

COVID-19'un Ülke Bazlı Analizi ve Veri Görselleştirmesi

Abdullah Erdal TÜMER^{1*}  Musab Kasım DOĞAN² 

¹ Necmettin Erbakan Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Konya, Türkiye

² Yıldırım Beyazıt Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Ankara, Türkiye

Makale Bilgisi

ÖZET

Geliş Tarihi: 15.03.2024
Kabul Tarihi: 07.07.2024
Yayın Tarihi: 31.12.2024

Anahtar Kelimeler:

COVID-19,
Koronavirüs,
Veri Görselleştirme,
Elasticsearch,
Logstash,
Kibana.

İnsan yaşamının her alanını etkileyen COVID-19 salgını, milyonlarca ölüme sonuçlanmıştır. Günümüzde adı tam olarak konulmamakla ve ölümcül olmamakla beraber COVID-19 günümüzde influenza olarak seyretmektedir. COVID-19 ve benzeri hastalıkların önlenmesinde hastalıklara ait verilerin görselleştirilmesi önemlidir. Özellikle hükümetlerin, iş yerlerine ve kurumlara güvenilir, anlaşılır ve kolay aktarılabilen bilgiler sunması, ilgili hastalıklarla mücadelede noktasında farkındalık sağlamaktadır. Bu çalışmanın amacı, veri görselleştirme yöntemi ile farklı kıta ve ülkelerdeki COVID-19'un etkisini karşılaştırmaktır. Doğrulanmış COVID-19 vakaları hakkındaki bilgilerin farkındalığı artıracağı düşünülmektedir. Bu salgını görselleştirmek için açık kaynaklı bir yazılım olan Elasticsearch kullanılmıştır. Veriler ABD, Çin, Türkiye gibi farklı ülkelerden elde edilmiştir. Veri başlıkları, doğrulanmış COVID-19 vaka sayısı, toplam ölüm sayısı ve kurtarılan toplam vaka sayısı olarak belirlenmiştir. COVID-19'un etkisi hakkında kapsamlı bir anlayış ortaya konulmaya çalışılmıştır. Ayrıca bu çalışma ile gerçek hayattaki birçok farklı uygulama ve hizmetten alınan büyük verilerin de görselleştirilmesine ışık tutacağı değerlendirilmiştir.

Country-based Analysis and Data Visualization of COVID-19

Article Info

ABSTRACT

Received: 15.03.2024
Accepted: 07.07.2024
Published: 31.12.2024

Keywords:

COVID-19,
Coronavirus,
Data visualization,
Elasticsearch,
Logstash,
Kibana.

The COVID-19 epidemic, which affects every aspect of human life, has resulted in millions of deaths. Today, COVID-19 is known as influenza, although its name is not fully known, and it is not fatal. Visualization of disease data is important in preventing COVID-19 and similar diseases. Governments provide reliable, understandable and easily transferable information to workplaces and institutions provide awareness in the fight against relevant diseases. The aim of this study is to compare the impact of COVID-19 in different continents and countries with the data visualization method. It is thought that information about confirmed COVID-19 cases will increase awareness. Elasticsearch, an open-source software, was used to visualize this outbreak. Data were obtained from different countries such as the USA, China and Türkiye. Data headlines are the number of confirmed COVID-19 cases, the total number of deaths, and the total number of cases recovered. An attempt has been made to provide a comprehensive understanding of the impact of COVID-19. In addition, it was evaluated that this study would shed light on the visualization of big data received from many different applications and services in real life.

To cite this article:

Tümer, A.E. & Doğan, M.K. (2024). COVID-19'un ülke bazlı analizi ve veri görselleştirmesi. *Necmettin Erbakan Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 6(3), 512-524. <https://doi.org/10.47112/neufmbd.2024.62>

*Sorumlu Yazar: Abdullah Erdal Tümer, tumer@erbakan.edu.tr



GİRİŞ (INTRODUCTION)

Sadece kelimeler yerine görsel araçlar sunulduğunda bireyler bunlara çok daha iyi tepki verir. Beyne iletilen bilgilerin %90'nının görsel olması ve görüntülerin beyinde metin hızının 60.000 katı hızda işlenmesi bilinmektedir [1]. İnsanlar sık sık resim ve video izlemektedir. İnsan zihni kelimelerden çok bunlarla meşguldür. Bu nedenle ister bir makale yazarken ister bir dinleyiciye sunum yaparken beyin görsel bilgiyi metinsel bilgiye göre daha kolay işler. Bilginin daha hızlı aktarılması kelimelerle değil görsellerle anlatılarak sağlanabilir. Bu örnek, veri görselleştirmenin değerini veya metinsel bilgiyi görsel bir temsile dönüştürme sürecini vurgulamaktadır. Bu yüzden bu dönemi "büyük veri çağı" olarak adlandırmak yanlış olmayacaktır. Teknolojideki ilerlemelerle birlikte, çok büyük miktarda veri kolayca üretilmekte ve çok çeşitli zengin veri kaynaklarından hızlı bir şekilde toplanabilmektedir [2]. Biyolojik veriler [3-5], ekonomik veriler [6-8], sosyal ağ verileri [9,10] ve salgın verileri gibi hastalık raporları [11-13] büyük veriye örnek olarak verilebilir.

Akut solunum sendromu (SARS), Orta Doğu solunum sendromu (MERS) gibi salgın hastalıklardan biri olan koronavirüs pandemisi (COVID-19) insan sağlığını, yaşamını, üretimini, sosyal etkileşimlerini ve uluslararası ilişkilerini ciddi şekilde tehdit etmiştir. COVID-19 virüs pandemisi 2019 yılının sonlarında Çin'in başkentlerinden olan Wuhan'da ortaya çıkmıştır [14]. Hızlı bir bulaşma özelliğine sahip olan bu virüs 2020 Nisan ayı sonlarında başta Avrupa, Kuzey ve Güney Amerika, Doğu Akdeniz, Batı Pasifik, Güney ve Doğu Asya ve Afrika olmak üzere 6 bölgeye yayılmıştır [14,15]. Bu durum 11 Mart 2020'de Dünya Sağlık Örgütü tarafından küresel salgın olarak ilan edilmesine neden olmuştur. 4 Eylül 2020 itibarıyla dünyada 26.191.461 vaka, 17.726.336 iyileşme, 865.558 ölüm rapor edilirken, Türkiye'de 276.555 vaka, 249.108 iyileşme, 6.564 ölüm rapor edilmiştir [15]. Büyük veri teknolojileri salgınla mücadelede önemli bir rol oynamıştır. Bu epidemiyolojik verilerden keşfedilen bilgiler, araştırmacıların, epidemiyologların ve politika yapıcıların hastalığı daha iyi anlamalarına yardımcı olmuştur [16].

