

COVID-19 Pandemisinde Salgın Tahmini İçin Geleneksel Makine Öğrenmesi Yöntemlerin Karşılaştırılması: Üç Ülke Perspektifi

Hazal Su BİÇAKCI YEŞİLKAYA^{1*}

¹Elektrik Elektronik Mühendisliği / Mühendislik ve Mimarlık Fakültesi, İzmir Bakırçay Üniversitesi, Türkiye

*(hazalsu.bicakci@bakircay.edu.tr)

(Received: 11 September 2025, Accepted: 21 September 2025)

(2nd International Conference on Pioneer and Academic Research ICPAR 2025, September 15-16, 2025)

ATIF/REFERENCE: Yeşilkaya Bıçakçı, H. S. (2025). COVID-19 Pandemisinde Salgın Tahmini İçin Geleneksel Makine Öğrenmesi Yöntemlerin Karşılaştırılması: Üç Ülke Perspektifi, *International Journal of Advanced Natural Sciences and Engineering Researches*, 9(9), 142-149.

Özet – Covid-19 pandemisinin matematiksel ve hesaplamalı yöntemlerle modellenmesi, mühendislik alanında önemli bir araştırma konusu haline gelmiştir. Özellikle makine öğrenmesi algoritmaları, salgının seyrini öngörmeye yüksek doğruluk oranları ile dikkat çekmektedir. Bu çalışmada iki farklı yaklaşım karşılaştırılmıştır: doğrusal regresyon ve karar ağacı modelleri. Çalışmada kullanılan veri seti Kaggle üzerinden elde edilmiş ve Türkiye, Çin ile ABD'ye ait günlük vaka verileri kullanılmıştır. Modellerin performansı ortalama mutlak hata (MAE), ortalama karekök hata (RMSE), R-kare (R^2) ve doğruluk değerleri ile değerlendirilmiştir. Elde edilen sonuçlar, doğrusal regresyon modelinin basit veri yapılarında kabul edilebilir performans sunduğunu ancak karmaşık dalgalı veri setlerinde yetersiz kaldığını göstermiştir. Karar ağacı yöntemi yüksek doğruluk değerleri ile öne çıkmaktadır. Bu çalışma, mühendislik odaklı bir yaklaşımla Covid-19 gibi epidemiyolojik süreçlerin modellenmesinde makine öğrenmesi yöntemlerinin etkinliğini ortaya koymaktadır.

Anahtar Kelimeler – Covid-19 Sınıflandırma, Doğrusal Regresyon, Karar Ağaçları, Salgın Tahmini, Makine Öğrenmesi.

I. GİRİŞ

Covid-19 2019'un son zamanlarında ilk olarak Çin'in Wuhan şehrinde belirtilerini ve etkisini göstermiş, daha öncesinde bilim insanlarınca tespiti yapılan SARS-CoV virüsü ile benzerliği dikkat çekmiş bu nedenle SARS-CoV-2 adını alıp yeni bir virüs türü olarak yayılmaya başlamıştır [1]. Dünya Sağlık Örgütü, hastalık artarak devam ettiği için pandemi kararını 11 Mart 2020 tarihinde vermiştir. Günümüz verileri itibarıyla dünya genelinde Covid-19 ile bağlantılı olarak, 777.720.205 resmi vaka, 7.094.477 ölüm vakası Dünya sağlık örgütü verilerine göre bildirilmiştir [2]. Covid-19 virüsüne yakalanma ve bu hastalığı en şiddetli düzeyde geçirme olasılığı yüksek risk taşıyan gruplar; diyabet, hipertansiyon, astım, kronik rahatsızlıkları, kanser ve 65 yaş üstü kişiler olarak belirtilmektedir [3]. Pandeminin gerçekleştiği dönemde sağlık kuruluşlarının ihtiyaçları ve talepleri vaka sayısına oranla yetersiz kaldığı için problemler ortaya çıkmıştır. Bunun üzerine bu gibi sorunların planlamasının yapılabilmesi için vaka sayılarına ihtiyaç duyulmuş, vaka sayılarının tahmin edilebilmesi önem kazanmıştır.

Epidemiyolojik verilerin tahmini ve modellenmesi, mühendislik disiplininde özellikle yapay zeka ve veri bilimi alanlarının ön plana çıkmasına yol açmıştır. Covid-19 pandemisi süresince, vaka artışlarının

öngörülmesi için kullanılan algoritmalar arasında makine öğrenmesi yöntemleri dikkat çekmiştir [4]. Bu yöntemler, yüksek boyutlu ve gürültülü veri setlerinde bile öğrenme kabiliyetine sahip olmaları nedeniyle mühendislik uygulamalarında geniş bir kullanım alanı bulmaktadır.

