

KÜRESEL COVID-19 SALGINININ DÜNYADA VE TÜRKİYE'DE DEĞİŞEN DURUMU VE KÜMELEME ANALİZİ

CHANGING STATUS OF GLOBAL COVID-19 OUTBREAK IN THE WORLD AND IN TURKEY AND CLUSTERING ANALYSIS

Elif KARTAL¹ , M. Erdal BALABAN² , Bülent BAYRAKTAR³ 

¹İstanbul Üniversitesi, Enformatik Bölümü, İstanbul, Türkiye

²İstanbul Gelişim Üniversitesi, İktisadi İdari ve Sosyal Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, İstanbul, Türkiye

³İstanbul Üniversitesi, İstanbul Tıp Fakültesi, Spor Hekimliği Anabilim Dalı, İstanbul, Türkiye

ORCID IDs of the authors: E.K. 0000-0003-4667-1806; M.E.B. 0000-0002-2210-0745; B.B. 0000-0001-8102-4896

Cite this article as: Kartal E, Balaban ME, Bayraktar B. Changing status of global COVID-19 outbreak in the world and in Turkey and clustering analysis. J Ist Faculty Med. Published online July 23, 2020. doi: 10.26650/IUITFD.2020.0077

ÖZET

Amaç: Bu çalışmanın amacı; küresel COVID-19 salgınının dünyada ve Türkiye'de değişen durumuna bağlı olarak ülkelere ait güncel COVID-19 verisine dayalı özet durum ve analiz sonuçlarına dinamik yapı kazandırılması, böylelikle hızlı ve proaktif kararlara destek verilebilmesidir. Bu kapsamda, COVID-19'u veriye dayalı olarak tanımlamak amacıyla öncelikle çevrimiçi bir R-Shiny uygulaması geliştirilmiştir (<https://elifkartal.shinyapps.io/covid19/>).

Gereç ve Yöntem: Bu çalışmada yöntem olarak Veri Madenciliği için Çapraz Endüstri Standart Süreç Modeli (*Cross-Industry Standard Process for Data Mining - CRISP-DM*) kullanılmıştır. Küresel ve ulusal boyutta COVID-19'un değişen durumu değerlendirilmiştir. Doğrusal Değişim Oranı (DDO), Üstel Büyüme Katsayısı (ÜBK) ve vaka sayısının ikiye katlanması için gereken gün sayısı gibi yeni değişkenler hesaplanmıştır. Böylece, yeni değişkenlerle güçlendirilen veriye k-Ortalama veri madenciliği algoritması uygulanarak kümeleme analizi yapılmış ve ülkelerin benzerlikleri belirlenmiştir. Küme ortalamasına en yakın ülkeler küme merkezi olarak kabul edilmiş, aynı kümedeki ülkeler küme merkezine olan uzaklıklarına göre sıralanmıştır.

Bulgular: Çalışmanın en önemli bulgularından biri ÜBK ve DDO eğilimlerinin aynı olmasıdır. Bu halyle COVID-19'un salgın özelliği olarak kabul edilen üstel bir davranış göstermediği veya kontrol altına alınabildiği söylenebilecektir. Geliştirilen uygulamayla ülkelerin, coğrafi konumlarından bağımsız ve zamana göre dinamik bir biçimde, hangi kümede yer aldığı, aynı kümedeki ülkelerin olası risk durumları ve benzerlikleri daha hassas biçimde belirlenmiştir.

Sonuç: Bu çalışma ve geliştirilen uygulama ile; küresel COVID-19 salgınının dünyada ve Türkiye'de değişen durumuna bağlı ola-

ABSTRACT

Objective: In this study, it is aimed to provide a dynamic structure to the summary status and analysis results based on the current COVID-19 data of the countries based on changing status of global COVID-19 outbreak in the world and in Turkey; thus, to support fast and proactive decisions. In this scope, to define COVID-19 based on data, an online R-Shiny application is developed (<https://elifkartal.shinyapps.io/covid19/>).

Material and Method: In this study, CRoss-Industry Standard Process for Data Mining - CRISP-DM is used as the study method. The changing situation of COVID-19 in global and national dimensions was evaluated. New variables are calculated such as Linear Change Rate (LCR), Exponential Growth Coefficient (EGC), and required days to double cases. Cluster analysis was performed by applying the k-Means data mining algorithm to the data reinforced with the new variables and similarities of countries were determined. The countries closest to the cluster average are accepted as cluster centers and the countries in the same cluster are ranked according to their distance from the cluster center.

Results: One of the most important findings of the study is that the trends of LCR and EGC are the same. As such, it can be said that COVID-19 does not display an exponential behavior or can be controlled. With the developed application, the countries in which the cluster is located, regardless of their geographical location and dynamically according to time, the possible risk situations and similarities of the countries in the same cluster have been determined more precisely.

Conclusion: With this study and the application developed; depending on changing status of global COVID-19 outbreak

İletişim kurulacak yazar/Corresponding author: elifk@istanbul.edu.tr

Başvuru/Submitted: 12.06.2020 • **Kabul/Accepted:** 24.06.2020 • **Online Yayın/Published Online:** 23.07.2020

©Telif Hakkı 2020 J Ist Faculty Med - Makale metnine jmed.istanbul.edu.tr web sayfasından ulaşılabilir.

©Copyright 2020 by J Ist Faculty Med - Available online at jmed.istanbul.edu.tr

rak ölkelere ait güncel COVID-19 verisine dayalı özet durum ve analiz sonuçlarına dinamik yapı kazandırılmış, böylelikle hızlı ve proaktif kararlara destek verilebilmesi sağlanmıştır.

Anahtar Kelimeler: COVID-19, Korona virüsü, kümeleme, veri madencilięi, tanımlayıcı istatistik

in the world and in Turkey, a dynamic structure has been given to the summary status and analysis results based on the current COVID-19 data of the countries, thus, it has been provided to support fast and proactive decisions.

Keywords: COVID-19, Corona virus, clustering, data mining, descriptive statistics

GİRİŞ

Aralık 2019 tarihinde Çin'in Wuhan (Hubei) şehrinde ortaya çıkan Yeni Korona Virüsü (COVID-19) kısa sürede tüm dünyaya yayılmış ve 11 Mart 2020 tarihinde Dünya Sağlık Örgütü - DSÖ (World Health Organization - WHO) tarafından küresel salgın olarak ilan edilmiştir (1). Şu ana kadar aşı ya da ilaç gibi kesin bir tedavi yönteminin bulunmadığı salgında, sosyal mesafenin korunması ve maske kullanımı salgının yayılmasının engellenmesi için alınması gereken önemli önlemler arasında gösterilmektedir (2). Tüm dünyada sağlık çalışanlarının hastalığın tanı ve teşhisi için gösterdiği üstün çabaların yanı sıra; COVID-19 salgınının olumsuz etkileri eğitim, ekonomi, tarım, turizm ve spor gibi pek çok sektörde hissedilmiştir. Bu nedenle, salgının en iyi şekilde yönetilebilmesi için mikro ve makro düzeyde öngörülerde bulunabilmede sağlık, enformatik, veri bilimi ve mühendislik gibi farklı disiplinlerden bir araya gelen araştırmacıların ortak çalışmaları ile mevcut COVID-19 verisinin iyi analiz edilmesi ve yorumlanması oldukça önemlidir. Bu noktada, matematik ve istatistiğin yanı sıra veri madencilięi teknikleri de tanımlayıcı (descriptive) ve tahmine yönelik (predictive) analizlerin gerçekleştirilmesinde farklı teknikler sunmaktadır.

