalıklarda çoğu makine öğrenimi çalışması araştırma, ilaç geliştirme veya klinik mikrobiyolojiye odaklanır. Bakteri genomunu analiz etmek ve direnç tahminini, HIV genotipini iyileştirmek ve antiretroviral ilaçlara duyarlılığı, sürveyans amacıyla salgın kalıplarını tahmin etmek veya yeni antibakteriyel ilaçlar veya aşılar keşfetmek için sistemler geliştirilmiştir (Peiffer-Smadja et al., 2020). Makine öğrenimi teknikleri, sağlık hizmetlerinde mevcut olan rutin elektronik verilerin büyük miktarda ve çeşitliliği nedeniyle sağlık epidemiyolojisi alanında son yıllarda giderek daha popüler hale gelmiştir. Ancak teorik makine öğrenimi araştırmaları ile klinik araştırmalar arasında hala bir boşluk bulunmaktadır (Roth, Battagay, Juchler, Vogt, & Widmer, 2018).

Bir makine öğrenimi modeli oluşturmak için çok büyük veri kümeleri ve sonuçlar, bir eğitim kümesi ve bir test kümesine bölünebilir ve bir bilgisayara girilebilir. Daha sonra bilgisayar veriler arasında bir

ilişki bulabilir ve buna göre bir algoritma oluşturabilir. Algoritmanın performansını artırmak için çeşitli hiperparametreler ayarlanabilir. Nihai algoritma gelecekteki görünmeyen veri kümelerinde kararlar üretmek için kullanılabilir (Chong et al., 2023). Çok sayıda çalışma, Logistic Regression' un tahmin performansı ile tahmin için çok iyi performans gösterdiğini, en azından MÖ teknikleri kadar iyi olduğunu göstermiştir (Nusinovici et al., 2020). Çalışmada kullanılan ve tahmin skoru yüksek olan modelin Logistic regression olması literatür ile de uyumludur. Çalışmanın tahmin gücünün 0.747 olması da örneklem genişliğinin görece az olmasıyla ilişkilidir. Algoritmaya yeni verilerin eklenmesiyle MÖ modelinin tahmin gücünün performansının arttığı da bilinmektedir.

Çalışmanın en dikkat çeken verilerinden biri yaş grubu ile ilgilidir. Çalışmanın meslek yüksekokullarında yapılmış olması çalışmaya katılanların %95.7'sinin 18 ile 24 yaş arasında birikmesine yol açmıştır. Bu durum, modele bu yaş aralığının dışında bir veri girişi olduğunda tahmin gücünü Şekil 1'de verilen makine öğrenmesi tahmininde özelliklerden yaşa önem vermekte ve doğruluk konusunda modelin yanılmasına sebep olabilmektedir. Dolayısıyla makine öğreniminde yaş yelpazesinin genişlemesinin süreç içerisinde yaş değişkeninin "XGBoost Feature Importance" skalasında önemi sırasını değiştirebileceğini düşündürmektedir.

Ancak model performansı, sonucu tahmin etmede ne kadar iyi olduğunu temsil eden çeşitli istatistiklere odaklanır. Performans istatistiklerini değerlendirirken dikkatli olunmalıdır çünkü çok sayıda istatistik vardır ve bunlar, özellikle de model doğruluğu genellikle uygunsuz şekilde kullanılır. Bunun bir örneği olarak, sepsisi öngörmek için bir model oluşturan bir ürün satıcısıdır. Ürettikleri ticari modelin %90 doğru olduğunu bildirmişlerdir. Ancak burada görünen değer açısından harika olmasına rağmen, doğruluk gibi büyük bir istatistik yerine duyarlılık, özgüllük, pozitif ve negatif tahmin değerlerine bakmak daha önemlidir. Bunun nedeni, doğruluğun, hatalı pozitif ve negatifleri ayırmadan, toplam doğru tahminlerin toplam tahminlere bölünmesiyle hesaplanmasıdır. Ayrıca, model uygun şekilde eğitilmemişse bu doğruluk istatistiği, modelin gerçek hayattaki performansı ile uyumsuz olabilir. Örneğin, model hastaların %90'ının sepsis olmadığı ve %10'unun sepsis hastası olduğu bir veri seti üzerinde eğitilmişse, model basitçe her hastanın sepsis olmadığını bildirebilir ve yine de bu rapor %90 ihtimalle doğru olur. Sepsis durumunda bu %10'u ve tahminin amacını yani "azınlık sınıfı"nı (daha az sıklıkta görülen sonuç grubunu) gözden kaçıracaktır (Wiemken&Rutschman, 2020).

