

Covid-19 Salgınına Yönerek Zaman Serisi Verileri ile Prophet Model Kullanarak Makine Öğrenmesi Temelli Vaka Tahminlemesi

Onur Sevli^{1*}, Vesile Gül Başer²

¹ Burdur Mehmet Akif Ersoy Üniversitesi, Mühendislik Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, Burdur, Türkiye (ORCID: 0000-0002-8933-8395)

² Burdur Mehmet Akif Ersoy Üniversitesi, Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Burdur, Türkiye (ORCID: 0000-0002-0752-9498)

(İlk Geliş Tarihi 8 Temmuz 2020 ve Kabul Tarihi 29 Ağustos 2020)

(DOI: 10.31590/ejosat.766623)

ATIF/REFERENCE: Sevli, O. & Başer, V. G. (2020). Covid-19 Salgınına Yönerek Zaman Serisi Verileri ile Prophet Model Kullanarak Makine Öğrenmesi Temelli Vaka Tahminlemesi. *Avrupa Bilim ve Teknoloji Dergisi*, (19), 827-835.

Öz

Koronavirüsler 21. yüzyılda keşfedilen ve dünyada yaygın olarak görülen, insan ve hayvanlarda hastalığa sebep olan bir virus türüdür. 2002 yılında SARS ve 2012 yılında MERS salgınlarından sonra, 2019 yılı sonunda Çin'in Wuhan eyaletinde ortaya çıkıp hızla yayılan Covid-19 bir pandemiye sebep olmuştur. Bu çalışmanın yapıldığı dönemde hala yayılmaya devam eden ve henüz kontrol altına alınamayan Covid-19 ile ilgili tıp alanı başta olmak üzere farklı disiplinlerde yoğun çalışmalar sürdürmektedir. Yapılan çalışmaların neredeyse tamamı virüsün yapısı, kontrol ve önlemeye dair tıbbi çalışmalardır. Bilgisayar bilimlerinde öngörücü disiplinler içerisinde yer alan makine öğrenmesi, mevcut verilerden hareketle geleceğe dönük tutarlı tahminlerde bulunabilmektedir. Bu amaçla geliştirilen farklı modeller içerisinde Facebook veri bilimi ekibi tarafından geliştirilen ve açık kaynak kodla kullanıma sunulan Prophet modeli zaman serisi verileri üzerinde, günlük, haftalık, dönemsel, yıllık tutarlı tahminler ortaya koymaktadır. Bu çalışmada, dünya genelindeki Covid-19 vakalarına ilişkin güncel veri seti üzerinde, Prophet modeli ile geleceğe dönük tahminler yapılmış ve gerçek vakalarla karşılaştırılarak büyük oranda tutarlı sonuç verdikleri ortaya konmuştur. Doğrulanmış Covid-19 vakaları, virüsün sebep olduğu hastalık kaynaklı ölümler ve hastalıktan kurtulan vakalar için gerçekleştirilen üç ayrı tahminleme çalışmasında, doğrulanmış vaka ve ölüm sayılarındaki artışın yapıldığı tarihi takip eden bir haftalık süreçte tahminlerden yüksek olabileceği, kurtulan vaka sayılarının ise tahminler doğrultusunda gerçekleşeceği öngörmüştür.

Anahtar Kelimeler: Covid-19, Makine öğrenmesi, Prophet model, Vaka tahminleme.

Machine Learning Based Case Estimation Using Prophet Model with Time Series Data for Covid-19 Outbreak

Abstract

Coronaviruses were discovered in the 21st century and are a common type of virus that causes disease in humans and animals worldwide. After the SARS outbreak in 2002 and the MERS outbreaks in 2012, Covid-19 appeared in the Wuhan province of China at the end of 2019, causing an outbreak. In the period when this study was carried out, intensive studies related to Covid-19 which is still spreading and not yet under control are continuing in different disciplines, especially in the field of medicine. Almost all of the studies carried out are medical studies on the epidemiology of the virus, control and prevention activities. Machine learning, which is one of the predictive disciplines in computer science, can make consistent predictions based on the available data. The Prophet model, developed by the Facebook data science team and offered as an open source project, can provide consistent daily, weekly, periodic and annual estimates on the time series data. In this study, future estimations made with the Prophet model on the current dataset of Covid-19 cases worldwide, and it was revealed that the predictions provide a highly consistent result by comparing with real cases. In three separate estimation studies for confirmed Covid-19 cases, disease-induced deaths and survivors, it was predicted that the increase in the number of confirmed cases and deaths will be higher than the estimates in the one-week period following the date of the study, and the number of recovered cases will be based on the estimates.

Keywords: Covid-19, Machine learning, Prophet model, Case estimation.

* Sorumlu Yazar: Burdur Mehmet Akif Ersoy Üniversitesi, Mühendislik Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, Burdur, Türkiye, ORCID: 0000-0002-8933-8395, onursevli@mehmetakif.edu.tr

1. Giriş

Koronavirüsler 21. yüzyılda keşfedilen, insanlarda solunum yolu hastalıklarına neden olan ve tüm dünyada yaygın görülen virüs türüdür [1]. Çeşitli koronavirüs türleri yakın dönemlerde salgınlarla neden olmuştur. 2002 yılında Çin'in Guandong eyaletinde ortaya çıktıktan sonra hızlı yayılan SARS [2] ve 2012 yılında Arap Yarımadası'nda ortaya çıktıktan sonra ölümçül salgınlarla sebep olan MERS [3] birer koronavirüs türüdür. SARS ve MERS salgınlarının ardından 2019 koronavirüs salgını 21. yüzyılın üçüncü büyük salgını olarak nitelendirilmektedir [4]. Çin'in Wuhan kentinde ortaya çıkarak kısa zamanda hızla yayılan ve Covid-19 olarak adlandırılan virüs ilk kez 24.01.2020 tarihinde Zu vd. tarafından bir çalışma raporu olarak sunulmuştur [5]. Dünya genelinde hızla yayılmaya devam eden ve bu nedenle Dünya Sağlık Örgütü (WHO) tarafından pandemik bir hastalık olarak ilan edilen Covid-19 daha önce tanımlanan koronavirüslerden farklılık göstermektedir.