Bu çalışmanın amacı; küresel COVID-19 salgınının dünyada ve Türkiye'de deęişen durumuna baęlı olarak ölkelere ait güncel COVID-19 verisine dayalı özet durum ve analiz sonuçlarına dinamik yapı kazandırılması, böylelikle hızlı ve proaktif kararlara destek verilebilmesidir. Bu kapsamda; "Küresel COVID-19 Salgınının Dünyada ve Türkiye'de Deęişen Durumu ve Kümeleme Analizi" adında çevrimiçi bir R-Shiny uygulaması geliştirilmiştir. Bu uygulama sayesinde; COVID-19'la ilgili tanımlayıcı istatistik olarak dünya çapında ve ölkede bazında günlük akışın, vaka ve vefatların hareketlerinin ve deęişimlerinin harita ve grafik üzerinde izlenmesi mümkün kılınmıştır. Ayrıca veri madencilięi tekniklerinden k-Ortalamlar Algoritması (k-Means Algorithm) kullanılarak ölkelerin COVID-19 salgınındaki benzerliklerinin bulunabilmesi ve buna göre ölkelerin gruplandırılabilmesi amacıyla veri yeni deęişkenlerle kuvvetlendirilmiştir. Sonuçta; coęrafi konumlarından baęımsız olarak ölkelerin COVID-19 çerçevesinde benzerlikleri/farklılıkları ortaya çıkarılmaya çalışılmıştır. Bu kapsamda geliştirilen Küresel Deęerlendirme, Ülkesel Deęerlendirme ve Ölkelerin Gruplandırılması (kümeleme) seçeneklerine sahip uygulamanın ayrıntıları çalışmada sunulmuştur. Bu çalışma sayesinde, özellikle COVID-19 ile ilgili gün gün dinamik deęişimin

ölkede grupları üzerinden izlenmesi ve salgınla ilgili alınacak kararlarda aynı grupta yer alan ölkelerin edindięi tecrübe ve salgın önlemleri, hasta tedavisi gibi konularda izledięi stratejilerden faydalanılması mümkün olacaktır.

GEREÇ VE YÖNTEM

Veri madencilięi, en basit haliyle veri içinde gizli olan işe yarar bilginin keşfi biçiminde tanımlanabilir. Büyük veri içinden deęerli bilginin elde edilebilmesi, daha önceden bilinmeyen, geçerli ve uygulanabilir bilginin geniş veritabanlarından elde edilerek, bir işletmenin karar sürecinde kullanılmasıdır (3,4). Veri madencilięi; büyük veriden yapay zekâ, makine öğrenmesi, istatistik, optimizasyon gibi farklı disiplinlerin sunduęu tekniklerle karar vericiler için etkin ve bilgiye dayalı karar vermeyi sağlamak için klasik yöntemlerle ortaya çıkarılması zor, gizli, örtük, faydalı ilişki, örüntü, baęıntı ya da trendlerin ortaya çıkarılmasıdır (5).

Bu çalışmada yöntem olarak Veri Madencilięi için Çapraz Endüstri Standart Süreç Modeli (CRoss-Industry Standard Process for Data Mining - CRISP-DM) kullanılmıştır (6,7). Bu yöntem, problemin tanımlanması (business understanding), veriyi anlama (data understanding), veri hazırlama (data preparation), model kurma (modeling), model deęerlendirme ve seçimi (evaluation) ile modelin uygulamaya geçirilmesi (deployment) olmak üzere altı adımdan oluşmaktadır. Çalışma problemi giriş bölümünde ele alındığından, bu bölüm veriyi anlama adımıyla başlamıştır.

Veriyi anlama

Bu çalışmada kullanılan COVID-19 verisi European Centre for Disease Prevention and Control'dan alınmıştır (8). Veri seti kapsamındaki tarih, vaka ve vefat sayıları, ölkede adı, ölkede kodu, ölkelerin 2018 yılı nüfus bilgisi ile kıta bilgisini içeren nitelikler çalışmada kullanılmıştır. Ayrıca; ölkelerin coęrafi konumlarından baęımsız olarak kümeleme analizinde COVID-19 salgını ile ilgili birbirine benzerliklerini daha iyi ortaya çıkarabilmek amacıyla aşağıda detayları yer alan ek nitelikler hesaplanmış ve veri setine eklenmiştir. Bu niteliklerin kümeleme analizini güçlendirmesi beklenmiştir. Veri incelenirken veri girişinin hatalı yapıldığı tespit edilen hatalı veri doğrusu ile güncellenmiş (Türkiye'ye ait 30.04.2020 tarihli vefat sayısı 89 yerine 809 olarak girildięi tespit edilerek gereken düzeltme yapılmıştır); ilerleyen günlerde ilgili verinin veri kaynağında da güncellendięi gözlemlenmiştir. Bu durum yazarların veri kaynağının güvenilirlięi konusundaki düşüncelerini kuvvetlendirmiştir.

$$Bugün(n_c) = Dün(n_c) * (Doğrusal Değişim Oranı) \quad (\text{Denklem 1})$$

$$Doğrusal Değişim Oranı (DDO) = \frac{Bugün(n_c)}{Dün(n_c)} \quad (\text{Denklem 2})$$

$$Dün(n_c)^{(\text{Üstel Büyüme Katsayısı})} = Bugün(n_c) \quad (\text{Denklem 3})$$

$$\text{Üstel Büyüme Katsayısı (ÜBK)} = \frac{\ln(Bugün(n_c))}{\ln(Dün(n_c))} \quad (\text{Denklem 4})$$

$$TodayBugün * (Doğrusal Değişim Oranı)^{\text{Gün Sayısı}(\text{vak'a sayısı} \times 2)} = 2 * Bugün(n_c) \quad (\text{Denklem 5})$$

$$Bugün(n_c) * \left(\frac{Bugün(n_c)}{Dün(n_c)}\right)^{\text{Gün Sayısı}(\text{vak'a sayısı} \times 2)} = 2 * Bugün(n_c) \quad (\text{Denklem 6})$$

$$\text{Gün Sayısı}(\text{vak'a sayısı} \times 2) = \frac{\ln(2)}{\ln\left(\frac{Bugün(n_c)}{Dün(n_c)}\right)} \quad (\text{Denklem 7})$$

$$\text{Vak'a Sayısı (bir milyon kişide)} = \frac{n_c}{n_{\text{üfus}}} * 1.000.000 \quad (\text{Denklem 8})$$

$$\text{Vefat Oranı} = \frac{n_d}{n_c} \quad (\text{Denklem 9})$$

Bir ülkenin popülasyonu n , toplam vaka sayısı n_c ve toplam vefat sayısı n_d ile gösterilsin. Buna göre vaka sayısı için doğrusal değişim oranı (denklem 1 ve 2), üstel büyüme katsayısı (denklem 3 ve 4), vaka sayısının iki katına çıkması için gereken gün sayısı (Gün Sayısı_(vak'a sayısı x 2)) (denklem 5, 6 ve 7), bir milyon kişideki vaka sayısı (denklem 8) ve vefat oranı (Denklem 9) aşağıdaki biçimde hesaplanmıştır (9). Denklem 1-8'dekilere benzer hesaplamalar vefat sayısı için de yapılmıştır. Ülkelerin vefat sayısının ikiye katlanması için gereken gün sayıları, doğrusal değişim oranının 1'e çok yakın çıkması sebebiyle hesaplamalarda reel bir sayı değeri olarak elde edilemediğinden kümeleme analizinde kullanılmamıştır.

Veri hazırlama

Geliştirilen uygulamada veri dinamik bir biçimde veri kaynağından çekilmekte, veri seti her gün güncellenmektedir. Veri setinde bazı ülkelerin 2018 yılına ait eksik kalan nüfus bilgisi sonradan tamamlanmıştır. Ayrıca veri setinde "Cases on an International Conveyance Japan" olarak yer alan gözleme ait veri Japonya'ya dahil edilmiştir. Kullanılan tüm sayısal nitelikler kümeleme analizi öncesinde min-max normalizasyon yöntemine göre normalize edilmiştir (10).

Modelleme

Veri madenciliği problemleri; sınıflandırma, kümeleme ve birliktelik analizi gibi üç temel başlık altında incelenebilmektedir. Ülkelerin COVID-19 verisine dayalı biçimde gruplandırılabilmesi amacıyla kümeleme analizinde ba-

şarılı ve çok yaygın olarak kullanılan k-Ortalamlar Algoritması (*k-Means Algorithm*) tercih edilmiştir. Kümeleme analizinde temel olarak gözlemler arasındaki yakınlık ve uzaklık (benzerlik ya da benzemezlik) durumlarına bakılarak gruplandırma işlemi yapılmaktadır. k-Ortalamlar algoritması kısaca şu şekilde açıklanmaktadır (11):

k adet küme merkezi (genellikle rastgele) belirlenir.

