

SOSYAL MEDYA (TWITTER) PAYLAŞIMLARININ VERİ MADENCİLİĞİ İLE DUYGU ANALİZİ: YALOVA İLİ ÖRNEĞİ

İlknur KOÇYİĞİT
Res. Asst., Yalova University
ilknur.kocyigit@yalova.edu.com.tr

Volkan POLAT
Asst. Prof. Dr., Yalova University
vpolat@yalova.edu.com.tr

Mustafa YILMAZ
Res. Asst., Yalova University
mustafa.yilmaz@yalova.edu.com.tr

Öz

Amaç

Bilgi teknolojilerindeki gelişim toplumların gündelik yaşam rutinlerini etkilemiş, bireylerin iletişim alışkanlıkları teknolojideki gelişim neticesinde değişime uğramıştır. İnternet teknolojisindeki gelişim ve akıllı telefon kullanımının artması sonucunda, bireylerin iletişimleri daha etkileşimli hale gelmiştir. Sosyal medya platformlarının kullanımının artışı ve gelişen bilgi teknolojileri sayesinde çok büyük miktarlarda veri elde edilebilmektedir. Elde edilen bu büyük miktardaki veriler işlenip analiz edilerek anlamlı hale getirilip, geleceğe yönelik çıkarımlar yapılabilmektedir. Böylece geleneksel veri toplama yöntemleri ile elde edilebilecek veri sayısına göre daha fazla veri incelenebilmekte ve daha objektif sonuçlara ulaşılabilmektedir. Bu çalışmanın amacı Twitter kullanıcılarının Yalova ili ile ilgili paylaşımlarının duygu analizi ile incelenerek sınıflandırılmasıdır. Çalışmanın Yalova ilinin şehir pazarlamasına ilişkin stratejilerini belirlemede, sosyal medya faaliyetlerini planlama ve uygulamada karar vericilere katkı sağlayacağı düşünülmektedir.

Model / metodoloji / yaklaşım

Çalışmada Twitter verilerini elde edebilmek için Python programlama dili kullanılmıştır. Ardından elde edilen veriler çeşitli sadeleştirme işlemlerinden sonra Orange veri madenciliği programı aracılığıyla duygu analizi gerçekleştirilmiştir.

Bulgular

Yapılan veri analizleri sonucunda tweetlerde sıklıkla ele alınan kelimeler kelime bulutu ile görselleştirilmiştir. Duygu analizi sonucunda ise Yalova ile ilgili ele alınan verilerde en fazla nötr tweetlerin, ardından pozitif tweetlerin, son olarak ise negatif tweetlerin yer aldığı görülmektedir.

Özgünlük

Çalışma güncel araştırma yaklaşımlarından olan veri madenciliği alanına katkıda bulunmaktadır. Ayrıca sosyal medya paylaşımlarının analizi aracılığıyla Yalova iline ait iç görü elde edilmesi açısından önemlidir.

Anahtar Kelimeler: Metin Madenciliği, Sosyal Medya, Duygu Analizi, Twitter, Yalova

SENTIMENT ANALYSIS OF SOCIAL MEDIA SHARES (TWITTER) WITH DATA MINING: CASE OF YALOVA

Abstract

Purpose

The development in information technologies has affected the daily routines of societies, and the communication habits of individuals have changed. As a result of the development in internet technology and the increase in the use of smartphones, the communication of individuals has become

more interactive. Social media refers platforms where individuals can communicate without the limitation of time and place. Due to the increase in the use of social media platforms and developing information technologies, huge amounts of data can be obtained. The data can be processed, analyzed and made meaningful, so inferences can be made for the future. Thus, more data can be analyzed and more objective results can be achieved compared to the number of data that can be obtained with traditional data collection methods. The aim of this study is to classify Twitter users' posts about Yalova province by analyzing them with sentiment analysis. It is thought that the study will contribute to the decision makers in determining the city marketing strategies, planning and implementing social media activities of Yalova province.

Design/methodology/approach

Python programming language was used to obtain Twitter data in the study. Then, after various simplification processes, sentiment analysis was carried out through Orange data mining program.

Findings

As a result of the data analysis, the words frequently used in the tweets were visualized with a word cloud. As a result of sentiment analysis, it shows that the most neutral tweets, followed by positive tweets, and finally negative tweets in the data about Yalova.

Originality/value

The study contributes to the field of data mining, which is one of the current research approaches. It is also important in terms of obtaining insight into Yalova province through the analysis of social media shares.