Makine öğrenmesi, veriler arasındaki karmaşık ilişkileri modelleyerek hem doğrusal hem de doğrusal olmayan eğilimleri yakalayabilmektedir [5]. Doğrusal regresyon, basitliği ve anlaşılır yapısı sayesinde başlangıç seviyesinde etkili bir yöntem sunmaktadır. Karar ağaçları, verilerin özelliklerine göre ayrıştırılması sayesinde daha esnek ve yüksek doğruluk oranları ile öne çıkmaktadır [6]. Örneğin; Özen vd. [7], 21 Ocak–6 Aralık 2020 dönemine ait The New York Times Covid-19 verilerini kullanarak ABD’deki vaka ve ölüm sayılarını makine öğrenmesi tabanlı tahmin modelleri ile incelemiştir. Prophet, Polinom Regresyon, ARIMA, Doğrusal Regresyon ve Random Forest algoritmaları karşılaştırılmış; doğrusal regresyon modeli için MAPE %27,99, RMSE 3672745 ve MAE 3467277 olarak raporlanmıştır. Benzer şekilde, Özel ve Demirsöz [8] Türkiye Sağlık Bakanlığı verileri (Mart 2020–Mart 2021) ile günlük vaka, iyileşen ve ölüm sayılarını Random Forest algoritması ile modellemiştir. Çalışmada %90 kesinlik, %89 duyarlılık ve %89 F1-skor elde edilmiştir. Çakmak vd. [9] ise Atlanta, Georgia’daki bir sağlık kuruluşuna ait hasta verilerini kullanarak Adaptive Boost, Karar Ağacı ve Random Forest algoritmalarını uygulamış, Karar Ağacı modelinde %81,4 doğruluk elde etmiştir. Elde edilen tahminler, hastane kaynak planlamasında (örneğin yatak ve ekipman ihtiyacı) etkin kullanılabilecek karar destek unsurları sunmaktadır.

Derin öğrenme modellerini kullanan çalışmalar incelendiğinde, Singh vd. [10] zaman serisi verilerinde LSTM tabanlı modellerin güçlü tahminler sunduğunu, Zeroual vd. [11] ise CNN, RNN ve LSTM gibi derin öğrenme yöntemlerinin klasik yöntemlere kıyasla daha başarılı olduğunu belirtmiştir. Ayrıca, Ribeiro vd. [12] Brezilya için makine öğrenmesi tabanlı vaka tahmini çalışmasında Random Forest ve Destek Vektör Makinelerinin etkinliğini ortaya koymuştur. Shastri vd. [13] ise farklı ülkelerde epidemiyolojik tahmin modellerini karşılaştırarak hibrit modellerin daha etkili sonuçlar verdiğini göstermiştir. Bu literatür bulguları, makine öğrenmesi yöntemlerinin farklı ülkelerde ve farklı veri setlerinde kullanıldığında öngörü gücünü artırdığını, ancak her yöntemin avantaj ve sınırlılıklarının bulunduğunu göstermektedir.

Bu çalışmada Türkiye, Çin ve ABD’ye ait vaka verileri kullanılarak iki farklı yöntem detaylı bir şekilde karşılaştırılmıştır. Amaç, mühendislik bakış açısıyla algoritmaların güçlü ve zayıf yönlerini ortaya koymak, salgın tahmini gibi kritik bir problemde hangi yöntemlerin daha etkili olduğunu belirlemektir. Ayrıca çalışmanın bulguları, literatürdeki benzer mühendislik uygulamaları ile karşılaştırılmış ve algoritmaların performansı farklı metrikler üzerinden detaylı olarak tartışılmıştır.

II. MATERYAL VE YÖNTEM

Bu bölümde çalışmada kullanılan veri seti, tahmin modelleri ve performans ölçütleri açıklanmıştır. Bu çalışmanın yürütülmesinde Matlab R2024b programı ve Lenovo 12th Gen Intel(R) Core(TM) i5-1235U 1.30 GHz, 16 GB RAM bilgisayar kullanılmıştır.