Sonuç olarak, dijital sağlık hizmetleri epidemiyolojisi, sağlık hizmetlerinde günlük rutin dokümantasyondan kaynaklanan büyük verilerin artan kullanılabilirliğiyle desteklenen, tıpta büyüyen bir alandır. Makine öğrenimi, enfeksiyon önleme ve kontrol, kalite iyileştirme ve hastane kaynaklarının optimal tahsisi için büyük verinin potansiyelinden daha iyi yararlanmak amacıyla sağlık hizmetleri epidemiyologlarının silahlanmasında önemli bir araç haline gelebilir. Karmaşıklıkları nedeniyle, makine öğrenimi projeleri genellikle en iyi uygulamalara dayalı olarak alan uzmanları ve makine öğrenimi uzmanları arasında yakın işbirliği içinde gerçekleştirilmelidir. Ancak şunun da farkında olmak gerekir, medyada pek çok sansasyonel rapora rağmen, makine öğrenimi, düşük kaliteli verileri altına dönüştürebilecek sihirli bir teknoloji değildir ve bir veri bilimcinin de yakın zamanda belirttiği gibi, "Sağlık hizmetlerinde makine öğreniminin hâlâ bir düzeni yoktur."(Roth, Battagay, Juchler, Vogt, & Widmer, 2018). Bu sebeple alandaki boşluğun daha büyük veri setleri ve daha ileri çalışmalarla doldurulması çağın gerekliliğidir.

KAYNAKÇA

- Bastidas, O. J., Garcia-Zapirain, B., Toticagüena, A. L., Zahia, S., & Carpio, J. U. (2021). Feature analysis and prediction of complications in ostomy patients based on laboratory analytical data using a machine learning approach. 2021 International Conference BIOMDLORE, 1–8. IEEE. doi: 10.1109/BIOMDLORE49470.2021.9594427
- Chiu, H.-Y. R., Hwang, C.-K., Chen, S.-Y., Shih, F.-Y., Han, H.-C., King, C.-C., ... Oyang, Y.-J. (2022). Machine learning for emerging infectious disease field responses. *Scientific Reports*, 12(1), 1–13. doi: <https://doi.org/10.1038/s41598-021-03687-w>
- Chong, Y. Y., Chan, P. K., Chan, V. W. K., Cheung, A., Luk, M. H., Cheung, M. H., ... Chiu, K. Y. (2023). Application of machine learning in the prevention of periprosthetic joint infection following total knee arthroplasty: A systematic review. *Arthroplasty*, 5(1), 1–13. doi: 10.1186/s42836-023-00195-2
- Keshavamurthy, R., Dixon, S., Pazdernik, K. T., & Charles, L. E. (2022). Predicting infectious disease for biopreparedness and response: A systematic review of machine learning and deep learning approaches. *One Health*, 100439. doi: 10.1016/j.onehlt.2022.100439
- Mahmoud, E., Al Dhoayan, M., Bosaeed, M., Al Johani, S., & Arabi, Y. M. (2021). Developing machine-learning prediction algorithm for bacteremia in admitted patients. *Infection and Drug Resistance*, 14, 757. doi: 10.2147/IDR.S293496
- Nusinovici, S., Tham, Y. C., Yan, M. Y. C., Ting, D. S. W., Li, J., Sabanayagam, C., ... Cheng, C.-Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. *Journal of Clinical Epidemiology*, 122, 56–69. doi: 10.1016/j.jclinepi.2020.03.002
- Özer, Z. Y., & Özcan, S. (2020). Birinci Basamak Sağlık Hizmetlerinde COVID-19 Enfeksiyonu: Korunma ve Kontrol Önerileri. *Arşiv Kaynak Tarama Dergisi*, 29(Özel Sayı), 67–72. doi: 10.17827/aktd.841220
- Peiffer-Smadja, N., Rawson, T. M., Ahmad, R., Buchard, A., Georgiou, P., Lescure, F.-X., ... Holmes, A. H. (2020). Machine learning for clinical decision support in infectious diseases: A narrative review of current applications. *Clinical Microbiology and Infection*, 26(5), 584–595. doi: 10.1016/j.cmi.2019.09.009
- Roth, J. A., Battagay, M., Juchler, F., Vogt, J. E., & Widmer, A. F. (2018). Introduction to machine learning in digital healthcare epidemiology. *Infection Control & Hospital Epidemiology*, 39(12), 1457–1462. doi: 10.1017/ice.2018.265.
- Wiemken, T. L., & Rutschman, A. S. (2020). Methodology minute: A machine learning primer for infection prevention and control. *American Journal of Infection Control*, 48(12), 1504–1505. doi: 10.1016/j.ajic.2020.09.009