Hızlı büyük veri toplama, pandemik bilgi görselleştirilmesi, salgın risk dökümü, doğrulanmış vaka tespiti, önleme düzeyinin izlenmesi ve uygun kontrol değerlendirmesi, uygulamalara birkaç örnek olarak verilebilir. [17]. Büyük veri görselleştirilmesi, sağlık hizmetlerinin erken müdahale, önleme ve optimal yönetime yönlendirmek için sağlık hizmetlerini daha sürdürülebilir ve verimli hale getirmesi mümkündür [18].

COVID-19'dan dolayı birçok araştırmacı, hastalığın farklı yönlerine odaklanmıştır. Bunlar arasında klinik ve tedavi bilgileri [19, 20] ile ilaç keşfi [21], tıp ve sağlık bilimleri araştırmalarıyla ilgilidir [2]. Buna karşılık, bilgisayar bilimcilerde genellikle hastalıkla ilgili tahmin çalışmaları yaptılar [22-26]. Bu çalışmanın amacı hastalığın görselleştirilerek; salgınla ilgili bilgi ve farkındalığın artırılması, epidemiyolojik takibin sağlanması, konu ile ilgili politika geliştirilmesine katkı verilmesi ve aşı dağıtımı vb. halk sağlığı müdahalelerinin kolaylaştırılması amaçlanmaktadır.

Veri görselleştirme sürecinin en önemli bileşeni, görselleştirilecek doğru bilgi, yani ham veridir. Ancak, verilerin tek başına bir anlamı ve işlevi yoktur. Veri, toplandıktan, gruplandırıldıktan ve özetlendikten sonra işlenip bilgiye dönüştürüldüğünde anlam kazanır. Ait olduğu bağlamı açıklama gücü kazanır. O zaman problem çözme veya karar verme gibi bir amaca hizmet edebilir [18].

Karmaşık bilginin analiz edilmesi ve anlaşılabilmesi için görselleştirmelerin kullanımı günümüz iletişim ortamında oldukça işlevsel bir yöntemdir. Veri görselleştirme, araştırmacıların sonuçları daha net ve zarif bir şekilde keşfetmesini, yorumlamasını ve sunmasını sağlayabilir [27-29]

Görselleştirme, kompleks verilerin çok daha hızlı bir şekilde analiz edilmesini ve geniş kitleler tarafından anlaşılabilir olması sağlamaktadır. Görselleştirme, grafik sunumları kullanarak bilginin iletişimi şeklinde tanımlanabilir. Tek bir görüntü yoğun bilgi içerebilmekte ve sayfalar dolusu kelimedenden

çok daha hızlı işlenebilmektedir. Çünkü imaj yorumlama, insanın algısal sistemine paralel olarak işlerken, metin analiz hızı ardışık okuma işlemi tarafından sınırlanmaktadır [30]. Ayrıca insanın görsel öğrenme becerisi ve görsel algı süreçlerinin büyük bir güce sahip olması, görselleştirmenin hafıza, öğrenme ve dikkat öncesi süreçlerde çok etkili bir rol oynamasını sağlamaktadır [31].

Bu makaledeki temel katkı, COVID-19 epidemiyolojik verileri için bir görselleştirme, görsel analiz aracı tasarlanması ve geliştirmesidir. Analiz aracı (a) veri analitiğini (özellikle sık örüntü madenciliği), (b) veri görselleştirmeyi ve (c) görsel analitiği (özellikle sık örüntülerin görselleştirilmesi ve analizini) içerir. Çalışmada geliştirilen araç her ne kadar COVID-19 epidemiyolojik verilerinin görselleştirilmesi ve görsel analitiği içerse de diğer birçok gerçek hayattaki uygulama ve hizmetten alınan büyük verilerin görselleştirilmesi ve görsel analitiği için de uygulanabilir.

Makalenin geri kalanı şu şekilde organize edilmiştir: Bir sonraki bölümde ilgili çalışmalara yer verilmiştir. Bir sonraki bölümde takip edilen araştırma yöntemi ortaya konulmuştur. Sonra sırasıyla araştırma bulguları ve sonuç bölümü sunularak makale sona ermektedir.

İLGİLİ ÇALIŞMALAR (RELATED WORKS)

Bu bölüm, 2019-2022 yıllarında yapılan çeşitli çalışmalara kısa bir genel bakış sunmaktadır. Yapılan çalışmaların bir kısmı COVID-19'un ekonomik etkilerini görselleştirmiştir [32]. Bazıları [33-36] COVID-19 araştırmalarını görselleştirdi. Burada COVID-19 araştırmalarına ait görselleştirme çalışmalarından bahsedilecektir.

Yazarlar Kaggle'daki COVID-19 verilerini kullanmışlardır. Çalışmada sıcaklığın COVID-19'un yayılması üzerindeki etkisi araştırılmıştır. Ayrıca, erkeklerin bu hastalığa daha yatkın olduğunu ve yaşlıların daha fazla risk altında olduğunu gösteren erkek/kadın enfeksiyon karşılaştırmasına ilişkin veri görselleştirmesi sağlanmaktadır. Verilere dayanarak, doğrulanmış vakaların artışındaki modelin doğası gereği üstel bir eğri olduğu bulunmuştur [37].

Johns Hopkins Üniversitesindeki araştırmacılar COVID-19'u gerçek zamanlı izlemek için web tabanlı online bir dashboard (pano) geliştirdiler. Herkese açık olarak paylaşılan panoda, COVID-19'dan etkilenen tüm ülkeler için doğrulanmış vakalarının, ölümlerin ve iyileşenlerin yerini ve sayısını göstermektedir [38].

Yazarlar çalışmalarında, COVID-19 epidemiyolojik verilerini görselleştirmek ve analiz etmek için bir büyük veri görselleştirme ve görsel analiz aracı sunmuşlardır. Aracın kullanıcıların doğrulanmış COVID-19 vakaları hakkındaki bilgileri daha iyi anlamalarına yardımcı olacağı ifade edilmiştir [2].

Dünya Sağlık Örgütü, Çin Hastalık Kontrol ve Önleme Merkezi, Ulusal Sağlık tarafından sağlanan COVID-19 verileri ile salgın bilgilerini derleme ve analiz etmeye çalışmışlardır [39].

