

Atf İçin: Okudan, B. ve Karcıoğlu, A.A. (2025). Makine Öğrenmesi Algoritmaları Kullanılarak Türkiye’de Bölgesel Salgın Risk Analizi ve Haritalandırılması. *Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 15(4), 1178-1192.

To Cite: Okudan, B. & Karcıoğlu, A.A. (2025). Regional Epidemic Risk Analysis and Mapping in Turkey Using Machine Learning Algorithms. *Journal of the Institute of Science and Technology*, 15(4), 1178-1192.

Makine Öğrenmesi Algoritmaları Kullanılarak Türkiye’de Bölgesel Salgın Risk Analizi ve Haritalandırılması

Bezanur OKUDAN¹, Abdullah Ammar KARCIOĞLU^{1*}

Öne Çıkanlar:

- İl bazlı sağlık risklerinin makine öğrenimi ile tahmini
- Coğrafi bilgi sistemleri ile haritalandırma
- Simülasyonlarla risk yayılımı modellenmesi
- XGBoost algoritması ile yüksek doğruluk oranı

Anahtar Kelimeler:

- Salgın risk analizi
- Makine öğrenimi
- Nüfus yoğunluğu
- Simülasyon

ÖZET:

Bölgesel salgın risklerinin belirlenmesi, halk sağlığını koruma, sağlık hizmetlerini planlama ve kaynak tahsisinin etkin yönetimi açısından büyük önem taşımaktadır. Salgın hastalıklar, hızlı yayılma potansiyelleri nedeniyle özellikle nüfus yoğunluğu yüksek bölgelerde ciddi sağlık ve ekonomik riskler oluşturmaktadır. Bu çalışmada, Türkiye’nin il bazlı salgın risk analizini gerçekleştirmek ve risk seviyelerini haritalandırmak amacıyla makine öğrenimi teknikleri kullanılmıştır. Çalışmada kullanılan veri seti, Türkiye İstatistik Kurumu (TÜİK), Google Cloud Console ve Wikipedia’den elde edilen demografik, coğrafi, sağlık altyapısı ve ulaşım verilerinin entegrasyonu ile oluşturulmuştur. Veri hazırlık sürecinde eksik veri tamamlama, ölçeklendirme ve kategorik değişkenlerin sayısal hale getirilmesi işlemleri uygulanmıştır. Öznitelik önem sıralaması, Rastgele Orman algoritması kullanılarak belirlenmiş ve en etkili değişkenler seçilerek analiz süreci optimize edilmiştir. Salgın risk tahmini için Lojistik Regresyon (LR), Rastgele Orman (RO), Destek Vektör Makineleri (DVM), Naive Bayes (NB) ve XGBoost (XGB) algoritmaları kullanılmıştır. Algoritmaların performansı doğruluk oranı, ROC eğrisi, AUC skoru ve karışıklık matrisi analiz edilerek karşılaştırılmıştır. Elde edilen sonuçlara göre, XGB algoritması %98 doğruluk oranı ile en iyi performansı sergilemiştir. RO ve DVM %96 doğruluk oranı ile başarılı sonuçlar elde etmiştir. NB algoritması %92 doğruluk oranına sahip olup hızlı tahminleme avantajı sağlamıştır. En düşük doğruluk oranına sahip algoritmalar ise %88 doğruluk ile LR olmuştur. Risk seviyelerinin coğrafi dağılımı görselleştirilmiştir ve Türkiye’nin il bazında salgın risk haritası oluşturulmuştur. Ayrıca, salgın yayılımı simülasyonlarla modellenmiştir. Bu çalışma, sağlık risk seviyelerinin tahmini ve haritalandırılması açısından karar alıcılara veri odaklı bir yaklaşım sunmayı amaçlamaktadır. Gelecekte, gerçek zamanlı veri entegrasyonu, ileri seviye simülasyon modelleri ve derin öğrenme teknikleri üzerine yoğunlaşılabilir.

Regional Epidemic Risk Analysis and Mapping in Turkey Using Machine Learning Algorithms

Highlights:

- Estimation of province-based health risks using machine learning
- Mapping with geographic information systems
- Simulation-based modeling of epidemic spread
- High accuracy with XGBoost algorithm

Keywords:

- Epidemic risk analysis
- Machine learning
- Population density
- Simulation

ABSTRACT:

Identifying regional epidemic risks is of great importance for protecting public health, planning healthcare services, and effectively managing resource allocation. Due to their potential for rapid spread, epidemic diseases pose serious health and economic risks, especially in regions with high population density. In this study, machine learning techniques were used to conduct a province-based epidemic risk analysis of Turkey and to map the risk levels. The dataset used in the study was created by integrating demographic, geographic, healthcare infrastructure, and transportation data obtained from Turkish Statistical Institute (TUIK), Google Cloud Console, and Wikipedia. During the data preparation process, missing data imputation, scaling, and numerical conversion of categorical variables were applied. Feature importance ranking was determined using the Random Forest algorithm, and the analysis process was optimized by selecting the most influential variables. Logistic Regression (LR), Random Forest (RF), Support Vector Machines (SVM), Naive Bayes (NB), and XGBoost (XGB) algorithms were used for epidemic risk prediction. The performance of the algorithms was compared by analyzing accuracy, ROC curve, AUC score, and confusion matrix. According to the results obtained, the XGB algorithm showed the best performance with 98% accuracy. RF and SVM achieved successful results with 96% accuracy. The NB algorithm provided the advantage of fast prediction with an accuracy of 92%. The algorithm with the lowest accuracy was LR, with 88%. The geographical distribution of risk levels was visualized, and a province-based epidemic risk map of Turkey was created. In addition, the spread of the epidemic was modeled through simulations. This study aims to provide a data-driven approach to decision-makers for the prediction and mapping of health risk levels. In the future, the focus may shift toward real-time data integration, advanced simulation models, and deep learning techniques.

¹ Bezanur OKUDAN ([Orcid ID: 0009-0001-3108-1857](https://orcid.org/0009-0001-3108-1857)), Abdullah Ammar Karcıoğlu ([Orcid ID: 0000-0002-0907-751X](https://orcid.org/0000-0002-0907-751X)), Atatürk Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği Bölümü, Erzurum, Türkiye.

*Sorumlu Yazar/Corresponding Author: Abdullah Ammar KARCIOĞLU, e-mail: ammar.karcioglu@atauni.edu.tr

Bu çalışma Bezanur Okudan’ın Yüksek Lisans tezinden üretilmiştir.

GİRİŞ

Bölgesel düzeyde salgın risklerinin belirlenmesi ve bu risklerin dinamik bir şekilde analiz edilmesi, yalnızca halk sağlığı açısından değil, aynı zamanda sağlık sistemlerinin sürdürülebilirliğinin sağlanması, kaynak planlamasının etkin yapılması ve kriz yönetimi süreçlerinin veri temelli olarak yürütülmesi açısından son derece kritik bir rol oynamaktadır. Günümüz toplumlarında artan nüfus yoğunluğu, kentleşme hızı, iç ve dış göç hareketliliği ile ulaşım ağlarının yoğunlaşması, salgın hastalıkların yayılımını hızlandıran temel faktörler arasında yer almaktadır. Özellikle metropol kentler gibi yoğun nüfuslu alanlarda virüslerin bulaşma riski daha yüksek olmakta, bu da yerel düzeyde önleyici politikaların oluşturulmasını gerekli kılmaktadır. COVID-19 pandemisi, bu gerçekliği net bir biçimde gözler önüne sermiş; geleneksel sağlık yönetimi yaklaşımlarının ötesinde, veri bilimi, yapay zekâ ve erken uyarı sistemlerinin entegre edildiği hibrit çözümlere olan ihtiyacı artırmıştır (Zhang ve Zhao, 2023).

