

COVID-19 Pandemisi ile İlgili Twitter Mesajlarının Metin Madenciliği Tekniği İle Değerlendirilmesi

Evaluation of Twitter Messages Related to COVID-19 Pandemic Using Text Mining Technique

• Ebru TEMİZHAN^a, • Mehmet MENDEŞ^a

^aÇanakkale Onsekiz Mart Üniversitesi Ziraat Fakültesi, Zootekni Bölümü Biyometri ve Genetik ABD, Çanakkale, TÜRKİYE

Bu çalışma, I. Uluslararası Uygulamalı İstatistik Kongresi'nde (1-4 Ekim 2020, Online-Tokat) bildiri olarak sunulmuştur ve kongre kitabında İngilizce özet olarak yayınlanmıştır.

ÖZET Amaç: Bu çalışma, yeni tip korona virüs hastalığı-2019 [coronavirus disease-2019 (COVID-19)] ile ilgili sosyal medyada yapılan paylaşımların ve atılan mesajların metin madenciliği tekniği (text mining technique) ile değerlendirilerek halkın bu salgına bakış açısını ve düşünceleri hakkında özetleyici bilgi üretmek amacıyla yürütülmüştür. **Gereç ve Yöntemler:** Bu amaçla, COVID-19 salgını ile ilgili atılan 40.000 Twitter mesajı metin madenciliği tekniği ile değerlendirilmiştir. **Bulgular:** Yapılan metin madenciliği analizi sonucunda, çalışmaya dâhil edilen Twitter kullanıcılarının tweetlerinde esas itibarıyla *maske*, *COVID-19*, *Ankara Wuhan* ve *Bilim Kurulu* olmak üzere 4 kelime üzerinde odaklandıkları ve yaptıkları paylaşımların bu 4 kelime etrafında şekillendiği belirlenmiştir. Aynı zamanda dikkate alınan Twitter kullanıcılarının, gündemlerini meşgul eden bu 4 kelimeye ilişkin duygu analizleri de (sentiment analysis) yapılmıştır. Bu vesile ile bireylerin gündemlerinde önemli bir yer teşkil eden maske, COVID-19, Ankara Wuhan ve Bilim Kurulu kelimelerinin duygularını nasıl etkilediklerine ilişkin bilgi üretilmesi de mümkün olmuştur. Yapılan duygu analizi sonucunda, bireylerin özellikle COVID-19 kelimesini duymaları ya da bundan bahsedilmesinin kendilerinde negatif bir duygu oluşturdukları sonucuna varılmıştır. Aynı zamanda Bilim Kurulunun kararlarının halkın üzerinde dikkate değer bir etki bıraktığı görülmüştür. **Sonuç:** Metin madenciliği ve duygu analizleri sonuçları birlikte değerlendirildiğinde halkın COVID-19 salgınının, halkın gündemini ciddi bir şekilde meşgul ettiği ve duyduklarında en fazla etkilendikleri terimlerin *maske*, *COVID-19*, *Ankara Wuhan* ve *Bilim Kurulu* olduğu sonucuna varılmıştır.

Anahtar kelimeler: COVID-19; metin madenciliği; duygu analizi; regresyon analizi; maske

ABSTRACT Objective: This study has been carried out to generate summarized information about the public's perspective and thoughts on coronavirus disease-2019 (COVID-19) pandemic by evaluating the posts and messages sent on social media about COVID-19 pandemic with the text mining technique. **Material and Methods:** For this purpose, 40.000 Twitter messages have sent regarding the COVID-19 outbreak were evaluated with the text mining technique. **Results:** As a result of the text mining analysis, it was determined that the tweets of the Twitter users included in the study focused mainly on four words, *mask*, *COVID-19*, *Ankara Wuhan* and the *scientific committee*, and their posts were shaped around these four words. At the same time, sentiment analysis was also performed after text mining analysis for four words that occupy the agendas of twitter users. Thus, it was possible to produce information about how the words *mask*, *COVID-19*, *Ankara Wuhan* and *scientific committee*, which constitute an important place in the agenda of Twitter users, affect their emotions. As a result of the sentiment analysis, it was concluded that individuals especially hearing the word *COVID-19* or mentioning this word led to have a negative emotion in themselves. It was also observed that the decisions of the science committee have left a remarkable impact on the people. **Conclusion:** When the results of text mining and sentiment analysis were evaluated together, it was concluded that the people's agenda were seriously occupied by the COVID-19 pandemic. It was concluded that the terms that people were most affected by when they hear are *mask*, *COVID-19*, *Ankara Wuhan* and the *scientific committee*.

Keywords: COVID-19; text mining; sentiment analysis; regression analysis; mask

Correspondence: Mehmet MENDEŞ
Çanakkale Onsekiz Mart Üniversitesi Ziraat Fakültesi, Zootekni Bölümü Biyometri ve Genetik ABD, Çanakkale, TÜRKİYE/TURKEY
E-mail: mmendes@comu.edu.tr



Peer review under responsibility of Türkiye Klinikleri Journal of Biostatistics.

Received: 09 Nov 2020 **Received in revised form:** 20 Nov 2020 **Accepted:** 23 Nov 2020 **Available online:** 15 Mar 2021

2146-8877 / Copyright © 2021 by Türkiye Klinikleri. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Korona virüs hastalığı-2019 [coronavirus disease-2019 (COVID-19) pandemisi, 1 Aralık 2019 tarihinde Çin'in Hubei bölgesinin başkenti olan Wuhan'da ortaya çıkmıştır.¹ Kişiden kişiye bulaşabilen virüsün, bulaşma oranı 2020 Ocak ayı ortasında büyüme göstererek, mart ayı sonlarında başta Avrupa, Kuzey ve Güney Amerika, Doğu Akdeniz, Batı Pasifik, Güney ve Doğu Asya ve Afrika olmak üzere 6 bölgeye yayılmıştır.^{1,2} 11 Mart 2020'de Dünya Sağlık Örgütü tarafından küresel salgın olarak ilan edildi. Avrupa'daki vakaların hızla artmasıyla, 13 Mart 2020'de Avrupa artık COVID-19 salgın merkezi hâline gelmiştir.¹ 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.^{2,3}

Türkiye pandemi sürecine diğer Avrupa ülkelerinden 1-2 ay sonra girmiştir. Bu ek zaman bize bu sürecin nasıl geçeceğini, yaşanan problemleri ve alınan tedbirleri önceden öğrenme imkânı sunmuş ve mümkün olduğunca tedbirli olmamızı sağlamıştır. Gerek Avrupa ülkelerinde gerekse Asya ülkelerinde, vaka sayılarının hızla artmasından dolayı Türkiye'de de kısa bir zaman sonra 11 Mart 2020 tarihinde ilk pozitif vaka görülmüştür.² 16 Mart 2020 tarihinde korona virüs kısıtlamalarının başlamasıyla, insanlar bir anda alıştikları sosyal hayattan ayrılıp sosyal hayatı kısıtlayan korona virüs tedbirlerine uymak zorunda kaldı. 1 Haziran'da normalleşme adımlarına geçilmesi ve tekrardan vaka sayılarının artışıyla, kısıtlamaların Ağustos ayında yeniden başlamasına sebep olmuştur. Bu zorlu süreçte, tüm dünyadaki insanlar duygu ve düşüncelerini çoğunlukla sosyal medya aracılığıyla birbirleri ile paylaştı. Her gün açıklanan istatistikler, vaka sayıları, maske, sosyal mesafe, kolonya, dezenfektan, sağlık çalışanları, işçiler, yatak sayısı, ilaç ve tedavi teknikleri, aşı, uzaktan eğitim, video konferans, hibrit eğitim, sosyal yardım ve ulaşım gibi başlıklar ülke gündeminde sıklıkla yer aldı ve bu gündem başlıkları her gün sosyal medyada tartışılır ve konuşulur oldu. Son yıllarda çoğu televizyon programı, sosyal medya hesaplarından takipçileriyle iletişim kurarak birçok gündeme ilişkin görüşleri izleyicileri ile paylaşmaktadır. Bazen de sosyal medyada oluşturulan, paylaşımları ve yorumları haber niteliğinde izleyiciye sunmaktadır. Özellikle Twitter'da açılan günlük gündem başlıkları, aynı anda birçok kişinin tartışmaya katıldığı, duygu ve düşüncelerin paylaşıldığı bir ortamı oluşturmaktadır. Günümüzde, sosyal medya diğer iletişim araçlarından daha hızlı ve popüler bir haber kaynağı hâline gelmektedir.

Bu çalışma, yeni tip (COVID-19) ile ilgili sosyal medyada yapılan paylaşımların ve atılan mesajların, metin madenciliği tekniği (text mining technique) ile değerlendirilerek halkın bu salgına bakış açısını ve düşünceleri hakkında özetleyici bilgi üretmek amacıyla yürütülmüştür.