Kanada'da araştırmacılar COVID-19'a ait demografik özellikler, konum, rapor tarihi, seyahat geçmişi ve maruz kalma kaynağı dahil olmak üzere doğrulanmış vakalarından oluşan bireysel düzeyde bir veri seti geliştirilmişlerdir. Bu veri seti görselleştirilerek kullanıcıların salgını takip etmesini sağlayan ve günlük olarak güncellenen bir web sayfasında yayınlanmıştır [40].

CoronaTracker adında bir topluluk tarafından bir model geliştirilmiştir. Model COVID-19 vakalarını, ölümlerini ve iyileşmelerini tahmin etmeyi amaçlamaktadır. Model, ilgili sağlık bilgilerinin yayılmasına ilişkin kamuoyu duyarlılığı kalıplarının yorumlanmasına ve virüsün yayılmasının siyasi ve ekonomik etkisinin değerlendirilmesine yardımcı olacağı ifade edilmiştir [41].

Yazarlar, COVID-19'un finansal olarak Türkiye inşaat sektörüne etkilerinin neler olduğunu araştırmışlardır. T.C. Sağlık Bakanlığı verileri kullanılmıştır. Verilerde COVID-19'dan enfekte olan ve ölenlerin sayısından yararlanılmıştır. Ayrıca inşaat sektörü fiyat ve maliyet endeksine ilişkin verilerde

Türkiye İstatistik Kurumundan elde edilmiştir. Vektör Otoregresyon modeli Zaman serisi ve GRANGER nedensellik analizi yapılmıştır. COVID-19'dan hastalanan ve ölen kişi sayısındaki değişim inşaat ve bina satış fiyatlarının farklılaşmasına neden olduğu bulunmuştur [42].

Literatürden de anlaşılacağı üzere çok sayıda çalışma farklı veri tabanlarından doğrulanmış verileri görselleştirmeye çalışmışlardır. Görselleştirme çalışmaları artırılmış gerçeklik [43] ve makine öğrenimi [44] gibi farklı alanlara da uygulanabilmektedir. Bu çalışma literatürdeki çalışmalardan farklı olarak COVID-19 verilerini Elasticsearch yöntemi ile bir görselleştirme sunulmuştur.

MATERYAL VE METOT (MATERIALS AND METHODS)

Bu çalışma 17 Kasım 2019- 9 Nisan 2020 arası teyit edilmiş [35] deki verilerden oluşmaktadır. COVID-19 interaktif dünya haritası panolarını oluşturmak için açık kaynak kodlu yazılımlar ve veri madenciliği teknikleri kullanılmıştır. Panolar koronavirüsteği değişimi anlık olarak izlemek için tasarlanmıştır. Bu sistem dört farklı ekrandan izlenebilmektedir.

- İlk ekran dünyanın genel durumu, ölüm oranları ve iyileşme oranları, ölüm sayısına göre boyut olarak değişen ülke isimleri, veri bilgilerine göre renklendirme gibi özelliklere sahiptir.

- İkinci ekran, dünya göre bugünün (GMT + 0) sayısal verileri: ne kadar değiştiği ne kadar arttığı, daha net hale getirilmesi için ülke bayrakları ile birlikte gösterilmektedir. Herhangi bir ülkenin bayrağına tıklandığında, ilgili ülke filtrelenmiş ve dashboardında yalnızca o ülkeye ait vaka sayısı gösterilecek şekilde ayarlanmıştır. Ayrıca ölüm sayısına göre yeşilden kırmızıya değişen renkler dünya haritasında gösterilmiştir. Bir bakışta en çok vakanın nerede olduğunu, ilginç ve anlaşılır kılan bu renklendirme sistemi, üçüncü ekranda farklı bir şekilde yeniden kullanılmıştır.

- Üçüncü ekranda, dünya haritasında ülke sınırları çizilmiş etkileşimli bir gösterge panosu sunulmaktadır. Bu interaktif harita üzerinde bir ülke ezdirildiğinde, o ülkeye ait vaka sayısı, iyileşme sayıları, ölümler gibi bilgilere bir "pop-up-açılır pencere" yardımıyla anında erişilebilmektedir.

- Dördüncü ve son ekran, "son güncelleme" ekranıdır. Tüm bu panoların oluşturulmasında güncel ham veriler kullanılmıştır.

Bu araştırmadaki görselleştirme aşağıdaki çalışma sorularına cevap verecek şekilde tasarlanmıştır:

- 17 Kasım 2019'dan 9 Nisan 2020 ye kadar kaç ülke enfekte oldu?
- Ölüm sayısı ülke bazında karşılaştırılıysaydı, neye benzerdi?
- Renkli bir dünya haritasında durum nasıl görünürdü?
- 17 Kasım 2019'dan 9 Nisan 2020'ye kadar Türkiye'deki son durum nedir?

Ham Veri Toplama (Raw Data Collection)

Görselleştirmede kullanılan ve [35] den elde edilen verileri çekmek için Shell Script kullanılmıştır. Veri çekme işlemi beş aşamadan oluşmaktadır.

İlk aşamada sürekli güncellenen ve [35] de yer alan ve Şekil 1'de gösterilen tablodaki veriler düzenli olarak indirilmiştir.

İkinci aşamada web sayfasındaki boş alanlar temizlenmiştir.

Üçüncü aşamada, ham veriler html2.text komutu ile metin biçimine dönüştürülmüştür.

Dördüncü aşamada metin dosyası for döngüsü ile ülke isimleri tek tek kaldırılarak sadece ülke bazlı veriler toplanmıştır.