A. Veri Seti

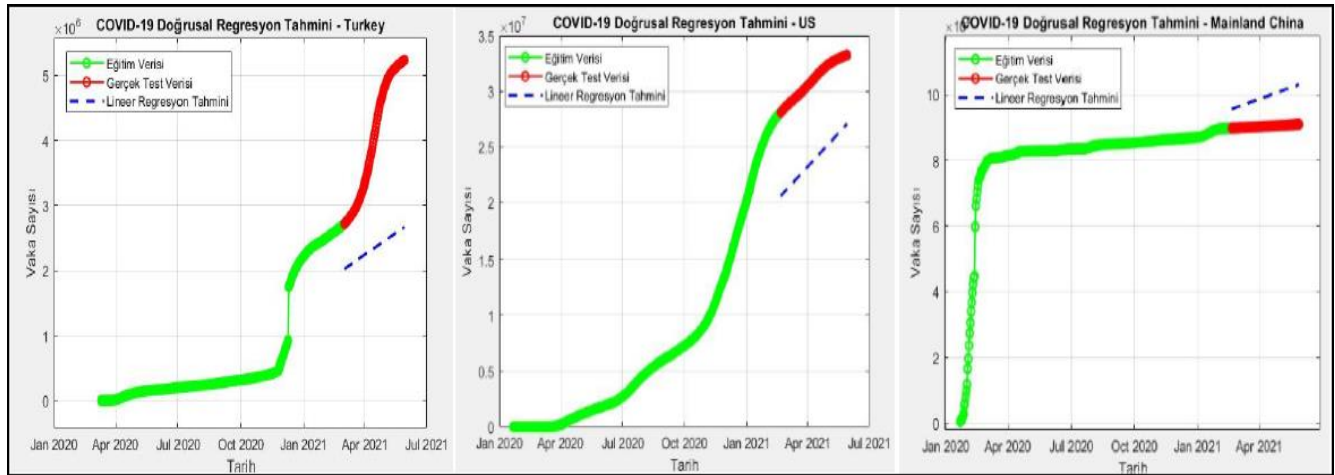
Novel Corona Virus 2019 Dataset [14], Covid-19 pandemisinin başlangıç dönemine ait küresel düzeyde vaka, ölüm ve iyileşme bilgilerini içeren zaman serisi tabanlı bir veri setidir. Veri, 22 Ocak 2020 tarihinden itibaren günlük olarak güncellenmiş olup 29 Mayıs 2021 tarihinde sonlandırılmıştır. Her bir gözlem satırı belirli bir ülke veya bölgeye ait o günkü kümülatif vaka durumunu yansıtmaktadır. Bu yönüyle veri seti, pandeminin zamansal ve coğrafi yayılımının analizini mümkün hale getirmektedir. Veri seti, Johns Hopkins Üniversitesi, Dünya Sağlık Örgütü (WHO), Çin CDC, ABD CDC gibi uluslararası sağlık kuruluşlarının sağladığı açık kaynaklı bilgilerden derlenmiştir. Ayrıca veri seti, Kaggle platformunda kamuya açık olarak sunulmuş ve birçok araştırmacı tarafından kullanılmıştır [14].

Veri seti içeriğinde: gözlem tarihi (Observation Date), eyalet/bölge (Province/State), ülke/bölge (Country/Region), son güncelleme zamanı (Last Update), onaylanmış vaka sayısı (Confirmed), ölüm sayısı (Deaths) ve iyileşen hasta sayısı (Recovered) bulunmaktadır. Bu çalışmada vakaların oldukça sık görüldüğü ABD, hastalığın başladığı Çin değerlendirilmiş ve Türkiye ile karşılaştırılmıştır. Toplam bildirilen hasta sayısına bakıldığında Türkiye'den 16671848 veri, ABD'den 93497149 veri ve Çin'den 1757073 veri sağlanmıştır. Veri hazırlığı aşamasında seçilen üç ülkenin verileri diğer ülke ve eyaletlerden ayrılmıştır. Elde edilen veri tablosundaki eksik veriler (Null) çıkarılmıştır.