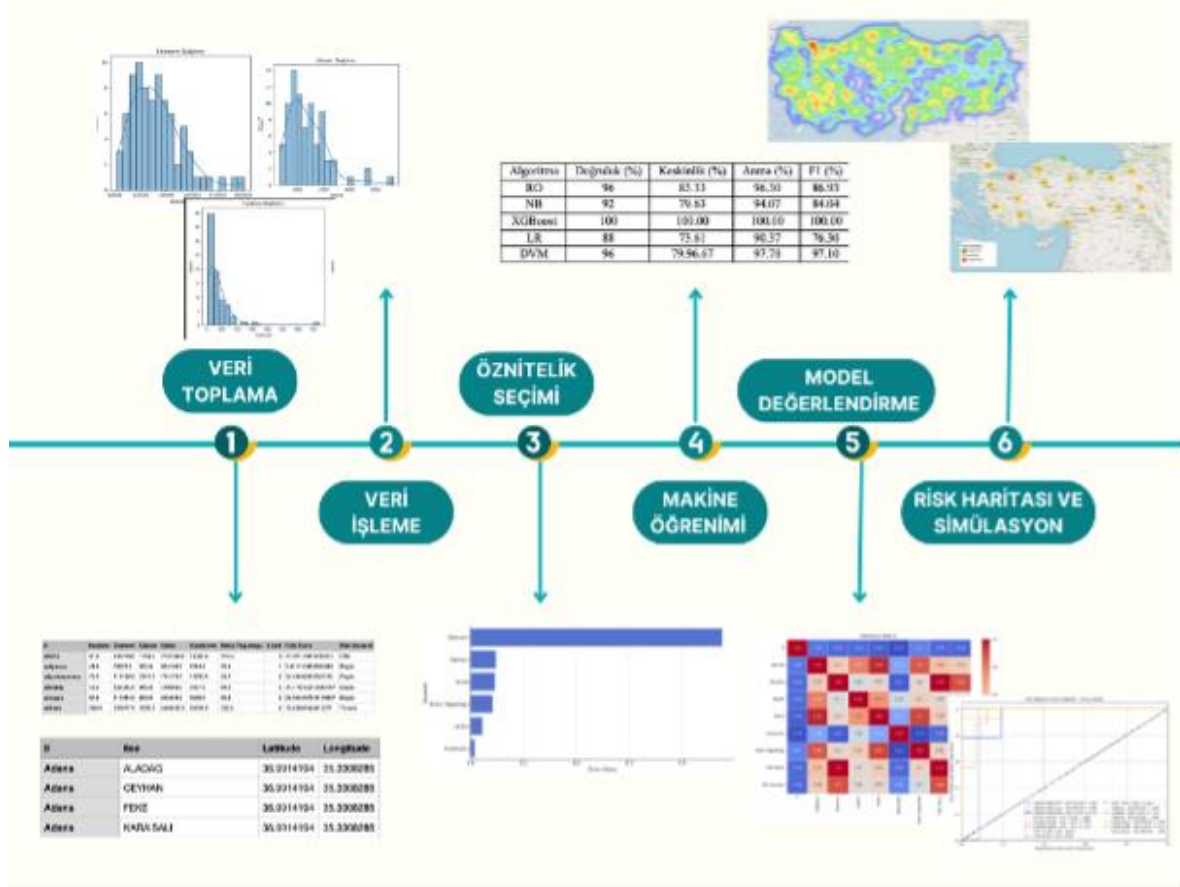
Geleneksel epidemiyolojik modeller, özellikle SEIR ve SIR gibi deterministik yapılar, salgınların temel dinamiklerini öngörmeye yaygın olarak kullanılmaktadır. Ancak bu modeller, bireysel hareketliliği, çevresel değişkenliği, sosyoekonomik farklılıkları ya da coğrafi bağlamları yeterince dikkate almamaktadır (Zhang ve Zhao, 2023). Ayrıca sabit parametre varsayımları nedeniyle değişken ve belirsizlik içeren gerçek dünya verilerini yansıtmakta sınırlı kalmaktadır (Khalid ve ark., 2024). Bu nedenle son yıllarda istatistiksel öğrenme ve yapay zekâ temelli modellere yönelim artmıştır (Madry ve ark., 2017), (Goodfellow ve ark., 2014).

Makine öğrenimi algoritmaları; doğruluk, ölçeklenebilirlik ve çok değişkenli veri analizi kabiliyetleri ile sağlık verilerinde son derece başarılı tahmin sonuçları sunabilmektedir. Yapılan pek çok çalışmada, özellikle Rastgele Orman (RO), Destek Vektör Makineleri (DVM), Naive Bayes (NB) ve XGBoost (XGB) gibi algoritmaların; hastalık sınıflandırması, bulaşma tahmini ve sağlık altyapısı yükünü modelleme gibi görevlerde yüksek başarı sağladığı gözlemlenmiştir (Pedregosa ve ark., 2011), (Zhang ve Zhao, 2023). COVID-19 özelinde yapılan araştırmalar, bu algoritmaların hem bireysel hem toplumsal düzeyde risk değerlendirmesinde geleneksel yöntemlerden daha etkin sonuçlar verdiğini ortaya koymuştur (Çakmak ve ark., 2022). Sağlık verilerinde sınıflandırma başarımı açısından, DVM'nin giderek daha fazla tercih edildiği ve performans açısından geliştirildiği belirtilmektedir (Guido ve ark., 2024).

Mekânsal epidemiyolojik analizler ise son yıllarda CBS (Coğrafi Bilgi Sistemleri) teknolojilerinin gelişmesiyle birlikte daha güçlü hale gelmiştir. CBS destekli sistemler, coğrafi konum verilerini kullanarak sağlık risklerinin bölgelere göre dağılımını analiz edebilmekte ve bu analizlerin çıktılarını, karar vericilere stratejik müdahale planları oluşturma fırsatı sunmaktadır (Naim, 2023). Ayrıca, özellikle kırsal bölgelerde sağlık hizmetlerine erişimde yaşanan mekânsal eşitsizlikler, coğrafi analizlerle somut biçimde ortaya konabilmektedir (Edward ve Biddle, 2017). Bu sistemlerin, makine öğrenimi algoritmalarıyla entegre edilmesi ise salgın yayılım tahminlerinin hem zamansal hem mekânsal boyutta detaylandırılmasını sağlamaktadır (Chumachenko ve ark., 2022). Türkiye bağlamında yapılan çalışmaların sayısı sınırlı olmakla birlikte, son dönemde bu alana yönelik akademik ilgide artış gözlemlenmektedir. COVID-19'un mekânsal ve zamansal yayılımına yönelik epidemiyolojik analizlerin birçok çalışmada çeşitli yöntemlerle ele alınarak sistematik derlemede detaylı şekilde ortaya konmuştur (Nazia ve ark., 2022). Bu kapsamda Karcioğlu ve arkadaşları (Karcioğlu ve ark., 2021), COVID-19'un Türkiye genelindeki yayılımını modellemek amacıyla ARIMA ve LSTM temelli zaman serisi analizleri gerçekleştirmiştir. Çalışmaları, makine öğrenimi tabanlı tahmin modellerinin ülke genelinde salgın öngörüsünde başarılı bir şekilde kullanılabileceğini

göstermesi bakımından dikkat çekicidir. Türkiye özelinde yapılan çalışmalarda ise bu tür analizlerin sayısında son yıllarda artış gözlemlense de hâlâ bölgesel ve yerel ölçekli, veri temelli karar destek sistemlerinin eksik olduğu görülmektedir. Yerli literatürde, genellikle il bazlı analizler sınırlı sayıdadır ve çoğu çalışma geniş coğrafi düzeyde ya da sabit model parametreleriyle sınırlı kalmaktadır. Ancak bazı öncü araştırmalarda, şehir bazlı sağlık risk haritaları oluşturularak, yerel sağlık politikalarının bilimsel temellere dayandırılması yönünde önemli adımlar atılmıştır (Yılmaz ve Demir, 2022). Ayrıca, karar verme süreçlerinde çok kriterli karar verme (ÇKKV) yöntemlerinin kullanımı da önemli bir yer tutmaktadır. Bu tür yaklaşımlar; sağlık altyapısı kapasitesi, ulaşım erişilebilirliği, ekonomik seviye, nüfus yoğunluğu gibi çok sayıda değişkeni aynı anda analiz ederek daha kapsamlı ve dengeli risk değerlendirmeleri yapılmasına olanak tanımaktadır (Khalid ve ark., 2024). Bu yapıların, makine öğrenimi modelleri ile kullanılması, karmaşık epidemiyolojik yapıların anlaşılmasında önemli katkı sunmaktadır. Buna ek olarak, zaman serisi modelleme ve özellikle LSTM (Long Short-Term Memory) gibi tekrarlayan sinir ağı mimarileri, salgın yayılım eğilimlerinin zamansal tahmininde giderek daha sık kullanılmaktadır. Özellikle mevsimsellik, bulaşma oranı ve davranışsal değişikliklerin etkilerini modelleyebilme kapasitesi nedeniyle derin öğrenme teknikleri, salgın yönetimi süreçlerinde tamamlayıcı bir araç olarak değerlendirilmektedir (Shastri ve ark., 2020).