Beşinci aşamada, toplanan veriler .csv formatına dönüştürülmüştür. Son aşama ile veriler işlenmeye hazır hale getirilmiştir. Bu beş aşama detaylandırılırsa;

All	Europe	North America	Asia	South America	Africa	Oceania											
#	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	New Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Total Tests	Tests/ 1M pop	Population			
	World	690,954,065	+8	6,895,908	0	663,550,833	+22,498	20,507,324	37,314	88,643	884.7						
1	USA	107,310,869		1,168,148		105,463,247		679,474	782	320,517	3,489	1,180,798,099	3,526,821	334,805,269			
2	India	44,994,228		531,906		44,460,809		1,513	N/A	31,987	378	930,797,975	661,721	1,406,631,776			
3	France	40,138,560		167,642		39,944,661		26,257	869	612,013	2,556	271,490,188	4,139,547	65,584,518			
4	Germany	38,428,685		174,352		38,240,600		13,733	N/A	458,119	2,078	122,332,384	1,458,359	83,883,596			
5	Brazil	37,671,420		703,964		36,249,161		718,295	N/A	174,928	3,269	63,776,166	296,146	215,353,593			
6	Japan	33,803,572		74,694		N/A	N/A	N/A	83	269,169	595	100,414,883	799,578	125,584,838			
7	S. Korea	32,131,606		35,017		31,890,379	+20,729	206,210	117	625,982	682	15,804,065	307,892	51,329,899			
8	Italy	25,897,801		190,868		25,593,568		113,365	56	429,748	3,167	273,634,829	4,540,695	60,262,770			
9	UK	24,618,436		226,278		24,392,158		0	N/A	359,404	3,303	522,526,476	7,628,357	68,497,907			
10	Russia	22,959,198		399,563		22,402,524		157,111	N/A	157,464	2,740	273,400,000	1,875,095	145,805,947			
11	Turkey	17,232,066		102,174		N/A	N/A	N/A		201,399	1,194	162,743,369	1,902,052	85,561,976			
12	Spain	13,914,811		121,760		13,762,417		30,634	231	297,840	2,606	471,036,328	10,082,298	46,719,142			
13	Australia	11,700,577		21,817		11,668,002		10,758	41	448,835	837	81,916,639	3,142,326	26,068,792			

Şekil 1

Ülke veya bölgeye göre bildirilen vakalar ve ölümler [35]

Ham Veri Çıkarma (Raw Data Extraction)

Arayüz verileri, linux işletim sistemine yazılmış bir Shell betiği ile gerçek zamanlı olarak <https://www.worldometers.info/coronavirus/> dan indirilmiştir. Bu komut dosyası önce date komutunu alarak günlük bir dizin oluşturur. İkinci bir date komutuyla dakika alınacak dosyalar için mkdir komutuyla iç içe geçmiş bir dizini oluşturulur. Sonra sırasıyla; cd komutu ile dizine giderek wget komutunu çalıştırır, ilgili web sitesinin html'si indirilir. Bu süreçte eklenen veriler, birçok ilgili ilgisiz veri içermektedir. .html dosyasındaki fazlalıkları temizlemek için sed komutu kullanılmıştır. Daha sonra aşağıdaki komut satırı ile ham veriler elde edilmiştir:

```
cat coronavirus | grep "</ * td> "| html2text -nobs | awk '1; / Total / {exit}' | grep -v Total > raw-data.txt
```

Burada kullanılan html2text komutu, adından da anlaşılacağı üzere bir html sayfasındaki okunamayan bilgileri temizlemek ve bu dosyayı text formatına dönüştürmek için kullanılır. Ham verilerde ülke adı, vaka sayısı, ölü sayısı, iyileşen sayısı vb. bilgiler bulunmaktadır. Bu verileri ayrı ayrı tanımlayabilmek ve csv formatına getirebilmek için öncelikle içerdikleri ülke isimleri country_name.txt isimli bir text dosyasına yazdırılır. Daha sonra for döngüsü oluşturularak i=country_name ile aynı ham verileri içeren tek tek metin dosyasına grep \$i ile sorgulama yapılmıştır. Toplanan veriler csv formatında virgüllerle ayrılarak dakika bazlı bir dosyaya yazılarak son halini almıştır. Elasticsearch'e gönderilirken veri kirliliğini önlemek için veriler değişmediyse aynı verilerin atılmaması için çeşitli yöntemler denenmiştir. Şekil 2 de en kararlı komut olarak kullanılan “grep -Fxfv \$ old_file \$ new_file > coronavirus_daily.txt” komutunun çıktısı görülmektedir.

Bu komut yardımıyla eski dosya ile yeni dosya arasındaki fark direkt olarak coronavirus_daily.txt dosyasına yazıldı. İstenirse md5sum değeri oluşturularak iki md5sum arasındaki değer karşılaştırılarak buna göre işlem yapılabilir. Bu işlemle birlikte, coronavirus-raw-data.txt dosyası artık hazırdır ve Elasticsearch'e düzenli olarak gönderilebilir.

Ham Veri İşleme (Raw Data Processing)

Ham veriler hazır olduktan sonraki adım, verileri Elasticsearch'e göndermektir. Bunun için önceden Elasticsearch'e veri türünün tanıtılması gerekir. Bu projede, düzenli olarak veri göndermek ve

```
Solomon Islands,17,0,0,0,10,7,4500,696503
Anguilla,16,0,0,0,15,1,7326,15080
Montserrat,13,0,1,0,12,0,626,4994
Western Sahara,10,0,1,0,8,1,0,605725
Wallis and Futuna,5,0,0,0,4,1,1202,11124
Marshall Islands,4,0,0,0,4,0,0,59421
Samoa,2,0,0,0,2,0,0,199176
Micronesia,1,0,0,0,1,0,0,115718
Vanuatu,1,0,0,0,1,0,470,311247
China,89378,52,4636,0,82940,1802,160000000,1439323776
Sierra Leone,3282,0,77,0,2260,945,106949,8069879
Poland,1496665,0,36443,0,1257352,202870,8517260,37822558
```

Şekil 2

Coronavirüs csv biçimi ham verileri

veri tipini belirlemek için Java tabanlı açık kaynaklı bir yazılım olan Logstash kullanılmıştır. Yapılandırma dosyasında üç parametre vardır. Bunlar girdi, filtre ve çıktıdır. Giriş kısmı, oluşturulan coronavirus_daily.txt dosyası ile dosyanın bulunduğu dizine girilmiştir. Yakalanan veriyi tekrar almamak için tasarlanmış olan sincedb_path bilgisi bu bölümde yer almaktadır. Şekil 3'te Filter kısmı ile verilerin string, integer, boolean vb. olarak tanımlanıp işlenebildiği kısım gösterilmiştir.

```
if "coronavirus" in [path] {
  csv {
    separator => ",",
    columns => ["country", "totalcases", "newcases", "totaldeaths", "newdeaths", "totalrecovered", "activecases", "totaltests", "population"]
  }
  convert => {
    "totalcases" => "integer"
    "newcases" => "integer"
    "totaldeaths" => "integer"
    "newdeaths" => "integer"
    "totalrecovered" => "integer"
    "activecases" => "integer"
    "totaltests" => "integer"
    "population" => "integer"
  }
}
```

Şekil 3

Ham veri işleme için Logstash yapılandırma dosyası

Ardından, gelecekteki verilerin sütun adları yazılır. Daha sonra kolona gelecek ifadelerin veri tipi belirlendi. Veri görselleştirmede kullanılacak verilerin eşleştirilmesi için translate komutu ile belirli ülke adlarının (Amerika Birleşik Devletleri => ABD gibi) otomatik olarak değiştirilmesi sağlanmıştır. Ülke bayraklarını kullanabilmek için sütunlardan biri olan ülke sütunu kopyalanarak country-flag komutu oluşturuldu. Küçük harf komutuyla tüm ülke bayrağı adları küçük harfe çevrilir. Verilerin alındığı internet sitesinde mortalite, iyileşme oranı, aktif vaka sayısı gibi bilgiler bulunmadığı için ruby kodu kullanıldı. Bahsedilen kod "ruby {code =>} event.set ('deathrate', event.get ('totaldeaths') * 100 / event.get ('totalcases'))" şeklindedir.