Keywords: Text Mining, Social Media, Sentiment Analysis, Twitter, Yalova

Giriş

İnternet teknolojisindeki gelişim ve akıllı telefon kullanımının artması sonucunda, bireylerin iletişimleri daha etkileşimli hale gelmiştir ve sosyal medyanın bu etkileşimde rolü büyüktür. Kullanıcıların oluşturduğu, kurumsal medya (Mandiberg, 2012) olarak tanımlanan sosyal medya bu etkileşimli iletişimin sağlandığı internet teknolojisi tabanlı alanlardır. Temelde sosyal medya; paylaşım, diyalog ve bilgi içeriklerinden oluşan (Elmas, 2019:29) ve bireylerin zaman ve mekân sınırlaması olmadan iletişim kurabildikleri platformlardır. Toplumda çeşitli rollere sahip birçok birey, kurum ve kuruluş sosyal medya platformlarında yer almayı etkileşim açısından bir kazanım (Akdeniz ve Cebeci, 2021) olarak değerlendirmektedir.

Hootsuite ve We Are Social tarafından yayınlanan 2021 Dijital Türkiye Raporu'na göre Ocak 2021 itibariyle Türkiye'de aktif 60 milyon sosyal medya kullanıcısı bulunmaktadır ve bu değer Türkiye toplam nüfusunun %70.8'ine denk gelmektedir. Raporla 16 ve 64 yaş arası bireylerin sosyal medyada günde ortalama 2 saat 57 dakika geçirdikleri belirtilmiştir. Türkiye'de her bir kullanıcı başına düşen sosyal medya hesabı ise 9.4'tür. En çok kullanılan sosyal medya platformları ise sırasıyla Youtube, Instagram, Whatsapp, Facebook ve Twitter'dır (Hootsuite ve We Are Social, Dijital 2021 Türkiye Raporu). Türkiye'deki yüksek sosyal medya kullanım oranı, ilgili platformların birer veri tabanı işlevi görmesini sağlamış ve araştırmalarda daha fazla veriye ulaşarak analizler yapmanın önünü açmıştır.

Türkiye'de de Twitter verileriyle duygu analizi yapılan çalışmalar bulunmaktadır. Akdeniz ve Cebeci (2021) kullanıcıların, Sakarya ve bazı ilçeleri ile ilgili tweetlerini duygu analizi ile inceleyerek, kullanıcıların belediyelere yönelik tutumlarını incelemiştir. Elmas (2019) ise twitter kullanıcılarının Sivas iline paylaşımlarını incelemiş ve duygu analizinde paylaşımları "mutlu, kızgın, şaşkın, üzgün ve tarafsız" olmak üzere beş sınıfta toplamıştır. Kasaba ve Yıldıztepe (2016) Türk filmlerine ilişkin tweetleri duygu analizi ile incelemiştir.

Bu çalışma, Twitter'da Yalova ili ile ilgili paylaşımların duygu analizi ile incelemeyi amaçlamaktadır ve kullanıcıların Yalova ile ilgili paylaştıkları mesajlar "pozitif", "negatif" ve "nötr" olarak sınıflandıracaktır. Çalışmanın Yalova ilinin şehir pazarlamasına ilişkin stratejilerini belirlemede hem

araştırmacılara hem de sosyal medya faaliyetlerini planlama ve uygulamada karar vericilere katkı sağlayacağı düşünülmektedir.

Kavramsal Çerçeve

Günümüzde dünya genelinde günde ortalama 2,5 milyar GB veri üretilmekte ve bunların yaklaşık %80'i metin halindedir (Atan, 2018). Metin madenciliği, metinlerin veri olarak yer aldığı veri madenciliği çalışmalarıdır. Başka bir tanımla, veri kaynağı olarak insanlar tarafından doğal dilde yazılmış metinlerin incelenmesi ve analiz edilebilmesini sağlayan yöntem ve araçlar bütünüdür (Atan,2018). Metin madenciliğin çalışma alanlarından bazıları enformasyon getirmesi, doğal dil işleme, adlandırılmış varlık tanıma, ilişki, kural, olay çıkarımları ve duygu analizidir (Şeker, 2015).