B. Tahmin Modelleri

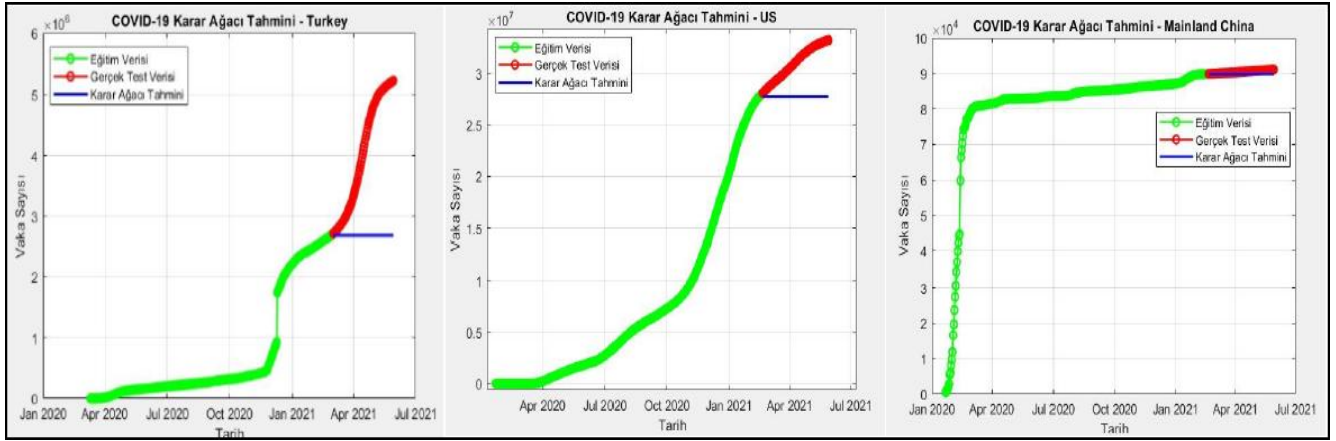
Klasik makine öğrenmesi yöntemlerinden doğrusal regresyon ve karar ağacı yöntemleri seçilip üzerinde çalışılmıştır. Covid-19 gibi ani dalgalanmalar içeren veri setlerinde yetersiz kalabilse de doğrusal regresyonun seçilme nedeni bağımlı ve bağımsız değişkenler arasında doğrusal bir ilişki olup olmadığının incelenmesidir [15]. Karar ağaçları ise regresyon problemlerinde daha esnek tahminler yapabildiği ve doğrusal olmayan verilerde modelleme avantajı sağladığı için tercih edilmiştir [16]. Tüm modellerde verilerin 80%'i eğitim ve 20%'si test için ayrılmıştır. Eğitim verileri, hastalığın başından itibaren olan verilerden seçilerek, modelin daha ileri aşamalar ile tahmin yapması amaçlanmıştır. Bu şekilde salgının nasıl devam edeceği konusunda modeller tahmin yapmıştır. Türkiye tahminleri için 445 örneklemin 356 tanesi eğitim setinde 89 tanesi ise test setinde kullanılmıştır. Çin ve ABD tahmin modelleri için 494 örneklemin 395 tanesi eğitim setinde ve 99 tanesi test setinde kullanılmıştır.

Şekil 1'de doğrusal regresyon modelinin salgın tahmini gösterilmiştir. Yeşil ve kırmızı renkler ile gösterilenler sırasıyla eğitim ve test verilerini ifade ederken mavi kesikli çizgi ile gösterilen doğru boyunca modelin tahminidir. Şekil 1'de görüldüğü gibi doğrusal regresyon modeli her üç ülke tahmininde de başarılı bir performans gösterememiştir.



Şekil 1. Doğrusal Regresyon modelinin üç ülke için tahmin edilen vaka sayısı grafiği

Şekil 2'de karar ağacı modelinin salgın tahmini gösterilmiştir. Sonuçlar incelendiğinde doğrusal regresyon modeline göre daha anlamlı sonuçlar elde edildiği görülmektedir. Özellikle Çin örneğini incelediğimizde modelin performansının diğer ülkelere göre çok yüksek olduğu görülmektedir.



Şekil 2. Karar Ağacı modelinin üç ülke için tahmin edilen vaka sayısı grafiği

C. Performans Metrikleri

1. Ortalama Mutlak Hata (MAE)

Ortalama Mutlak Hata (MAE), regresyon problemlerinde modelin tahmin performansını değerlendirmek için kullanılan temel hata ölçütlerinden biridir. Matematiksel olarak Denklem (1)'de ifade edilmiştir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

Burada: n toplam örnek sayısını, y_i gerçek değeri ve $\hat{y}_i$ tahmin edilen değeri göstermektedir [17]. MAE, hataların büyüklüğünü doğrudan ölçer ve birim olarak hedef değişkenle aynıdır. Bu nedenle yorumlaması kolaydır. MAE'nin düşük olması, modelin tahminlerinin gerçek değerlere daha yakın olduğunu gösterir.

2. Ortalama Karekök Sapması (RMSE)

RMSE, tahmin modelinin doğruluğunun ölçümünü yapan istatistik üzeri bir metriktir. RMSE, tahmin edilen değerler ile gerçek değerler arasındaki farkların karesinin ortalamasının kareköküdür [18]. Hataların büyüklüğünü ölçer ve büyük hatalara daha fazla ağırlık verir. Matematiksel olarak Denklem (2)'de ifade edilmiştir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

RMSE, özellikle büyük hataların önemli olduğu durumlarda tercih edilir. MAE'ye göre daha duyarlıdır çünkü hataları karesini alarak büyütür.