Tüm bu literatür ışığında, bu çalışma Türkiye’nin 81 ili özelinde, salgın risklerini makine öğrenimi temelli yöntemlerle analiz etmeyi ve risk seviyelerinin coğrafi olarak haritalandırılmasını amaçlamaktadır. Kullanılan geniş kapsamlı veri seti hem demografik hem de coğrafi değişkenleri içermekte olup; risklerin hem mevcut dağılımının hem de olası yayılım senaryolarının modellenmesine olanak sağlamaktadır. Böylece karar vericilere veri odaklı, görselleştirilmiş ve senaryoya dayalı bir karar destek altyapısı sunulması hedeflenmektedir. Yapılan önceki çalışmalarda genellikle sabit model parametrelerine dayalı analizler yapılmışken, bu çalışmada 2022 yılına ait ekonomik göstergeler ile güncel coğrafi ve demografik veriler bütünleştirilerek, il düzeyinde dinamik bir salgın risk sınıflandırması gerçekleştirilmiştir. Ayrıca hem coğrafi hem simülasyon tabanlı görselleştirmelerin birlikte sunulması çalışmayı özgün kılmaktadır. Bu yapı, karar alıcılara hem mevcut durumu değerlendirme hem de müdahale öncesi senaryo üretme imkânı tanımaktadır. Çalışmada kullanılan yöntemler ve elde edilen sonuçlar, gelecekte gerçek zamanlı veri entegrasyonu ve gelişmiş simülasyon teknikleri ile daha geniş kapsamlı uygulamalara yönelik bir temel oluşturabilir. Bu çalışmanın genel çerçevesi Şekil 1’de görsel olarak sunulmuştur. Şekilde sırasıyla veri toplama, ön işleme, öznitelik seçimi, makine öğrenimi ile model oluşturma, performans değerlendirme ve risk haritalama adımları yer almaktadır. Ayrıca her adım, ilgili alt görsellerle desteklenmiş ve çalışmanın bütünsel akışı şematik olarak özetlenmiştir.



Şekil 1. Bu çalışmanın genel çerçevesi

Materyal ve yöntem bölümü veri kaynaklarını, ön işleme adımlarını ve modelleme sürecini içermektedir. Bulgular ve tartışma bölümünde model performansları, risk haritaları ve elde edilen sonuçların yorumları sunulmakta; sonuç bölümünde ise elde edilen bulgular özetlenmekte ve gelecekteki çalışmalar için öneriler verilmektedir.

MATERYAL VE METOT

Bu çalışmada, Türkiye'deki illere ait salgın risk seviyelerinin belirlenmesi için veri analizi ve makine öğrenimi teknikleri kullanılmıştır. Çalışmada kullanılan veri seti, Türkiye'nin 81 iline ve bu illere bağlı toplam 973 ilçeye ait demografik, coğrafi ve ekonomik verileri içermektedir. Verilerin büyük kısmı 2022 yılına aittir; bazı coğrafi veriler ise statik nitelikte olup güncel verilerdir. Veriler, TÜİK, Google Cloud Console ve Wikipedia kaynaklarından derlenmiştir. Çalışma, beş ana aşamadan oluşmaktadır: veri kaynaklarının belirlenmesi, veri hazırlık süreci, makine öğrenimi algoritmalarının uygulanması, model değerlendirme ve sonuçların risk haritası ile simülasyon teknikleri kullanılarak görselleştirilmesi.

Veri Seti Kaynakları

Çalışmada kullanılan veri seti, Türkiye'nin illerine ait çeşitli demografik, coğrafi, sağlık altyapısı ve ulaşım verilerini içermektedir. Demografik veriler Türkiye İstatistik Kurumu'ndan (TÜİK), coğrafi veriler ise Google Cloud Console ve Wikipedia'dan elde edilmiştir. Sağlık altyapısı, ulaşım ağı ve ekonomik göstergeler ise TÜİK ile yerel yönetim kaynaklarından toplanmıştır. İllere ait enlem ve boylam bilgilerini içeren coğrafi veri seti, risk seviyelerinin haritalandırılması ve simülasyonların oluşturulması sürecinde kullanılmıştır. Ayrıca nüfus, ulaşım erişimi, sağlık altyapısı ve nüfus

yoğunluğu gibi değişkenleri içeren veri kümesi, risk seviyelerinin sınıflandırılması için analiz edilmiştir. Bu kapsamda analizlerde kullanılan iki temel veri seti aşağıda özetlenmiştir:

- 1. Enlem-boylam veri seti:** Türkiye’deki 973 ilçeye ait konum bilgilerini içermektedir. Veri setinde toplam dört değişken yer almaktadır. İl ve ilçe isimleri TÜİK kaynaklıdır; enlem ve boylam bilgileri ise Google Cloud Console üzerinden alınmıştır. Bu veri setinin örnek içeriği Şekil 2’de gösterilmektedir.

| il | ilce | Latitude | Longitude |
|-------|----------|------------|------------|
| Adana | ALADAG | 36.9914194 | 35.3308285 |
| Adana | CEYHAN | 36.9914194 | 35.3308285 |
| Adana | FEKE | 36.9914194 | 35.3308285 |
| Adana | KARISALI | 36.9914194 | 35.3308285 |

Şekil 2. Enlem-boylam veri seti içeriği

- 2. Risk skoru veri seti:** Türkiye’deki 81 il için sağlık risk skorlarını ve ilgili değişkenleri içermektedir. Veri setinde bulunan değişkenler çeşitli platformlardan erişilerek toplanmıştır. Veri setinde bulunan nüfus, yüzölçümü, nüfus yoğunluğu bilgileri Wikipedia üzerinden elde edilmiştir. Hastane sayısı, ulaşım altyapısı, bölgedeki ekonomik seviye ise Google Cloud Console ve TÜİK üzerinden elde edilmiştir. Risk skoru ve risk seviyesi ise elde edilen diğer değişkenler üzerinden öznitelik hesabı yapılarak sonuçlandırılmıştır. Risk skoru veri setinin içeriği Şekil 3’te gösterildiği gibidir.

| il | Hastane | Ekonomi | Ulasim | Nufus | Yuzolcumu | Nufus Yogunlugu | il_kod | Risk Skoru | Risk Seviyesi |
|----------------|---------|----------|--------|-----------|-----------|-----------------|--------|--------------------|---------------|
| adana | 90.0 | 135798.0 | 1794.0 | 2237940.0 | 14045.0 | 159.3 | 0 | 33.02116919199013 | Orta |
| adiyaman | 41.0 | 79223.0 | 985.0 | 635169.0 | 7614.0 | 83.4 | 1 | 9.411155828095348 | Düşük |
| afyonkarahisar | 72.0 | 114168.0 | 2141.0 | 744179.0 | 14230.0 | 52.3 | 2 | 24.59582850050132 | Düşük |
| aksaray | 44.0 | 134484.0 | 960.0 | 429069.0 | 7617.0 | 56.3 | 4 | 31.176744211832187 | Düşük |
| amasya | 38.0 | 112044.0 | 803.0 | 335494.0 | 5520.0 | 60.8 | 5 | 22.054552919163907 | Düşük |
| ankara | 239.0 | 230677.0 | 3000.0 | 5663322.0 | 25437.0 | 222.5 | 6 | 72.62954555612271 | Yüksek |

Şekil 3. Risk skoru veri seti içeriği

Veri ön işleme aşaması

Veri seti, analiz edilebilir hale getirilmeden önce çeşitli ön işleme adımlarından geçirilmiştir. Öncelikle, eksik veya hatalı verilerin tespiti ve düzeltilmesi sağlanmıştır. Eksik değer oranı genel olarak %3–5 arasında değişmektedir ve bu değerler, benzer demografik ve coğrafi özelliklere sahip illerin ortalamaları kullanılarak doldurulmuştur. Bu doldurma stratejisinin doğruluğu, küçük bir test veri kümesi üzerinden model performansına etkisi incelenerek geçerlilik açısından değerlendirilmiştir. Ayrıca, bazı yeni değişkenler (örneğin, nüfus yoğunluğu) türetilmiş ve veri setine eklenmiştir. Sayısal veriler, StandardScaler yöntemi ile normalize edilmiştir; bu sayede tüm değişkenler ortalaması 0 ve standart sapması 1 olacak şekilde aynı ölçeğe getirilmiştir. Kategorik değişkenler sayısal formata dönüştürülmüş; salgın risk seviyeleri “düşük”, “orta” ve “yüksek” olmak üzere üç sınıfa ayrılarak sırasıyla 0, 1 ve 2 olarak etiketlenmiştir.