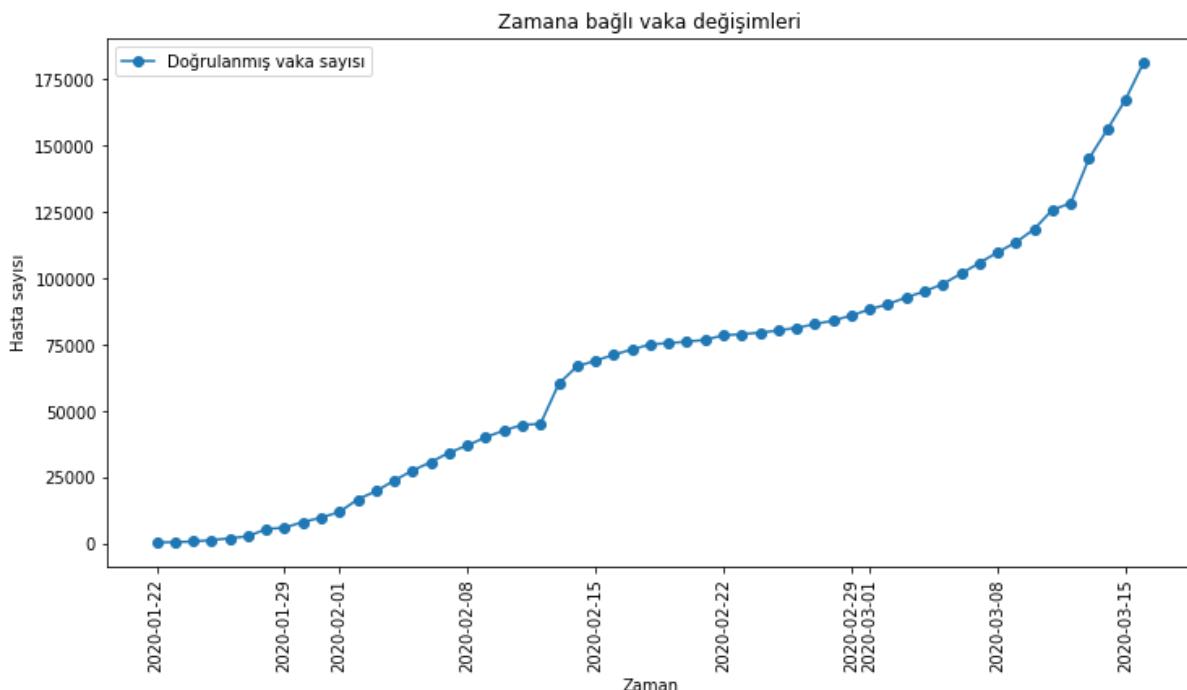
Koronavirüsler geniş bir aile olup bazıları insanlarda bazıları ise deve, kedi, yarasa gibi hayvanlarda hastalığa sebep olmaktadır. Hayvansal kaynaklı olduğu düşünülen Covid-19'un, Çin'in Huanan deniz ürünleri pazarından bir insana geçmiş olduğu tahmin edilmekte ve günümüzde insandan insana hızlı bir şekilde yayılmaktadır. Covid-19 virüsü insan vücudunda; ateş, öksürük, boğaz ağrısı ve nefes darlığı gibi belirtiler ile kendini göstermeye teşhisini bu yeni virüse özgü geliştirilmiş testlerle yapılmaktadır.

John Hopkins Üniversitesi'nin (JHU) dünya genelinden anlık derlediği güncel verilere göre, çalışmanın gerçekleştirildiği 2020 Mart ayının ilk yarısında dünya genelindeki doğrulanmış koronavirüs vaka sayısı yaklaşık 240 bin, virüsten kaynaklı ölüm sayısı 10 bin, virüsten kurtulanların sayısı ise 85 bin civarındadır. Virüsün yaygın olarak görüldüğü ilk beş ülke yaklaşık 81 bin vaka ile Çin, 41 bin vaka ile İtalya, 18 bin vaka ile İran, 18 bin vaka ile İspanya ve 15 bin vaka ile Almanya'dır. Bu çalışmanın gerçekleştirilmesi esnasında dünya genelindeki vaka sayıları artmaya devam etmektedir ve araştırmancının yaptığı gün itibarıyle Türkiye'deki doğrulanmış vaka sayısı ise 400'e yaklaşmaktadır [6]. Araştırmancının yaptığı gün dünya genelinde Covid-19'un yayıldığı bölgeler ve yoğunlukları gösteren harita Şekil 1'de verilmiştir.



Sekil 1. 2020 Mart ayının ilk yarısında Covid-19 vaka yayılımını gösterir harita [7]

Covid-19 dünya genelinde hızlı bir şekilde yayılmakta ve üstel bir artış göstermektedir (Şekil 2). Tüm bilim insanların ortak amacı büyümeye eğrisinin düzleşmesini sağlamaktır. Vaka sayılarının zaman içerisinde artması olasıdır ancak eğrinin düzleşmesi birim vaka sayısının daha geniş bir zaman düzlemine yayıldığını gösterir ve vakanın kontrol altına alındığına dair bir göstergedir.



Şekil 2. Zamana bağlı vaka değişimleri grafiği

Henüz tam anlamıyla kontrol altına alınamayan Covid-19 salgını ile ilgili yapılan yayın çalışmalarının pek çoğu hastalığın epidemiyolojisi [8] [9] [10], klinik [11] [12] [13] ve karakteristik özellikleri [14] [15], yayılımı [16], kontrolü ve önlenmesi [17] gibi tıbbi bilimlerle ilgilidir. Bunun yanında henüz yayımlanmamış, taslak şeklinde, tıbbi çalışmalara da rastlanmaktadır.