3. Determinasyon Katsayısı (R^2)

R^2 , bağımlı değişkendeki toplam varyansın ne kadarının model tarafından açıklandığını gösterir. 0 ile 1 arasında bir değer alır [19]. 1 değeri alması mükemmel uyumdur ve model tüm varyansı açıklar. 0 değerinde ise model hiçbir varyansı açıklayamaz. Matematiksel olarak Denklem (3)'de ifade edilmiştir.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (3)$$

Burada SS_{res} artık kareler toplamını yani açıklanmayan varyans değerini ifade ederken SS_{tot} ise toplam kareler toplamı yani verinin toplam varyansını ifade etmektedir.

4. Karmaşıklık Matrisi

Karmaşıklık matrisi, sınıflandırma algoritmalarının performansını değerlendirmek için kullanılan temel araçlardan biridir. Bu matris, modelin tahmin ettiği sınıflar ile gerçek sınıflar arasındaki ilişkiyi iki boyutlu bir tablo şeklinde gösterir. Satırlar genellikle gerçek sınıfları, sütunlar ise modelin tahmin ettiği sınıfları temsil eder [23]. Karmaşıklık matrisi dört temel bileşenden oluşur. Bunlar [20]:

- True Positive (TP): Gerçek pozitif ve doğru tahmin edilen örneklerdir.
- True Negative (TN): Gerçek negatif ve doğru tahmin edilen örneklerdir.
- False Positive (FP): Gerçek negatif ama yanlışlıkla pozitif tahmin edilen örneklerdir.
- False Negative (FN): Gerçek pozitif ama yanlışlıkla negatif tahmin edilen örneklerdir.

Bu bileşenler üzerinden doğruluk oranı bulunmaktadır. Doğruluk oranı Denklem (4)'de ifade edilmiştir.

$$\text{Doğruluk oranı} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (4)$$

III. BULGULAR

Bölüm II' de belirtilen yöntemler kullanıldığında elde edilen bulgular Tablo 1'de gösterilmektedir.