Herhangi bir noktanın (gözlemin) küme ataması değiştiği sürece;

- ❖ Veri setindeki her bir gözlem için;
 - Her küme merkezi için;
 - Her nokta ile küme merkezi arasındaki uzaklık hesaplanır.
 - Nokta, en yakınındaki kümeye atanır.
- ❖ Her küme için, küme içindeki tüm noktaların ortalaması hesaplanır.
 - Ortalama, yeni küme merkezi olarak atanır.

Küme sayısının kesin belirli olmadığı durumlarda en iyi küme sayısının bulunabilmesi, gözlemlerin kümelere ne kadar isabetli atandığının bulunabilmesi için çeşitli yöntemler kullanılmaktadır (12). Bunlardan biri de Silhouette İndeksidir (12-14):

Veri setindeki bir i noktası alınsın. Her bir i için bir $s(i)$ Silhouette değeri tanımlanabilir. A , i noktasının atandığı

Tablo 1: Toplam ve ortalama COVID-19 vaka ve vefat sayıları (17.05.2020)

	Vaka		Vefat		Vefat oranı	Sıra	Ülke sayısı	Ortalama	
	n	%	n	%				Vaka (n)	Vefat (n)
Amerika	2.049.566	44,57	123.091	39,49	6,01	1	49	41.827,88	2.512,06
Avrupa	1.687.098	36,69	161.525	51,82	9,57	2	54	31.242,56	2.991,20
Asya	771.984	16,79	24.252	7,78	3,14	3	43	17.953,12	564,00
Afrika	81.435	1,77	2.699	0,87	3,31	4	55	1.480,64	49,07
Okyanusya	8.464	0,18	126	0,04	1,49	5	8	1.058,00	15,75
Dünya	4.598.547	100	311.693	100	6,78	-	209	22.002,62	1.491,35
Türkiye	148.067	3,22	4.096	1,31	2,77	-	-	148.067,00	4.096,00

küme olsun. $a(i)$, A kümesindeki her i için ortalama benzerliği, $C (\neq A)$ gibi A'dan farklı bir başka küme olmak üzere; $d(i, C)$, i noktasının C'deki tüm noktalara olan ortalama benzerliği olmak üzere; $C (\neq A)$ olan tüm kümeler için $d(i, C)$ hesaplanır ve bunlar içinde en küçüğü seçilir:

$$b(i) = \min_{C \neq A} d(i, C)$$

Buradan bir i noktasının Silhouette İndeksi aşağıdaki gibi hesaplanır:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}; -1 \leq s(i) \leq 1$$

Gözleme ait Silhouette İndeksinin 1'e yakın çıkması, gözlemin atandığı kümeye uygun olduğunu göstermektedir; aksi halde bu durumun tam tersini ifade ettiği söylenebilir. Bir kümedeki her bir i noktası için $s(i)$ silhouette değerlerinin ortalaması, o kümenin Silhouette İndeksini (ortalama Silhouette genişliğini) vermektedir. $i=1,2,\dots,n$ için $s(i)$ silhouette değerlerinin ortalaması ise tüm veri setinin ortalama Silhouette genişliğini vermektedir.

Bu çalışmada yapılan kümeleme analizinde ülkelerin COVID-19 durumlarının iyi, kritik ve kötü olmak üzere üç kümeye ayrılması hedeflenmiştir; ancak ayrıca yapılan analizler için küme sayısına göre Silhouette İndeksleri de raporlanmıştır. Tüm analizler R programlama dili kullanılarak RStudio'da gerçekleştirilmiştir (15,16). Çalışmada kullanılan R paketleri şu şekilde sıralanabilir: cluster, clusterSim, dplyr, DT, formattable, ggplot2, googleVis, plyr, shiny, shinyjs (17–26). Bazı analizlerde veri görselleştirme MS Excel'den de faydalanılmıştır. Geliştirilen Shiny uygulaması shinyapps.io hesabı üzerinden erişime açılmıştır (27–29). Geliştirilen uygulamanın tüm detaylarına dördüncü bölümde yer verilmiştir.

BULGULAR

Bu bölümde; geliştirilen uygulamada 17.05.2020 tarihi dikkate alınarak COVID-19'un dünyada ve Türkiye'de değişen durumu tablo ve grafiklerle sunulmuştur. Ardından

veri setindeki niteliklerin (değişkenlerin) birbirleri arasındaki ilişkilere değinilerek, kümeleme analizinden elde edilen bulgulara yer verilmiştir.

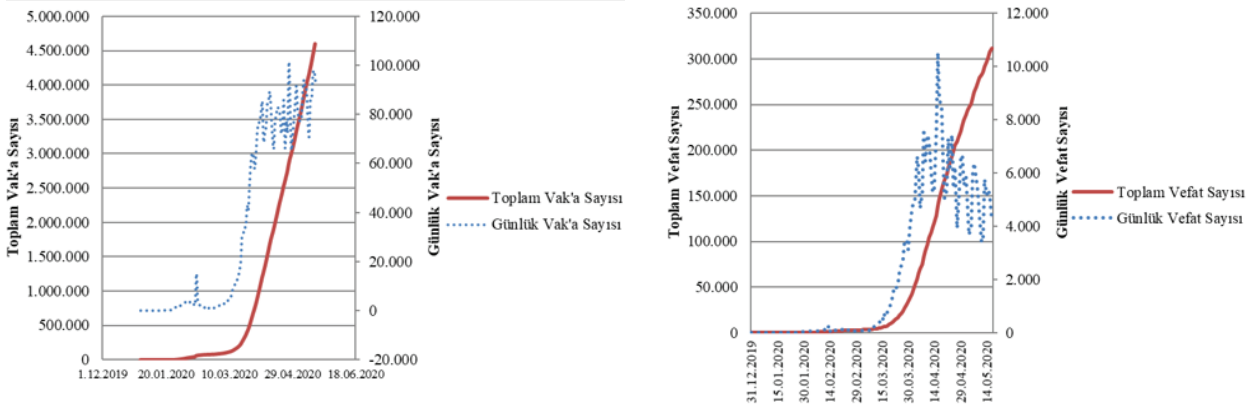
Küresel değerlendirme

Dünyanın genel COVID-19 durumuna bakıldığında (Tablo 1), Amerika kıtası %44,57 ile vaka sayılarında dünyanın tüm kıtalarının önünde yer almaktadır. Okyanusya ülkelerinde ise bu oranın %0,18 olduğu görülmektedir; ancak vefat sayılarında %51,82 ile Avrupa kıtası önde yer almaktadır. Avusturalya, Yeni Zelanda gibi Okyanusya ülkelerinin vaka sayılarında olduğu gibi ($n=8.464$), %0,04'lük vefat yüzdesiyle de en iyi durumda olduğu gözlemlenmiştir. Ayrıca Tablo 1'de Türkiye'nin genel COVID-19 durumu da kıta ve dünya ortalamalarına göre karşılaştırılabilir. Dünya genelinde, 17.05.2020 tarihine kadar vaka ve vefat sayısı eğilimine bakıldığında özellikle günlük vaka ve vefatların eğiliminde azalma olduğu veya salgının kontrol altına alınabildiği genelde söylenebilir (Şekil 1).

Vaka sayılarına göre dünyadaki ilk 10 ülke sürekli yer değiştirmiş, 17.05.2020 tarihinde Amerika Birleşik Devletleri (ABD) en fazla vaka sayısı 1.467.884 (%45) ve vefat sayısı 88.754 (%36) ile tüm dünya ülkelerinin başına geçmiştir (Tablo 2, Şekil 2); ancak bir milyon nüfusa düşen vaka sayısı sıralamasında 11, vefat sayısı sıralamasında ise 13. sırada yer almaktadır. ÜBK ve DDO dikkate alındığında salgının kontrol altına alınmaya başlandığı söylenebilir. Özellikle İspanya, İtalya, Almanya, Fransa ve Türkiye'de artışların kontrol altına alındığı görülmektedir. Türkiye vaka sayısı sıralamasında 8. sırada yer alırken vefat oranlarında %2,77 ile 113. sıradadır. Şekil 2'de, Tablo 2'de yer alan ülkelerin kendi içinde vaka ve vefat sayılarının yüzdesel değerlendirmesi pasta grafiği olarak sunulmuştur.

