

(Geliş Tarihi / Received Date: 18.11.2020, Kabul Tarihi / Accepted Date: 23.01.2023)

Metin Madenciliği Yöntemleri ile Twitter Verilerinden Bilgi Keşfi

Ayşe BEŞKİRLİ^{*1,2}, Eyyüp GÜLBANDILAR¹, İdiris DAĞ¹

¹ Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Böl., 26480, Eskişehir

² Burdur Mehmet Akif Ersoy Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 15000, Burdur

Anahtar Kelimeler:

Metin Madenciliği,
Duygu Analizi,
Twitter Verileri,
Bilgi Keşfi.

Özet: Son zamanlarda bilişim teknolojilerinin gelişmesiyle birlikte veri miktarı gün geçtikçe artmaktadır. Bir verinin bilgiye dönüşmesi için verilerin analiz edilmesi, yorumlanması, işlenmesi ve bunlara bağlı olarak verilerden anlamlı bilgilere erişmek için kullanılan bir yöntem olan metin madenciliğinin önemi de artmaktadır. Metin madenciliği yönteminde, veri kaynağı olarak metinler ele alınmaktadır. Metin madenciliği, metinlerden istenilen bilgiye erişmek amacıyla kullanılmaktadır. Bu çalışmada, ilk olarak literatürde önemli bir yere sahip olan metin madenciliği yöntemi ile twitter üzerinden aşı (vaccine) ile ilgili iki farklı veri elde edilmiştir. Daha sonra ise verilerin duygu analizleri yapılmıştır. İlk veri seti covid19 aşısı 3. faz denemeleri esnasındaki elde edilen twitter verilerinden oluşmakta iken ikinci veri seti covid19 aşısının 3. faz denemeleri esnasında seri üretim duyurusu yapıldıktan sonra elde edilen twitter verilerinden oluşmaktadır. Bu verilere göre yapılan duygu analiz sonuçlarında önemli farklılıkların olduğu tespit edilmiştir.

Information Discovery from Twitter Data with Text Mining Methods

Keywords:

Text Mining,
Sentiment Analysis,
Twitter Data,
Information Discovery.

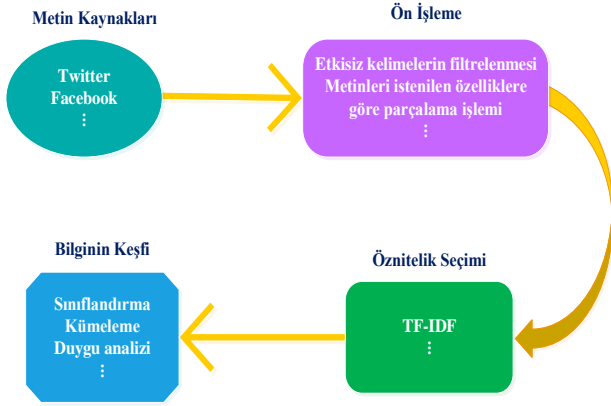
Abstract: Recently with the development of information technologies, the amount of data is increasing day by day. Text mining, which analyze, interpret, process data and access meaningful information from the data, is increasing in importance. As a result of text mining, it is possible to transform data into information. In the text mining method, texts are considered as data sources. Text mining is used to access the information requested from the texts. The data are collected from twitter. First set of the data is collected during the 3rd phase trials of the covid19 vaccine, whereas second set of data is collected after the serial production announcement of the covid19 vaccine during the 3rd phase. Then, emotional analysis of the data was made. Significant differences were found in the emotion analysis, according to the text mining results.

1. GİRİŞ

Günümüzde kullanımı oldukça yaygınlaşmaya başlayan sosyal medya, genellikle bilgi paylaşımının yapıldığı bir platformdur. Bilgi paylaşımı arttıkça üretilen bilgi miktarı da artmaktadır. Bu durum verilerin analiz edilmesini ve erişilmek istenen bilgiye ulaşılmasını zorlaştırmaktadır. Araştırmacılar, sosyal medya platformlarından bilgiye erişebilmek için verilerin işlenmesi ve analizinin gerçekleştirilmesi aşamasında genellikle metin madenciliği yöntemini kullanmaktadırlar. Herhangi bir dokümandaki kelime analizlerinin yapılması, metinleri gruplama, sınıflama ya da kategorilerine ayırma gibi birçok uygulama alanları vardır [1]. Metin madenciliği tıpkı veri madenciliğinde olduğu gibi verinin işlenerek anlamlı hale getirilmesi işlemidir. Metin madenciliğinde veri kaynağı olarak metinler ele alınmaktadır. İlk olarak metin