Tablo 1. Üç ülke perspektifinden performans metriklerinin karşılaştırılması

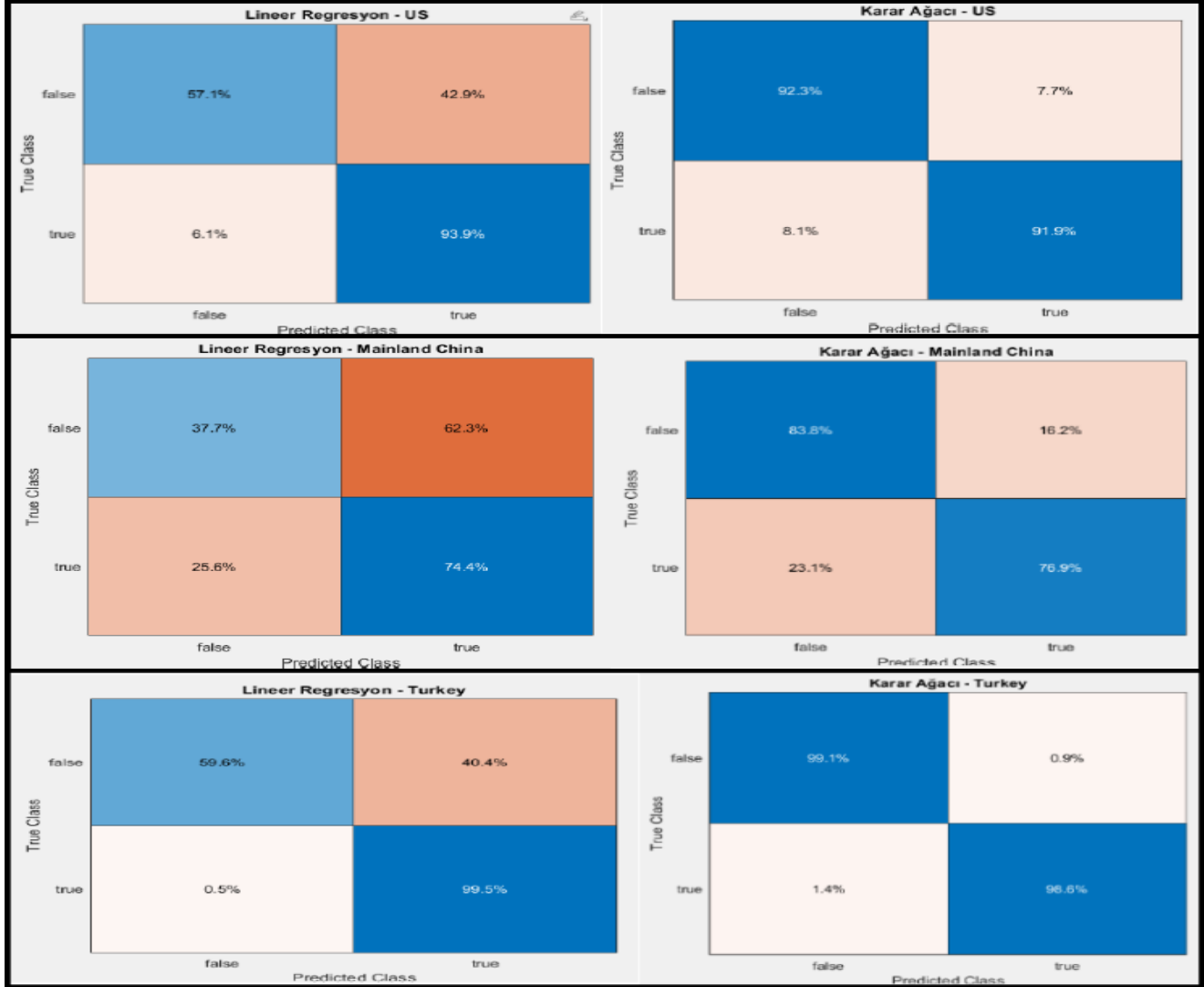
Sınıflandırıcılar	Ülkeler	MAE	RMSE	R ²	Doğruluk Oranı
Doğrusal Regresyon	Türkiye	516170	656850	0.83281	79.55%
	ABD	2466900	2924700	0.94035	75.50%
	Çin	6372.5	12026	0.33204	55.06%
Karar Ağacı	Türkiye	18328	34395	0.99954	98.88%
	ABD	111030	172050	0.99979	92.10%
	Çin	287.16	1150.6	0.99389	80.57%

Elde edilen bulgular göz önüne alındığında Karar Ağacı, tüm ülkelerde Doğrusal Regresyona göre çok daha düşük MAE değerleri üretmiştir. RMSE büyük hatalara daha duyarlıdır. Yine Karar Ağacı, RMSE açısından da çok daha başarılı performans göstermiştir. Determinasyon Katsayısı ise modelin verideki değişkenliği ne kadar açıkladığını gösterir. 1'e ne kadar yakınsa, model o kadar başarılı olarak yorumlanır. Karar Ağacı modelleri neredeyse mükemmel uyum gösterirken Doğrusal Regresyon ise özellikle Çin'de zayıf kalmıştır. Tahminlerin ne kadarının doğru olduğunu gösterir. Yüksek doğruluk, iyi performans demektir. Karar ağacı tüm ülkelerde Doğrusal Regresyondan daha iyi sonuçlar elde etmiştir. Doğrusal Regresyon modeli özellikle Çin'de 55.06% ile düşük performans göstermiştir.

Genel bir bakışla Karar Ağacı modeli hem hata metriklerinde hem de doğrulukta Doğrusal Regresyona göre çok daha başarılı sonuçlar elde etmiştir. Özellikle Türkiye ve ABD için Karar Ağacı oldukça yüksek performanslar elde etmiştir. Sınıflandırıcılar Çin verilerinde düşük performanslar göstermiştir. Çin verisi daha zor tahmin edilebilir olabilir (düşük R² ve doğruluk), bu da veri kalitesi veya salgının dinamikleriyle ilgili olarak açıklanabilir.

Tablo 1'deki sonuçlara ek olarak elde edilen performanslara ek olarak modellerin Karmaşıklık matrisi Şekil 3'de gösterilmiştir. Şekil 3'de satırlar ülkeleri ifade ederken: sırasıyla ABD, Çin ve Türkiye, sütunlar kullanılan modelleri göstermektedir: sırasıyla Doğrusal Regresyon ve Karar Ağacı modelleri.

Karmaşıklık matrisi yüzdelik olarak başarı oranlarını ifade etmektedir. Özellikle Doğrusal Regresyon modellerinde FP oranlarının oldukça yüksek olduğu gözlemlenmiştir. Bu durumda modeller, gerçekte negatif olan değerleri pozitif olarak sınıflandırmakta ve hata oranını arttırmaktadırlar. Ancak Karar Ağacı modelleri incelendiğinde Tablo I' de olduğu gibi daha başarılı sınıflandırma oranları elde edildiği gözlemlenmektedir.