Türkiye'nin toplam vaka sayısı açısından incelenmesi

Türkiye'nin dünyadaki durumu ve diğer ülkeler ile karşılaştırması yapıldığında vefat sayısı ile 14. sıradayken vefat oranlarında 209 ülke arasında 113. sırada olduğu gözlemlenmiştir. Ülkemiz vaka sayılarında üstel büyüme katsayısı (ÜBK) dikkate alındığında sıralamada 98. sırada yer alırken, vaka sayısının ikiye katlanması için gereken gün sayısı

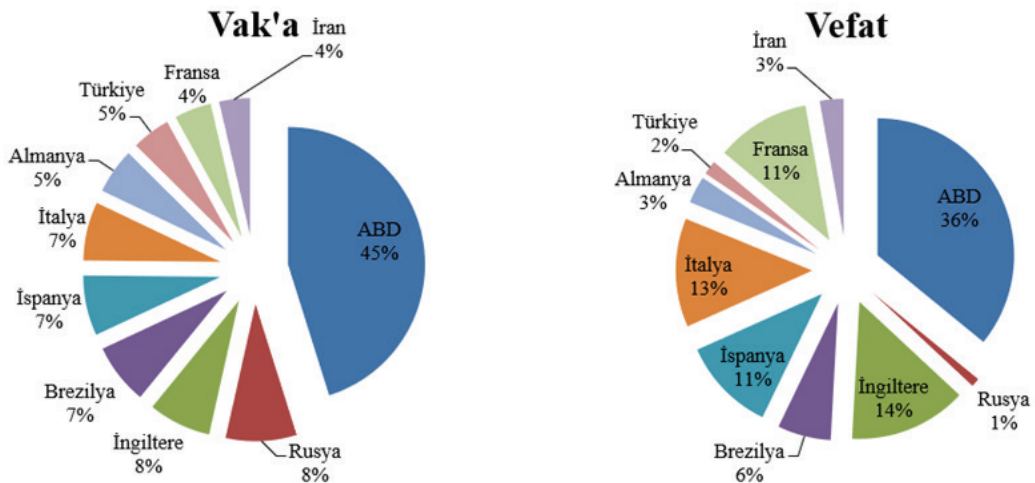


Şekil 1: Dünyada COVID-19 vaka ve vefat sayılarının eğilimi (17.05.2020).

Tablo 2: COVID-19 vaka sayılarına göre ilk 10 ülke (17.05.2020)

Ülke	Vaka (n)	Vefat (n)	DDO (vaka)	ÜBK (vaka)	Gün sayısı (vaka x2)	Vefat oranı (%)	DDO (vefat)	ÜBK (vefat)	Vaka sayısı (1m)	Vefat sayısı (1m)
1 ABD	1.467.884	88.754	1,017	1,001	41,2	6,05	1,014	1,001	4.487	271
2 Rusya	272.043	2.537	1,035	1,003	20,1	0,93	1,049	1,006	1.883	18
3 İngiltere	240.161	34.466	1,015	1,001	47,9	14,35	1,014	1,001	3.612	518
4 Brezilya	233.142	15.633	1,068	1,005	10,5	6,71	1,055	1,006	1.113	75
5 İspanya	231.350	27.650	1,005	1	137,1	11,95	1,007	1,001	4.951	592
6 İtalya	224.760	31.763	1,004	1	177,7	14,13	1,005	1	3.719	526
7 Almanya	174.355	7.914	1,003	1	206,9	4,54	1,004	1	2.102	95
8 Türkiye	148.067	4.096	1,011	1,001	63,4	2,77	1,010	1,001	1.799	50
9 Fransa	142.291	27.643	1,003	1	264,8	19,43	1,003	1	2.124	413
10 İran	118.392	6.937	1,015	1,001	46,4	5,86	1,005	1,001	1.447	85

DDO: Doğrusal Değişim Oranı, ÜBK: Üstel Büyüme Katsayısı, 1m: Bir milyon kişide



Şekil 2: Vaka ve vefat sayılarına göre ilk 10 ülkenin kendi içinde yüzdesel değerlendirmesi (17.05.2020).

Tablo 3: Türkiye'nin vaka ve vefat sayılarına göre üç farklı tarih aralığı açısından incelenmesi

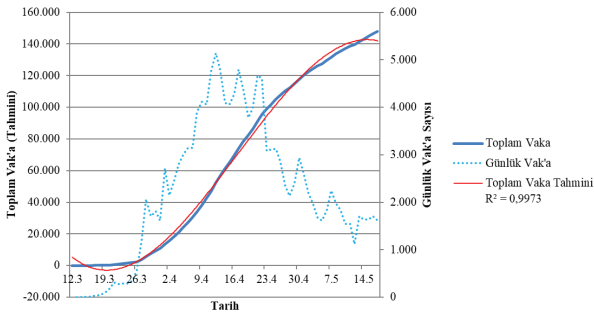
	Tarih aralıkları					
	12.03.2020 17.05.2020		12.03.2020 30.04.2020		12.03.2020 31.03.2020	
	Mevcut değerler	Sıra	Mevcut değerler	Sıra	Mevcut değerler	Sıra
Türkiye						
Vaka (n)	148.067	8	117.589	7	10.827	12
Vefat (n)	4.096	14	3.081	12	168	12
ÜBK (vaka)	1,001	98	1,002	92	1,018	61
Gün sayısı (vakax2)	63	41	27	65	4	101
Vefat oranı (%)	2,77	113	2,62	113	1,55	80
Vaka (1m)	1.799	45	1.428	38	132	61
Vefat (1m)	50	35	37	33	2	37

Gün Sayısı (vakax2): Vaka sayısının ikiye katlanması için gereken gün sayısı, 1m: Bir milyon kişide

Tablo 4: Türkiye'nin COVID-19 vaka sayısındaki üstel ve doğrusal büyümenin incelenmesi

	Toplam vaka	Günlük vaka	ÜBK (toplam vaka)	DDO (toplam vaka)	Gün sayısı (toplam vakax2)	Toplam vaka tahmin (ÜBK)	ÜBK mutlak hata	Toplam vaka tahmin (DDO)	DDO mutlak hata
Minimum	1	1	1,001	1,008	0,32	169	5	123	5
Maksimum	148.067	5.138	4,170	9,000	86,62	148.187	1.783	148.185	1.762
Ortalama	66.589	2.622	1,074	1,319	21,19	71.179	378	71.149	360
Standart Sapma	-	1525	0,397	1,017	22,16	52.687	381	52.710	363
Gün sayısı	65	65	65,000	65,000	65,00	65	65	65	65

Gün sayısı (vakax2): Vaka sayısının ikiye katlanması için gereken gün sayısı, DDO: Doğrusal Değişim Oranı, ÜBK: Üstel Büyüme Katsayısı



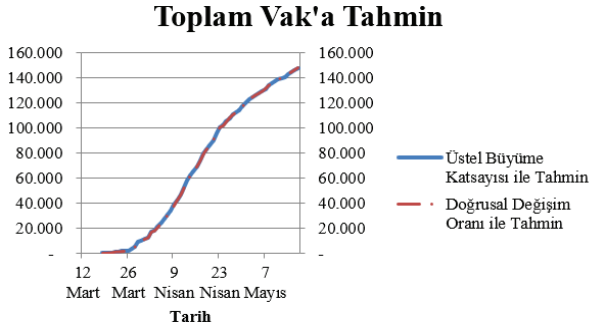
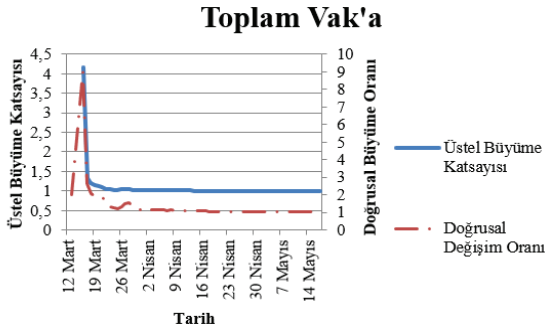
Şekil 3: Türkiye'nin vaka sayısındaki eğilim (17.05.2020).

Mart sonunda 4, Nisan sonunda 27 ve 17.05.2020 tarihi itibarı ile 63 güne çıkmıştır. Türkiye'nin gerek vefat oranlarının diğer ülkelere göre düşük seyretmesi, gerekse vaka sayısının ikiye katlanması için gereken gün sayısının giderek artması ile kritik durumdan iyi duruma doğru eğilim gösterdiği söylenebilir (Tablo 3).