Öznitelik seçimi aşamasında, model performansını optimize etmek ve gereksiz değişkenlerin model üzerindeki etkisini azaltmak amacıyla öznitelik önem sıralaması belirlenmiştir. RO algoritması

kullanılarak değişkenlerin önem dereceleri hesaplanmış ve model üzerinde en fazla etkisi olan değişkenler belirlenmiştir. Elde edilen sonuçlara göre, ekonomi, sağlık altyapısı ve nüfus değişkenleri en önemli faktörler olarak tespit edilmiştir. Daha az etkili olan değişkenler modelden çıkarılarak analiz süreci sadeleştirilmiş ve modelin genelleme yeteneği artırılmıştır. Son yıllarda yapılan çalışmalar, RO algoritmasının performansını artırmak amacıyla karar ağaçlarının doğruluk ve korelasyon ölçümlerine göre optimize edildiğini göstermektedir (Sun ve ark., 2024).

Makine öğrenmesi algoritmaları

Bu çalışmada, salgın risk seviyelerinin belirlenmesi amacıyla beş farklı makine öğrenimi algoritması kullanılmıştır. Lojistik Regresyon (LR), ikili veya çoklu sınıflandırma problemlerinde kullanılan bir modeldir. Hedef değişkenin olasılıklarını tahmin etmek için doğrusal bir ilişki kurar ve tahmin edilen değerleri 0 ile 1 arasında sıkıştıran sigmoid fonksiyonu kullanır ve Formül 1’de gösterildiği gibi hesaplanmaktadır. Burada kullanılan β katsayıları değişkenlerin ağırlıklarını temsil etmektedir (Tanaka ve ark., 2024). LR ve makine öğrenmesi tabanlı modeller, klinik çıktıları öngörmeye etkili bir yöntem olarak çeşitli sağlık çalışmalarında başarıyla uygulanmıştır (Miyazaki ve ark., 2024).

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}} \quad (1)$$

RO, birden fazla karar ağacından oluşan bir topluluk modelidir. Her bir ağaç bağımsız olarak eğitim verisinde tahmin yapar ve nihai tahmin, bu ağaçların çoğunluk oyu veya ortalamasından alınır ve Formül 2’de gösterildiği gibi hesaplanmaktadır. Büyük veri setlerinde genelleme yeteneği yüksektir. Burada kullanılan $T_i(X)$, i. karar ağacını temsil etmektedir (Liu ve ark., 2023).

$$f(x) = \frac{1}{N} \sum_{i=1}^N T_i(X) \quad (2)$$

DVM, veriyi sınıflara ayırmak için bir hiper düzlem (border) kullanır ve sınıflar arasındaki en geniş marjini oluşturur. Veri doğrusal olmasa bile kernel fonksiyonları kullanarak verileri daha yüksek boyutlu uzaylara projekte eder ve doğrusal bir sınıflandırma yapar. DVM, sınıflar arasındaki ayrımı optimize etmek için margin genişliğini maksimize eder ve Formül 3’te gösterildiği gibi hesaplanmaktadır (Sharma ve ark., 2024).

$$\max \frac{2}{\|w\|}, \text{ koşuluyla } y_i(w \cdot x_i + b) \geq 1, \forall i \quad (3)$$

NB, Bayes teoremini kullanan bir sınıflandırma algoritmasıdır ve özelliklerin birbirinden bağımsız olduğu varsayımına dayanır. Model, her bir sınıf için olasılıkları hesaplar ve en yüksek olasılığa sahip sınıfı seçer ve Formül 4’te gösterildiği gibi hesaplanmaktadır. Kullanılan matematiksel gösterimde C_k ve X özellikler kümesini temsil etmektedir (Zhang ve Zhao, 2023).

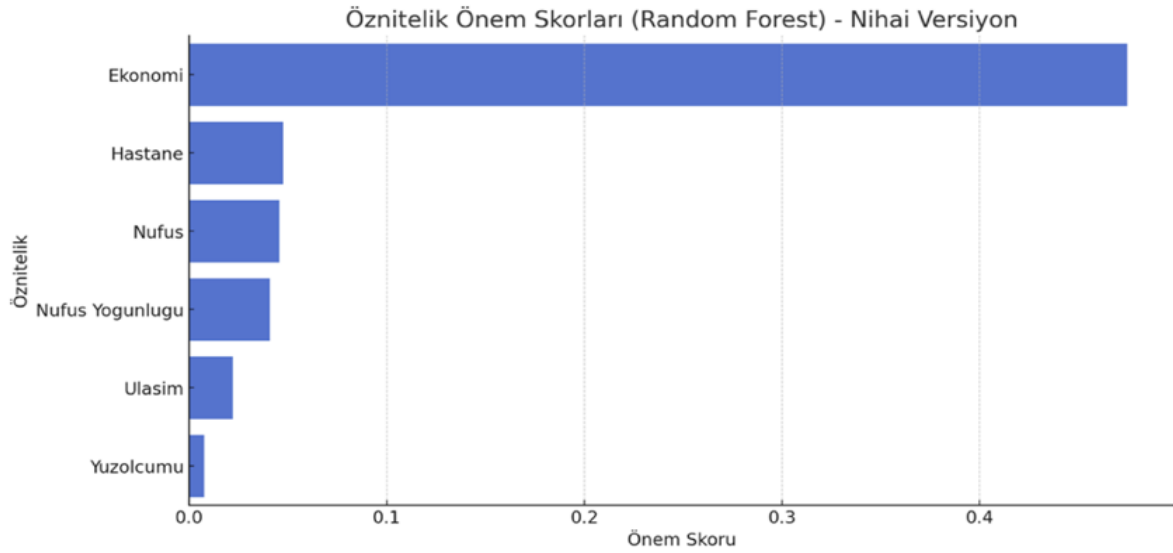
$$P(C_k|X) = \frac{P(X|C_k)P(C_k)}{P(X)} \quad (4)$$

XGB, gradyan artırma (gradient boosting) yöntemi ile çalışan güçlü bir topluluk modelidir. Her yeni model, önceki modelin hatalarını düzelterek tahmin yapar ve hataları minimize etmeye çalışır. XGB, düzenleme (regularization) teknikleriyle aşırı öğrenmeyi önler ve büyük veri setlerinde yüksek doğruluk sağlar ve Formül 5’te gösterildiği gibi hesaplanmaktadır (Goodfellow ve ark., 2014). Burada $L(\theta)$, kayıp fonksiyonunu, $\Omega(f_k)$ ise modelin düzenleme terimini ifade eder (Chen ve Guestrin, 2016).

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (5)$$

Öznitelik seçimi ve modelin değerlendirilmesi

Modelin başarısını artırmak için yapılan özellik seçimi sürecinde, RO algoritması kullanılarak özniteliklerin önem dereceleri hesaplanmıştır ve Şekil 4’te gösterilmiştir. Ekonomi ve sağlık altyapısı en yüksek öneme sahipken, Yüzölçümü ve ulaşım verileri gibi değişkenler daha düşük önem derecelerine sahip olmuştur. Bu süreç, modelin daha doğru tahminler yapabilmesini sağlamak için gereksiz veya düşük etkiye sahip olan değişkenlerin analizden çıkarılmasını sağlamıştır.