Şekil 3. Tüm modellerin karmaşıklık matrisi. İlk sırada ABD, ikinci sırada Çin ve son sırada Türkiye'ye ait veriler sunulmuştur. İlk sütun Doğrusal Regresyon modeli sonuçlarını gösterirken, ikinci sütun ise Karar Ağacı modelinin sonuçlarını ifade etmektedir.

IV. TARTIŞMA

Elde edilen sonuçlara bakıldığında; üç ülke (Türkiye, Çin, ABD) için de Karar Ağacı modeli, Doğrusal Regresyon modeline göre daha başarılı sonuçlar vermektedir. Karar ağacının en doğru tahmin yapabildiği ülke Türkiye'dir. Çin için doğrusal regresyon modelinin doğruluk değeri 55.06% çıkarken, karar ağacı modelinde 80.57% çıkmaktadır. Bu doğruluk artışı Çin gibi vaka sayısında ani yükseliş olan ve doğrusal bir artış sergilemeyen ülkeler için doğrusal regresyon modelinin yetersiz kaldığını göstermektedir. Karar ağacı doğrusal regresyona göre daha esnek bir yapıya sahiptir bu yüzden de daha iyi sonuçlar vermektedir.

Özen vd. [7] tarafından yapılan çalışmada; 21 Ocak 2020 ve 6 Aralık 2020 tarihleri arasını kapsayan The New York Times, 2020 veri seti kullanılarak ABD'nin vaka ve ölüm sayısını tahmin edebilmek için

makine öğrenmesi yöntemlerinden olan doğrusal regresyon kullanılmıştır. Performans metrikleri de kullanılarak karşılaştırmalar yapılmıştır. Bu karşılaştırmalar sonucunda; MAE: 3467277, RMSE: 3672745 olarak bulunmuştur. Bu çalışmada ise Novel Corona Virus 2019 Dataset verisi kullanılarak ABD'nin vaka ve ölüm sayısını tahmin edebilmek için makine öğrenmesi yöntemlerinden olan doğrusal regresyon kullanılmıştır. Performans metrikleri de kullanılarak karşılaştırmalar yapılmıştır. Bu karşılaştırmalar sonucunda MAE: 2466900, RMSE: 2924700 olarak bulunmuştur.

Özel ve Demirsöz [8] tarafından yapılan çalışmada; Mart 2020 ve Mart 2021 tarihleri arasını kapsayan Türkiye Sağlık Bakanlığından alınan veri seti kullanılarak Türkiye'nin günlük vakalarını, iyileşenlerini ve ölüm sayısını tahmin edebilmek için makine öğrenmesi yöntemlerinden olan Rastgele Orman kullanılmıştır. Kullanılan performans metriklerinden kesinlik %90, duyarlılık %89 ve f ölçüsü ise %89 olarak belirlenmiştir. Bu çalışmada ise kullanılan veri seti farklı olmakla birlikte Türkiye'nin günlük vakalarını, iyileşenlerini ve ölüm sayısını tahmin edebilmek için makine öğrenmesi yöntemlerinden olan Karar Ağacı modeli kullanılmıştır. Kullanılan performans metriklerinden kesinlik %99, duyarlılık %98 ve f ölçüsü ise %98 olarak elde edilmiştir.

V. SONUÇLAR

Bu çalışmada, Covid-19 pandemisinin yayılımını Türkiye, Çin ve ABD ülkeleri için vaka sayısı, ölüm sayısı, iyileşen hasta sayısı ve tarih incelenmiştir. Makine öğrenmesi yöntemlerinden olan doğrusal regresyon ve karar ağacı kullanılarak karşılaştırmalar yapılmıştır. Bu karşılaştırmaların sonucunda Karar Ağacı modelinin, Doğrusal Regresyon modeline göre bu konu bazında daha başarılı sonuç verdiği gözlemlenmiştir. Değerlendirmelerde performans metriklerinden RMSE, MAE, R^2 , karmaşıklık matrisi ve doğruluk değerlerinin modelin başarısını ölçmekte önemli bir rol oynadığı anlaşılmıştır. Çalışmadan elde edilen çıkarım ise makine öğrenmesi yöntemlerinin farklı veri setlerine göre yapıları gereği avantaj ve dezavantaj sağladıkları görülmüş, yalnızca bir modele bağlı kalmayıp çoklu modeller ele alındığı taktirde daha gerçekçi, doğrusal ve sağlıklı sonuçlar verdiği anlaşılmıştır. Gelecekte bu çalışma, nüfus yoğunluğu, test sayısı, hasta verileri gibi öznetelikler artırılarak salgına karşı nasıl bir politika izleneceğini, ne gibi önlemler alınabileceğini daha erken harekete geçebileceğimizi sağlayacaktır.