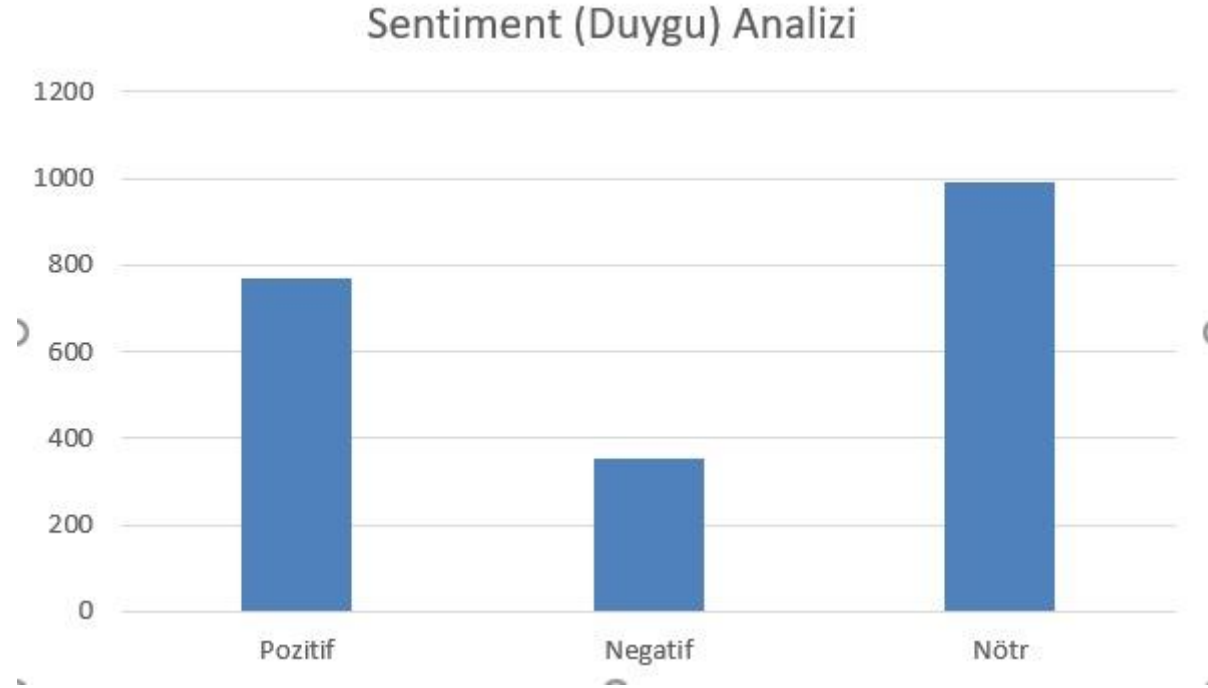
Sosyal medyada metin analizi için en çok tercih edilen yöntemlerden biri duygu analizidir (Giachanou ve Crestani, 2016). Duygu analizi, belirli bir metindeki ifade edilmek istenen duyguyu belirlemeyi amaçlar. Bing (2012) duygu analizini insanların ürünler, olaylar ve organizasyonlar hakkında farklı görüşlerini inceleyen bir araştırma alanı olarak tanımlar. Örneğin duygu analiziyle belirli bir konuya ilişkin tutum ölçülebilir ve pozitif, negatif ve nötr olarak metinler sınıflandırılabilir (Antonakaki vd., 2021). Ayrıca duygu analizi; pazar büyüklüğü (Bollen vd., 2010), rakip analizi, sektöre ilişkin fırsat ve tehditler (Ravi ve Ravi, 2015) dahası işsizlik oranları (Antenucci vd., 2014) tahminlerinde de kullanılmaktadır. Günümüzde bireyler günlük yaşantıları, kullandıkları ürünler, izledikleri film ve diziler, gezdikleri yerler, yedikleri yemekler, mevcut gündem vb. birçok konu hakkında düşüncelerini forumlarda, bloglarda ve mikrobloglarda dile getirmektedir. Birmingham ve Smeaton (2010) duygu analizi çalışmalarını incelemiş ve mikrobloglarda duygu analizin bloglara göre daha kolay olduğunu ortaya koymuşlardır. Buradan yola çıkarak, bu çalışmada duygu analizi uygulanacak sosyal medya platformu olarak bir mikroblog olarak nitelendirilen Twitter seçilmiştir.