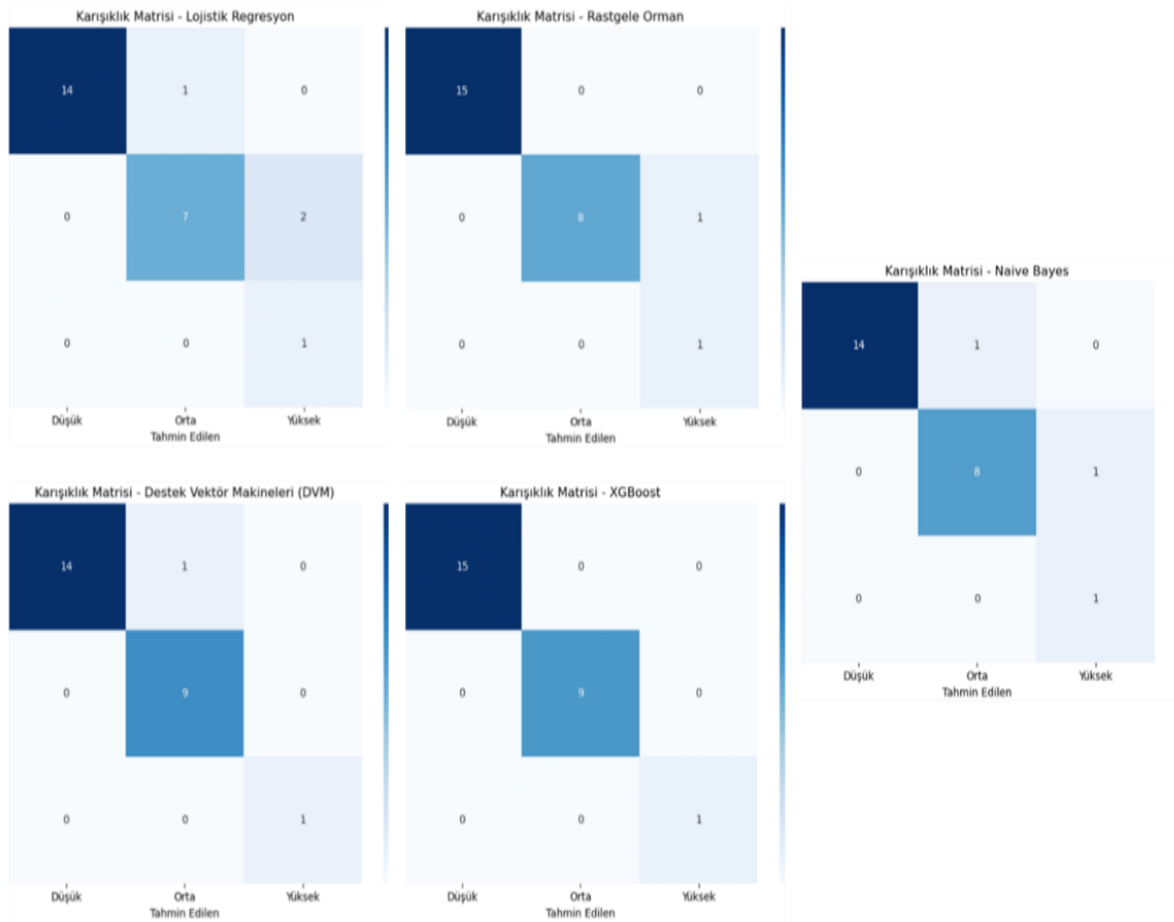
Şekil 4. Öznitelik önem skorları

Performans değerlendirme metrikleri

Makine öğrenimi modellerinin performansını değerlendirmek amacıyla çeşitli metrikler kullanılmıştır. Doğruluk oranı, modellerin başarılı tahmin yüzdelerini belirlemek için kullanılmıştır. ROC eğrisi, modellerin pozitif ve negatif sınıfları ayırt etme kapasitesini göstermiştir. AUC skoru, modellerin genel başarısını değerlendirmek için hesaplanmıştır. Karışıklık matrisi ise tahmin edilen ve gerçek değerler arasındaki uyumu görselleştirmek amacıyla oluşturulmuştur. Modeller arasındaki karşılaştırmalar sonucunda, RO, DVM ve XGB algoritmalarının en yüksek doğruluk oranına sahip olduğu belirlenmiştir.

Bu çalışmada kullanılan makine öğrenimi algoritmalarının performansları çeşitli performans metriklerinden yararlanılmıştır. Sınıflandırma problemlerinde en yaygın kullanılan değerlendirme metrikleri arasında doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru ve AUC (Eğri Altındaki Alan – Area Under Curve) yer almaktadır. Ancak, model performansını değerlendirmede yalnızca doğruluk oranına dayanmak yetersiz kalabilir; bu nedenle, F1 skoru ve AUC gibi daha bütüncül ölçütlerin birlikte kullanılması önerilmektedir (Tharwat, 2021). Özellikle dengesiz sınıf dağılımı olan veri setlerinde, Matthews korelasyon katsayısının (MCC), F1 skoru ve doğruluk oranına kıyasla daha tutarlı ve anlamlı bir performans ölçütü sunduğu belirtilmiştir (Chicco ve Jurman, 2020). Bu metriklerin her biri, modelin sınıflandırma başarımını farklı yönleriyle ele almakta ve karşılaştırmalar yapılmasını kolaylaştırmaktadır. Sınıflandırma modellerinin değerlendirilmesinde kullanılan temel yapı taşlarından biri de karmaşıklık matrisidir (confusion matrix). Bu matris, modelin tahmin ettiği sınıflar ile gerçek sınıfların karşılaştırılmasıyla elde edilir. İkili sınıflandırmada dört temel bileşen içerir: TP (True Positive): Modelin doğru pozitif tahminleri, TN (True Negative): Modelin doğru negatif tahminleri, FP (False Positive): Modelin yanlış pozitif tahminleri, FN (False Negative): Modelin yanlış negatif tahminlerini göstermektedir. Aşağıdaki grafiklerde, LR, RO, DVM,

XGB ve NB algoritmalarının her biri için elde edilen karmaşıklık matrisleri sunulmaktadır. Karmaşıklık matrisi, modelin hangi sınıflarda doğru ve yanlış tahminler yaptığını gösterir. Şekil 5’te, her bir modelin karmaşıklık matrisleri gösterilmektedir. Bu matrisler, her bir modelin düşük, orta ve yüksek risk sınıflarındaki doğruluğunu, yanlış pozitif ve yanlış negatif oranlarını görselleştirerek değerlendirme fırsatı sunmaktadır. Şekil incelendiğinde XGB modeli tüm sınıflarda doğru sınıflandırma oranı açısından en yüksek başarıyı göstermiştir. Özellikle düşük risk sınıfında NB modelinin duyarlılığı öne çıkmaktadır.



Şekil 5. Karmaşıklık matrisleri

Doğruluk (Accuracy): Modelin tüm tahminlerinin ne kadarının doğru olduğunu gösteren en yaygın kullanılan performans ölçütüdür (Tharwat, 2021). Formül 6’de gösterildiği gibi, doğruluk oranı doğru tahminlerin toplam tahminlere oranı olarak hesaplanır. Kesinlik (Precision): Modelin pozitif olarak tahmin ettiği örneklerin ne kadarının gerçekten doğru olduğunu gösterir (Chicco ve Jurman, 2020). XGB, %98 kesinlik değeri ile en yüksek kesinlik oranına sahip olmuştur. Bu, modelin yanlış pozitif tahminleri minimize ederek doğru sınıflandırmalar yaptığı anlamına gelir. Duyarlılık (Recall): Modelin gerçekten pozitif olan örnekleri doğru şekilde tahmin etme oranıdır (Lever ve ark., 2016). NB, düşük risk sınıfında en yüksek duyarlılık değerini elde etmiştir. Bu, modelin yanlış negatif tahminleri minimize ederek, tüm gerçek pozitifleri doğru şekilde tespit ettiğini gösterir. F1-Skor: Kesinlik ve duyarlılık değerlerini birleştirerek her iki metriği dengelemeye çalışır (Powers, 2011). RO ve XGB, F1-skor açısından en dengeli sonuçları elde etmiştir. Bu modeller hem kesinlik hem de duyarlılıkta iyi performans göstererek optimize edilmiş sonuçlar elde etmiştir.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{F1 – Skor} = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (9)$$

Risk haritası geliştirme

Salgın risk seviyelerinin coğrafi olarak görselleştirilmesi için Folium kütüphanesi kullanılarak Türkiye haritası üzerinde risk haritaları oluşturulmuştur. Bu süreçte, enlem ve boylam bilgilerini içeren veri seti kullanılarak Türkiye’deki illerin coğrafi konum bilgileri haritaya işlenmiş olup risk skorunu içeren veri seti ile illerin risk seviyeleri ve salgın durumu belirtilmiştir. İllerin risk seviyelerine göre renklendirilmiş harita oluşturulmuş, düşük risk yeşil, orta risk turuncu, yüksek risk ise kırmızı renk ile temsil edilmiştir. Harita, risk seviyelerinin coğrafi dağılımını analiz etmek için kullanılmıştır. Ayrıca, salgının farklı senaryolara göre yayılımını analiz etmek amacıyla simülasyon çalışmaları gerçekleştirilmiştir. Olası Türkiye risk haritası Şekil 6’da gösterildiği gibidir.



Şekil 6. Türkiye risk haritası

Marmara Bölgesi, İstanbul, Bursa ve Kocaeli gibi büyük sanayi ve ticaret merkezleri nedeniyle orta ve yüksek risk seviyesine sahipken, kırsal alanlarda risk daha düşüktür. Ege Bölgesi’nde, İzmir ve Manisa gibi sanayileşmiş şehirler orta riskli iken kıyı kesimlerinde daha düşük risk gözlemlenmiştir. Akdeniz Bölgesi, turizm ve göç hareketliliği nedeniyle Antalya ve Mersin gibi şehirlerde orta risk taşırken, kırsal kesimlerde risk seviyesi düşmektedir. İç Anadolu Bölgesi’nde, özellikle Ankara yüksek riskli olarak belirlenmişken, Konya gibi diğer büyük şehirlerde orta risk, kırsal alanlarda ise daha düşük risk seviyeleri gözlemlenmiştir. Karadeniz Bölgesi’nde, Samsun, Trabzon ve Ordu gibi kıyı şehirlerinde orta risk seviyeleri gözlemlenirken, dağlık bölgelerde risk daha düşüktür. Doğu Anadolu Bölgesi, genel olarak düşük risk taşısa da Van ve Erzurum gibi şehirlerde orta risk seviyeleri

görülmektedir. Güneydoğu Anadolu Bölgesi’nde ise sanayi ve nüfus yoğunluğunun yüksek olduğu Gaziantep ve Diyarbakır gibi şehirlerde risk seviyeleri orta ve yüksek arasında değişmektedir. Bu bölgesel analiz, Türkiye'nin farklı alanlarındaki risk faktörlerini görselleştirerek daha kapsamlı bir değerlendirme sunmaktadır.