KAYNAKLAR

- [1] J. M. Henwood, "Coronavirus disinfection in histopathology," *Journal of Histotechnology*, vol. 43, no. 2, pp. 102–104, 2020.
- [2] World Health Organization, "WHO Director-General's opening remarks at the media briefing on COVID-19," 11 March 2020. [Online]. Available: <https://www.who.int>.
- [3] Centers for Disease Control and Prevention, Underlying Conditions and the Higher Risk for Severe COVID-19, CDC, Jun. 2025. [Online]. Available: <https://www.cdc.gov/covid/risk-factors/index.html>.
- [4] S. Ghafouri-Fard, H. Mohammad-Rahimi, P. Motie, M. A. S. Minabi, M. A. S. Taheri, ve S. Nateghinia, "Application of machine learning in the prediction of COVID-19 daily new cases: A scoping review," **Heliyon**, cilt. 11, no. 4, mak. e42943, Feb. 2025.
- [5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [6] L. Rokach and O. Maimon, *Data Mining with Decision Trees: Theory and Applications*, World Scientific, 2014.
- [7] N. S. Özen, "Prediction of COVID-19 Cases in the United States of America with Machine Learning Algorithms," *Avrupa Bilim ve Teknoloji Dergisi*, 2021. [Online]. Available: <https://dergipark.org.tr/en/download/article-file/1492256>.
- [8] Z. Özel and M. Demirsöz, "Makine Öğrenmesi Yöntemleri İle COVID-19 Verilerinin İncelenmesi: Türkiye Örneği," *Sağlık Bilimlerinde Yapay Zeka Dergisi*, vol. 1, no. 2, pp. 1–7, 2021.
- [9] M. A. Çakmak, M. E. Kurt, and C. Çakmak, "Makine Öğrenmesi Algoritmaları ile Covid-19 Hastalarının Mortalite Risklerinin Hesaplanması," *Vizyoner*, vol. 13, no. 35, pp. 994–1011, 22 Aug. 2022. DOI: 10.21076/vizyoner.1074212.
- [10] R. Singh, A. Adhikari, and S. R. Rathi, "Time series forecasting of COVID-19 using deep learning models," *Chaos, Solitons & Fractals*, vol. 140, p. 110121, 2020.
- [11] A. Zeroual, et al., "Deep learning methods for forecasting COVID-19 time-series data: A comparative study," *Chaos, Solitons & Fractals*, vol. 140, p. 110121, 2020.

- [12] M. H. D. Ribeiro, et al., “Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil,” *Chaos, Solitons & Fractals*, vol. 135, p. 109853, 2020.
- [13] S. Shastri, et al., “Time series forecasting of COVID-19 using epidemiological models and machine learning methods: A comparative study,” *Applied Intelligence*, vol. 51, no. 5, pp. 2790–2807, 2021.
- [14] S. Rajkumar, “Novel Corona Virus 2019 Dataset,” Kaggle, 2020. [Online]. Available: <https://www.kaggle.com/datasets/sudalairajkumar/novel-corona-virus-2019-dataset>.
- [15] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer, 2013.
- [16] L. Rokach and O. Maimon, *Data Mining with Decision Trees: Theory and Applications*, World Scientific, 2014.
- [17] G. Naidu, T. Zuva, and E. M. Sibanda, “A review of evaluation metrics in machine learning algorithms,” in *Computer Science On-line Conference*, Cham, Switzerland: Springer International Publishing, 2023, pp. 15–25. doi: https://doi.org/10.1007/978-3-031-35314-7_2
- [18] J. F. Hair, M. Sarstedt, C. M. Ringle, and S. P. Gudergan, *Advanced Issues in Partial Least Squares Structural Equation Modeling*, 2nd ed., Thousand Oaks, CA: SAGE Publications, 2021.
- [19] T. Hayes, “R-squared change in structural equation models with latent variables and missing data,” *Behavior Research Methods*, vol. 53, no. 5, pp. 2127–2157, 2021. doi: 10.3758/s13428-020-01532-y.
- [20] K. M. Ting, “Confusion Matrix,” in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Cham, Switzerland: Springer, 2011, pp. 209–209.