GEREÇ VE YÖNTEMLER

GEREÇ

Araştırmanın materyalini 4 Eylül 2020 tarihinde Twitter API ile çekilen COVID-19 ile ilgili gündem başlıkları altında 40.000 adet Türkçe tweet oluşturmuştur (Tablo 1). #Covid_19, #Maske ve #bilimkurulu etiketleri Twitter'da devamlı olarak gündem oluştururken, **Ankara Wuhan** etiketi 3 Eylül 2020 tarihli Bilim Kurulu toplantısında, sağlık bakanlığının Ankara'daki vakaların İstanbul'daki vakaların 2 katı olduğunu açıklamasından sonra Twitter'da "Ankara Wuhan" başlığı altında Ankara gündemi tartışılmaya başlamıştır. COVID-19 pandemisi bu 4 gündem başlığı etrafında değerlendirilmiştir. İstatistiksel analizlerde R (4.0.2) yazılım programı geliştirme ortamı olan R-Studio (1.2.5033) ve SPSS (Ver. 19) paket programı kullanılmıştır.⁴⁻⁶

TABLO 1: Twitter’den çekilen korona virüs ile ilgili gündemler ve tweet sayıları.

Gündemler	Tweet Sayıları (n)
#Covid_19	10.000
#Maske	10.000
#bilimkurulu	10.000
"Ankara Wuhan"	10.000

YÖNTEMLER

Veri analizleri için kullanılan teknikler, işlem adımları ve kütüphaneler [Tablo 2](#)'de verilmiştir. Veri setlerinin oluşturulması, metin ön işleme, kelime bulutu, duygu analizi ve terim doküman matrisi işlemleri için metin madenciliği tekniği kullanılırken, değişkenlerin tahmin denklemlerinin oluşturulmasında otomatik linear model istatistik tekniği kullanılmıştır.

METİN MADENCİLİĞİ (TEXT MINING)

Metin madenciliği, veri kaynağı metin olan veri madenciliği (data mining) çalışması olup yapısal olmayan verinin yapısal hâle dönüştürülmesini sağlamaktadır.⁷ Diğer bir deyişle metin madenciliği, metin veri setinin sayısal veri setine çevrilmesi sırasında uygulanan doğal dil işleme tekniklerinin tümünü ifade eder. Twitter’den COVID-19 pandemisi ile ilgili gündem başlıklarına ait tweetler twitterR ve ROAuth kütüphaneleri `setup_twitter_oauth()` ve `searchTwitter()` fonksiyonları ile Twitter API aracılığıyla çekilerek metin veri setleri oluşturulmuştur ([Tablo 2](#)). Bu işleme aynı zamanda, sosyal ağ madenciliği (social network mining) de denilmektedir.⁸ Daha sonra metin içeriğinin yapısal hâle dönüştürülmesi için bir takım metin ön işleme adımları uygulanır ([Tablo 2](#)). `Tolower()` fonksiyonu ile metin veri setlerinin küçük harflere çevrilmesi sağlanır, `gsub()` fonksiyonu ile metin içeriğinde bulunan gereksiz karakterler, simgeler ve kelimeler çıkartılarak metin içeriği temizlenir, `tm` ve `NLP` kütüphaneleri `tm_map()` fonksiyonu ile Türkçe stopwords kelimeler kaldırılarak ağırlıklandırılmış optimum terim doküman matrisinin oluşturulmasıyla metin ön işleme işlemleri tamamlanır. Böylece metin madenciliği işlem adımları tamamlanarak metin veri setleri kelime bulutu (word cloud), duygu analizi (sentiment analysis) ve otomatik linear model (automatic linear modeling) yöntemleri için hazır hâle getirilmiş olur.

KELİME BULUTU (WORD CLOUD)

Kelime bulutu, metin veri setleri için kullanılan veri görselleştirme yöntemidir. Kelime bulutu frekansa dayalı bir görselleştirme yöntemi olduğundan metinlerin görsel bir özetini oluşturur.⁸ Kelime bulutunu oluşturmak için `word cloud` ve `R Color Brewer` kütüphaneleri ile `word cloud ()` fonksiyonu kullanılmıştır. Kelime bulutları en sık kullanılan 100 kelime üzerinden oluşturulmuştur.

DUYGU ANALİZİ (SENTIMENT ANALYSIS)

Fikir madenciliği olarak da adlandırılan duygu analizi; insanların fikirlerini, duygularını, değerlendirmelerini ve tutumlarını yazılı dil aracılığıyla analiz eden çalışma alanıdır. Duygu analizi görüşlerin ölçülmesini sağlayarak, fikirlerden yaralanmamızı sağlamaktadır.⁸ Duygu analizinde, `syuzhet` kütüphanesinde `get_nrc_sentiment()` fonksiyonu aracılığıyla toplamda 10 adet duygu skoru (öfke, beklenti, bıkmak, korku, sevinç, negatif, pozitif, üzüntü, sürpriz, güven) hesaplanabilmektedir.

TABLO 2: Veri analizleri için kullanılan teknikler, işlem adımları ve kütüphaneler.

Teknikler	İşlem Adımları	Kütüphaneler
Metin madenciliği (Text mining)	Veri setlerinin oluşturulması Metin ön işleme <ul style="list-style-type: none"> • Metni küçük harfe dönüştürme • Gereksiz karakterlerin ve kelimelerin kaldırılması <ul style="list-style-type: none"> ○ Retweet ○ @kullanıcı adı ○ Noktalama işaretleri ○ Rakamlar ○ Linkler ○ Tabs ○ Cümle başındaki ve sonundaki boş alanlar ○ Stop words 	Library (twitteR) Library (ROAuth) Library (NLP) Library (tm) Library (stringi) Library (topicmodels)
	Kelime bulutu (Word cloud)	Library (RColorBrewer) Library (wordcloud)
	Duygu analizi (Sentiment analysis)	Library (syuzhet) Library (ggplot2) Library (NLP) Library (gridExtra)
	Terim doküman matrisi <ul style="list-style-type: none"> • Metni vektörleştirme • Kelimelerin köklere ayrılması • Ağırlıklandırılmış frekans hesaplama • Optimum terim doküman matrisi oluşturma • Özellik çıkartma 	Library (openNLP) Library (NLP) Library (tm) Library (SnowballC)
İstatistiksel yöntem	Otomatik linear model (Automatic linear modeling) <ul style="list-style-type: none"> • Forward stepwise • Akaike bilgi kriteri 	

OTOMATİK LİNEAR MODEL (AUTOMATIC LINEAR MODELING)

Otomatik linear model (1), birden fazla bağımsız değişken içeren bir çoklu doğrusal regresyon modelidir. Otomatik linear model, forward stepwise model seçimi ile akaike bilgi kriterine göre gerçekleştirilmiştir. Bilgi kriteri modellerin karşılaştırılmasında kullanılmaktadır. Bilgi kriter değeri küçük olan modeller, diğer modellere göre daha iyi bir model olarak değerlendirilmektedir. Model performansının değerlendirilmesinde kullanılan bir diğer ölçüt, modelin doğruluk derecesi ya da açıklanan varyasyonudur. Otomatik linear model tahmin edicilerin istatistiksel ve pratik önemini, modele olan etkilerini, tahmin edilen (bağımsız) değişken ile doğrusal (linear) ya da doğrusal olmayan (nonlinear) ilişkilerinin rapor edilmesini sağlamaktadır.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (1)$$