Yüksek kullanım oranı ve kullanıcılar tarafından oluşturulan gönderilere kolay erişim sebebiyle Twitter günümüzün önemli veri tabanları arasında yer almaktadır (Giachanou ve Crestani, 2016). Bununla birlikte en dinamik çevrimiçi sosyal ağlardan biridir ve dünya genelinde 650 milyon kullanıcısıyla Instagram ve Facebook'tan sonra en çok kullanılan sosyal medya platformudur (Antonakaki vd., 2021). Türkiye'de ise Twitter %72,5'lük bir kullanım oranına sahiptir (Hootsuite ve We Are Social, Dijital 2021 Türkiye Raporu). Duygu analizinde sınıflandırmaya yardımcı olan Twitter'a özgü bazı kavramlar Giachanou ve Crestani (2016) tarafından şöyle açıklanmıştır;

- Tweet: Twitter'da yayınlanan her bir mesajdır. Başlangıçta 140 karakterle sınırlandırılmış fakat sonrasında karakter sayısı 280'e çıkarılmıştır. Bu mesajların içeriği çoğunlukla kişisel bilgilerden, ürünler veya çeşitli olaylar hakkındaki kişisel görüşlerden, haber, fotoğraf, videolardan oluşmaktadır.
- Mention/Bahsetme: Bir tweette başka bir kullanıcıdan bahsedilmesidir. Kullanıcılar başka kişiye atıfta bulunmak için "@" sembolünü kullanırlar.
- Follower/Takipçi: Bir kullanıcının tweetlerini takip eden diğer kullanıcıları ifade eder. Takip etme Twitter'da diğer kullanıcılar etkileşimde bulunmanın temel yoludur.
- Retweet: Tweetlerin yeniden diğer kullanıcılar tarafından gönderilmesidir.
- Hashtag/etiket: Bir tweetin belirli bir konuyla alaka düzeyini belirtmek için kullanılmaktadır. "#" işareti ile belirtilir.

İlgili literatür incelendiğinde Twitter verileri kullanılarak yapılmış birçok çalışma bulunduğu görülmektedir. Twitter'da duygu analizine ilişkin ilk çalışma Go vd. (2009) tarafından gerçekleştirilmiştir ve yazarlar makine öğrenimi algoritmalarını kullanarak duygu sınıflandırması yapmışlardır. Diakopoulos ve Shamma (2010) 2008 yılında ABD Başkanlık seçimlerine ilişkin bir tartışma yayını sırasında atılmış 1820 tweet ile duyguların zamansal değişimini incelemişlerdir. Mitchell vd. (2013) ise duygu farklılıklarını coğrafi bölgelere göre incelemiştir. Jiang vd. (2011) tweetlerin genellikle kısa ve belirsiz olması sebebiyle duygu sınıflandırmasında sadece mevcut tweeti dikkate almanın yeterli olmadığını öne sürmüş ve hedefe bağlı Twitter duygu sınıflandırmasını önermiştir. Ayrıca alan yazında twitter kullanıcıların çeşitli konularla ilgili kutuplaşmaları inceleyen duygu analizi çalışmaları da bulunmaktadır. Bu konulardan bazıları iklim krizi (Newman, 2017)),

Son olarak verilere duygu analizi uygulanmıştır. Şekil 4 Orange proramı duygu analizi ekranını göstermektedir. Çok dilde duygu analizi seçeneği işaretlenerek Türkçe dili seçildikten sonra veriler duygu analizine tabi tutulmuştur. Duygu analizinde eksi(-) değerler negatif, 0 değer nötr, artı(+) değerler ise pozitif duygu içeren tweetleri ifade etmektedir. Duygu analizi sonucunda nötr 991 tweet,pozitif 771 tweet ve negatif 352 tweet bulunmuştur. Duygu analizi sonuçları tablosu aşağıdaki gibidir.



Şekil 5. Duygu Analizi Sonuçları

Sonuç

Bilgi teknolojilerinin gelişmesi, internet kullanımının artması ve sosyal medya platformlarının insan hayatın içerisindeki öneminin artması sonucunda bireyler ve toplumlar daha etkileşimli hale gelmiştir. Bu kullanım yoğunluğu ve etkileşim sonucunda elde edilen veri miktarı da büyük oranda artış göstermiştir. Son yıllarda ortaya çıkan büyük veri kavramı ve elde edilen bu veri analizleri aracılığıyla etkili stratejiler oluşturulabilmesi için veri madenciliği kavramının da önemi artmaktadır.

Bu çalışmada bireylerin sıklıkla kullandığı sosyal medya platformu olan Twitter üzerinden veriler alınarak veri madenciliği işlemleri uygulanmıştır. Twitter kullanıcılarının Yalova iline ait yaptığı paylaşımlar (tweet) analiz edilerek belirli çıkarımlar yapılmaya çalışılmıştır. Analizler sonucunda Yalova ili ile ilgili nötr tweetlerin daha yoğun olduğu, pozitif tweetlerin ise negatif tweetlerden daha fazla olduğu sonucu elde edilmiştir.