Covid-19 ile ilgili tıbbi çalışmalara ek olarak bilgisayar teknolojisi yardımıyla hastalığın yayılımı, seyrine ve geleceğine ilişkin tahminleme çalışmalarının yapılabileme potansiyeli vardır. Bu anlamda yapay zekânnın bir alt dalı olan makine öğrenmesi, mevcut veriler üzerinden öğrenen algoritmalar ile geleceğe dönük tahminler yapabilme yeteneği ile ön plana çıkmaktadır. Bu çalışmanın gerçekleştirildiği zaman aralığında Covid-19 konusunda makine öğrenmesi ile ilgili mevcut bir çalışmaya rastlanmamıştır. Makine öğrenmesinin Covid-19 salgının seyri hakkında faydalı öngörüler sağlayabileceği düşüncesinden hareketle JHU tarafından paylaşılan ve anlık olarak güncellenen, tüm dünyadaki Covid-19 vakalarını gösteren veri seti üzerinden bir makine öğrenmesi modeli olan ve Facebook tarafından geliştirilen Prophet ile dünya genelindeki doğrulanmış vaka, ölüm ve hastalıktan kurtulan vakaları tahminleyen bir çalışma gerçekleştirılmıştır. Verinin elde edildiği tarih itibarıyle takip eden bir haftalık dönem için de tahminler üretilmiş ve gerçek değerlerin ilerlemeleri ile karşılaştırılarak yorumlanmıştır.

2. Materyal ve Metot

JHU tarafından derlenen ve çalışmanın yapıldığı gün itibarıyle, 22 Ocak 2020 ile 18 Mart 2020 tarih aralığındaki verileri kapsayan veri seti üzerinde, Facebook tarafından geliştirilen Prophet model kullanılarak vaka tahminlemeleri gerçekleştirılmıştır.

Yapılan çalışmada kullanılan model Python ile kodlanmıştır. Çalışma 32 GB RAM, Intel Core i7-9750H işlemci, NVIDIA GeForce RTX 2070 grafik kartına sahip donanım üzerinde gerçekleştirilmiştir.

Eldeki veri seti %80 eğitim ve %20 test seti şeklinde ayrılmıştır. Eğitim verileri kullanılarak doğrulanmış hasta vakalar, ölümlü vakalar ve hastalıktan kurtulan vaka sayılarına yönelik olarak tahminlemeler yapılmıştır. Sistemin tahmin ettiği değerler gerçek vaka değerleri ile karşılaştırılmış ve ortalama hata değerleri üzerinden sistem başarısı yorumlanmıştır. Ayrıca veri setinde mevcut olmayan sonraki bir hafta için de tahminler üretilmiştir. Veri setindeki gerçek vakaların seyrine göre sistemin ürettiği tahminler, grafikteki yönelimler açısından yorumlanmıştır. Herhangi ek bir regresör ya da ek tahmin parametresi kullanılmadan gerçekleştirilen çalışmanın grafikleri ve tahmin hatalarına ilişkin değerler verilerek elde edilen sonuçların doğruluğu tartışılmıştır.

2.1. Veri Seti

Bu çalışmada John Hopkins Üniversitesi tarafından derlenen ve dünya genelindeki güncel Covid-19 vakalarını içeren, sürekli güncellenen ve kamunun erişimine açık bir şekilde paylaşılan "Coronavirus COVID-19 Global Cases" veri seti kullanılmıştır [7]. Veri setine dünya genelindeki günlük vakalar eklenmekte ve veri miktarı sürekli artmaktadır. Halen süregelen pandemi dönemi içerisinde gerçekleştirilen bu çalışma belirli bir tarih aralığı ile sınırlanmıştır. Veri seti içerisinde yer alan ilk vaka 22 Ocak 2020 tarihine aittir ve analiz çalışmasının gerçekleştirildiği 18 Mart 2020 tarihine kadar olan güncel veriler kullanılmıştır. Veri setinde bulunan ve çalışmada kullanılan özellikler şunlardır:

- Date: Vakanın görüldüğü tarih
- Country: Vakanın görüldüğü ülke

- Confirmed: Doğrulanmış hasta vaka sayısı
- Deaths: Hastalık kaynaklı ölüm sayısı
- Recovered: Hastalıktan kurtulanların sayısı

Çalışmanın yapıldığı tarih itibarıyle veri seti içerisinde 6722 adet kayıt yer almaktır ve hastalığın etki etiği 179 ülkeye ilişkin veriler bulunmaktadır. Çalışmanın yapıldığı tarih itibarıyle en çok vaka sayısı Çin'de bulunmakta olup zaman içerisinde değişkenlik göstermesi muhtemeldir. Veriler bir zaman serisi formunda olup 1'den başlayıp binlere kadar uzanan değerler içermektedir. Veri seti tarih ve ilgili tarihe kadar olan toplam vaka sayısı üzere temel iki nitelik üzerinden analiz edilmiştir.

2.2. Makine Öğrenmesi ve Prophet Modeli

Makine öğrenmesi yapay zekâının bir alt dalı olup makinelerin daha önceki çözümlerden yola çıkarak yeni çözümler üretmesini sağlayan bir tekniktir. Üretilen dijital veri miktarının her geçen gün daha da artmasına bağlı olarak ortaya çıkan büyük veri yığınları üzerinden matematik ve istatistik yöntemlere dayalı çeşitli algoritmalar ile tahminler yapmayı sağlayan yöntemdir. Makine öğrenmesi, denetimli ve denetimsiz öğrenme olarak iki alt grupta incelenebilir. Denetimli öğrenme veri seti içerisinde girdi özelliklerine karşılık hedef özellik değerinin de yer aldığı durumları ifade eder. Denetimsiz öğrenmede ise veri setinde girdi özelliklerine karşılık ulaşılacak hedef özellik direkt olarak yer almamaktadır.