Ham Veri İletimi (Raw Data Transmission)

Çıktı bölümünde verinin alındıktan ve işlendikten sonra nereye yazılacağı bilgisi yer almaktadır. Logstash; Elasticsearch, kafka, file, java_stdout, mongoddb, rabbitmq, zabbix, email ve benzeri birçok yere çıktı verebilir. Ülke bazlı unique_id oluşturularak yeni gelen veriler eski verilere güncellenmiştir. Daha sonra çıktı bölümüne Elasticsearch'ün Ip:port ve user:pass bilgileri girilmiştir. Gerekli güvenlik duvarı erişimi sağlandıktan sonra veriler Elasticsearch'e gönderilmiştir. Şekil 3'te veriler görselleştirilmeden önce coronavirüs verileri indekslenmiştir.

Veri Depolama (Data Storage)

Elasticsearch'e ham veri gönderilmeden önce şablon kısmı ayarlanmalıdır. Bunun nedeni, index oluşturulduğunda gelecek verilerin indexleme, adlandırma, shard ve replika yapısının değiştirilememesidir. İlk olarak şablon bölümünde oluşturulacak indeksin adı corona* formatında

hr_@timestamp	country	country-flag	totalcases	newcases	totaldeaths	newdeaths	activecases	totalrecovered	totaltests	population
Jan 29, 2021 @ 12:06:16.620	United States		26,338,497	+0	140,726	+0	382,471	16,076,326	386,452,138	332,124,463
Jan 29, 2021 @ 12:06:16.620	India		10,720,971	+0	15,851	+0	72,572	16,836,525	195,981,079	1,387,827,850
Jan 29, 2021 @ 12:06:16.620	Brazil		9,060,786	+0	221,671	+0	313,316	33,577,000	286,000,000	213,432,842
Jan 29, 2021 @ 12:06:16.620	Russia		3,813,848	+19,238	7,105	+534	385,400	16,525,504	101,100,000	145,970,742
Jan 29, 2021 @ 12:06:16.620	United Kingdom		3,743,734	+0	103,101	+0	196,672	10,799,000	788,724,021	68,892,198
Jan 29, 2021 @ 12:06:16.620	France		3,130,629	+0	74,000	+0	283,259	2,265,700	429,739,978	65,357,167
Jan 29, 2021 @ 12:06:16.620	Spain		2,792,360	+0	97,800	+0	1	1	320,277,228	46,765,278
Jan 29, 2021 @ 12:06:16.620	Italy		2,515,507	+0	97,800	+0	37,461	1,053,500	318,862,021	60,410,226
Jan 29, 2021 @ 12:06:16.620	Turkey		2,457,118	+0	25,800	+0	31,200	1,400,000	291,811,225	84,862,817

Şekil 4

Görselleştirilmeden önce indekslenmeden koronavirus verileri

belirtilir. Daha sonra shard ve replika sayısı belirlenir ve şablon tamamlanır. Bu çalışmada 2 düğüm elastik arama kümesi bulunmaktadır. Hızlı arama ve fazlalık için her dizin 1 parça ve 1 kopya olarak ayarlanmıştır. Şekil 4 de görüleceği üzere endeksler günlük olarak oluşturulmuştur. Bu, aylık yıllık grafiği oluşturmak ve analiz etmek için yapılmıştır. Böylece herhangi bir hatalı indeksleme durumunda silme işlemi kolaylıkla yapılabilmektedir. Dizinde yer alan veriler düzenli olarak güncellendiği için dizinde herhangi bir veri birikimi olmamaktadır. Gerekirse yaşam döngüsü eklenebilir. Bu yaşam döngüsü süreci sayesinde indeks, 12 ay gibi istenilen süre kadar saklandıktan sonra sistem tarafından otomatik olarak silinebilmektedir.

Veri Görselleştirme (Data Visualization)

Veriler Elasticsearch'e düzgün bir şekilde yüklendikten sonra, görselleştirme işlemi artık başlatılabilir.

Bunun için kullanılabilecek Kibana ve Grafana gibi açık kaynaklı analitik ve etkileşimli görselleştirme web uygulamaları vardır. Bu çalışmada hızlı ve kullanışlı bir arayüz sunduğu için ücretsiz olan Elasticsearch ürünü Kibana tercih edilmiştir. Kibana, Elasticsearch için açık kaynaklı bir veri görselleştirme panosudur. Bir Elasticsearch kümesinde indekslenen içeriğin üzerinde görselleştirme yetenekleri sağlamaktadır. Kullanıcılar, büyük hacimli verilerin üzerinde çubuk, çizgi ve dağılım grafikleri veya pasta grafikler ve haritalar oluşturabilir.

Index Management

[Index Management docs](#)

[Indices](#) [Data Streams](#) [Index Templates](#) [Component Templates](#)

Update your Elasticsearch indices individually or in bulk. [Learn more.](#)

Include rollout indices

Include hidden indices

Lifecycle status

Lifecycle phase

<input type="checkbox"/> Name ↓	Health	Status	Primaries	Replicas	Docs count	Storage size	Data stream
<input type="checkbox"/> coronavirus-2021-01-29	● green	open	1	1	218	284.2kb	
<input type="checkbox"/> coronavirus-2020-12-23	● green	open	1	1	220	356.8kb	
<input type="checkbox"/> coronavirus-2020-12-22	● green	open	1	1	220	571.3kb	
<input type="checkbox"/> coronavirus-2020-12-21	● green	open	1	1	220	609.8kb	
<input type="checkbox"/> coronavirus-2020-12-20	● green	open	1	1	220	346.6kb	
<input type="checkbox"/> coronavirus-2020-12-19	● green	open	1	1	220	282.4kb	
<input type="checkbox"/> coronavirus-2020-12-18	● green	open	1	1	220	417kb	
<input type="checkbox"/> coronavirus-2020-12-17	● green	open	1	1	220	609.9kb	
<input type="checkbox"/> coronavirus-2020-12-16	● green	open	1	1	220	287.6kb	
<input type="checkbox"/> coronavirus-2020-12-15	● green	open	1	1	220	619.6kb	