Bu çalışmada analize tabi tutulan veriler 2021 yılına ait verilerdir. Yalova ili ile ilgili yapılacak sonraki çalışmalarda son birkaç yılın verileri ele alınarak daha kapsamlı analiz yapılabilir. Ayrıca bu çalışmada duygu analizi Orange veri madenciliği programı aracılığıyla yapılmıştır. Bundan sonra yapılacak olan çalışmalarda doğal dil işleme işlemleri ve farklı analizler de entegre edilerek daha güvenilir sonuçlar elde edilebilecektir.

Kaynakça

- Akdeniz, F. N. U., & Cebeci, H. İ. (2021) Belediye Hizmetlerin Değerlendirilmesinde Duygu Analizi Yaklaşımı: Sakarya İli Örneği. *Zeki Sistemler Teori ve Uygulamaları Dergisi* 4(2) 127-135
- Antenucci, D., Cafarella, M., Levenstein, M. C., R, C., and Shapiro, M. (2014). Using social media to measure labor market flows. <http://www.nber.org/papers/w20010>. Accessed: 2015-04-10
- Antonakaki, D., Fragopoulou, P., & Ioannidis, S. (2021). A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. *Expert Systems with Applications*, 164, 114006.
- Atan S. (2020). Metin Madenciliği: İmkânlar, Yöntemler ve Kısıtlar. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (31), 220-239.
- Bermingham, A., & Smeaton, A. F. (2010, October). Classifying sentiment in microblogs: is brevity an advantage?. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 1833-1836).
- Bing, L. (2012). Sentiment analysis: A fascinating problem. In *Sentiment Analysis and Opinion Mining*, pages 7–143. Morgan and Claypool Publishers.
- Bollen, J., Mao, H., and Zeng, X. (2010). Twitter mood predicts the stock market. In *CoRR*, volume abs/1010.3003.
- Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan B (2013) Orange: Data Mining Toolbox in Python, *Journal of Machine Learning Research* 14(Aug): 2349–2353.
- Diakopoulos, N. A., & Shamma, D. A. (2010, April). Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1195-1198).
- Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*, 49(2), 1-41.
- Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N project report, Stanford, 1(12), 2009.
- Hootsuite ve We Are Social, Dijital 2021 Türkiye Raporu <https://wearesocial.com/uk/> adresinden 05.10.2021 tarihinde erişilmiştir.
- Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011, June). Target-dependent twitter sentiment classification. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies* (pp. 151-160).
- Kasaba, E., & Yıldıztepe, E. (2016). Destek vektör makinesi yöntemi ile bir duygu çözümlemesi. *Akademik Bilişim*.
- Mandiberg, M. (Ed.). (2012). *The social media reader*. Nyu Press.
- Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one*, 8(5), e64417.
- Newman, T. P. (2017). Tracking the release of IPCC AR5 on Twitter: Users, comments, and sources following the release of the Working Group I Summary for Policymakers. *Public Understanding of Science*, 26(7), 815-825.
- Ravi, K.ve Ravi, V., 2015. ‘A survey on opinion mining and sentiment analysis: tasks, approaches and applications’, Article in *Knowledge-Based Systems*, pp.14-46
- Ross, B., Rist, M., Carbonell, G., Cabrera, B., Kurowsky, N., & Wojatzki, M. (2017). Measuring the reliability of hate speech annotations: The case of the european refugee crisis. arXiv preprint arXiv:1701.08118.
- Şeker, E.Ş. (2019) Sosyal Medya Mesajlarının Veri Madenciliği Yöntemi İle Duygu Analizi (Sivas İli Örneği) Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü Yüksek Lisans Tezi, Sivas
- Seker, S. E. (2015). Metin Madenciliği (Text Mining). *YBS ansiklopedi*, 2(3), 30-32.

- Wang, Y., Feng, Y., Hong, Z., Berger, R., & Luo, J. (2017, September). How polarized have we become? a multimodal classification of trump followers and clinton followers. In *International Conference on Social Informatics* (pp. 440-456). Springer, Cham.
- Weber, I., Garimella, V. R. K., & Batayneh, A. (2013). Secular vs. islamist polarization in egypt on twitter. In *Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 290-297).