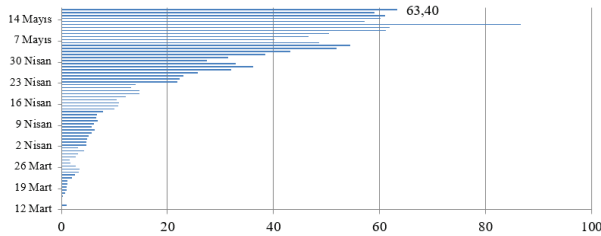
Diğer taraftan; toplam vaka artışında tam olarak yataylaşma görülmesi de günlük vaka sayılarındaki değişim azalma yönünde eğilim göstermektedir (Şekil 3). Toplam vaka sayısının üçüncü dereceden polinom kullanılarak gerçekleştirilen tahminlerinin ise büyük ölçüde örtüştüğü yine Şekil 3'ten görülebilir ($R^2=0,9973$).

Çalışmada elde edilen en önemli bulgulardan biri de ÜBK ve DDO eğilimlerinin aynı olmasıdır (Tablo 4, Şekil 4). Bu haliyle COVID-19'un salgın özelliği olarak kabul edilen üstel bir davranış göstermediği veya kontrol altına alınabileceği söylenebilir. Tahmindeki mutlak hatalarının yakınlığı dikkat çekmekte, hatta doğrusal değişim oranına ait mutlak hata daha düşük olduğundan doğrusal değişim oranına göre yapılacak tahminlerin kullanılabilirliği söylenebilir.

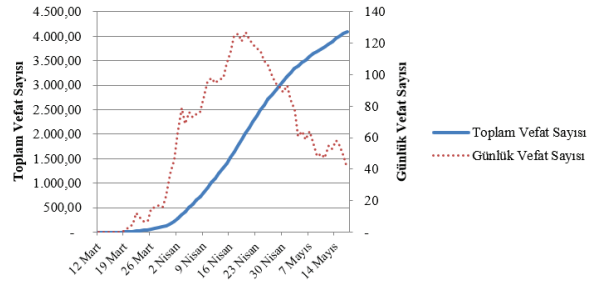
Diğer bir gösterge ise toplam vaka sayılarındaki ikiye katlanma gün sayısı 60. günde 87 güne çıkarken 65. gün (17.05.2020) itibarı ile 63,40 (~63) gün olarak belirlenmiştir. Böylelikle, ülkemiz dünya sıralamasında 111. sıraya



Şekil 4: Vaka sayısına ait ÜBK ve DDO'nun karşılaştırılması (17.05.2020).



Şekil 5: Toplam vaka sayılarındaki ikiye katlanma gün sayısı.



Şekil 6: Türkiye'nin vefat sayısındaki eğilim.

gelmiştir. Bu bulgu da yine ülkemizin COVID-19 salgınının da daha iyi bir duruma geldiğini göstermektedir (Şekil 5).

Türkiye'nin toplam vefat sayısı açısından incelenmesi

Vefat sayılarında da vaka sayılarında olduğu gibi her ne kadar eğilimdeki yataylaşma net olarak görülme de, günlük vefat sayılarındaki değişim eğilimi azalma yönündedir (Şekil 6).

Yine vaka sayılarında olduğu gibi, vefat sayısına ait ÜBK ve DDO) etkisi tahmin üzerinde benzerlik göstermiş, vefat sayıları ile ilgili de doğrusal değişim oranıyla yapılacak tahminlerin çok daha iyi olabileceği ortaya çıkmıştır (Tablo 5, Şekil 7).

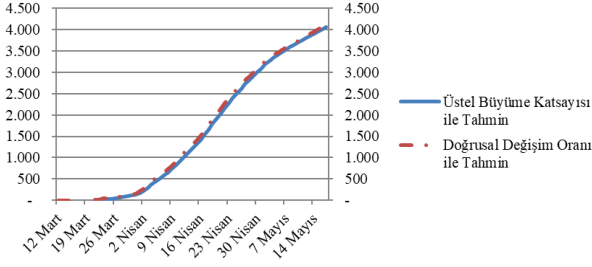
Aynı şekilde vefat sayılarının ikiye katlanması için gereken gün sayılarının ise 65. günde 68,90 (~69) güne çıkmasından, kritik durumdan iyi duruma eğilim gösterdiği anlaşılmaktadır (Şekil 8). Özellikle önceki gün ve aylara bakıldığında gerek vaka gerekse vefat sayılarındaki değişim oranlarının iyiye gidişatı görülmektedir; ancak DSÖ tarafından salgın olarak kabul edilen COVID-19 üstel artış gösteren bir küresel salgın hastalığıdır. Kontrol edilmediğinde koruyucu ve tedavi yönünde gerekli tedbirlerin alınmadığı durumlarda her an hareketlenebileceği de bilinmektedir. Tüm bu azalışlarda alınan önlemlerin ve tedavi yöntemlerinin yanısıra COVID-19'un iyi yönde mutasyona uğramış olabilmesinin de etkili olabileceği olasıdır; ancak bu konuda kesinleşmiş bir bilgi henüz mevcut değildir.

Tablo 5: Türkiye'nin vefat sayısı incelemesi

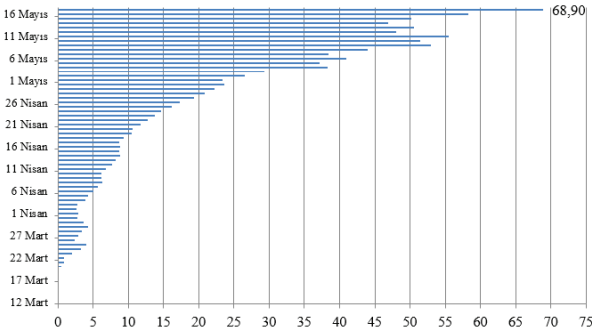
	Toplam vefat	Günlük vefat	ÜBK (toplam vefat)	DDO (toplam vefat)	Gün Sayısı (toplam vefatx2)	Vefat (%)	Toplam vaka tahmin (ÜBK)	Übk mutlak hata	Toplam vefat tahmin (DDO)	Ddo Mutlak hata
Minimum	0	0	1	1	1	0,00	14	1	0	0
Maksimum	4.096	127	2	4	69	2,77	4.061	110	4.104	33
Ortalama	1.690	63	1	1	19	2,09	1.866	61	1.749	7
Standart Sapma	-	40	0,090	0,438	18	0,75	1.418	31	1.470	6
Gün Sayısı	65	65	65	65	65	-	65	65	65	65

Gün Sayısı (vefatx2): Vefat sayısının ikiye katlanması için gereken gün sayısı, DDO: Doğrusal Değişim Oranı, ÜBK: Üstel Büyüme Katsayısı

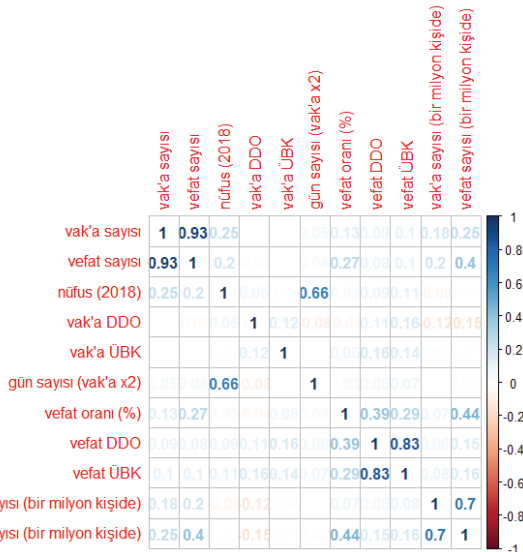
Toplam Vefat Tahmin



Şekil 7: Vefat sayısı için ÜBK ve DDO'nun karşılaştırılması.



Şekil 8: Toplam vefat sayılarındaki ikiye katlanması için gereken gün sayısı.