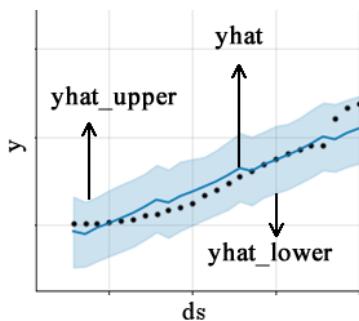
Makine öğrenmesi ile analiz edilecek veriler zamandan bağımsız değerler olabileceği gibi zamana bağlı değişim gösteren değerler de olabilir. Veri seti içerisinde işlenecek bir verinin zamana bağlı değişim göstermesi durumunda zaman serisi analizi kullanılır [18]. Zaman serisi analizinde nicel yöntemler kullanılır. Zaman serisi analizi zaman içerisinde düzenli aralıklarla ölçümlenen değerlerin istatistik olarak analiz edilip geleceğe dönük öngörüler üretilmesini hedefler. Büyük veri kavramı ile birlikte ortaya çıkan çok sayıda tarihsel verinin etkin bir şekilde analiz edilerek geleceğe dönük isabetli tahminler üretilmesi konusunda makine öğrenmesi teknikleri klasik istatistik tekniklere göre daha yüksek başarı sergilemektedir [19]. Makine öğrenmesi ile zaman serisi analizlerinin yapılabilmesi için eldeki verilerin denetimli öğrenme modeline uygun hale getirilmesi gereklidir.

Makine öğrenmesi ile zaman serisi analizi konusunda uygulanagelen farklı teknikler olmakla birlikte Facebook tarafından geliştirilip açık kaynak kodla kullanıma sunulan Prophet modeli başarılı tahminlemeler yapabilmeyi sağlamaktadır.

Prophet modeli Facebook veri bilimi ekibi tarafından geliştirilen açık kaynak kodlu bir tahminleme uygulamasıdır [20]. Doğrusal olmayan zaman serisi verileri üzerinden yıllık, dönemlik, haftalık, günlük tahminlerde bulunmayı sağlayan prosedürler içerir. Prophet modeli veri kayıtları ve uç değerleri başarılı bir şekilde ele alma kabiliyetine sahiptir. Model, günümüzde Facebook bünyesindeki pek çok uygulamada aktif olarak kullanılmaktadır. Prophet modeli Python ve R dilleri ile birlikte kullanılabilen fonksiyonlar barındırır.

Bu çalışmada Prophet modelin Python dili modülleri kullanılmıştır. Prophet, makine öğrenmesi alanında yaygın olarak kullanılan sklearn modelinin prosedürlerini takip eder. Prophet iki özellikli bir veri kümesini girdi olarak alır. Bu özelliklerden ilki olan ds zaman damgasıdır ve Python dilindeki veri işleme kütüphanesi olan pandas tarafından ele alınabilecek zaman formatlarını destekler. Diğer özellik olan y ise tahmine konu olan nümerik ölçüm değeridir. İçerisinde ds ve y değerleri içeren veri seti ile eğitilen model farklı periyotlarda tahminler üretebilir. Prophet modelinin işleyişindeki sezgisel yaklaşım verinin detaylarına bogulmadan etkili tahminler oluşturabilmeyi sağlar [21].

Model üzerinden yapılan tarih bazlı gelecek tahminleri bir değer aralığı olarak üretilir. Muhtemel tahmin değeri yhat, tahmin üst sınırı yhat_upper ve tahmin alt sınırı yhat_lower değerleri ile bir değer bandı olarak üretilir (Şekil 3).



Şekil 3. Prophet model tahmin bandı

Verilerin regrese edilmesi ile gerçek verilerin dağılımına bağlı olarak tahmin bandının genişliği farklılık gösterir. Düzenli ilerleme gösteren veriler için tahmin bandı dar ve üst alt tahmin değerleri birbirine yakınlık, veri seyrinin düzensiz olması halinde Prophet modeli daha geniş bir tahmin bandı üretir.

3. Deneyel Uygulamalar ve Bulgular

Bu çalışmada Prophet model kullanılarak dünya genelindeki onaylanmış vakalar, ölümler ve kurtulan vakalara yönelik tahminlemeler yapılmıştır. Yapılan tahminler süreçteki gerçek sayılarla karşılaştırılarak tahmin başarısı yorumlanmıştır.

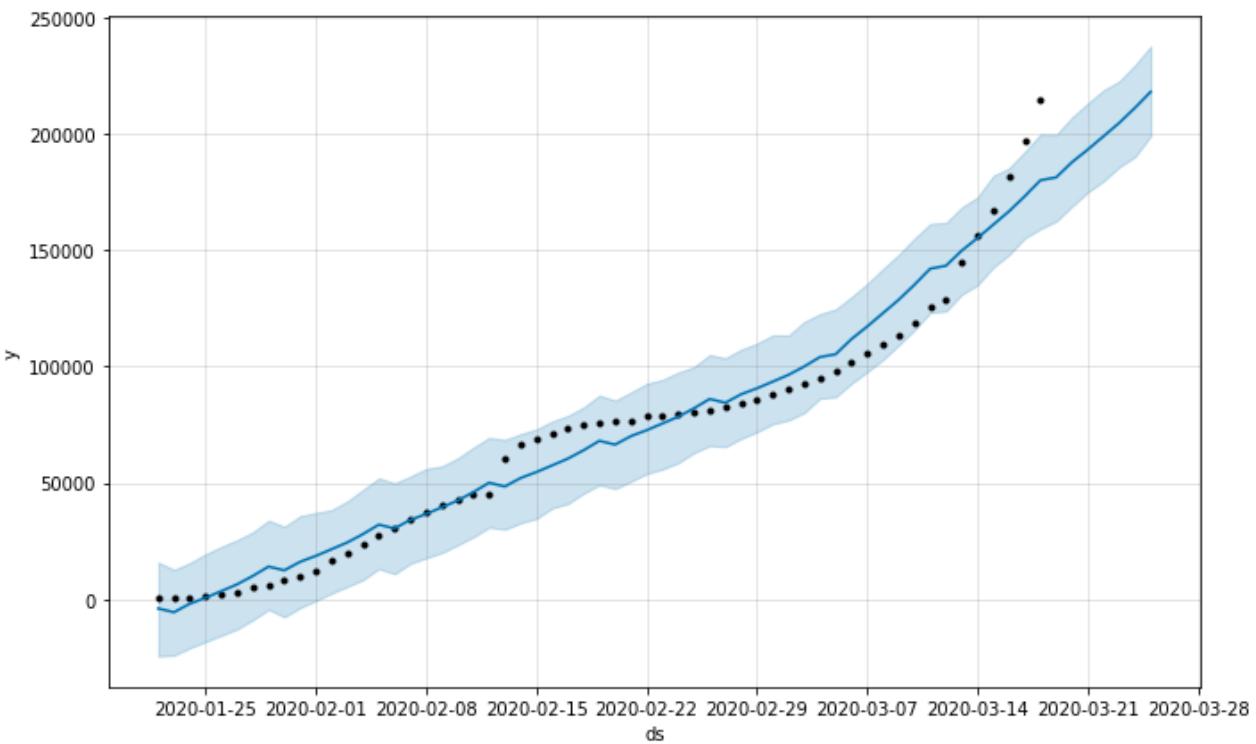
Modelin performansı değerlendirilirken gerçek değerler ile modelin ürettiği tahminlerin uyumuna bakılmıştır. Tahmin edilen değerlerin gerçek değerlerden uzaklıklarını temel alınarak farklı metrikler üretilebilir. Bu metriklerden biri olan ortalama mutlak yüzde hata (mean absolute percent error –MAPE), zaman serisi modellerinde tahminlerin doğruluğunu ölçmek için sıkça kullanılmaktadır. Bu nedenle bu çalışmada modelin başarısını yorumlamak için MAPE metriği kullanılmıştır. MAPE metriği Denklem 1'de yer alan formül ile hesaplanır.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{g_i - t_i}{g_i} \right| \quad (1)$$