kaynaklarından amaca uygun veri seçimi yapılır. Daha sonra tekrar eden ve etkisiz kelimeler filtrelenir ve metinler istenilen özelliklere göre parçalara ayırma işlemi gerçekleştirilir. Aynı zamanda metinler noktalama işaretlerinden, sayısal verilerden arındırılmakta ve metinlerdeki büyük harfler küçük harflere dönüştürülmektedir. Böylelikle metin analizi için bir ön işleme gerçekleştirilmiş olur. Metin verileri ön işleme aşamasını geçtikten sonra gösterim aşamasında bir kelimenin metin içerisinde ne kadar önemli olduğunu istatistiki olarak değerlendirmek için terim sıklığı ve ters doküman sıklığı (TF-IDF) olarak adlandırılan ağırlıklandırma işlemi gerçekleştirilir [2]. Son olarak birçok aşamadan geçip metin madenciliği için kullanıma hazır hale getirilen veriler metin kaynaklarından bilgi keşfi için sınıflandırma, kümeleme, duygu analizi gibi yöntemlere uygulanmaktadır. Metin madenciliğinde bilgiye erişim adımları Şekil 1'de gösterilmiştir.

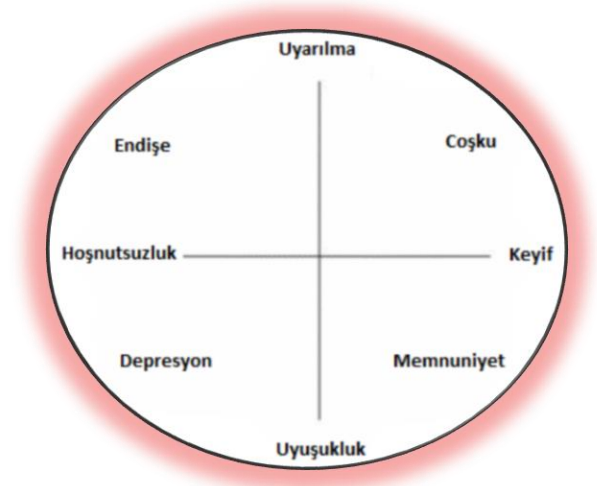
*İlgili yazar: abeskirli@mehmetakif.edu.tr



Şekil 1. Metin madenciliğinde bilgiye erişim [3]

2. DUYGU ANALİZİ

Günümüzde twitter, facebook, instagram gibi sosyal medya platformları bilgi elde etmek ve aynı zamanda bilgi paylaşmak için kullanılmaktadır. Sosyal medya kullanıcıları metin, resim gibi çeşitli şekillerde veri paylaşımı yaparak bilgilendirme yapmakta, duygularını paylaşmakta ya da bir konu hakkındaki fikirlerini sunmaktadırlar. Bundan dolayı metin madenciliği ya da fikir madenciliği için sosyal medya platformları zengin bir veri kaynağı olmaktadır. Sosyal medya içerikleri sadece meydana gelen olay ile ilgili bilgileri değil, onları oluşturan kullanıcıların duygularını da içerir. Kullanıcıların belirli bir olay hakkındaki düşünceleri metin analizi yoluyla çıkarılır ve daha sonra elde edilen analize göre duygular tahmin edilir. Ancak buradaki verilerden bilgi keşfi yapılması için çeşitli metin madenciliği yöntemleri kullanılmaktadır. Bunlardan bir tanesi de duygu analizi yöntemidir. Duygu analizi literatürde fikir madenciliği olarak da adlandırılmaktadır [4-5]. Metinler içerisindeki gizli olan duygu ve düşüncelerin pozitif, nötr ve olumsuz olup olmadığının analizi için kullanılmaktadır. Duygu, “duygu tarafından harekete geçirilen bir tutum, düşünce veya yargı” olarak tanımlanır [6-7]. Duygu analizi ise tahminlere olan güvenin bir ölçüsü olmakla birlikte metinde somutlaşan temel duygusal özellikleri sunan bir görselleştirmedir [8]. Literatürde, duygusal durumları tanımlamak ve karşılaştırmak için duygusal modeller önerilmektedir. Bu modellerde genellikle duyguları 2B düzlemde yatay ve dikey olarak konumlandırmak için Russell’in duygu etki modeli kullanılarak duygusal boyutlar oluşturulur. Şekil 2’de verilen Russell’in duygu etki modelinde coşku olabilmesi keyif ve uyarılmanın olması ile gerçekleşmektedir [9-10].