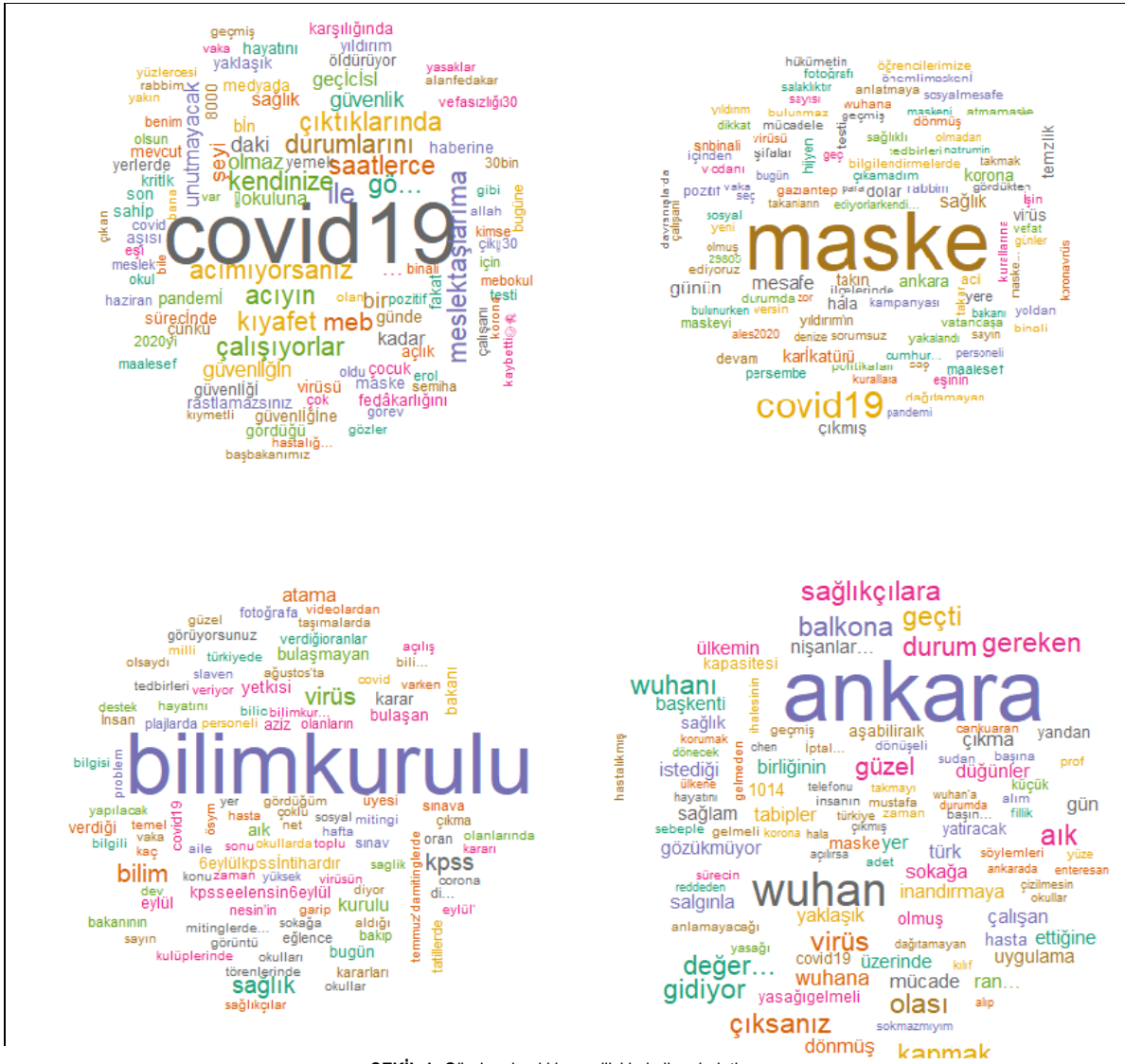
BULGULAR

Gündem başlıklarına ilişkin kelime bulutları [Şekil 1](#)'de, duygu analizi sonuçları [Şekil 2](#)'de, tanıttıcı istatistikler [Tablo 3](#)'te ve otomatik linear model analiz sonuçları [Şekil 3](#), [Şekil 4](#), [Şekil 5](#), [Şekil 6](#), [Şekil 7](#), [Şekil 8](#), [Şekil 9](#)'da verilmiştir. “#Covid_19”, “#Maske”, “#bilimkurulu” ve “Ankara Wuhan” etiketlerine

ilişkin kelime bulutlarında en çok tekrar eden kelimeler bu gündem başlıkları etrafında toplanarak her bir gündemin görsel olarak özetini oluşturmuştur (Şekil 1).

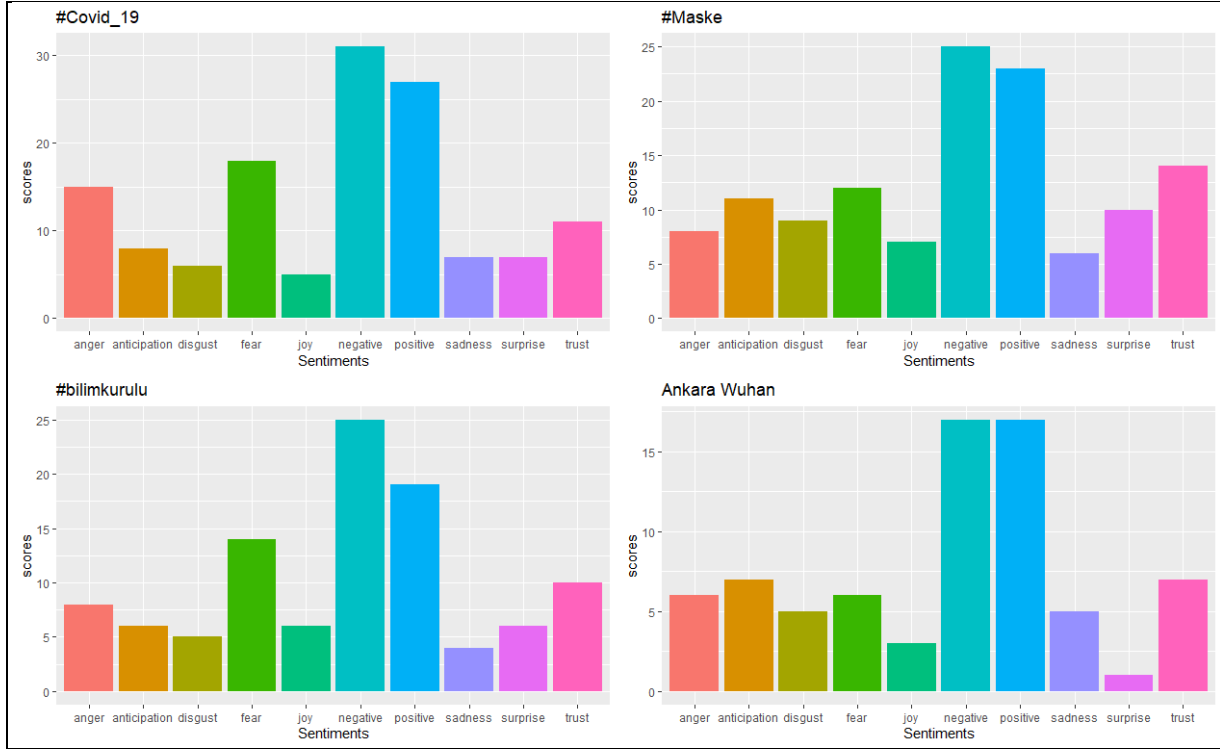
Gündem başlıklarına ilişkin duygu analiz sonuçları incelendiğinde, gündemlerin duygu skorları birbirlerine benzer bir dağılım göstermektedir (Şekil 2). Özellikle negatif, korku, kızgınlık ve beklenti gibi duygular COVID-19 sürecinde yaşanan baskın duyguları oluşturmaktadır. #bilimkurulu gündeminde, negatif ve pozitif duygu skorları arasındaki farkların diğer gündemlere göre daha belirgin olduğunu görmekteyiz. Demek ki halkın, Bilim Kurulu kararlarından etkilendiğinden ve bu kararların dikkate değer olduğundan bahsedebiliriz. Diğer taraftan Ankara Wuhan gündemindeki şaşırma duygusunun diğer gündemlere göre düşük olması, halkın Ankara'da yaşanan salgın artışını bekliyor olacağı söylenebilir.

#Covid_19, #Maske, #bilimkurulu ve Ankara Wuhan gündem başlıklarına ilişkin otomatik linear model sonuçları Şekil 3, Şekil 4, Şekil 5, Şekil 6, Şekil 7, Şekil 8 ve Şekil 9'da verilmiştir. İstatistiksel önemliliğe sahip otomatik linear modellerin AIC ve Adjusted R² değerleri sırasıyla #Covid_19 -18.811,03 %42,95, #Maske -870,97 %3,98, #bilimkurulu 11.803,19 %9,93 ve Ankara Wuhan -35.483,95 %97,20 olarak bulunmuştur.



ŞEKİL 1: Gündem başlıklarına ilişkin kelime bulutları.

#Covid_19 gündemine ilişkin en önemli tahmin ediciler, *çalışan* ve *güvenlik* değişkenleri olarak bulunmuştur ve gündeme ilişkin tahmin edicilerin modele etkileri hemen hemen benzerdir (Şekil 3, Şekil 4). COVID-19 bağımlı değişkenine ait otomatik linear model incelendiğinde çalışan, güvenlik, pozitif vaka, eş ve yemek değişkenleri modeli **pozitif yönde** etkilerken pandemi, maske, fedakarlık, hayat ve ölüm değişkenleri modeli **negatif yönde** etkilemektedir (AIC= -18.811,03, Adjusted R²= %42,95) (Şekil 5). Bağımlı değişken ile bağımsız değişkenler arasındaki linear ve linear olmayan ilişkiler Şekil 6'da verilmiştir.