Verilen formülde g gerçek değerleri, t modelin tahmin ettiği değerleri temsil eder.

3.1. Dünya Geneli Doğrulanmış Vaka Tahminlemesi

Veri seti üzerinden Prophet model ile gelecek bir haftayı da içine alacak şekilde yapılan dünya genelindeki doğrulanmış vaka sayılarına ilişkin tahmin ve gerçek değerlere ait grafikler Şekil 4'te görülmektedir.



Şekil 4. Dünya geneli doğrulanmış vakalar ve tahmin grafiği

Grafikte modelin ürettiği tahmin aralığı bir bant şeklinde görülmektedir. Kesikli noktalar ile gösterilen eğri gerçek vaka değerlerine ilişkin graftır.

Dünya genelindeki vaka sayıları belirli dönemlerde farklı yönelikler göstermiştir. Tahmin edilen değer bandı genel anlamda gerçek değer grafiği ile uyum göstermesine karşın belirli aralıklarda sapmalar olduğu görülmektedir. Grafikte son dönemde yakın gerçek vaka grafiği seyrinin değiştiği ve daha hızlı bir artışa geçtiği, bunun sonucu olarak tahmin bandının önceki dönem seyrine göre ilerlediği çizginin doğru dışına çıktıgı görülmektedir. Gerçek değerlere ait grafiğin son bulduğu noktadan itibaren, veri seti içerisindeki tarih arlığını takip eden 7 günlük sonrasındaki tahminlerin, gerçek değerlerin çıkışına göre daha alt değerlerde ilerlediği görülmektedir. Diğer bir değişle takip eden dönemde gerçekleşmesi olası vakaların, önceki dönem ilerlemesi göz önüne alındığında oluşturulan tahminlere göre daha yüksek olacağı yorumu yapılabilir.

Dünya geneli vaka sayılarının tahmininde modelin başarısını değerlendirmek, gerçek ve tahmin değerleri grafiği 15 eş dönemde ayrılmış, her bir kesitteki MAPE değerleri hesaplanmıştır. Kesit sayısı artırılarak daha küçük ölçekli incelemeler yapılabilir ancak bu durumda genel görünüm ve değerlendirmeden uzaklaşılması söz konusudur. Elde edilen MAPE değerleri Tablo 1'de verilmiştir.

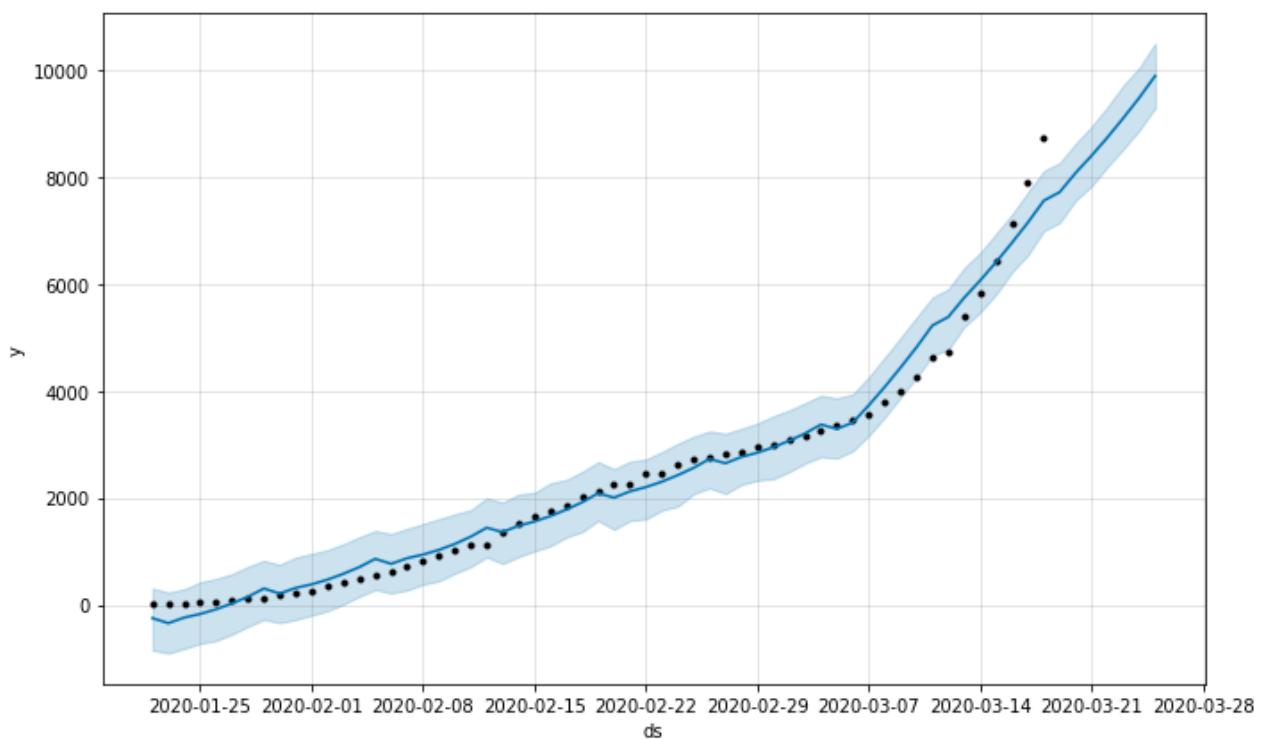
Tablo 1 incelendiğinde MAPE değerlerinin son dönemlerde yükselişе geçtiğе görülmektedir. Bu durum, ilgili dönemde gerçek değerlerin tahmin edilen değerlerden farklı bir seyir izlemesinden kaynaklanmaktadır. Belirli dönemlerdeki sabit eğimli ilerlemeler yanında bazı dönemlerde normal seyrin aksine, vaka düşüş ya da yükselişleri meydana gelmiştir. Bu durumlarda tahminin eğrisi gerçek değerlerin daha önceki seyrine bağlı ilerlerken, gerçek değerler ile tahminler arasındaki farklılık artış göstermiştir. Şekil 3'teki grafiğin son kısımları incelendiğinde, gerçek değerler grafiğinin seyrinin değişmesi, tahmin edilen değerlerden uzaklaşmaya neden olmuş ve bunun sonucu Tablo 1'deki son dönem hata değerleri artış göstermiştir. Son dönemlerdeki hata değerinin artışı, takip eden dönemde ortaya çıkacak vaka sayılarının tahminlerden fazla olacağı şeklinde yorumlanabilir.