Şekil 2. Russell’in duygu etki modeli [11]

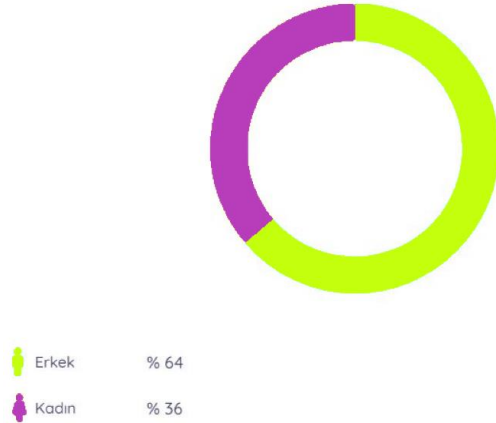
Literatürde duygu analizi ile ilgili birçok çalışma yer almaktadır. Bu çalışmalardan bazıları; Sailunaz ve Alhaji [12], çalışmalarında twitter üzerinde paylaşım yapılan metinlerden kullanıcıların ifade ettikleri duyguların analizi için belirli bir konuda tweet ve bu tweetlere verilen yanıtların oluşturduğu veri kümesi oluşturmuşlar ve duygu analizini gerçekleştirmişlerdir. Albayrak vd. [13], çalışmalarında twitterde konuşulan güncel bir konu üzerine veri seti oluşturmuşlardır. Metin madenciliği yöntemleri ile işlenmiş olan verilerin aynı zamanda duygu analizlerini gerçekleştirerek elde edilen sonuçları görselleştirmişlerdir. Küçükkartal [14], çalışmasında twitterde konuşulan bir konu üzerine veri seti oluşturmuştur. Metin madenciliği yöntemleri ile bu verileri işlemiştir. Kullanıcıların paylaştıkları tweetlerin pozitif, nötr ya da negatif olup olmadığını duygu analizi yöntemi ile gerçekleştirmiştir. Yüksel ve Tan [15], çalışmalarında foursquare platformunda bulunan mekanlar ile ilgili veri seti oluşturmuşlardır. Oluşturulan veri setini metin madenciliği teknikleri ile işlemişlerdir. Aynı zamanda kullanıcıların mekanlara yaptıkları yorumların olumlu, olumsuz ya da nötr olup olmadığını karşılaştırmasını duygu analizi yöntemi ile gerçekleştirmişlerdir. Böylelikle kullanıcılar tarafından foursquare platformunda yer alan mekanların nasıl değerlendirildiklerini tespit etmişlerdir. Mishra vd. [16], çalışmalarında güncel bir konu hakkındaki twitter kullanıcılarının duygu ve düşüncelerin tespiti için twitter platformu üzerinden verileri toplamış ve metin madenciliği yöntemleri ile işleyerek toplanan fikirlerin pozitif, nötr ve olumsuz olup olmadığını tespiti için duygu analizi yöntemini kullanmışlardır. Caballero vd. [17], çalışmalarında web tabanlı duygu analizi aracı ile twitter platformundan verileri toplamışlardır. Toplanan verilere göre duygu analizini gerçekleştirmişler ve aynı zamanda Russell’in duygu etki modeline göre görselleştirmişlerdir.

3. DENEYSEL SONUÇLAR

Bu çalışma, twitter sosyal platformu üzerinde yapılan konuşmalar içerisinde geçen belirli bir kelimenin duygu analizi gerçekleştirilmiştir. Twitter’da şu anda gündemde olan aşı kelimesinin İngilizcesi olan “vaccine” kelimesi aranmış ve konuşmalar içerisinde vaccine kelimesi içeren 10.000 tweet seçim işlemi gerçekleştirilmiştir. Yapılan analizler ile twitter kullanıcılarının orijinal konuşma, konuşmalara verilen cevaplar ve retweet paylaşımı yüzdeleri oranları, konuşmalar içerisinde geçen duygu belirtileri olan pozitif, negatif ve nötr belirtilerinin yüzdeleri, kullanıcıların cinsiyet yüzdeleri, kullanıcıların hangi tür cihazlardan mesaj attıklarının yüzdeleri görselleştirilerek sunulmuştur. Bu çalışmada iki farklı zamanda elde edilen veriler kullanılmıştır. Birinci veriler, covid19 aşı çalışmaları yapılırken elde edilmiştir. İkinci veriler ise covid19 aşısının 3. faz denemeleri devam ederken seri üretim duyurusu yapıldıktan sonra elde edilmiştir. Her iki veri üzerinde analizler gerçekleştirilmiştir.