BULGULAR VE TARTIŞMA

Bu bölümde, çalışmada kullanılan makine öğrenimi algoritmalarının performansını değerlendirmek amacıyla gerçekleştirilen deneysel süreçlere ilişkin sonuçlar sunulmaktadır. Deneyler, Türkiye'deki 81 ili kapsayan, demografik ve coğrafi bilgileri içeren bir veri seti üzerinden yürütülmüştür. Kullanılan veri seti; nüfus, yüzölçümü, nüfus yoğunluğu, sağlık altyapısı, ulaşım ağı ve ekonomik seviye gibi çeşitli değişkenleri içermektedir. İş sağlığı ve güvenliği gibi yüksek riskli alanlarda, çok kriterli karar verme yöntemlerinin risk değerlendirme süreçlerine katkısı sıkça vurgulanmaktadır (Dabbagh ve Yousefi, 2019). Veri setinde eksik olan değerler, benzer illerin ortalamaları ile tamamlanmış ve tüm değişkenler StandardScaler yöntemiyle ölçeklendirilmiştir. Ayrıca, öznitelik seçimi aşamasında RO algoritması kullanılarak değişkenlerin önem dereceleri hesaplanmış ve düşük etkili değişkenler analizden çıkarılmıştır. Modelin genelleme kabiliyetini artırmak amacıyla 5 katlı çapraz doğrulama ($k=5$) yöntemi uygulanmıştır. Bu sayede, veri her iterasyonda farklı bir alt küme üzerinde test edilerek aşırı öğrenmenin (overfitting) önüne geçilmiştir. Deneysel çalışmalarda, LR, NB, RO, DVM ve XGB algoritmaları kullanılmıştır. Her bir model için hiperparametre optimizasyonu GridSearchCV yöntemiyle gerçekleştirilmiş ve en iyi parametre kombinasyonları belirlenmiştir. Örneğin, LR için çözüm yöntemi ve düzenleme katsayısı; RO için ağaç sayısı, maksimum derinlik ve minimum örnek bölme sayısı; DVM için kernel tipi ve düzenleme parametresi; XGB için öğrenme oranı, ağaç sayısı ve derinlik gibi parametreler optimize edilmiştir. NB modeli için de varyans yumuşatma parametresi ayarlanarak eğitim gerçekleştirilmiştir. Seçilen parametrelerle tüm modeller yeniden eğitilmiş ve test verileri üzerinde değerlendirilmiştir.

Değerlendirme sonucunda, XGB algoritması %98 doğruluk oranı ile en yüksek başarıyı göstermiştir. RO ve DVM algoritmaları %96 doğruluk oranı ile güçlü performans sergilerken, NB %92 doğruluk oranıyla hızlı tahmin avantajı sunmuştur. LR ise %88 doğruluk oranı ile kabul edilebilir ancak görece daha düşük performans göstermiştir. XGB modeli, aynı zamanda kesinlik, duyarlılık ve F1 skorlarında da en yüksek değerleri sunmuş; RO ve DVM ise bu metriklerde dengeli sonuçlar elde etmiştir. NB modeli, özellikle düşük riskli sınıflarda yüksek duyarlılık göstermiştir. LR'nin özellikle yüksek riskli sınıflarda performans kaybı yaşadığı gözlemlenmiştir. LR modelinin doğruluk oranının görece düşük olmasının temel nedeni, doğrusal sınıflandırma yapısına sahip olmasıdır. Bu yapı, çok boyutlu ve doğrusal olmayan sınıf ayrımlarını öğrenmede yetersiz kalabilir. NB algoritması ise, öznitelikler arasında bağımsızlık varsayımı yapması nedeniyle karmaşık öznitelik ilişkilerini tam olarak yansıtamaz. Bu nedenle, NB modeli hızlı tahmin yeteneğine sahip olsa da özellikle çok değişkenli veri setlerinde sınırlı genelleme başarımı göstermektedir. Deneysel sonuçların özetlendiği Çizelge 1’de algoritmalara ilişkin doğruluk, kesinlik, duyarlılık ve F1 skoru değerleri yer almaktadır. Bu tablo incelendiğinde, XGB'nin tüm metriklerde en yüksek değerleri sunduğu, RO ve DVM'nin de oldukça başarılı sonuçlar verdiği, NB'nin hız açısından avantajlı ancak pozitif tahminlerde sınırlı olduğu görülmektedir. LR modeli ise nispeten daha düşük metrikler sunmakla birlikte kabul edilebilir performans sergilemiştir.

Çizelge 1. Deneysel Çalışma Sonuçları

| Algoritma | Doğruluk | Kesinlik | Duyarlılık | F1 |
|-----------|----------|----------|------------|-------|
| RO | 96 | 83.33 | 96.30 | 86.93 |
| NB | 92 | 79.63 | 94.07 | 84.04 |
| XGB | 98 | 97.56 | 98.00 | 97.78 |
| LR | 88 | 73.61 | 90.37 | 76.30 |
| DVM | 96 | 79.67 | 97.78 | 97.10 |

Model başarısını etkileyen özneliklerin katkısını daha net ortaya koymak amacıyla ablation analizi gerçekleştirilmiştir. Bu analizde, her bir öznelik tek tek modelden çıkarılarak doğruluk ve F1 skorlarındaki değişimler gözlemlenmiştir. Ekonomi değişkeni çıkarıldığında, XGB algoritmasının doğruluk oranı %98’den %94’e düşmüş; sağlık altyapısı çıkarıldığında RO modelinin F1 skoru %86.93’ten %82.1’e gerilemiştir. DVM ve NB modelleri bu değişimlere karşı daha stabil performans sergilerken, LR modelinin özellikle yüksek risk sınıflarındaki duyarlılığı belirgin şekilde azalmıştır. Bu bulgular, öznelik seçiminin model başarısı üzerinde önemli bir etkisi olduğunu göstermektedir.

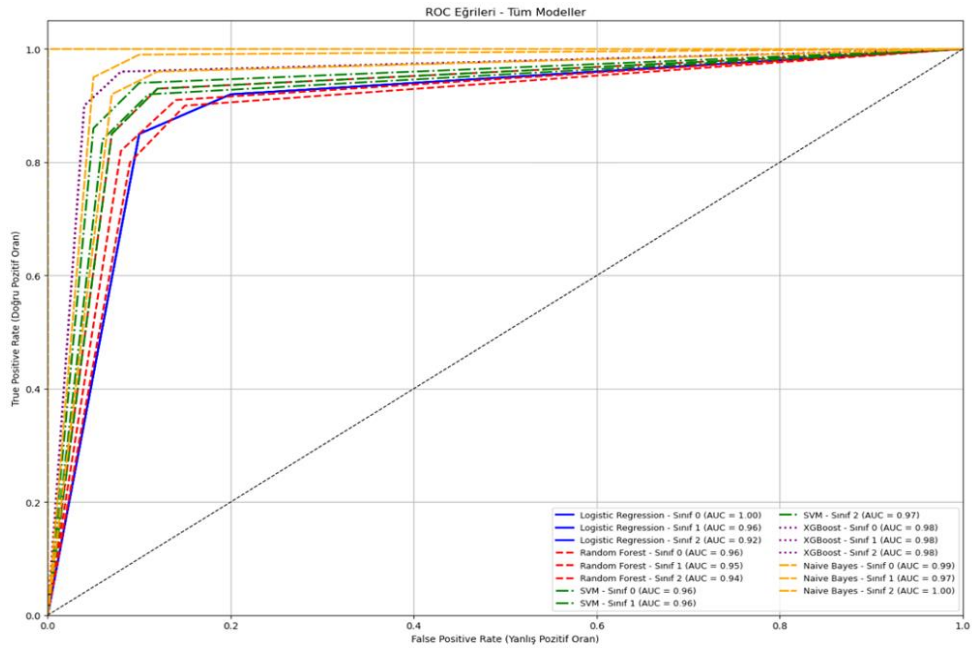
Ayrıca, modellerin ROC eğrileri incelenmiş ve modellerin farklı risk seviyelerinde gösterdiği ayırım gücü değerlendirilmiştir. Şekil 7’de sunulan ROC eğrisi, her bir makine öğrenimi algoritmasının sınıflandırma performansını görsel olarak ortaya koymaktadır. Bu eğriler, modelin pozitif sınıfı ayırt etme başarısını ölçen AUC (Eğri Altındaki Alan) değerlerine dayalı karşılaştırma imkânı sunmaktadır. XGB algoritması, tüm sınıflar için %98 AUC değeri ile en başarılı model olurken, RO ve DVM algoritmaları yaklaşık %96 AUC ile oldukça güçlü bir ayırım gücüne sahiptir. Buna karşılık, Lojistik Regresyon ve Naive Bayes algoritmalarının özellikle düşük ve yüksek riskli sınıflarda ayırım başarısı daha düşüktür.