Tablo 1. Dünya geneli doğrulanmış vaka sayısı tahminlerine ait MAPE değerleri

Dönem	MAPE	Dönem	MAPE	Dönem	MAPE
1	0.032779	6	0.054031	11	0.238940
2	0.036100	7	0.077866	12	0.280545
3	0.013790	8	0.105800	13	0.329363
4	0.012677	9	0.098306	14	0.372845
5	0.033422	10	0.190662	15	0.409744

3.2. Dünya Geneli Ölümle Sonuçlanan Vaka Tahminlemesi

Veri seti üzerinden Prophet model ile gelecek bir haftayı da içine alacak şekilde yapılan dünya genelindeki ölümle sonuçlanan vakalara ilişkin tahmin grafiği Şekil 5'te yer almaktadır.



Şekil 5. Dünya geneli ölümle sonuçlanan vaka ve tahmin grafiği

Grafiklerde görüldüğü gibi yapılan tahminler gerçek değerler ile yüksek uyumluluk göstermektedir. Ancak doğrulanmış vaka sayısı tahminlerinde olduğu gibi burada da son dönemde sapmalar olduğu görülmektedir. Son dönemlerde ölümle sonuçlanan vaka sayıları önceki dönemlerdeki vaka ilerleme seyrine göre yapılan tahminlerden daha hızlı bir artış göstermiştir. Grafiğe bakarak, veri setindeki mevcut tarih aralığının sonrasındaki dönem için yapılan tahminlere göre gerçek vakaların daha yüksek seyreceği yorumu yapılabilir. Tahminler üzerinden 15 ayrı dönem için hesaplanan MAPE değerleri Tablo 2'de yer almaktadır.

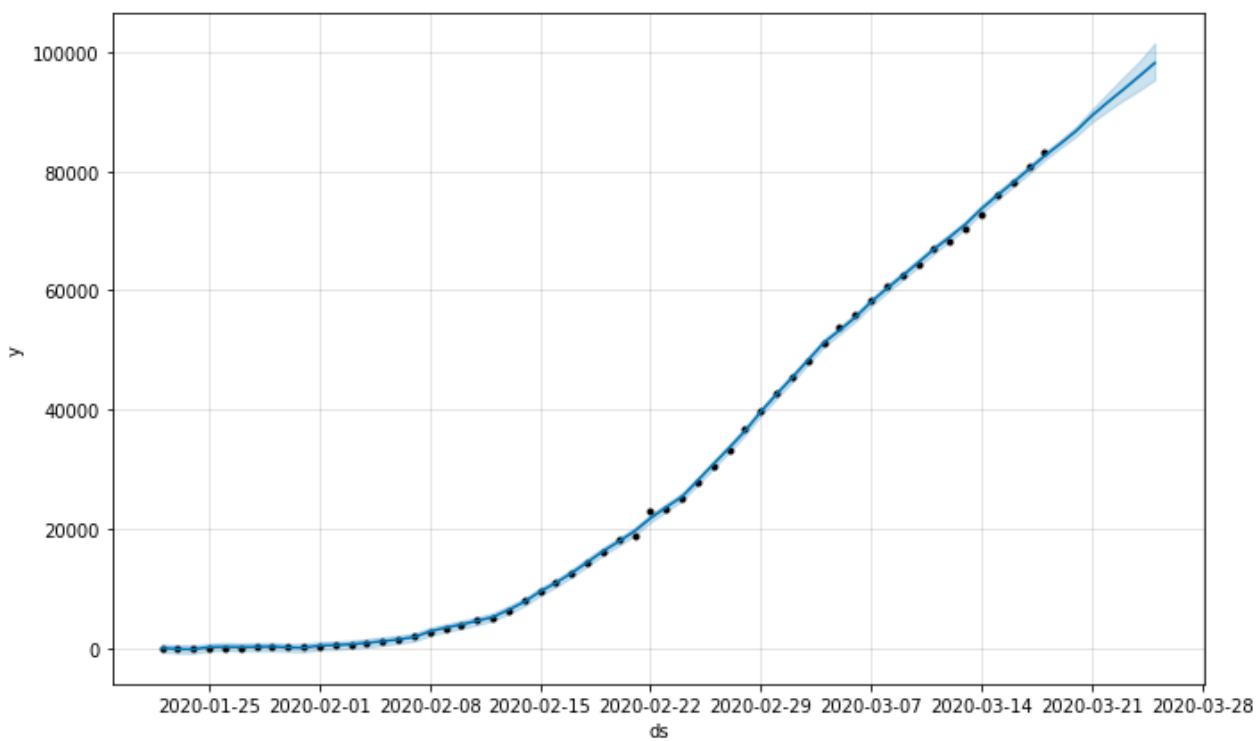
Tablo 2. Dünya geneli ölümlü vaka sayısı tahminlerine ait MAPE değerleri