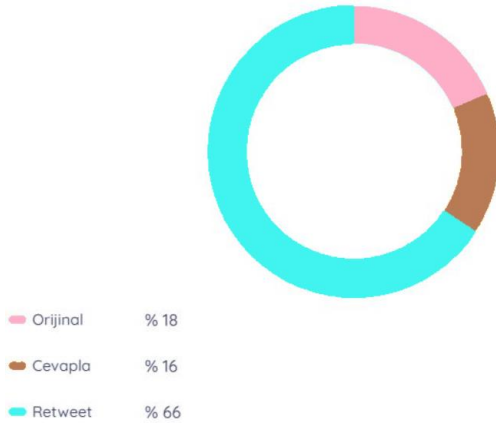
3.1. Birinci veriler ile elde edilen duygu analizleri

Birinci verileri oluşturan kişilerin yüzdeleri dağılımları Şekil 3’te verilmiştir. Toplanan verilerin %64’ünün erkek kullanıcılara ait olduğu görülmüştür.



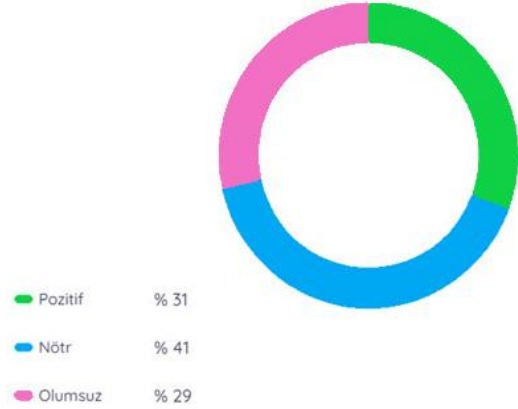
Şekil 3. Cinsiyet analizleri

Bu veriler içerisinde orijinal olarak yazılan sadece %18 iken cevap yazanlar ise %16’dır. Bu yazılanları retweet olarak paylaşanlar ise %66 olduğu Şekil 4’te görülmüştür.



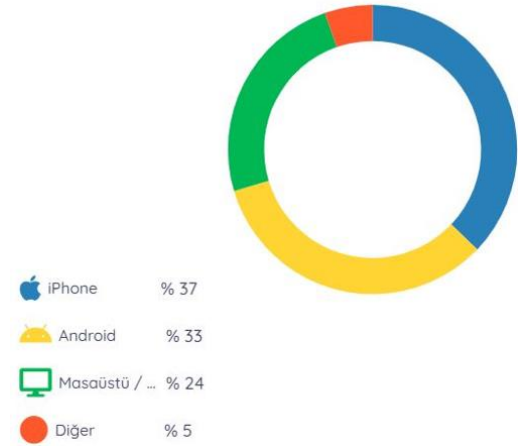
Şekil 4. Konuşma tipleri analizi

Vaccine kelimesinin konuşmalar içerisindeki duygu yansımalarının analizinde pozitif düşünce %31 olmuş, negatif düşünce %29 olmuş, nötr düşünce ise %41 olmuştur ve Şekil 5’te verilmiştir.



Şekil 5. Düşünce analizleri

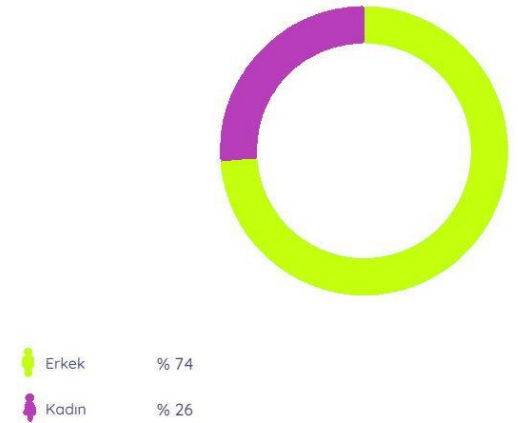
Konuşmaları cep telefonu ile gerçekleştirenlerin oranı %70 iken bilgisayar ile gerçekleştirenlerin ise %24’tür. Analizler Şekil 6’da grafiksel olarak verilmiştir.