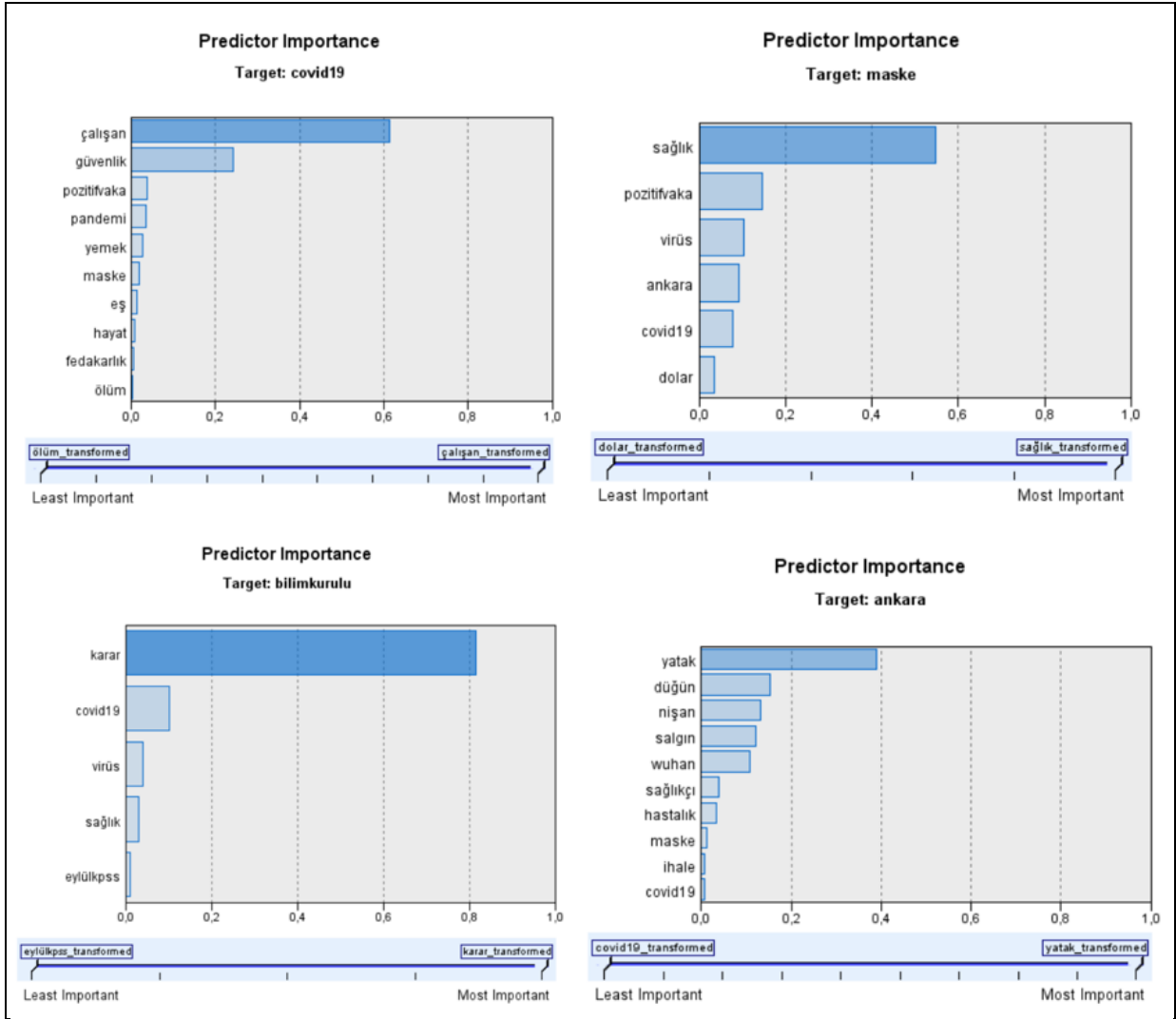
Not: x eksen: sırasıyla soldan sağa anger (kızgınlık), anticipation (beklenti), disgust (bıkmak), fear (korku), joy (neşe), negative (negatif), positive (pozitif), sadness (üzgünlük), surprise (şaşkınlık), trust (güvenmek). y eksen: scores (skorlar).

ŞEKİL 2: Gündem başlıklarına ilişkin duygu skorları.

#Maske gündemine ilişkin en önemli tahmin edici, *sağlık* değişkeni olarak bulunmuştur ve gündeme ilişkin tahmin edicilerin modele etkileri hemen hemen benzer olarak bulunmuştur (Şekil 3, Şekil 4). Maske bağımlı değişkenine ait otomatik linear model incelendiğinde; sağlık, pozitif vaka, Ankara, COVID-19 ve dolar değişkenleri modeli **pozitif yönde** etkilerken, virüs değişkeni modeli **negatif yönde** etkilemiştir (AIC= -870,97, Adjusted R²= %3,98) (Şekil 5). Bağımlı değişken ile bağımsız değişkenler arasındaki, linear ve linear olmayan ilişkiler Şekil 7'de verilmiştir.

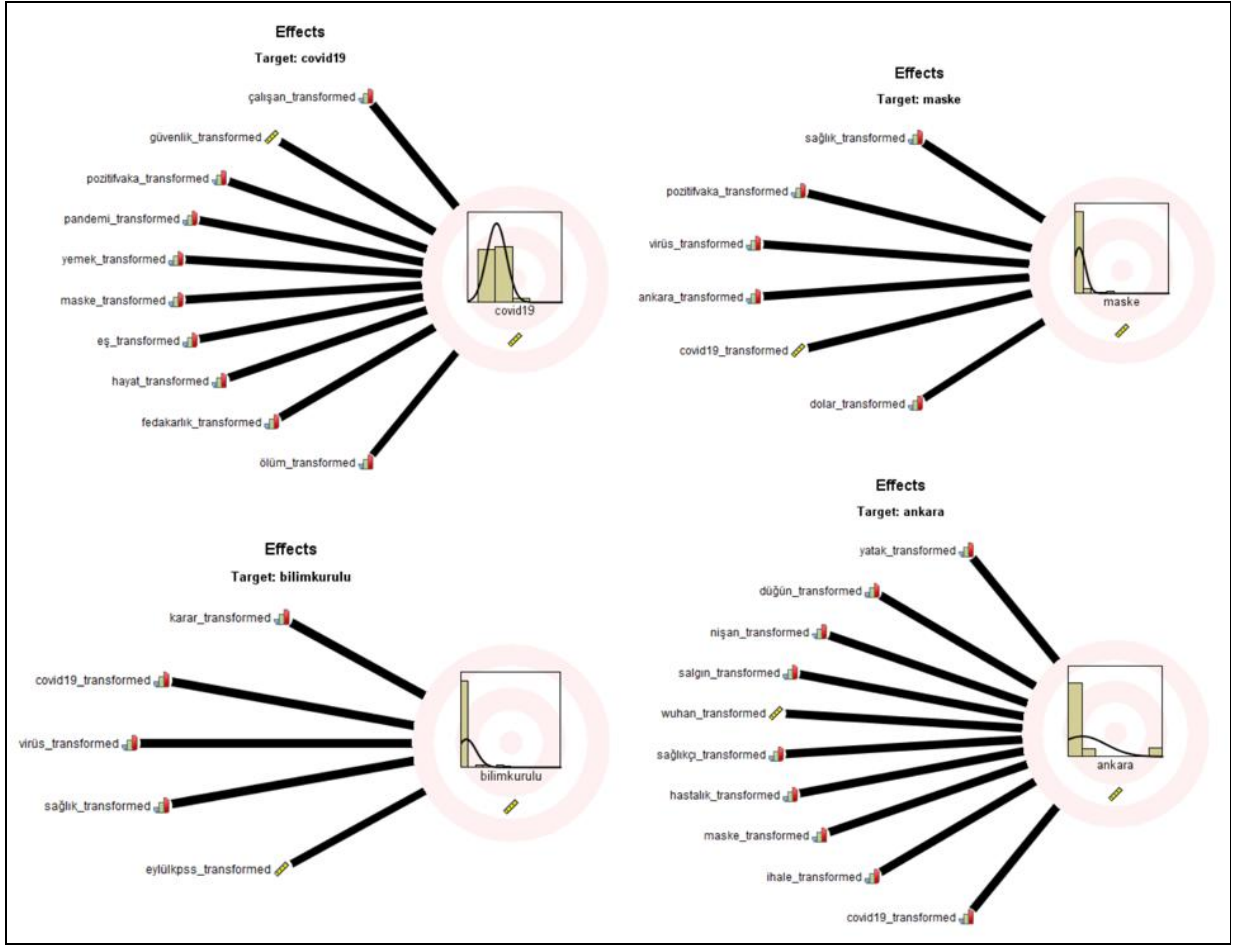
TABLO 3: #Covid_19, #Maske, #bilimkurulu ve Ankara Wuhan gündemlerine ilişkin tanıtıcı istatistikler.