Dönem	MAPE	Dönem	MAPE	Dönem	MAPE
1	0.008512	6	0.098874	11	0.318934
2	0.010443	7	0.137431	12	0.376171
3	0.026255	8	0.192274	13	0.424326
4	0.027047	9	0.191867	14	0.471522
5	0.075439	10	0.283769	15	0.515744

Tablo incelendiğinde ilk dönemdeki hata değerlerinin çok küçük olduğu görülmektedir. Bu durum gerçek vakalar ile tutarlı tahminlerin yapıldığını gösterir. Son dönemlere gelindiğinde MAPE değerlerinin arttığı görülmektedir. Son dönemde ölümlü vaka sayılarındaki ani yükseliş tahmin edilen değerlerden uzak sonuçlar ortaya çıkarmış ve hata oranı yükselmiştir. Hataladaki artış ve grafikteki ilerlemelere bakıldığında sonraki dönem için yapılan tahminlere göre gerçek değerlerin daha yüksek olacağı öngörlülebilir.

3.3. Dünya Geneli Hastalıktan Kurtulan Vaka Tahminlemesi

Veri seti üzerinden Prophet model ile gelecek bir haftayı da içine alacak şekilde yapılan, dünya genelindeki hastalıktan kurtulan vakalara ilişkin tahmin grafiği Şekil 6'da gösterilmektedir.



Şekil 6. Dünya geneli hastalıktan kurtulan vaka ve tahmin grafiği

Hastalıktan kurtulanların sayısındaki düzenli ilerleme model ile elde edilen tahmin bandının daha dar olmasını, daha net tahminler yapılabilmesini sağlamıştır. Grafikten görüleceği üzere gerçek değerler ile tahmin edilen değerler arasında son derece yüksek bir uyumluluk söz konusudur. Belirli günlerde küçük sapmalar yaşansa da genel olarak gerçek değerler tahmin aralığında seyretmiştir. Veri setindeki tarih aralığının sonrasında 7 günlük tahminlerin de grafikteki ilerlemeye bağlı olarak gerçek değerler ile örtüşeceğinin öngörlülmektedir.

Yapılan tahminler üzerinden 15 ayrı dönem için hesaplanan MAPE değerleri Tablo 3'te yer almaktadır. Tablodaki değerler incelendiğinde hata değerlerinin son derece düşük olduğu görülmektedir. Belirli dönemlerde küçük yükselişler söz konusu olsa da genel anlamda tahmin edilen ve gerçek değerler birbiri ile uyumludur. Tablodaki değerler göz önüne alındığında sonraki dönemler içerisinde, hastalıktan kurtulan vaka sayılarının beklenen şekilde seyredeceği sonucuna varılabilir.

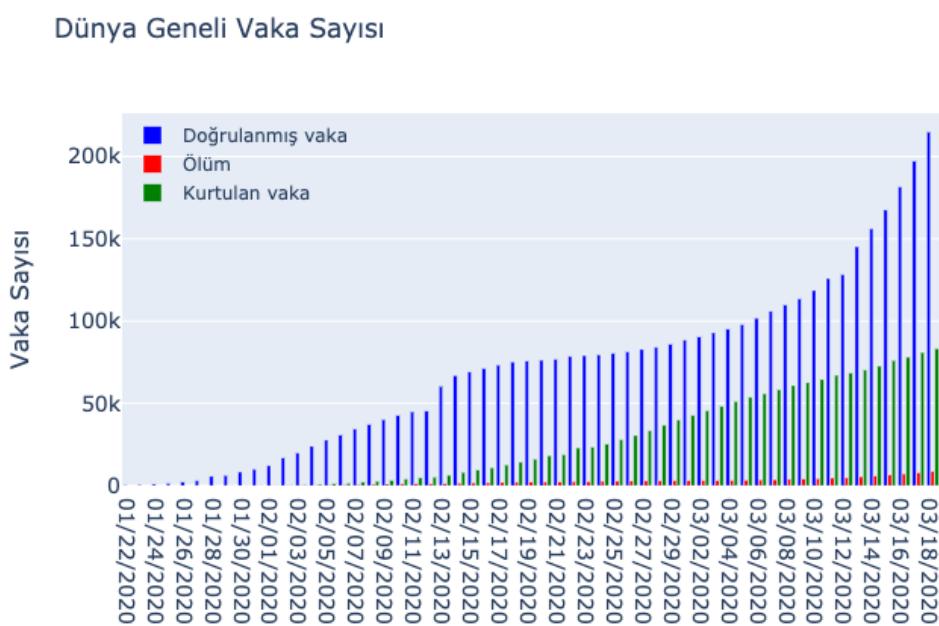
Tablo 3. Dünya geneli kurtulan vaka sayısı tahminlerine ait MAPE değerleri

Dönem	MAPE	Dönem	MAPE	Dönem	MAPE
1	0.014432	6	0.031592	11	0.084837
2	0.008573	7	0.043573	12	0.069185
3	0.005040	8	0.042231	13	0.074262
4	0.017583	9	0.064573	14	0.071380
5	0.019761	10	0.075396	15	0.071053

4. Sonuç

Koronavirüsler yaygın olarak görülen ve belirli dönemlerde salgınlara neden olan bir virüs türüdür. Birer koronavirüs türü olan SARS ve MERS salgınlarından sonra 2019 yılı sonunda Çin'in Wuhan kentinde ortaya çıkarak hızla yayılan Covid-19 bir pandemiye sebep olmuştur. Çalışmanın gerçekleştirildiği sırada yayılımı devam eden bir virüs ile ilgili bilim insanları farklı alanlarda yoğun çalışmalar sürdürmektedir. Bilgisayar bilimleri alanında, yapay zekâının bir alt dalı olan makine öğrenmesi teknikleri ile virüsün dünya üzerindeki yayılımı, önlenmesi ve tedavisi konusunda çalışmalar yapılabılır. Bu doğrultuda gerçekleştirilen çalışmada dünya genelindeki güncel Covid-19 salgın verileri kullanılarak, bir makine öğrenmesi modeli olan Prophet yardımıyla vaka tahminlemeleri yapılmıştır.