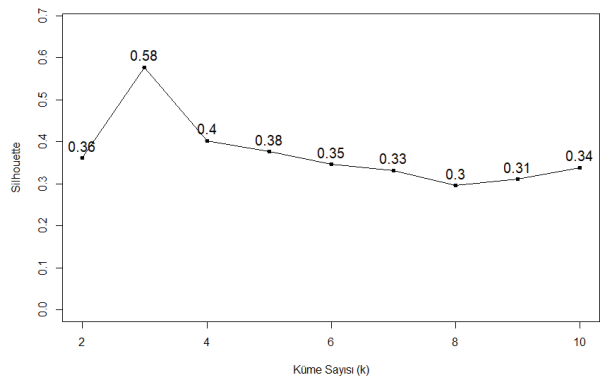
Şekil 9: Nitelikler arasındaki ilişkinin incelenmesi (10.06.2020).

Veri setindeki nitelikler arasındaki ilişki (korelasyon), Pearson korelasyon katsayısı hesaplanarak Şekil 9'da sunulmuştur. İlişkinin yönü pozitiften negatife maviden kırmızıya doğru renklendirilmiştir. Vaka ile vefat sayısı arasında (0,93), vefat sayısındaki DDO ile ÜBK arasında (0,83) ve bir milyon kişideki vaka ve vefat sayıları arasında (0,70) pozitif yönde kuvvetli bir ilişki olduğu tespit edilmiştir.

Kümeleme analizinden elde edilen bulgular

Bu çalışmanın en önemli analizlerinden biri olan ülkelerin benzerliklerine göre gruplandırılması (kümelenmesi) ve buradan çıkabilecek sonuçlar ile ülkeler alınacak tedbirler ve yöntemler konusunda işbirliklerine ve ortak kararlara gidebilecekleri üzerinedir. Kümeleme analizinin özet tablosu olan Tablo 6'ya bakıldığında kümelerin dinamik olarak değiştiği ve ülkelerin kritik durumdan iyi duruma geçebileceği gibi daha kötü bir duruma da geçebileceği görülebilmektedir. Kötü, kritik ve nispeten iyi durumda olan ülkeleri ayırmak için 3 küme varsayımı ile sonuçlar ele alındığında, Mart sonu, Nisan sonu ve 17.05.2020 tarihi itibari ile küme merkezlerinin (ortalamalarına en yakın ülkelerin) değiştiği görülmüş ve bu kümedeki ülkeler merkeze olan uzaklıkları sıralandığı zaman mevcut kümedeki sıralamaları da anlam kazanmıştır. Buna göre 17.05.2020 itibari ile küme merkezlerine en yakın ülkeler sırası ile İspanya, Ukrayna ve Moğolistan olarak belirlenmiştir. Bu kümelere ait ortalama vaka, vefat sayıları dikkate alınarak, kümelerin COVID-19 durumu (iyi, kritik veya kötü) çıkarılmıştır. Buna göre Türkiye'nin her zaman kritik ülkeler arasında yer aldığı ve salgında kötü durumda bir ülke pozisyonuna düşmediği görülmektedir. 17.05.2020 tarihinde bulunduğu küme içinde sırası dikkate alındığında, aynı kümede bulunduğu 152 ülke arasında ülkemize en yakın benzer ülkelerin coğrafi konumdan bağımsız olarak Yunanistan (52. sırada) ve Japonya (54. sırada) olduğu belirlenmiştir. Şüphesiz bu durum zaman içinde değişebilecektir; ancak bu benzerlikte rol alan tedbirlerin ve tedavi yöntemlerinin araştırılması ve ortak çalışmaların yapılması önem kazanmaktadır.

Yine 17.05.2020 tarihi dikkate alınarak gerçekleştirilen kümeleme analizi sonucunda k=2, 3, ..., 10 için elde edilen Silhouette değerleri Şekil 10'da verilmiştir. Şekil 10'a göre özellikle üç kümeden sonra, Silhouette İndeksinin genelde azalan bir eğilimde olduğu söylenebilir. Analiz sonuçlarına göre Silhouette İndeksi en yüksek elde edilen küme sayısı 3 olmuştur (0,58).



Şekil 10: Küme sayısı 2 ile 10 arasında elde edilen Silhouette değerleri.

GELİŞTİRİLEN COVID-19 UYGULAMASI

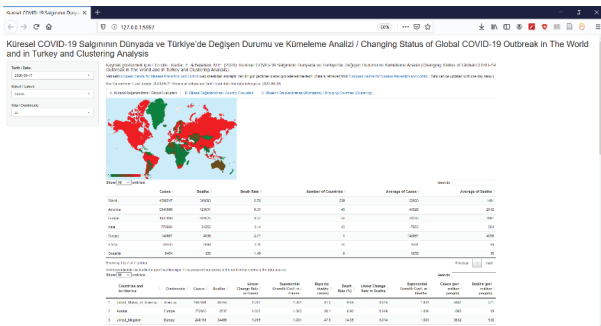
"Küresel COVID-19 Salgınının Dünyada ve Türkiye'de Değişen Durumu ve Kümeleme Analizi" (<https://elifkartal.shinyapps.io/covid19/>), R programlama dili kullanılarak geliştirilen, web üzerinden erişilebilen bir Shiny uygulamasıdır (27). Analizlerde kullanılan veri, veri kaynağında güncellendiğinde, bu güncelleme analizlere de yansımaktadır (bir gün gecikmeli olarak). Örneğin; Türkiye'de ilk COVID-19 vakası 11 Mart 2020 tarihinde görülmüştür; ancak veri raporlanma ve güncellenme durumu göz önünde bulundurulduğunda bu bilgiye 12 Mart 2020 tarihi itibarıyla ulaşmak mümkündür. Geliştirilen Shiny uygulaması ise 27 Mart 2020 tarihinden itibaren erişime sunulmuştur. COVID-19'un ülkelere yayılma durumuna göre uygulamada yer alan ülke sayısı da değişmektedir. Uygulamanın ilk geliştirildiği tarihte 196 olan ülke sayısı Mayıs 2020 itibarıyla 200'ün üzerindedir.

Veri kaynağında veri setindeki sütun adlarının değiştirilmesi ya da yeni sütun eklenmesi gibi değişiklikler yakından takip edilmiş ve geliştirilen uygulamanın olumsuz etkilememesi sağlanmaya çalışılmıştır; ancak ileride veri seti üzerinde geçmişe yönelik yapılabilecek değişiklikler, uygulamadan elde edilecek sonuçları da değiştirebilecektir. Veri kaynağında kimi zaman bazı ülkelerde (İspanya gibi) günlük veri güncelleme gecikmesi uygulamanın bütünlüğü açısından görülen bir başka olumsuz durum olmuştur. Bu durumda uygulamada ilgili ülkenin dün ve bugüne dayalı hesaplanan verisi yerine bir önceki gün ve dün ait verisi kullanılarak hesaplanmış ve uygulamada da ilgili ülkenin adının bulunduğu kısımda parantez içinde belirtilmiştir.

Uygulamada; "A. Küresel Değerlendirme", "B. Ülkesel Değerlendirme" ve "C. Ülkelerin Gruplandırılması (kümeleme)" olmak üzere üç ayrı seçenek bulunmaktadır. Uygulamaya ilişkin bu bölümler sırasıyla açıklanmıştır.

KÜRESEL DEĞERLENDİRME

Küresel değerlendirme bölümünde (Şekil 11) her ülkenin; kıta, vaka ve vefat sayısı, vaka ve vefat sayısındaki doğrusal değişim oranı ve üstel büyüme katsayısı, vaka sayısının

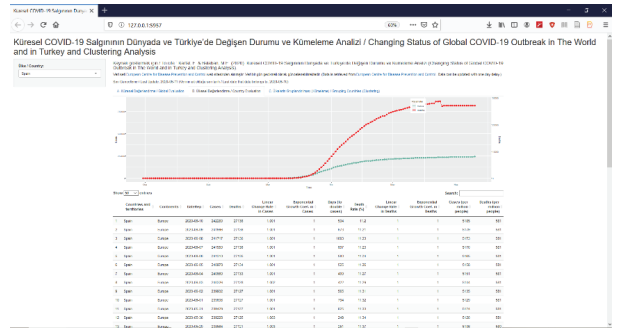


Şekil 11: Uygulamanın küresel değerlendirme kısmından bir ekran görüntüsü.

iki katına çıkması için gereken gün sayısı, vefat oranı, bir milyon kişideki vaka ve vefat sayıları listelenmiştir. Kullanıcılar, verilen dünya haritası üzerinde tarih, etiket (vaka ya da vefat sayısı) ya da kıta seçimi yaparak harita ve tabloyu seçimlerine göre güncelleyebilmektedir. Haritanın renklendirmesinde yapılan etiket seçimine göre uç nokta olarak tespit edilen ülkeler kırmızı renge yakın olacak biçimde hazırlanmıştır. Ayrıca, COVID-19'un kıtalara dağılımı ve dünya ortalamaları karşılaştırmalı olarak gösterilmiştir. Türkiye de özel olarak dünya ve kıta ülkeleri ortalamaları ile karşılaştırmalar arasında eklenmiştir.