Rows per page: 10

< 1 2 3 4 5 ... 25 >

Şekil 5

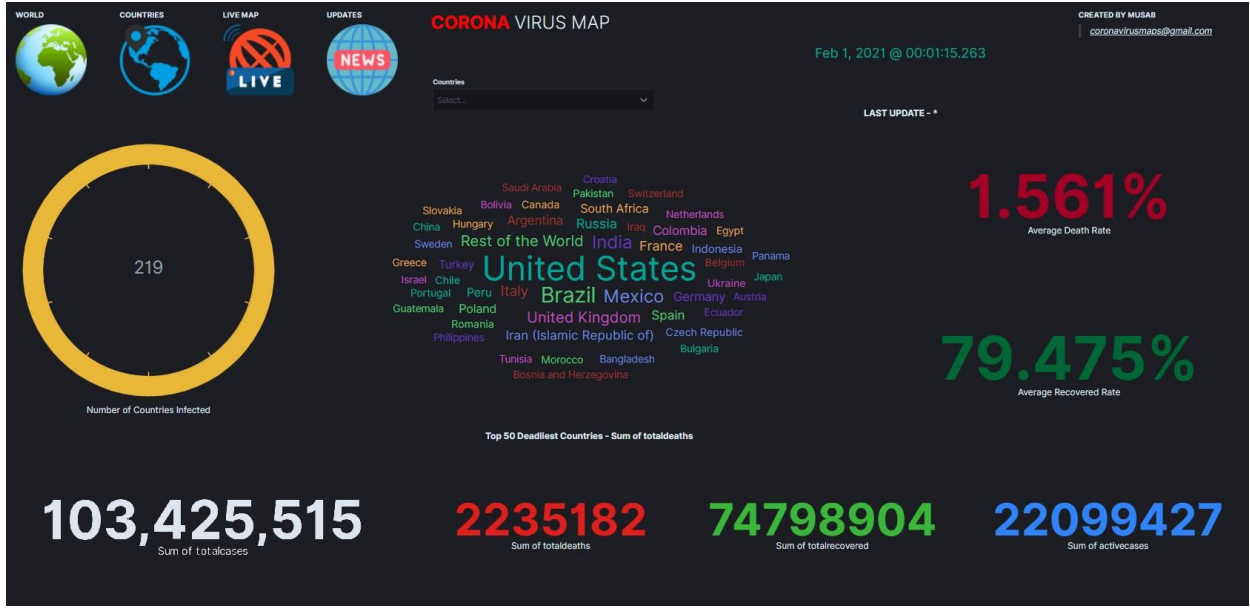
Koronavirus interaktif dünya haritası

BULGULAR (RESULTS)

Bu bölümde COVID-19'un başladığı 17 Kasım 2019- 9 Nisan 2020 tarihine kadar enfekte olan ülke sayıları, ülke bazlı ölümler ve renklendirilmesi ve Türkiye'deki durumun görsel sonuçları yer almaktadır.

9 Nisan 2020 de kaç ülke enfekte oldu? (How many countries are infected now?)

COVID-19 salgınında dünya çapında 9 Nisan 2020'de 219 ülkenin enfekte olduğu tespit edilmiştir. Enfekte veriler, benzersiz ülke adı bilgileri toplanarak elde edilmiştir. Arka planda çalışan yazılımlar siteye sürekli sorgular göndermektedir. Şekil 6'deki ekran her dakika güncellenir. Sağ üstteki saat bölümü ekranın güncellenmesi hakkında bilgi veren son güncellemeyi göstermektedir.



Şekil 6

Coronavirus interaktif Dünya haritası – world tab

Ülke bazlı Ölüm Sayısı Karşılaştırma (Country-based Death Number Comparison)

Şekil 6 'da görüldüğü gibi, ülke adlarının bazıları büyük, bazıları küçüktür. Ülke isimlerinin yazı tipi değerlerinin farklılık göstermesinin nedeni, oluşturulan görselleştirme mimarisinin ülke bazlı ölüm sayılarını alıp 28-67 yazı tipinde ekrana yazdırmasıdır.

Dünya Haritasında Ülke Bazlı Ölümler (Deaths by Country on the World Map)

Ülke bazlı ölüm sayıları beyazdan kırmızıya doğru arttığını göstermektedir. Şekil 8 bu durumu göstermektedir. Bu görselin herhangi bir metinden daha anlaşılır olduğu değerlendirilmektedir.

09 Nisan 2020'de Türkiye'deki Son Durum (Latest Situation in Türkiye)

Şekil 8'deki ara yüz Türkiye'de şehir bazlı görselleştirme için hazırlanmıştır. Veriler, <http://trmetrik.com/corona/> adlı siteden oluşturulan komut dosyalarıyla indirilmiştir. Harita, Sağlık Bakanlığı tarafından şehir şehir sayılar açıklandıkça güncellenmiştir. Ancak panik ve kaos oluşturduğu için birçok ülke gibi Sağlık Bakanlığı da 9 Nisan 2020 de şehir şehir vaka sayısını açıklamaktan vazgeçmiştir. Gerekeceği vaka, ölüm ve iyileşme gibi sayı değerlerine ulaşamadığı beyan edilmiştir (<https://covid19.saglik.gov.tr/>).

artacaktır. Bu çalışmadaki COVID-19 salgınının görselleştirilmesi, gelecekte benzer salgınlar için erken uyarı sistemlerinin oluşturulmasına, yoğun bakım üniteleri ve sağlık personeli gibi kaynakların verimli yönetilmesine, halkın bilgilendirilmesi ve yanlış bilgilerin önlenmesi bakımından kriz iletişimi gibi çalışmalara ışık tutabilecektir. Ayrıca bu çalışmaya benzer çalışmalar farklı sektörlerde de uygulanabilir. Eğitimde öğrenci başarıları ve eğitim trenlerinin analizi, tarımda ürün verimliliği ve iklim değişikliklerinin etkileri, finansta piyasa analizlerinde kullanılması örnek olarak verilebilir.