Şekil 6. Kullanılan platformların analizleri

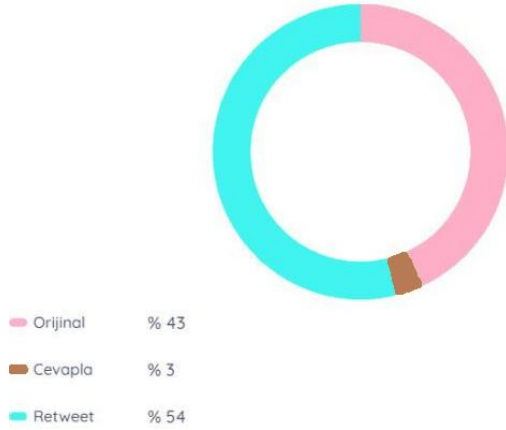
3.2. İkinci veriler ile elde edilen duygu analizleri

İkinci verileri oluşturan kişilerin yüzdeleri dağılımları Şekil 7’de verilmiştir. Toplanan verilerin %74’ünün erkek kullanıcılara ait olduğu görülmüştür.



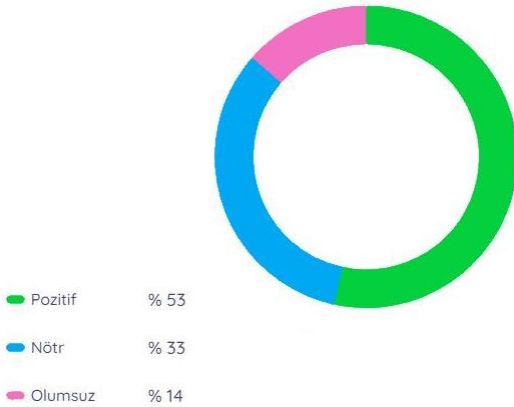
Şekil 7. Cinsiyet analizleri

Bu veriler içerisinde orijinal olarak yazılan sadece %43 iken cevap yazanlar ise %3'tür. Bu yazılanları retweet olarak paylaşanlar ise %54 olduğu Şekil 8'de görülmüştür.



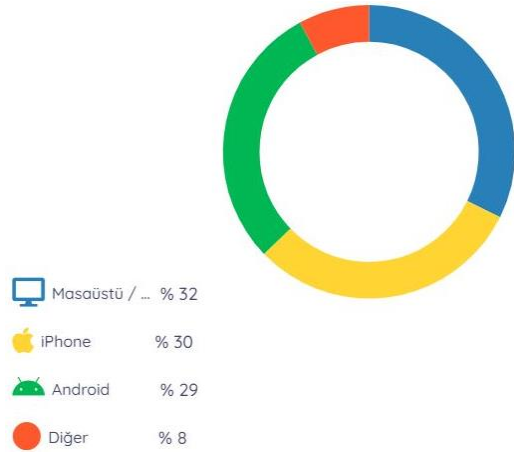
Şekil 8. Konuşma tipleri analizi

Vaccine kelimesinin konuşmalar içerisindeki duygu yansımalarının analizinde pozitif düşünce %53 olmuş, negatif düşünce %33 olmuş, nötr düşünce ise %14 olmuştur ve Şekil 9'da verilmiştir.



Şekil 9. Düşünce analizleri

Konuşmaları cep telefonu ile gerçekleştirenlerin oranı %62 iken bilgisayar ile gerçekleştirenlerin ise %29'dur. Analizler Şekil 10'da grafiksel olarak verilmiştir.



Şekil 10. Kullanılan platformların analizleri

3.3. Her iki verinin karşılaştırmalı analizleri

Covid19 aşısının seri üretim çalışmalarından önce elde edilen verilerin duygu analizlerinin yüzdeleri ile covid19 aşısının seri üretim duyurusundan sonra elde edilen verilerin duygu analizlerinin yüzdeleri Tablo 1'de kıyaslanmıştır.

Tablo 1. Tüm verilerin duygu analizleri

Aşı seri üretim duyurusu	Olumlu (%)	Nötr (%)	Olumsuz (%)
Öncesi	31	41	29
Sonrası	53	33	14

Tablo 1 incelendiğinde aşının seri üretim duyurusundan önce yapılan konuşmaların duygusal içeriği %31 olumlu, %29 olumsuz olurken nötr düşüncedekiler ise %41 olmuştur. Ancak aşı çalışmalarının seri üretim duyurusu yapıldıktan sonra ise bu oranlarda değişimler olduğu fark edilmiştir. Olumlu düşünceler %53'e yükselirken olumsuz düşüncelerin oranı ise %14'e gerilemiştir. Nötr düşüncedekiler ise %33'e gerileme gösterdiği görülmüştür.