Ülkesel değerlendirme

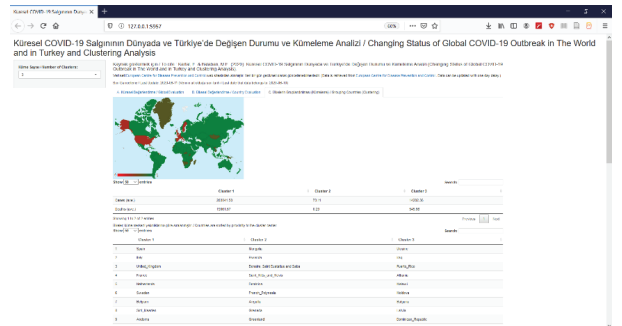
Ülkesel değerlendirmede (Şekil 12); ziyaretçiler dilediği ülkeyi seçerek kümülatif vaka ve vefat sayı hareketlerini izleyebilmekte, grafik üzerinde değişimlerini ve artış hızlarını görebilmektedir. Bu ekranda da varsayılan ülke olarak Türkiye seçilmiştir.



Şekil 12: Uygulamanın ülkesel değerlendirme kısmından bir ekran görüntüsü.

Ülkelerin gruplandırılması (kümeleme)

Bu kısımda (Şekil 13); kullanıcının küresel değerlendirme kısmında seçmiş olduğu tarihe göre veriye dayalı öğrenme algoritmalarından biri olan k-Ortalamalar Algoritması ile istenilen küme sayısına göre ülkeler hem harita üzerinde hem de tabloda gruplandırılmaktadır. Varsayılan küme sayısı ülkelerin COVID-19 durumlarına göre "iyi", "kritik" ve "kötü" olmak üzere gruplandırılabilmesi için 3 seçilmiştir;



Şekil 13: Uygulamanın ülkelerin gruplandırılması (kümeleme) kısmından bir ekran görüntüsü.

ancak kullanıcıların küme sayısını 2-10 arasında küme sayısını deęiştirebilme olanaęı bulunmaktadır. Ayrıca k-Ortalama Algoritması ile kümeleme analizi sonucunda elde edilen kümelerin durumlarının kullanıcılar tarafından anlaşılabilmesi için kümelerdeki ortalama vaka ve vefat sayıları da listelenmektedir. Ayrıca tabloda her kümede yer alan ölkeler; küme merkezlerine yakınlıklarına göre küçükten büyüğe doęru sıralanmıştır. Bir başka ifade ile küme merkezine en yakın ölkeler, en üst sırada listelenmektedir.

TARTIŞMA

Günümüzde COVID-19 salgınının takibi için çeşitli kurum ve kuruluşlar tarafından veri setleri paylaşılmaktadır (8,30,31). Araştırmacılar da bu veri setlerinden yola çıkarak salgının takibi için çeşitli çevrimiçi uygulamalar geliştirmektedir (32–34). Bu çalışma ile küresel COVID-19 salgınının dünyada ve Türkiye’de deęişen durumuna baęlı olarak ölkelere ait güncel COVID-19 verisine dayalı özet durum ve analiz sonuçlarına dinamik yapı kazandırılması, böylelikle hızlı ve proaktif kararlara destek verilebilmesi amaçlanmıştır. Bu kapsamda çevrimiçi erişilebilen “Küresel COVID-19 Salgınının Dünyada ve Türkiye’de Deęişen Durumu ve Kümeleme Analizi” adlı bir R-Shiny uygulaması geliştirilmiştir. Literatürde COVID-19 hastalığına ilişkin gerçekleştirilen kümeleme analizi çalışmaları bulunmaktadır (35,36). Bu çalışmanın ise, analizlerde kullanılan nitelikler ve uygulama yöntemi ile literatürde farklılık yaratacağına inanılmaktadır.

Geliştirilen Shiny uygulamasının en önemli özelliklerinden biri veri kaynağına göre dinamik olarak çalışabilmesi ve raporun otomatik olarak güncellenmesidir. Veri, harita ve grafik okur yazarları için faydalı bilgi sunmasını ve karar vericiler için rasyonel ve proaktif kararların verilmesinde faydalı olabilmesidir. Bu nedenle, 27 Mart tarihinden itibaren geliştirilen uygulamaya erişim sağlanarak güncel bilgiler paylaşılmaktadır. Uygulamaya ilişkin bir dięer önemli özellik ise; doęrusal deęişim oranı, üstel büyüme katsayısı, vaka sayısının iki katına çıkması için gereken gün sayısı gibi ölkelerin coęrafik konumlarından bağımsız biçimde daha iyi gruplandırılabilmesi için tanımlanan ek nitelikler ile salgının takibi için farklı ve daha etkin bir bakış açısı sunmasıdır. Uygulama dili hem Türkçe hem de İngilizcedir. Böylelikle Türkçe alandaki çalışmalara katkı yapılarak, ölkemizden uygulamayı ziyaret edecek kişiler için daha anlaşılabilir ve herkes tarafından kullanılacak bir uygulamanın geliştirilmesi amaçlanmıştır.

Uygulamanın, belki de mevcut COVID-19 uygulamaları arasında en ayırt edici özellięi ise içerdiği ölkelerin gruplandırılması (kümeleme analizi) kısmıdır. Bu kısımda; veri madencilięi algoritmalarından kümeleme analizi yapılarak ölkelerin benzerlikleri ve tarihe göre deęişimleri analiz edilmiştir. Böylece mevcut veri kaynağına göre, coęrafik konumlarından bağımsız olarak ölkelerin COVID-19 verilerine göre kümeleme yapılarak daha iyi tanımlanabilme-

si ve sınıflandırılabilmesi sağlanmıştır. Yukarıda da ifade edildięi gibi kümeleme analizde vaka ve vefat sayılarının ölkeler nüfusuna göre payı ve artış göstergeleri hesaplanarak verinin güçlendirilmesi sağlanmıştır. Kümeleme analizinin sonucunda; küme içindeki ölkelerin belirlenmesinin yanısıra ölkelerin ilgili küme merkezine (o kümenin ortalama deęerine) uzaklıkları sıralanarak aynı küme içindeki yakın benzerlikler de ortaya çıkarılmıştır. Tarih seçimine göre küme merkezi olarak kabul edilen ölkeler ve kümelerdeki ölkelerin sıralanması da deęişebilmektedir. Uygulamanın dinamik web sitesinden bu durumun izlenmesi ve bu doęrultuda bir karar destek sistemi olarak kullanılabilmesi mümkün olabilmektedir. Çalışmada, 209 ölkelerin küresel boyutta deęerlendirilmesi yapılmış, ayrıca Türkiye’nin ölkeler bazında durumu ve deęişimi incelenmiştir; yine geliştirilen uygulama sayesinde her bir ölkeler için de benzer küresel deęerlendirmelerin yapılabileceęi unutulmamalıdır.

Çalışma sonucunda elde edilen önemli bulgulardan biri de vaka ve vefat sayıları yüksek olan ilk 10 ölkelerde de salgının artış eğiliminin azaldığıdır. Özellikle İspanya, İtalya, Fransa, Almanya ve Çin’de vaka ve vefat sayılarının yüksek olmasına karşın, artış hızları sıfıra yaklaşmaktadır. Bu durumu etkileyen koruyucu tedbirlerin ve tedavi yöntemlerinin incelenmesinin yanısıra tıbbi yönden araştırılması da önem kazanmaktadır.