#Covid_19			#Maske		
Kelime	Ağırlıklandırılmış frekans	Yüzde	Kelime	Ağırlıklandırılmış frekans	Yüzde
Güvenlik	8.996,79	0,12	COVID-19	7.836,67	0,21
Acı duymak	7.264,71	0,1	Maske	6.289,25	0,17
Zaman	5.574,03	0,07	Mesafe	3.085,25	0,08
Çalışan	5.224,14	0,07	Sağlık	2.967,77	0,08
COVID-19	5.192,55	0,07	Ankara	2.485,88	0,07
Okul	3.968,41	0,05	Temizlik	2.195,31	0,06
Meslektaş	3.629,89	0,05	Dolar	2.146,53	0,06
Kıyafet	3.624,94	0,05	Virüs	2.114,46	0,06
MEB	3.502,87	0,05	Pozitif vaka	1.933,35	0,05
Sağlık	2.432,29	0,03	Test	1.846,15	0,05
Pandemi	1.970,51	0,03	Acil	1.792,79	0,05
Virüs	1.962,52	0,03	Hijyen	1.607,46	0,04
Çocuk	1.881,31	0,03	Eş	1.560,72	0,04
Ölüm	1.879,86	0,03	Ankara Wuhan		
Yemek	1.879,86	0,03	Wuhan	1.4073,28	0,19
Maske	1.862,48	0,02	Yatak	6.096,22	0,08
Medya	1.860,74	0,02	Ankara	5.819,98	0,08
Aşı	1.851,13	0,02	Virüs	4.791,46	0,07
Hayat	1.828,58	0,02	Sağlıkçı	4.689,45	0,06
Haber	1.825,35	0,02	Yasaklar	4.591,2	0,06
Açlık	1.818,87	0,02	Salgın	3.566,42	0,05
Fedakarlık	1.685,98	0,02	Çalışan	3.557,94	0,05
Pozitif vaka	1.582,81	0,02	Tabipler	3.556,24	0,05
Eş	1.542,86	0,02	Düğün	3.182,79	0,04
#bilimkurulu			Nişan	3.166,84	0,04
			Maske	2.897,92	0,04
Bilim Kurulu	1.0565,94	0,37	Sağlık	2.867,66	0,04
Eylül kpss	8.022,9	0,28	COVID-19	2.848,9	0,04
Virüs	3.091,38	0,11	Hasta	2.776,5	0,04
Sağlık	3.046,47	0,11	Zaman	1.699,55	0,02
COVID-19	2.003,74	0,07	Hastalık	1.672,34	0,02
Karar	1.819,15	0,06	İhale	1.655,19	0,02



ŞEKİL 3: #Covid_19, #Maske, #bilimkurulu ve Ankara Wuhan gündemlerine ilişkin önemli tahmin ediciler.

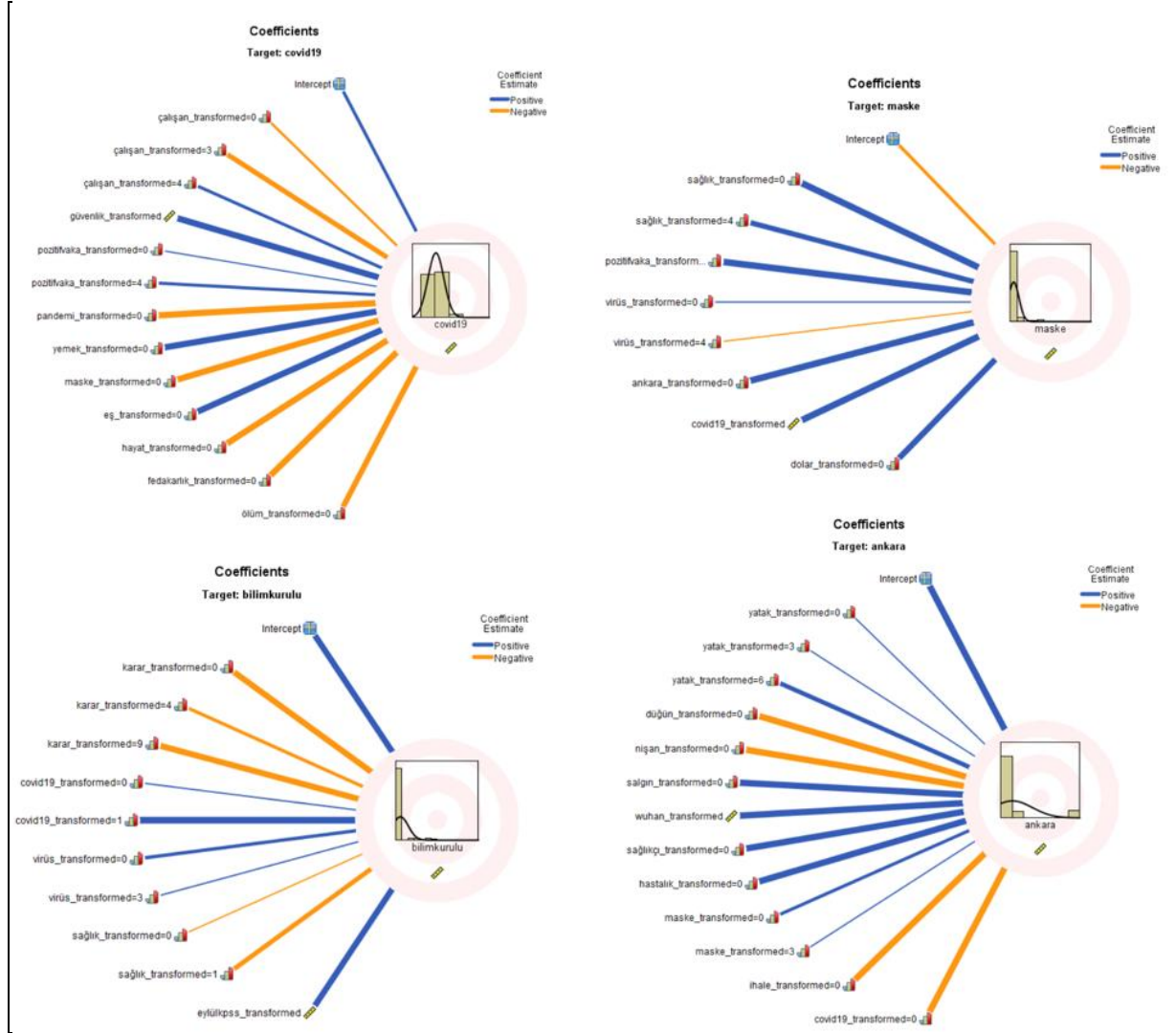
#bilimkurulu gündemine ilişkin en önemli tahmin edici *karar* değişkeni olarak bulunmuştur ve gündeme ilişkin tahmin edicilerin modele etkileri hemen hemen benzer olarak bulunmuştur (Şekil 3, Şekil 4). Bilim Kurulu bağımlı değişkenine ait otomatik linear model incelendiğinde, COVID-19, virüs ve eylül kpss değişkenleri modeli **pozitif yönde** etkilerken karar ve sağlık değişkenleri modeli **negatif yönde** etkilemiştir (AIC= 11.803,19, Adjusted R² = %9,93) (Şekil 5). Bağımlı değişken ile bağımsız değişkenler arasındaki, linear ve linear olmayan ilişkiler Şekil 8'de verilmiştir.



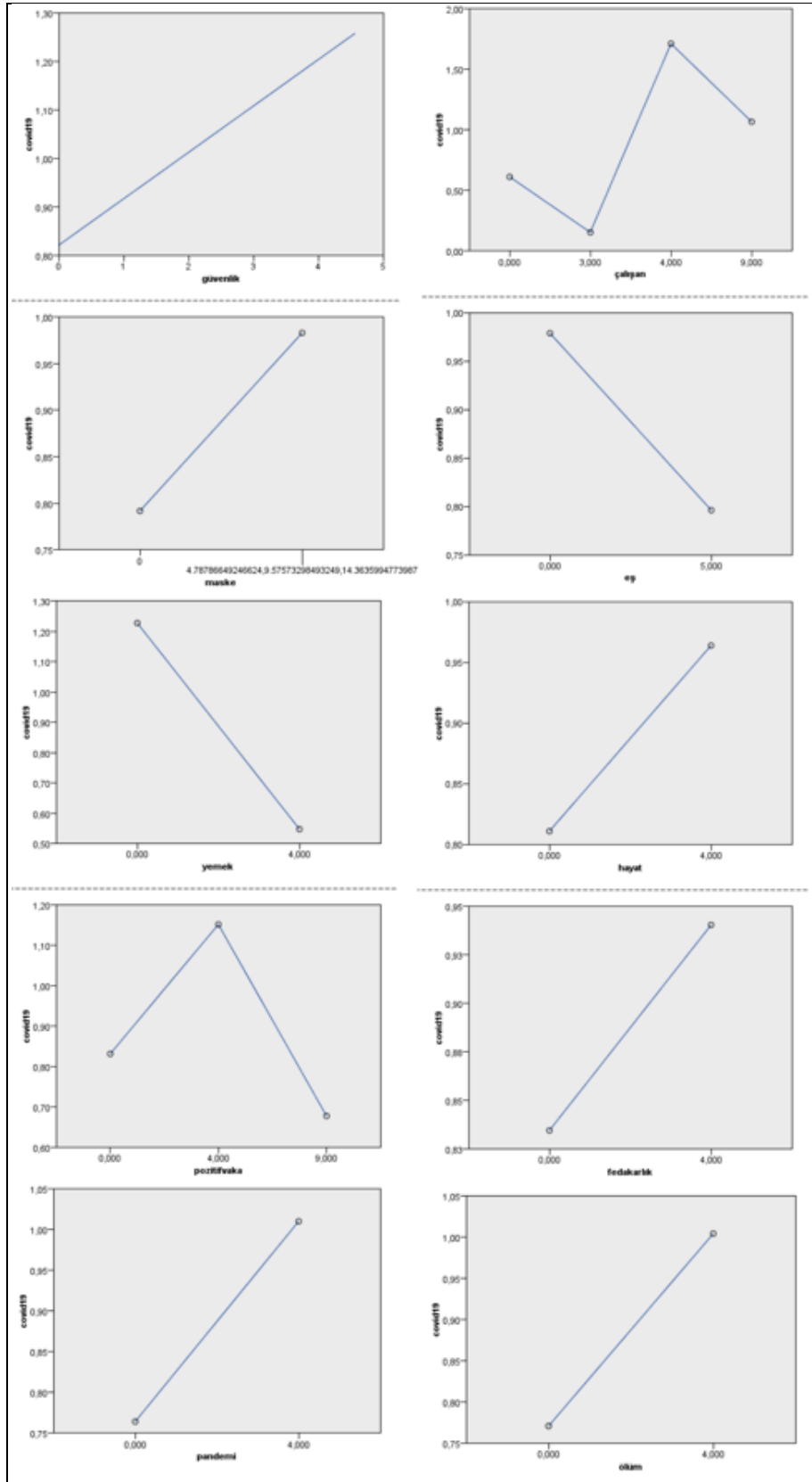
ŞEKİL 4: #Covid_19, #Maske, #bilimkurulu ve Ankara Wuhan gündemlerine ilişkin tahmin edicilerin modele etkileri.