4. SONUÇ

Bu çalışmada, aşı kelimesinin İngilizcesi olan vaccine kelimesi ile ilgili twitter platformu üzerindeki konuşmalardan veriler toplanarak vaccine ile ilgili aşı duyurusu öncesi ve sonrası olmak üzere iki farklı veri seti oluşturulmuştur. Oluşturulan veri setleri üzerinden cinsiyet analizi, konuşma tipleri analizi, düşünce analizi, kullanılan platformların analizi yapılmıştır. Yapılan analiz sonuçları şekiller ile görselleştirilip tablolarda tüm verilerin duygu analizleri kıyaslanmıştır. Kıyaslama sonucunda aşı çalışmaları sürecinde iken sosyal medya konuşmaları olumlu ve olumsuz düşünceler neredeyse aynı oranda iken nötr düşüncelerin daha fazla oranda olduğu görülmüştür. Covid19 için aşı seri üretimine geçilecek denildikten sonra sosyal medya kullanıcılarının aşı ile ilgili konuşmaları olumlu yönde önemli bir artış olduğu fark edilmiştir. Olumsuz konuşmaların oranında önemli bir azalma olduğu elde edilen analizlerden görülmüştür.

KAYNAKÇA

- [1] Gürsakal, N. 2014. Büyük Veri. Baskı, Bursa: Dora.
- [2] Aninditya, A., Hasibuan, M. A., Sutoyo, E. 2019. Text Mining Approach Using TF-IDF and Naive Bayes for Classification of Exam Questions Based on Cognitive Level of Bloom's Taxonomy. Paper presented at the 2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS).
- [3] Hu, X., Liu, H. 2012. Text analytics in social media. In Mining text data (pp. 385-414): Springer.
- [4] Seker, S. E. 2016. Duygu Analizi (Sentimental Analysis). YBS Ansiklopedi, 3 (3), 21-36.

- [5] Fang, X., Zhan, J. 2015. Sentiment analysis using product review data. *Journal of Big Data*, 2 (1), 5. doi:10.1186/s40537-015-0015-2
- [6] Christi, J., Jain, G. 2019. Sentiment Categorization through Natural Language Processing: A Survey.
- [7] Smith, A. N. 2019. Studies of sentiment: John White Alexanders depictions of gilded age men and women. (UMI thesis),
- [8] Clavel, C., Callejas, Z. 2015. Sentiment analysis: from opinion mining to human-agent interaction. *IEEE Transactions on affective computing*, 7 (1), 74-93.
- [9] Liljander, V., Bergenwall, M. 1999. Consumption-based emotional responses related to satisfaction: Swedish School of Economics and Business Administration Helsinki: Helsinfors.
- [10] Liljander, V., Strandvik, T. 1997. Emotions in service satisfaction. *International Journal of service industry management*.
- [11] Russell, J. A. 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39 (6), 1161.
- [12] Sailunaz, K., Alhajj, R. 2019. Emotion and sentiment analysis from Twitter text. *Journal of Computational Science*, 36, 101003.
- [13] Albayrak, M., Topal, K., Altıntaş, V. 2017. Sosyal Medya Üzerinde Veri Analizi: Twitter. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22 (Kayfor 15 Özel Sayısı), 1991-1998.
- [14] Küçükkartal, H. K. 2020. Twitter'daki Verilere Metin Madenciliği Yöntemlerinin Uygulanması. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, 1 (2), 10-13.
- [15] Yüksel, A. S., Tan, F. G. 2018. Metin madenciliği teknikleri ile sosyal ağlarda bilgi keşfi. *Mühendislik Bilimleri ve Tasarım Dergisi*, 6 (2), 324-333.
- [16] Mishra, P., Rajnish, R., Kumar, P. 2016. Sentiment analysis of Twitter data: Case study on digital India. Paper presented at the 2016 International Conference on Information Technology (InCITE)-The Next Generation IT Summit on the Theme-Internet of Things: Connect your Worlds.
- [17] Caballero, A., Niguidula, J. D., Caballero, J. M. 2017. Analysis and Visualization of University Twitter Feeds Sentiment. Paper presented at the International Conference on Big Data Technologies and Applications.