SONUÇ

Bu çalışma ve geliştirilen uygulama (<https://elifkartal.shinyapps.io/covid19/>) ile; küresel COVID-19 salgınının dünyada ve Türkiye’de deęişen durumuna baęlı olarak ölkelere ait güncel COVID-19 verisine dayalı özet durum ve analiz sonuçlarına dinamik yapı kazandırılmış, böylelikle hızlı ve proaktif kararlara destek verilebilmesi sağlanmıştır. Uygulamada yer alan ölkelerin gruplandırılması özellięi sayesinde, özellikle COVID-19 ile ilgili gün gün dinamik deęişimin ölkeler grupları üzerinden izlenmesi ve salgınla ilgili alınacak kararlarda aynı grupta yer alan ölkelerin izledięi stratejiler dikkate alınarak salgına karşı etkili ve proaktif bir yaklaşım sergilenebilir. Bu çalışmanın uzantısında COVID-19 hasta verilerine dayalı olarak kümeleme, benzerlik analizleri ile virüsün taşıdığı ortak özelliklerin belirleneceęi çalışmalar yapılmalıdır. Ayrıca takip edilen hasta verisi üzerinde yapay zekâ uzantısında farklı veri madencilięi ve makine öğrenmesi teknikleri kullanılmalıdır. Özellikle, çok disiplinli bir çalışma anlayışıyla COVID-19 tahmin modellerinin geliştirilmesi ve hastanelerin kullanımına sunulması önemlidir.

Hakem Deęerlendirmesi: Dış bağımsız.

Yazar Katkıları: Çalışma Konsepti/Tasarım- M.E.B., E.K.; Veri Toplama- M.E.B., E.K.; Veri Analizi/Yorumlama- M.E.B., E.K., B.B.;

Yazı Taslağı- M.E.B., E.K.; İçeriğin Eleştirel İncelemesi- M.E.B., E.K., B.B.; Son Onay ve Sorumluluk- M.E.B., E.K., B.B.; Malzeme ve Teknik Destek- M.E.B., E.K.; Süpervizyon- E.K.

Çıkar Çatışması: Yazarlar çıkar çatışması beyan etmemişlerdir.

Finansal Destek: Yazarlar finansal destek beyan etmemişlerdir.

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- M.E.B., E.K.; Data Acquisition- M.E.B., E.K.; Data Analysis/Interpretation- M.E.B., E.K., B.B.; Drafting Manuscript- M.E.B., E.K.; Critical Revision of Manuscript- M.E.B., E.K., B.B.; Final Approval and Accountability- M.E.B., E.K., B.B.; Technical or Material Support- M.E.B., E.K.; Supervision- E.K.

Conflict of Interest: Authors declared no conflict of interest.

Financial Disclosure: Authors declared no financial support.

KAYNAKLAR/REFERENCES

1. T.C Sağlık Bakanlığı. Yeni Korona Virüsü. 2020 [cited 2020 Jun 8]. Available from: <https://covid19.saglik.gov.tr/>
2. Koca F. "Tedbirlere Uyum; Kendimize, Devletimize ve Milletimize Karşı Bir Ödevdir". T.C. Sağlık Bakanlığı. 2020 [cited 2020 Jun 7]. <https://www.saglik.gov.tr/TR,65876/tedbirlere-uyum-kendimize-devletimize-ve-milletimize-karsi-bir-odevdir.html>
3. Özkan Y. Veri madenciliği yöntemleri. Papatya Yayıncılık Eğitim; 2008. 214 p.
4. Silaharoğlu G. Kavram ve Algoritmalarıyla Temel Veri Madenciliği. İstanbul: Papatya Yayıncılık Eğitim; 2008.
5. Şentürk A. Veri Madenciliği Kavram ve Teknikler. Bursa: Ekin Yayınevi; 2006.
6. Balaban ME, Kartal E. Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili ile Uygulamaları. 2nd ed. Beyoğlu, İstanbul: Çağlayan Kitabevi; 2018.
7. Shearer C. The CRISP-DM model: the new blueprint for data mining. Journal of data warehousing. 2000;5(4):13-22.
8. European Centre for Disease Prevention and Control. Download today's data on the geographic distribution of COVID-19 cases worldwide. 2020. <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>
9. Balaban ME. Temel Matematik ve İşletme Uygulamaları. 4th ed. İstanbul: Türkmen Kitabevi; 2010.
10. Han J, Kamber M. Data Mining: Concepts and Techniques. 2nd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers; 2006.
11. Harrington P. Machine Learning in Action. 1st ed. Shelter Island, NY: Manning Publications Co.; 2012. 384 p.
12. Koçoğlu FÖ, Özcan T. Veri Madenciliğinde Kümeleme Algoritmaları ile Müşteri Segmentasyonu. In: Balaban ME, Kartal E, editors. R ile Veri Madenciliği Uygulamaları. 1st ed. İstanbul: Çağlayan Kitabevi; 2016. p. 187-220.
13. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. Vol. 344. John Wiley & Sons; 2009.
14. Zoubi MBA-, Rawi M al. An efficient approach for computing silhouette coefficients. Journal of Computer Science 2008;4(3):252-5. [CrossRef]
15. RStudio. RStudio | Open source & professional software for data science teams. 2020 [cited 2020 Jun 8]. <https://rstudio.com/>
16. The R Foundation. R: What is R? 2020 [cited 2020 Jun 8]. <https://www.r-project.org/about.html>
17. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: Cluster Analysis Basics and Extensions. 2019.
18. Walesiak M, Dudek A. clusterSim: Searching for Optimal Clustering Procedure for a Data Set. 2019. <https://CRAN.R-project.org/package=clusterSim>
19. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. 2020. <https://CRAN.R-project.org/package=dplyr>
20. Xie Y, Cheng J, Tan X. DT: A Wrapper of the JavaScript Library "DataTables". 2020. <https://CRAN.R-project.org/package=DT>
21. Ren K, Russell K. formattable: Create "Formattable" Data Structures. 2016. <https://CRAN.R-project.org/package=formattable>
22. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. <https://ggplot2.tidyverse.org> [CrossRef]
23. Gesmann M, Castillo D de. googleVis: Interface between R and the Google Visualisation API. The R Journal 2011;3(2):40-4. [CrossRef]
24. Wickham H. The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software 2011;40(1):1-29. [CrossRef]
25. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. shiny: Web Application Framework for R. 2020. <https://CRAN.R-project.org/package=shiny>
26. Attali D. shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds. 2020. <https://CRAN.R-project.org/package=shinyjs>
27. Kartal E, Balaban ME. Küresel COVID-19 Salgınının Dünyada ve Türkiye'de Değişen Durumu ve Kümeleme Analizi. 2020. <https://elifkartal.shinyapps.io/covid19/>
28. RStudio. Shiny. 2020 [cited 2020 Jun 8]. <https://shiny.rstudio.com/>
29. RStudio. shinyapps.io. 2020 [cited 2020 Jun 8]. <https://rstudio.com/products/shinyapps/>
30. Johns Hopkins University COVID-19 dashboard. Coronavirus COVID-19 (2019-nCoV). 2020 [cited 2020 Jun 4]. <https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>
31. World Health Organization (WHO). WHO Coronavirus Disease (COVID-19) Dashboard. 2020 [cited 2020 Jun 4]. <https://covid19.who.int/>
32. Paker E, Leclerc Q. COVID-19 tracker [Internet]. COVID-19 tracker. 2020 [cited 2020 Jun 4]. https://vac-lshtm.shinyapps.io/ncov_tracker/?_ga=2.127325129.557643434.1591276997-241478561.1591276997
33. Schönbrodt F, Stefan A, Zumbstein P, Bloom PA. Experience Statistics. 2020 [cited 2020 Jun 4]. <http://shinyapps.org/apps/corona/>
34. Thams NT, Jakobsen ME, Mogensen PB, All PhD-students in Statistics at the University of Copenhagen. COVID19. 2020 [cited 2020 Jun 4]. <http://shiny.science.ku.dk/pbm/COVID19/>

35. Zarikas V, Pouloupoulos SG, Gareiou Z, Zervas E. Clustering analysis of countries using the COVID-19 cases dataset. Data in Brief 2020;31:105787. [CrossRef]
36. Zoumpakas T. COVID-19 Cluster Analysis. Medium. 2020 [cited 2020 Jun 11]. <https://towardsdatascience.com/covid-19-cluster-analysis-405ebbd10049>