Ankara Wuhan gündemine ilişkin en önemli tahmin ediciler *yatak*, *düşün*, *nişan*, *salgın* ve *Wuhan* değişkenleri olarak bulunmuştur ve gündeme ilişkin tahmin edicilerin modele etkileri hemen hemen benzerdir (Şekil 3, Şekil 4). Ankara bağımlı değişkenine ait otomatik linear model incelendiğinde yatak, salgın, Wuhan, hastalık, sağlıkçı ve maske değişkenleri modeli **pozitif yönde** etkilerken düşün, nişan, ihale ve COVID-19 değişkenleri modeli **negatif yönde** etkilemiştir (AIC= -35.483,95, Adjusted R²= %97,20) (Şekil 5). Bağımlı değişken ile bağımsız değişkenler arasındaki, linear ve linear olmayan ilişkiler Şekil 9'da verilmiştir.

COVID-19 pandemisine ilişkin gündem analizinde, Ankara gündemine ait otomatik linear model çok iyi bir tahmin edicidir (AIC= -35.483,95, Adjusted R²= %97,20). Özellikle de halkın, bu salgına bakış açısını ve düşünceleri hakkında istenilen düzeyde özetleyici bilgiyi üretmektedir. Diğer gündem başlıklarına ait oluşturulan otomatik linear modeller, pratik açıdan gündemin değerlendirilebilmesine katkı sağlayabilir.

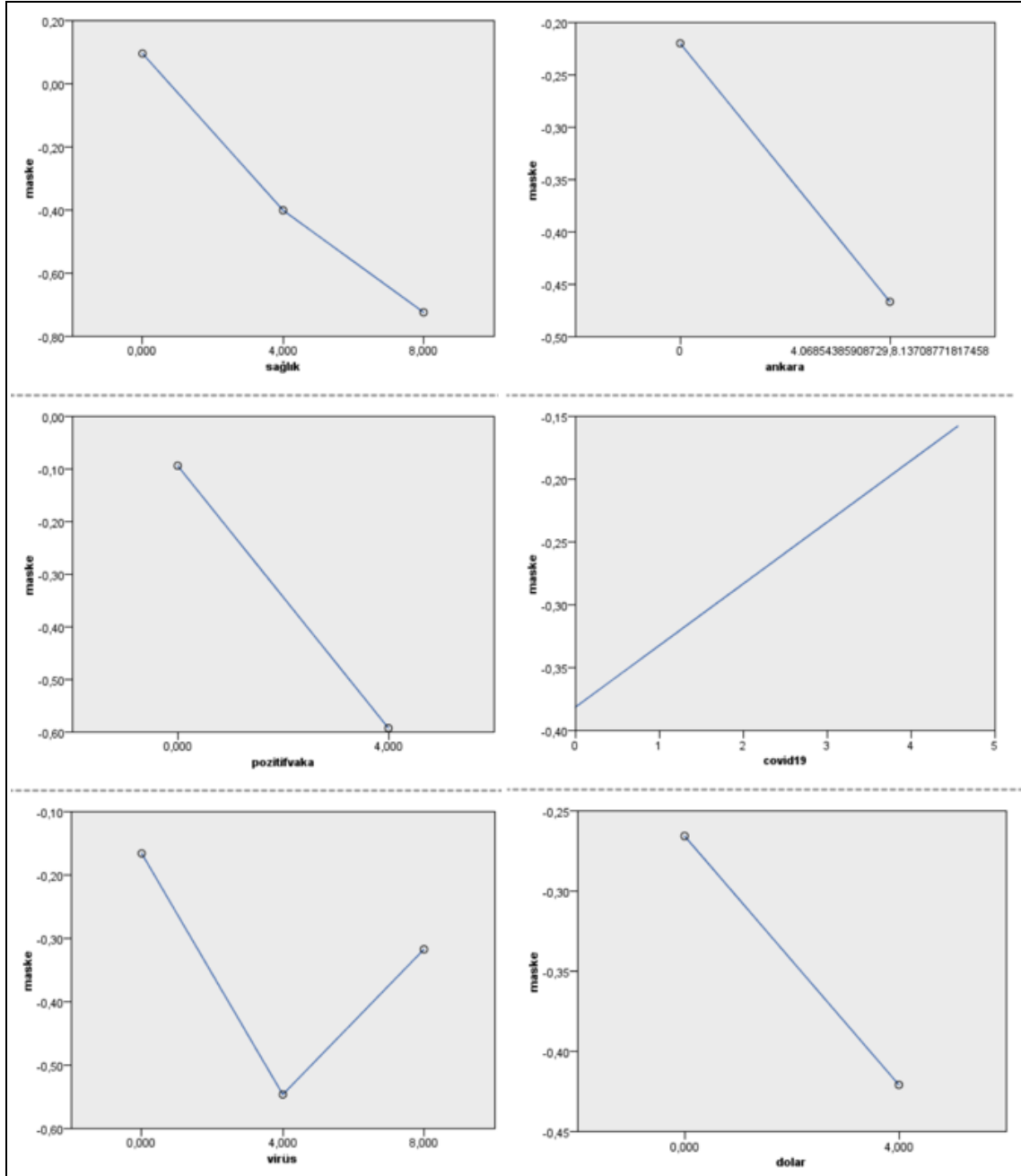


ŞEKİL 5: #Covid_19, #Maske, #bilimkurulu ve Ankara Wuhan gündemlerine ilişkin otomatik lineer model sonuçları.