Etik Beyan (Ethical Statement)

Bu çalışma, yazarların bilimsel iş birliği ve ortak entelektüel katkısı sonucunda hazırlanmış müşterek bir akademik üründür. Çalışmanın tüm aşamalarında yazarlar eşit düzeyde sorumluluk ve emek paylaşımında bulunmuştur.

Etik Kurul Onayı (Ethics Committee Approval)

Bu çalışma, insan veya hayvan katılımcıları içermeyen, ikincil veri kaynaklarına dayalı teorik/analitik bir araştırma niteliğindedir. Dolayısıyla etik kurul onayı gerektirmemektedir. Çalışmada kullanılan tüm veriler açık kaynaklardan elde edilmiş olup, atıf yapılan çalışmaların akademik etik ilkelerine riayet edilmiştir.

Yazar Katkıları (Author Contributions)

Araştırma Tasarımı (CRediT 1) Yazar 1 (%50) – Yazar 2 (%50)

Veri Toplama (CRediT 2) Yazar 1 (%80) – Yazar 2 (%20)

Araştırma - Veri Analizi - Doğrulama (CRediT 3-4-6-11) Yazar 1 (%50) – Yazar 2 (%50)

Makalenin Yazımı (CRediT 12-13) Yazar 1 (%35) – Yazar 2 (%65)

Metnin Tashihi ve Geliştirilmesi (CRediT 14) Yazar 1 (%30) – Yazar 2 (%70)

Finansman (Financing)

Bu çalışma herhangi bir dış kurum veya kuruluştan finansal destek almadan gerçekleştirilmiştir. Araştırmanın tüm süreçleri yazarların kendi akademik ve entelektüel çabaları ile yürütülmüştür.

Çıkar Çatışması (Conflict of Interest)

Yazarlar bu çalışma kapsamında herhangi bir çıkar çatışması olmadığını beyan ederler

Sürdürülebilir Kalkınma Amaçları (SDG)

Sürdürülebilir Kalkınma Amaçları 3: Sağlıklı ve kaliteli yaşam

KAYNAKÇA (REFERENCES)

- [1] 3M Visual Systems Division. Polishing your presentation, Austin, TX: 3M Visual Systems Division, 2020, https://3rd-force.org/pubs/meetingguide_pres.pdf
- [2] C. K. Leung, Y. Chen, C.S. Hoi, S. Shang, Y. Wen, A. Cuzzocrea, Big data visualization and visual analytics of COVID-19 data, *24th International Conference Information Visualisation (IV)*, IEEE, Australia, 2020, 415-420. doi:10.1109/IV51561.2020.00073
- [3] M. Hesami, M. Alizadeh, A.M.P. Jones, D. Torkamaneh, D, Machine learning: Its challenges and opportunities in plant system biology, *Applied Microbiology and Biotechnology*. 106(9-10) (2022), 3507-3530. doi:10.1007/s00253-022-11963-6
- [4] J. Yan, X. Wang, X, Unsupervised and semi-supervised learning: the next frontier in machine learning for plant systems biology, *The Plant Journal*. 111(6) (2022), 1527-1538. doi:10.1111/tpj.15905
- [5] M. Naeem, T. Jamal, J. Diaz-Martinez, S.A. Butt, N. Montesano, M.I. Tariq, E. De-La-Hoz-Valdiris, Trends and future perspective challenges in big data, *In Advances in Intelligent Data Analysis and Applications: Proceeding of the Sixth Euro-China Conference on Intelligent Data Analysis and Applications*, Springer, Arad, Romania, 2019, 309-325.
- [6] Y. Tang, J.J. Xiong, Y. Luo, Y.C. Zhang, How do the global stock markets Influence one another? Evidence from finance big data and granger causality directed network, *International Journal of Electronic Commerce*. 23(1) (2019), 85-109. doi:10.1080/10864415.2018.1512283
- [7] M. Pejić Bach, Ž. Krstić, S. Seljan, L. Turulja, Text mining for big data analysis in financial sector: A literature review, *Sustainability*. 11(5) (2019), 1277. doi:10.3390/su11051277
- [8] H. Sun, M.R. Rabbani, M. S., Sial, S. Yu, J.A. Filipe, J. Cherian, Identifying big data's opportunities, challenges, and implications in finance, *Mathematics*. 8(10) (2020), 1738. doi:10.3390/math8101738
- [9] J.L. Jimenez-Marquez, I. Gonzalez-Carrasco, J.L. Lopez-Cuadrado, B. Ruiz-Mezcua, Towards a big data framework for analyzing social media content. *International Journal of Information Management*. 44 (2019), 1-12. doi:10.1016/j.ijinfomgt.2018.09.003
- [10] X. Kong, Y. Shi, S. Yu, J. Liu, F. Xia, Academic social networks: Modeling, analysis, mining and applications, *Journal of Network and Computer Applications*. 132 (2019), 86-103. doi:10.1016/j.jnca.2019.01.029
- [11] L. Nemes, A. Kiss, Social media sentiment analysis based on COVID-19, *Journal of Information and Telecommunication*. 5(1) (2021), 1-15. doi:10.1080/24751839.2020.1790793
- [12] I. Ahmed, M. Ahmad, G. Jeon, F. Piccialli, A framework for pandemic prediction using big data analytics, *Big Data Research*. 25 (2021), 100190. doi:10.1016/j.bdr.2021.100190
- [13] M. Elsouhny, G. Jain, A. Shrivastava, Disaster Management during pandemic: A big data-centric approach, *International Journal of Innovation and Technology Management*. 18(04) (2021), 2140003. doi:10.1016/j.bdr.2021.100190
- [14] World Health Organization, Number of COVID-19 cases reported to WHO, 2020, covid19.who.int, 2020
- [15] S. Kunt, COVID-19 Pandemisinin turizme etkisi konusunda yapılan çalışmaların carrot² analizi ile değerlendirilmesi, *Güncel Turizm Araştırma Dergisi*. 5(1) (2021), 30-47. doi:10.32572/guntad.794537
- [16] M. Bektaş, A. Yavuz, F. Bulut, Salgın hastalıklarla mücadelede açık kaynak kodlu çözümler, *İstanbul Sabahattin Zaim Üniversitesi Fen Bilimleri Enstitüsü Dergisi*. 3(1), (2021), 99-105.
- [17] I. Ahmed, M. Ahmad, G. Jeon, F. Piccialli, A framework for pandemic prediction using big data analytics, *Big Data Research*. 25 (2021), 100190. doi:10.1016/j.bdr.2021.100190