Kullanılan veri seti içerisindeki verilerin dağılımları Şekil 7'de yer almaktadır. Doğrulanmış vaka, ölüm ve kurtulan vakaların sayısı her geçen gün artış göstermektedir. Facebook tarafından geliştirilen Prophet model kullanılarak yapılan tahminler, veri seti içerisindeki 22 Ocak – 18 Mart 2020 tarih aralığındaki vakalar yanında takip eden bir haftayı da kapsayacak şekilde kurgulanmıştır.



Şekil 7. Veri seti içerisinde yer alan dünya geneli vaka sayıları

Model üzerinden elde edilen tahminler her üç vaka türü için de gerçek değerler ile büyük oranda uyumluluk göstermiştir. Tahminler ve gerçek değerler arasındaki uyumu ölçmek için MAPE hata metriği kullanılmıştır. Şekil 6'daki grafikte doğrulanmış vaka sayılarında sürekli bir artış olmakla birlikte artışın düzenli olmadığı görülmektedir. Özellikle son dönemdeki hızlı yükselişler tahminler ile gerçek değerler arasındaki farkın açılmasına sebep olmuştur. Benzer şekilde ölüm vakalarındaki düzensiz artış da özellikle son dönemdeki tahminlerle uyumsuzluğu artırmıştır. Elde edilen tahmin bulguları ve gerçek değerlerin seyri göz önüne alınarak gelecek dönem seyrinin tahminlerden yüksek olacağı yorumu yapılabilir. Hastalıktan kurtulan vaka sayısının daha düzenli bir ilerleme gösterdiği Şekil 6'da görülmektedir. Bu veriler ışığında model ile yapılan kurtulan vaka tahminleri de gerçek değerler ile çok yüksek uyum göstermiştir. Bu nedenle gelecek döneme ilişkin tahmin değerlerinin de gerçek değerler ile yüksek uyum göstereceği öngörülmektedir.

Bu çalışma ve benzeri makine öğrenmesi temelli çalışmalar ile vakaların seyri ve yayılımı ile ilgili farklı çalışmalar yapılabilir. Bu doğrultuda erken önlemler ve protokollerin geliştirilmesi sağlanabilir.

Kaynakça

- [1] K. McIntosh and S. Perlman, "Coronaviruses, including severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS)," *Mand. Douglas Bennetts Princ. Pract. Infect. Dis. Updat. Ed. 8th Ed Phila. PA Elsevier Saunders*, 2015.
- [2] P. K. Chan and M. C. Chan, "Tracing the SARS-coronavirus," *J. Thorac. Dis.*, vol. 5, no. Suppl 2, p. S118, 2013.
- [3] R. J. de Groot et al., "Commentary: Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group," *J. Virol.*, vol. 87, no. 14, pp. 7790–7792, 2013.
- [4] E. R. Ahmet Görkem and S. ÜNAL, "2019 Koronavirüs Salgını—Anlık Durum ve İlk İzlenimler," *FLORA*, vol. 25, p. 8, 2020.
- [5] N. Zhu et al., "A novel coronavirus from patients with pneumonia in China, 2019," *N. Engl. J. Med.*, 2020.
- [6] Johns Hopkins University (JHU), "Coronavirus COVID-19 (2019-nCoV) Global Cases." <https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html> (accessed Mar. 18, 2020).
- [7] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infect. Dis.*, p. S1473309920301201, Feb. 2020, doi: 10.1016/S1473-3099(20)30120-1.
- [8] H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," *J. Autoimmun.*, p. 102433, 2020.
- [9] V. Surveillances, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020," *China CDC Wkly.*, vol. 2, no. 8, pp. 113–122, 2020.
- [10]R. M. Anderson, H. Heesterbeek, D. Klinkenberg, and T. D. Hollingsworth, "How will country-based mitigation measures influence the course of the COVID-19 epidemic?," *The Lancet*, vol. 395, no. 10228, pp. 931–934, 2020.
- [11]A. J. Rodriguez-Morales et al., "Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis," *Travel Med. Infect. Dis.*, p. 101623, 2020.
- [12]T. Xu et al., "Clinical features and dynamics of viral load in imported and non-imported patients with COVID-19," *Int. J. Infect. Dis.*, 2020.
- [13]W. Yang et al., "Clinical characteristics and imaging manifestations of the 2019 novel coronavirus disease (COVID-19): A multi-center study in Wenzhou city, Zhejiang, China," *J. Infect.*, 2020.
- [14]M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, and R. Siddique, "COVID-19 infection: origin, transmission, and characteristics of human coronaviruses," *J. Adv. Res.*, 2020.
- [15]S. Tian et al., "Characteristics of COVID-19 infection in Beijing," *J. Infect.*, 2020.
- [16]P. Li et al., "Transmission of COVID-19 in the terminal stage of incubation period: a familial cluster," *Int. J. Infect. Dis.*, 2020.
- [17]J. Hellewell et al., "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts," *Lancet Glob. Health*, 2020.
- [18]S. E. Seker, C. Mert, K. Al-Naami, N. Ozalp, and U. Ayan, "Time series analysis on stock market for text mining correlation of economy news," *Int. J. Soc. Sci. Humanity Stud.*, vol. 6, no. 1, pp. 69–91, 2013.
- [19]G. Bontempi, S. B. Taieb, and Y.-A. Le Borgne, "Machine learning strategies for time series forecasting," in European business intelligence summer school, 2012, pp. 62–77.
- [20]S. J. Taylor and B. Letham, "Prophet: forecasting at scale," *Facebook Res.*, 2017.
- [21]S. J. Taylor and B. Letham, "Forecasting at scale," *Am. Stat.*, vol. 72, no. 1, pp. 37–45, 2018.