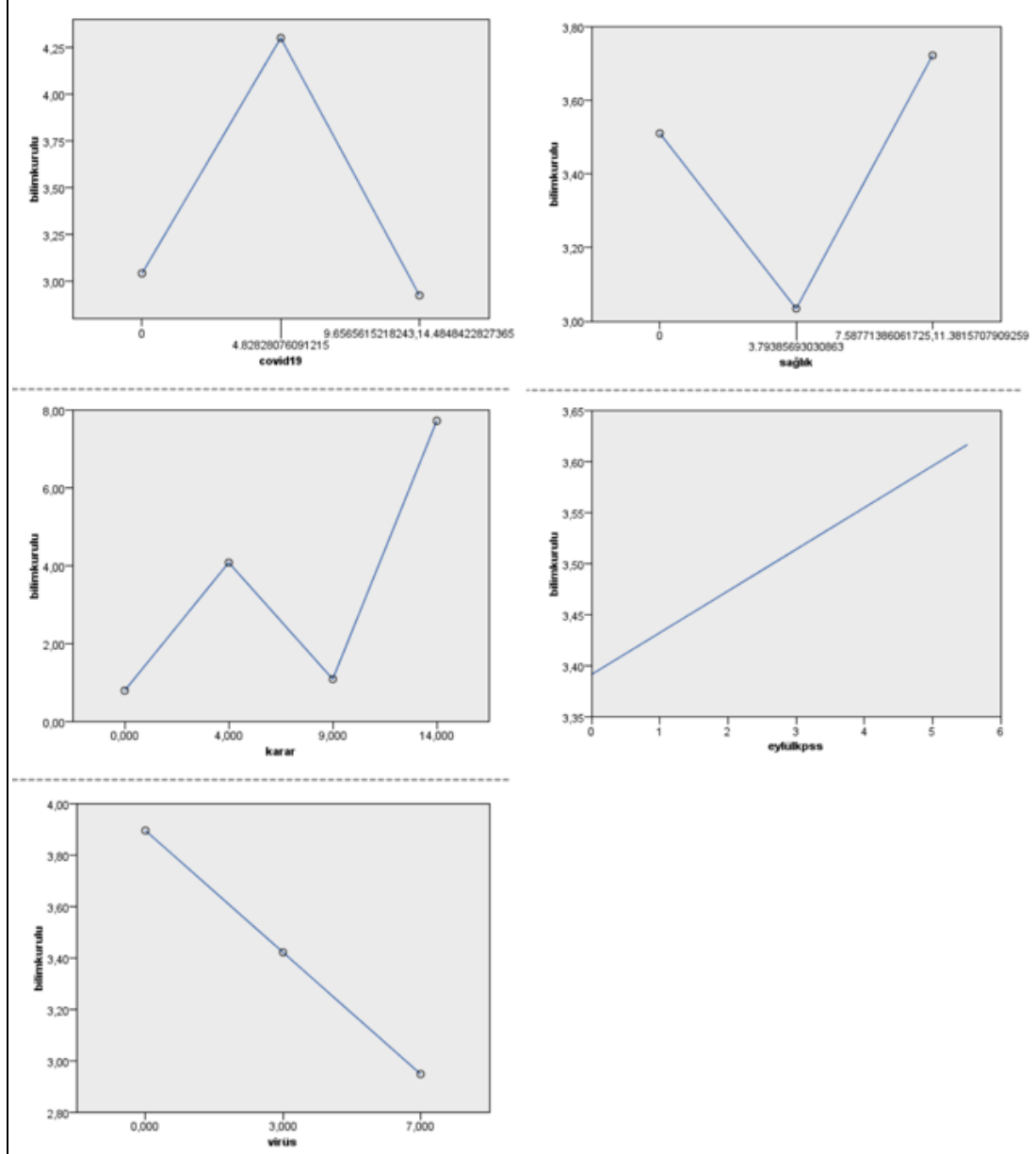
Not: x eksenini: sırasıyla yukarıdan aşağıya güvenlik, maske, yemek, pozitif vaka, pandemi, çalışan, eş, hayat, fedakarlık ve ölüm. y eksenini: COVID-19

ŞEKİL 6: COVID-19 bağımlı değişkeni ile bağımsız değişkenler arasındaki linear ve linear olmayan ilişkiler.



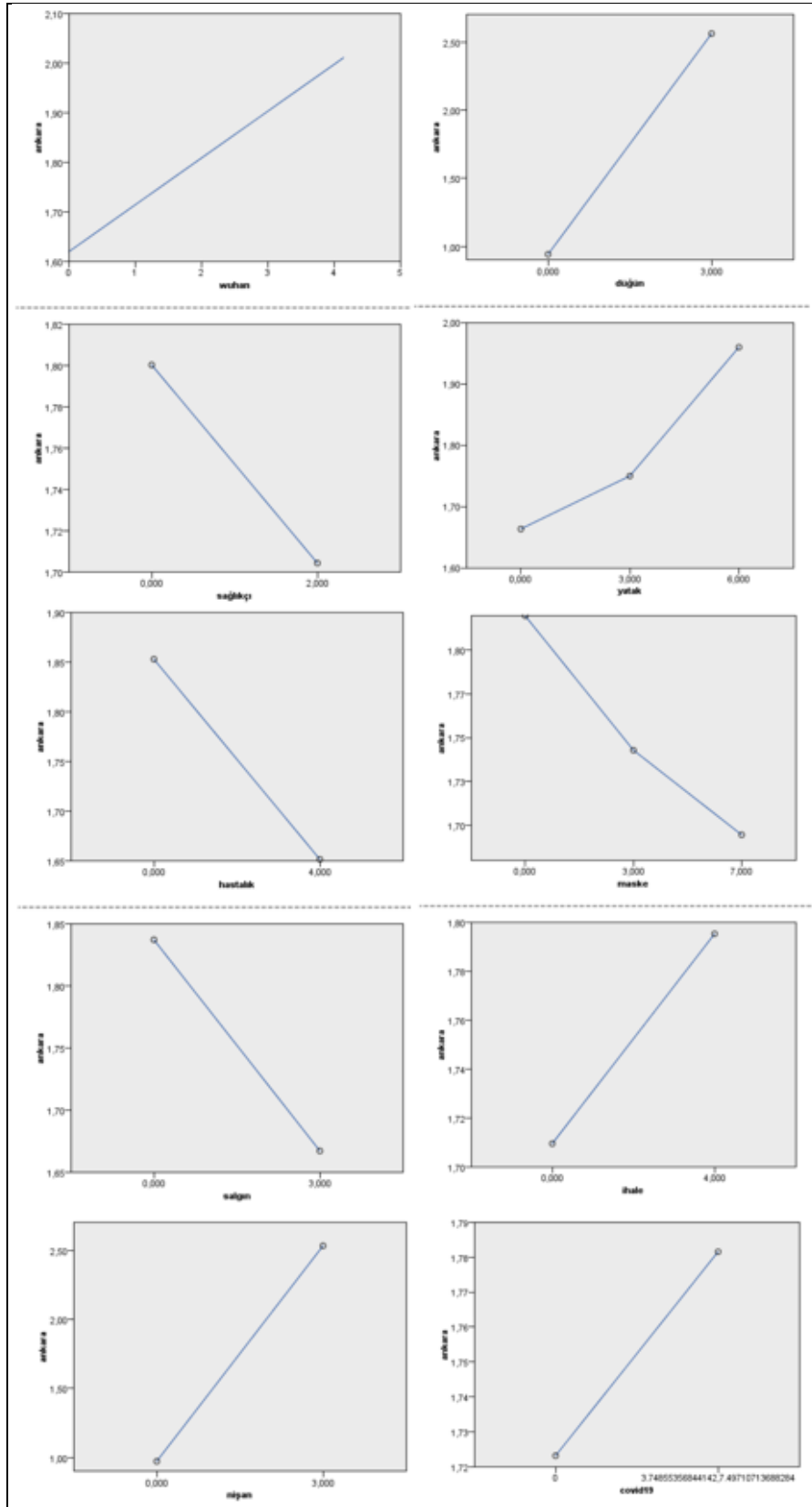
Not: x eksen: sırasıyla yukarıdan aşağıya sağlık, pozitif vaka, virüs, ankara, COVID-19 ve dolar. y eksen: maske

ŞEKİL 7: Maske bağımlı değişkeni ile bağımsız değişkenler arasındaki linear ve linear olmayan ilişkiler.



Not: x eksen: sırasıyla yukarıdan aşağıya COVID-19, karar, virüs, sağlık ve eylül kpss. y eksen: Bilim Kurulu

ŞEKİL 8: Bilim Kurulu bağımlı değişkeni ile bağımsız değişkenler arasındaki lineer ve lineer olmayan ilişkiler.



Not: x eksenleri: sırasıyla yukarıdan aşağıya Wuhan, sağlıkçı, hastalık, salgın, nişan, düğün, yatak, maske, ihale ve COVID-19. y eksenleri: Ankara

ŞEKİL 9: Ankara bağımlı değişkeni ile bağımsız değişkenler arasındaki linear ve linear olmayan ilişkiler.