- [18] A. Kurşun, Büyük veri ve sağlık hizmetlerinde büyük veri işleme araçları, *Hacettepe Sağlık İdaresi Dergisi*. 24(4), 921-940.
- [19] A.A. Ardakani, A. R. Kanafi , U. R. Acharya, N. Khadem , A. Mohammadi, Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks, *Computers in Biology and Medicine*. 121 (2020), 103795:1-103795:9. doi:10.1016/j.compbiomed.2020.103795
- [20] M.B. Jamshidi, et al., Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment, *IEEE Access*. 8 (2020), 109581-109595. doi:10.1109/ACCESS.2020.3001973
- [21] B. Robson, COVID-19 coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed achilles' heel conserved region to minimize probability of escape mutations and drug resistance, *Computers in Biology and Medicine*. 121 (2020), 103749. doi:10.1016/j.compbiomed.2020.103749
- [22] Z. A. A. Alyasseri, M.A. Al-Betar, I. A., Doush, M. A., Awadallah, A. K., Abasi, S. N. Makhadmeh, R.A. Zitar, Review on COVID-19 diagnosis models based on machine learning and deep learning approaches, *Expert Systems*. 39(3) (2022), e12759. doi:10.1111/exsy.12759
- [23] K. Moulaei, M. Shanbehzadeh, Z. Mohammadi-Taghiabad, H. Kazemi-Arpanahi, Comparing machine learning algorithms for predicting COVID-19 mortality. *BMC Medical Informatics and Decision Making*. 22(1) (2022), 1-12. doi:10.1186/s12911-021-01742-0
- [24] D. Mhlanga, The role of artificial intelligence and machine learning amid the COVID-19 pandemic: What lessons are we learning on 4IR and the sustainable development goals, *International Journal of Environmental Research and Public Health*. 19(3) (2022), 1879. doi:10.3390/ijerph19031879
- [25] C. Comito, C. Pizzuti, Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review, *Artificial Intelligence in Medicine*. 128 (2022), 102286. doi:10.1016/j.artmed.2022.102286.
- [26] P.G. Asteris, E. Gavriilaki, T. Touloumenidou, E.E. Koravou, M. Koutra, P.G. Papayanni, A. Anagnostopoulos, Genetic prediction of icu hospitalization and mortality in COVID-19 patients using artificial neural networks, *Journal of Cellular and Molecular Medicine*. 26(5) (2022), 1445-1455. doi:10.1111/jcmm.17098
- [27] S.I. O'Donoghue, Grand challenges in bioinformatics data visualization, *Frontiers in Bioinformatics*. 17 (2021), 669186. doi:10.3389/fbinf.2021.669186
- [28] S.I. O'Donoghue, B.F. Baldi, S.J. Clark, A.E. Darling, J.M. Hogan, S. Kaur, L. Maier-Hein, D. J. McCarthy, W.J. Moore, E. Stenau, J.R. Swedlow, J. Vuong, J.B. Procter, Visualization of biomedical data, *Annual Review of Biomedical Data Science*. 1 (2018), 275-304. doi:10.1146/annurev-biodatasci-080917-013424
- [29] B. Wong, Visualizing biological data, *Nature Methods*. 9: 1131. 2012. doi:10.1038/nmeth.2258
- [30] O. ByMatthew Ward, G. Grinstein, D. Keim, Interactive Data Visualization, *Foundations, Techniques, and Applications, Second Edition*, New York, A K Peters/CRC Press 2015. doi:10.1201/b18379
- [31] C. Ware, Information Visualization: Perception for Design. Third Edition., *Morgan Kaufmann Publishers Inc.*, San Francisco, CA, United States, 2012.
- [32] F. Zuo, J. Wang, J. Gao, K. Ozbay, X.J. Ban, Y. Shen, H. Yang, S. Iyer, An interactive data visualization and analytics tool to evaluate mobility and sociability trends during COVID-19, arXiv preprint arXiv:2006.14882, 2020.
- [33] P.L. Bras, A. Gharavi, D.A. Robb, A.F. Vidal, S. Padilla, M.J. Chantler, Visualising COVID-19 research, *CoRR* abs/2005.06380, 2020.

- [34] J. Tu, M. Verhagen, B. Cochran, J. Pustejovsky, Exploration and discovery of the COVID-19 literature through semantic visualization, *CoRR* abs/2007.01800, 2020.
- [35] Worldometer, COVID-19 Coronavirus Pandemic, 2020. <https://www.worldometers.info/coronavirus/> (13 April 2020)
- [36] F. Wolinski, Visualization of diseases at risk in the COVID-19 literature, *CoRR* abs/2005.00848, 2020.
- [37] F. Khanam, I. Nowrin, M. Mondal, Data visualization and analyzation of COVID-19, *Journal of Scientific Research and Reports*. 42-52 (2020). DOI: 10.9734/JSRR/2020/v26i330234.
- [38] E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time, *The Lancet Infectious Diseases*. 20(5) (2020), 533-534. doi:10.1016/S1473-3099(20)30120-1
- [39] C.K. Leung, Y. Chen, C.S. Hoi, S. Shang, Y. Wen, A. Cuzzocrea, Big data visualization and visual analytics of COVID-19 data. In 2020 24th International Conference Information Visualisation (IV) (pp. 415-420). IEEE.
- [40] I. Berry, J.P.R. Soucy, A. Tuite, D. Fisman, Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada, *Cmaj*. 192(15) (2020), E420-E420. doi:10.1503/cmaj.75262
- [41] F.B. Hamzah, C. Lau, H. Nazri, D.V. Ligot, G. Lee, C.I. Tan, M.H. Chung, Corona Tracker: worldwide COVID-19 outbreak data analysis and prediction, *Bull World Health Organ*. 1(32) (2020), 1-32. doi:10.2471/BLT.20.251561.
- [42] F. Bahadır, F.S. Balık, H.S. Yalçınkaya, The impact of COVID-19 on the financial structure of the construction industry in Turkey. *Necmettin Erbakan University Journal of Science and Engineering*. 5(2) (2023), 134-149. doi:10.47112/neufmbd.2023.17
- [43] Y. Uzun, H. Ergün, E Şeker, Augmented reality approach for stories. *Necmettin Erbakan University Journal of Science and Engineering*. 4(2) (2022), 1-7. doi:10.47112/neufmbd.2022.1
- [44] A. Pektaş, O. İnan. application of tree seed algorithm on clustering problems, *Necmettin Erbakan University Journal of Science and Engineering*. 4(1) (2022), 1-10.