COVID-19 salgını, özellikle İstanbul, Ankara ve İzmir gibi metropol kentlerde hızlı bulaş örnekleriyle kent yoğunluğu ve hareketliliğin etkisini açıkça ortaya koymuştur. Bu doğrultuda, yoğun nüfuslu bölgelerde gerçek zamanlı izleme sistemlerinin geliştirilmesi gerekliliği ortaya çıkmıştır. Gündüz ve Işık (2023a) bu ihtiyaca yönelik YOLO tabanlı kalabalık tespiti geliştirilmiştir. Ayrıca Gündüz ve Işık (2023b) sosyal mesafe analizi sistemleri güncel çözümler sunmaktadır. Bu sistemler, şehir içi alanlarda anlık yoğunluk takibi sağlayarak, epidemiyolojik karar destek süreçlerinde yapay zekâ uygulamalarının etkinliğini göstermektedir. Çalışmamızda önerilen bölgesel risk haritalandırma yaklaşımı da bu tür teknolojik sistemlerle entegre edilebilecek yapıda olup, karar vericilere dinamik ve görsel odaklı bir değerlendirme zemini sunmaktadır.

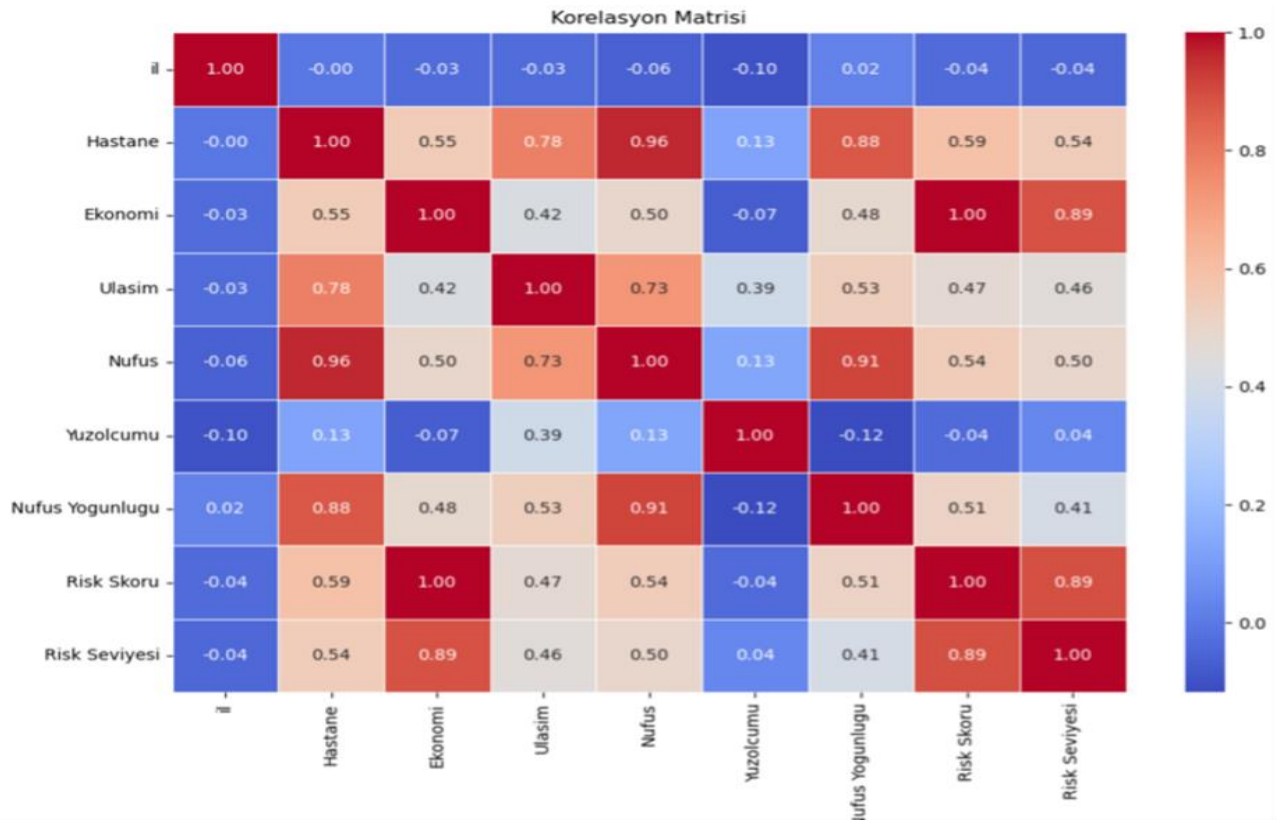
Ayrıca, korelasyon matrisi oluşturularak değişkenler arasındaki ilişkiler analiz edilmiştir. Şekil 8’de yer alan korelasyon matrisi, veri setinde yer alan değişkenler arasındaki pozitif ve negatif ilişkileri göstermektedir. Nüfus yoğunluğu, ulaşım altyapısı ve sağlık hizmetlerine erişim gibi faktörlerin sağlık risk seviyeleri ile yüksek korelasyon gösterdiği açıkça görülmektedir. Buna karşın yüzölçümü gibi bazı değişkenlerin etkisinin düşük olduğu görülmüş ve bu tür değişkenler modelden çıkarılmıştır. Bu analiz, öznelik mühendisliği sürecinde doğru değişken seçiminin model başarısı üzerinde ne kadar etkili olduğunu ortaya koymaktadır.

Son olarak, bu çalışmada kullanılan simülasyon yaklaşımı, salgının yüksek riskli illerden çevresine yayılma sürecini modellemeye yöneliktir. Simülasyon senaryoları, il bazında risk seviyeleri ile coğrafi komşuluk ilişkilerine dayanılarak oluşturulmuştur. Nüfus, nüfus yoğunluğu, yüzölçümü, sağlık altyapısı (hastane sayısı), ulaşım erişilebilirliği, ekonomik seviye, enlem-boylam konumu ve risk seviyesi gibi değişkenler, ağırlık katsayıları olarak tanımlanmış ve bu katsayılar doğrultusunda risk geçiş ihtimalleri hesaplanmıştır. Her adımda, yüksek riskli illerden başlayarak komşu illere doğru yayılım modellemesi yapılmış ve zamansal risk artış dinamikleri analiz edilmiştir. Elde edilen

simülasyon çıktıları, Türkiye haritası üzerinde ısı haritası (Şekil 9) biçiminde görselleştirilmiştir. Bu, karar vericilere müdahale önceliklerini belirlemede zamana bağlı risk dönüşümünü inceleme sağlar.



Şekil 7. ROC eğrisi sonuçları



Şekil 8. Korelasyon matrisi



Şekil 9. Türkiye salgın riskine ait ısı haritası

SONUÇ

Bu çalışmada, Türkiye'nin illerine ait sağlık risk seviyelerinin belirlenmesinde ve bu seviyelerin coğrafi olarak haritalandırılmasında makine öğrenimi algoritmaları kullanılmıştır. Risk analizi için iki temel veri seti oluşturulmuş olup bütünleştirilmiştir. Veri setleri, 973 ilçeye ait coğrafi konum bilgilerini ve 81 il için nüfus, yüzölçümü, nüfus yoğunluğu, hastane sayısı, ulaşım altyapısı ve ekonomik seviye faktörleri olmak üzere iki addettir. Bu değişkenler risk skorlarını tahmin etmek amacıyla analiz edilmiştir.