TARTIŞMA

Son zamanlarda metin madenciliği ile ilgili yapılan çalışmalara baktığımızda, Ayan ve ark. 2018 Şubat ayı boyunca 20.438 adet tweet üzerinden Twitter kullanıcılarının kozmetik, pazaryeri ve elektronik sektöründe faaliyet gösteren, rakip firmalar hakkında yaptıkları paylaşımlar ve bu firmaların kurumsal Twitter hesaplarından yapmış oldukları paylaşımlar sosyal medya madenciliği yöntemi ile analiz etmiştir.⁹ Firmaların, Twitter etkinliğinin ölçülmesi amacıyla tweet sayısı, tweet değeri (olumlu, olumsuz ve nötr), takipçi kazanımı, yanıt sayısı, retweet sayısı ve beğeni sayısı başlıklarından oluşan değişkenlere ait bir başarı sıralaması yapılmış ve en başarılı firmanın kozmetik, en düşük sıralamaya sahip firmanın ise elektronik sektöründen olduğunu tespit etmişlerdir. Yapılan duygu analizi sonuçlarına göre olumlu tweet oranının, kozmetik firmaları için daha fazla olduğu sonucuna ulaşılmıştır. Küçükkartal, Twitter üzerinden vaccine (aşı) anahtar kelimesini içeren 10.000 adet İngilizce tweet çekmiştir.¹⁰ Veriler, Python yazılım programı aracılığıyla metin madenciliği teknikleri ile analiz edilmiştir. Analiz sonuçlarına göre kelime bulutunda; vaccine (aşı), flu (grip), hpv (HPV), tumor (tümör), ebola (Ebola), dementia (bunaklık) ve polio (çocuk felci) kelimeleri en çok geçen kelimelerdir. Duygu analizi sonuçlarına göre de tweetlerin yaklaşık %42'si pozitif, %14'ü negatif ve %43'ü ise nötr olduğu tespit edilmiştir. Kızılkaya ve Oğuzlar, Twitter'dan ÖSYM hakkında atılan 1.824 adet tweet çekerek, sosyal ağın yapısını detaylı olarak incelenmiş ve görselleştirme yaparak önemli algoritmalar (Fruchterman-Reingold ve Harel-Koren Fast Multiscale algoritmaları) hesaplamıştır.¹¹ Atılan tweetlerde, 713 adet negatif öğelere sahip olan tweet ve 618 adet pozitif öğe içeren tweet tespit etmiştir. Kelime üçlülere incelendiğinde, en sık kullanılan 3 kelime; öğrenci-seyyar-yerleştirme 115 kez, seyyar-yerleştirme-merkezi 115 kez, kimlerin-hakkı-yendi 90 kez, çok-vahim-iddia 90 kez ve ÖSYM-sınıfta-kaldı 83 kez birlikte kullanılmıştır. Yapılan analiz sonuçları, son zamanlarda ÖSYM'nin imajının sarsıldığını ve kuruma güven duyulmadığını ortaya çıkarmıştır. Sarıkaya ve ark. 1.7.2016 ve 1.7.2017 tarihleri arasında paylaşılan, içinde "samsung" ve "apple" kelimelerinin geçtiği, 1.000 adet Twitter mesajı üzerinden R yazılım programı geliştirme ortamı R Studio aracılığıyla metin madenciliği analiz tekniklerini uygulamıştır.¹² Metin veri setinin metin ön işleme sonrası, kelime bulutu ve dendrogram grafiği oluşturularak metinden bilgi çıkarımı yapılması sağlanmıştır. Yapılan işlemler sayesinde, kelimeler arasındaki ilişkilerin daha açık bir şekilde gözlemlendiği tespit edilmiştir. Albayrak ve ark. Twitter'da 10.10.2017 ve 12.10.2017 tarihleri arasında paylaşılan "bedelli askerlik geliyor" anahtar sözcüğü içeren 12.739 adet tweet çekmiştir.¹³ Python dili aracılığıyla, metin ön işleme ve metin madenciliği analiz tekniklerinden duygu analizini gerçekleştirmiştir. Yapılan duygu analizi sonucunda, tweetlerin %16 oranında pozitif, %5 oranında negatif ve %79 oranında ise nötr olduğu tespit edilmiştir. Sonuç olarak, karar alıcı mercilerin tutumunu, sosyal medyadaki tepkiler şekillendirebilecek ve kitleler sesini bu şekilde çok daha hızlı ve etkili bir şekilde duyurabileceği ifade edilmiştir.

Biz de bu çalışmada, metin madenciliği teknikleri ile işlenmiş veri seti üzerinden gerçekleştirdiğimiz Kelime bulutu, duygu analizi ve otomatik linear model analizleri ile metin veri setinde gizli kalan bilginin ve kelimeler arasındaki örtük ilişkilerin açığa çıkartılması ve metinden bilgi çıkarımı yapılmasını sağladık. Böylelikle, COVID-19 gündemine ilişkin önemli başlıklar altında atılan tweetlerle dikkate alınması gereken önemli faktörlerin tespit edilmesi sağlanmıştır.

SONUÇ

Yapılan analizler sonucunda, çalışmaya dâhil edilen Twitter kullanıcılarının tweetlerinde esas itibarıyla maske, COVID-19, Ankara Wuhan ve Bilim Kurulu olmak üzere 4 kelime üzerinde odaklandıkları ve yaptıkları paylaşımların bu 4 kelime etrafında şekillendiği belirlenmiştir. Duygu analizi sonuçlarında, bireylerin özellikle COVID-19 kelimesini duymaları ya da bundan bahsedilmesinin kendilerinde negatif bir duygu oluşturduğu sonucuna varılmıştır. Ayrıca Bilim Kurulu kararlarının, halkın üzerinde dikkate değer bir etki bıraktığı ve Ankara'da salgın artışının Ankara'yı Türkiye'nin salgın merkezi hâline getirmesi beklenen bir durum olduğunu ortaya koymuştur. Otomatik linear model sonuçlarına göre Ankara gündemine ilişkin

tahmin denklemi %97,20'lik bir açıklanan varyasyon ile en iyi tahmin edicidir. Ayrıca bu sonuçların ortaya çıkmasında Ankara'nın politika, hükümet ve karar merkezi olması ve bundan dolayı Twitter kullanıcılarının bu yönde daha fazla içerikli tweet atma eğiliminde olması belirleyici bir etmendir. Bu nedenle Ankara gündemine ait Otomatik Linear Model denklemine etki eden **yatak, salgın, Wuhan, hastalık, sağlıkçı, maske, düğün, nişan, ihale** ve **COVID-19** değişkenleri halkın bu salgına bakış açısı ve düşünceleri hakkında özetleyici bilgi üretmektedir.

Finansal Kaynak

Bu çalışma sırasında, yapılan araştırma konusu ile ilgili doğrudan bağlantısı bulunan herhangi bir ilaç firmasından, tıbbi alet, gereç ve malzeme sağlayan ve/veya üreten bir firma veya herhangi bir ticari firmadan, çalışmanın değerlendirme sürecinde, çalışma ile ilgili verilecek kararı olumsuz etkileyebilecek maddi ve/veya manevi herhangi bir destek alınmamıştır.

Çıkar Çatışması

Bu çalışma ile ilgili olarak yazarların ve/veya aile bireylerinin çıkar çatışması potansiyeli olabilecek bilimsel ve tıbbi komite üyeliği veya üyeleri ile ilişkisi, danışmanlık, bilirkişilik, herhangi bir firmada çalışma durumu, hissedarlık ve benzer durumları yoktur.

Yazar Katkıları

Fikir/Kavram: Ebru Temizhan, Mehmet Mendes; **Tasarım:** Ebru Temizhan, Mehmet Mendes; **Denetleme/Danışmanlık:** Mehmet Mendes; **Veri Toplama ve/veya İşleme:** Ebru Temizhan, Mehmet Mendes; **Analiz ve/veya Yorum:** Ebru Temizhan, Mehmet Mendes; **Kaynak Taraması:** Ebru Temizhan, Mehmet Mendes; **Makalenin Yazımı:** Ebru Temizhan, Mehmet Mendes; **Eleştirel İnceleme:** Ebru Temizhan, Mehmet Mendes; **Kaynaklar ve Fon Sağlama:** Ebru Temizhan, Mehmet Mendes.

KAYNAKLAR

1. Vikipedi [Internet]. [cited 06.10.2020]. COVID-19 pandemisi. Available from: [\[Link\]](#)
2. WHO, COVID-19 Dashboard. WHO. Available from: [\[Link\]](#) Erişim Tarihi: 06/10/2020
3. HDX [Internet]. [cited 5, 2020]. Novel coronavirus (COVID-19) cases data. Available from: [\[Link\]](#)
4. Course Hero [Internet]. ©2021. Course Hero, Inc. [cited 22.6.2020]. R (4.0.2). R version 4.0.2 -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit). Available from: [\[Link\]](#)
5. R Studio (1.2.5033). RStudio Version 1.2.5033 - ©2009-2020 RStudio, Inc. [\[Link\]](#)
6. SPSS (Ver. 19). IBM SPSS Statistics Version 19. Licensed Materials. Property of IBM Corp. Copyright IBM Corporation and its licensors 1989, 2010.
7. Şeker ŞE. [Metin madenciliği]. YBS Ansiklopedi. 2015;2(3):30-32. [\[Link\]](#)
8. Makhabel B, Mishra P, Danneman N, Heimann R. Module 1: Learning data mining with R. Chapter 9: Graph mining and network analysis. Social network mining. Module 2: R data mining blueprints. Chapter 3: Visualize diamond dataset. Word cloud. Module 3: Social media mining with R. Chapter 1: Going viral. Social media mining using sentiment analysis. R: Mining Spatial, Text, Web and Social Media Data. 1st ed. Birmingham: Packt Publishing; 2017. p.246, p.99, p.1.
9. Ayan B, Can M, Gürsoy UT. [Measuring the effectiveness of social media: an investigation of companies' twitter use]. BAİBÜ Sosyal Bilimler Enstitüsü Dergisi. 2020;20(1):121-146. [\[Crossref\]](#)
10. Küçükkartal HK. [Applying text mining methods to twitter data]. ESTUDAM Bilişim Dergisi. 2020;1(2):10-3. [\[Link\]](#)
11. Kızılkaya YM, Oğuzlar A. [Effectiveness of tweets with social networks]. U.Ü. Sosyal Bilimler Enstitüsü Dergisi. 2017;10(2):135-150. [\[Link\]](#)
12. Doğan B, Sarıkaya U, Aktaş A. [Social network mining with R]. Marmara Fen Bilimleri Dergisi. 2017;29(3):94-101. [\[Link\]](#)
13. Albayrak M, Topal K, Altıntaş V. [Data analysis on social media: twitter]. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi. 2017;22(Kayfor15 Özel Sayısı):1991-8. [\[Link\]](#)