LR, NB, RO, DVM ve XGB algoritmaları, sağlık risklerini tahmin etmek için uygulanmış ve elde edilen sonuçlar çeşitli performans metrikleri ile değerlendirilmiştir. Deneysel çalışma sonuçlarına göre, XGB algoritması ile tüm metriklerde en yüksek başarı elde edilerek %98 doğruluk oranı elde edilmiştir. Bu sonuç, XGB'nin sağlık risklerinin doğru bir şekilde sınıflandırılmasında son derece etkili olduğunu göstermektedir. RO ve DVM algoritmaları, sırasıyla %96 doğruluk oranlarıyla güçlü performans sergilemiş, ancak kesinlik ve duyarlılık değerlerinde XGB'ye kıyasla daha düşük sonuçlar elde edilmiştir. NB ve LR modelleri ise sırasıyla %92 ve %88 doğruluk oranlarına sahip olmuş, düşük doğruluk oranlarına rağmen NB beklenenden yüksek doğruluk göstermiştir. F1-skoru, modellerin kesinlik ve duyarlılık arasında denge kurarak her iki metriği optimize etme yeteneğini ölçmektedir. XGB ve RO, F1-skoru açısından en dengeli sonuçları elde ederek her iki metrikte de yüksek performans sergilemiştir. Bu modeller, özellikle yüksek risk sınıflarında doğru sınıflandırmalar yaparken, düşük risk sınıflarında ise yanlış pozitifleri minimize etme konusunda başarılı olmuştur.

NB, düşük hesaplama maliyeti ile hızlı tahminler yapabilen bir model olarak öne çıkmaktadır. Ancak, yanlış pozitiflerin oranı yüksek olabileceğinden, düşük risk sınıfı için daha uygun bir seçenek olarak değerlendirilebilir. Diğer yandan, Lojistik Regresyon ve DVM, daha basit ve pratik modeller arayanlar için uygun seçeneklerdir. Ancak, özellikle yüksek risk sınıflarında performans kayıpları yaşanabilmektedir. Bu çalışma, sağlık risklerini tahmin etmek için makine öğrenimi tekniklerinin güçlü bir araç olduğunu ve bu tekniklerin sağlık politikalarını şekillendirme ve bölgesel sağlık stratejilerini oluşturma sürecinde önemli bir rol oynayabileceğini ortaya koymaktadır. Simülasyon analizleri sayesinde, iller arasındaki sağlık risklerinin coğrafi dağılımı ve yüksek risk bölgeleri görselleştirilmiş, bu da sağlık yönetiminde karar alıcılar için önemli bir kaynak sağlamıştır. Ayrıca, oluşturulan sağlık risk haritası, Türkiye'nin illerindeki sağlık risk seviyelerinin coğrafi dağılımını görselleştirmiştir. Bu ısı haritası, yüksek riskli bölgelerin kırmızı, orta riskli bölgelerin turuncu, düşük riskli bölgelerin ise yeşil renkte temsil edilmesini sağlamıştır. Sağlık riski yüksek olan bölgelerin

görsel olarak belirlenmesi, sağlık hizmetlerinin kaynaklarının daha etkin bir şekilde tahsis edilmesine ve stratejik kararlar alınmasına olanak tanımaktadır.

Gelecekteki çalışmalar, gerçek zamanlı veri entegrasyonu, derin öğrenme ve daha geniş veri setleri kullanılarak bu modellerin daha kapsamlı ve genellenebilir hale getirilmesini sağlayabilir. Ayrıca, sağlık risklerinin dinamiklerini daha iyi anlayabilmek için daha ileri düzey simülasyon teknikleri ve zaman serisi analizleri üzerinde durulabilir. Özellikle LSTM veya GRU gibi zaman serisi temelli derin öğrenme mimarileri, salgın yayılım eğilimlerinin dinamik biçimde modellenmesini sağlayabilir. Bu mimariler, mevcut yapı ile entegre edilerek anlık veri akışına dayalı olarak bölgesel risk tahminlerinin güncellenmesine olanak tanıyabilir. Böylece önerilen sistem, yalnızca statik analiz değil, aynı zamanda gerçek zamanlı karar destek aracı olarak da kullanılabilir hâle gelecektir.

Çıkar Çatışması

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

Yazar Katkısı

Yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan eder.

KAYNAKLAR

- Chen, T., ve Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
- Chicco, D., ve Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21, 1-13. <https://doi.org/10.1186/s12864-019-6413-7>
- Chumachenko, D., Meniaailov, I., Bazilevych, K., Chumachenko, T., ve Yakovlev, S. (2022). Investigation of statistical machine learning models for COVID-19 epidemic process simulation: Random forest, K-nearest neighbors, gradient boosting. *Computation*, 10(6), 86. <https://doi.org/10.3390/computation10060086>
- Çakmak, M. A., Kurt, M. E., ve Çakmak, C. (2022). Makine Öğrenmesi Algoritmaları ile Covid-19 Hastalarının Mortalite Risklerinin Hesaplanması. *Süleyman Demirel Üniversitesi Vizyoner Dergisi*, 13(35), 994-1011. <https://doi.org/10.21076/vizyoner.1074212>
- Dabbagh, R., ve Yousefi, S. (2019). A hybrid decision-making approach based on FCM and MOORA for occupational health and safety risk analysis. *Journal of safety research*, 71, 111-123. <https://doi.org/10.1016/j.jsr.2019.09.021>
- Edward, J., ve Biddle, D. J. (2017). Using geographic information systems (GIS) to examine barriers to healthcare access for Hispanic and Latino immigrants in the US south. *Journal of racial and ethnic health disparities*, 4, 297-307. <https://doi.org/10.1007/s40615-016-0229-9>
- Goodfellow, I. J., Shlens, J., ve Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*. <https://doi.org/10.48550/arXiv.1412.6572>
- Guido, R., Ferrisi, S., Lofaro, D., ve Conforti, D. (2024). An overview on the advancements of support vector machine models in healthcare applications: a review. *Information*, 15(4), 235. <https://doi.org/10.3390/info15040235>
- Gündüz, M. Ş., ve Işık, G. (2023a). A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models. *Journal of Real-Time Image Processing*, 20(5). <https://doi.org/10.1007/s11554-023-01276-w>

- Gündüz, M. Ş., ve Işık, G. (2023b). A new YOLO-based method for social distancing from real-time videos. *Neural Computing and Applications*, 35, 15261–15271. <https://doi.org/10.1007/s00521-023-08556-3>
- Karcıoğlu, A. A., Tanışman, S., ve Bulut, H. (2021). Türkiye'de COVID-19 Bulaşısının ARIMA Modeli ve LSTM Ağı Kullanılarak Zaman Serisi Tahmini. *Avrupa Bilim ve Teknoloji Dergisi*, (32), 288-297. <https://doi.org/10.31590/ejosat.1039394>
- Khalid, J., Chuanmin, M., Altaf, F., Shafqat, M. M., Khan, S. K., ve Ashraf, M. U. (2024). AI-Driven Risk Management and Sustainable Decision-Making: Role of Perceived Environmental Responsibility. *Sustainability*, 16(16), 6799. <https://doi.org/10.3390/su16166799>
- Lever, J., Krzywinski, M., ve Altman, N. (2016). Points of significance: classification evaluation. *Nature methods*, 13(8), 603-604. <https://doi.org/10.1038/nmeth.3945>
- Madry, A., Makelov, A., Schmidt, L., Tsipras, D., ve Vladu, A. (2017). Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*. <https://doi.org/10.48550/arXiv.1706.06083>
- Miyazaki, Y., Kawakami, M., Kondo, K., Hirabe, A., Kamimoto, T., Akimoto, T., ... ve Tsuji, T. (2024). Logistic regression analysis and machine learning for predicting post-stroke gait independence: a retrospective study. *Scientific reports*, 14(1), 21273. <https://doi.org/10.1038/s41598-024-72206-4>
- Naim, A. (2023). Application of machine learning techniques to identify the business financial risks. *Academy of Marketing Studies Journal*, 27(S5), 1-11.
- Nazia, N., Butt, Z. A., Bedard, M. L., Tang, W. C., Sehar, H., ve Law, J. (2022). Methods used in the spatial and spatiotemporal analysis of COVID-19 epidemiology: a systematic review. *International journal of environmental research and public health*, 19(14), 8267. <https://doi.org/10.3390/ijerph19148267>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... ve Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, 12, 2825-2830.
- Powers, David ve Ailab., (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness ve correlation. *J. Mach. Learn. Technol.* 2. 2229-3981. <https://doi.org/10.9735/2229-3981>
- Shastri, S., Singh, K., Kumar, S., Kour, P., ve Mansotra, V. (2020). Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. *Chaos, Solitons ve Fractals*, 140, 110227. <https://doi.org/10.1016/j.chaos.2020.110227>
- Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., ve Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, 237, 121549. <https://doi.org/10.1016/j.eswa.2023.120515>
- Tharwat, A. (2021). Classification assessment methods. *Applied computing and informatics*, 17(1), 168-192. <https://doi.org/10.1016/j.aci.2018.08.003>