

Research Article

Received: date: 05.11.2022
Accepted: date: 22.11.2022
Published: date: 31.12.2022

Spor Pazarlamasında Veri Madenciliği: Twitter'da Veri Madenciliği ve Derbi Örneği

Habib Mehmet Akpınar ¹

¹Istanbul University, Institute of Social Sciences, Department of Marketing, Phd candidate, Istanbul Turkey; mehmet@mehmetakpinar.com

¹Orcid: 0000-0002-7035-4168

*Correspondence Author: mehmet@mehmetakpinar.com

Öz: Teknolojinin gelişmesi, internetin kullanımının akıllı telefonlar sayesinde yaygınlaşması sosyal medya platformlarında ortaya çıkmasında büyük rol oynamıştır. Sosyal medya günlük yaşamın iletişim açısından değerlendirildiğinde en önemli araçlarından biri haline gelmiştir. Özellikle markalar için kullanıcıların ürettiği içerikleri anlamak ve pazarlama iletişimi stratejilerini buna göre belirlemek son derece önemlidir. Araştırma kapsamında spor tüketiminin önemli paydaşlarından biri olan taraftarları anlamak adına Galatasaray – Beşiktaş derbi karşılaşması için belirlenen #GSvBJK hashtagiyle atılmış 49258 tweet incelenmiştir. Verilerin toplanmasında R programlama dili ve RStudio isimli yazılım ile academictwitte, isimli kod paketi kullanılmıştır. Hashtag (etiket) ve emoji verilerinde ise ilgili kısımlara yönelik analiz yapılmıştır. Ngram adı verilen birbirleriyle bir arada tekrar eden kelime grupları için 2'li ngram ve 3'lü ngramlara çalışmada yer verilmiştir. Araştırmanın sonuçlarına göre derbide atılan tweetlerin büyük çoğunluğunu kaçak yayın ve yasadışı bahis sitelerinin oluşturduğu sonucuna ulaşılmıştır. Ayrıca araştırma sonuçları kelime bulutu ile görselleştirilerek sunulmuştur.

Anahtar Kelimeler: sosyal medya, r programlama dili, twitter,

Data Mining in Sports Marketing: The Example of Data Mining and Derby Match on Twitter

Abstract: The development of technology and the widespread use of the internet by smartphones have played a significant role in the emergence of social media platforms. Social media has become one of the essential tools of daily life in terms of communication. Especially for brands, it is vital to understand users' content and determine marketing communication strategies accordingly. Within the scope of the research, 49258 tweets with the #GSvBJK hashtag, determined for the Galatasaray - Beşiktaş derby match, were examined to understand the fans, one of the crucial stakeholders of sports consumption. R programming language and software named RStudio and a code package named academictwitte were used to collect the data. The hashtag and emoji data analysis was performed for the relevant parts. For groups of words, called ngrams, that repeat together, 2-ngrams and 3-ngrams are included in the study. According to the results of the research, it has been concluded that the majority of tweets posted in the derby are composed of illegal broadcasting and illegal betting sites. In addition, the research results are presented by visualizing a word cloud.

Keywords: social media, r programming language, twitter

1. Giriş

İnternetin ortaya çıkmasıyla birlikte insanlar kolayca içerik üretebilecekleri ve ürettikleri içerikleri diğer insanlar ile paylaşacakları bir ortam bulmuşlardır. İnternetin gelişmesi ve sosyal medya platformlarının ortaya çıkmasıyla birlikte tek yönlü geleneksel iletişim yerini tüm paydaşların bu iletişime katılabileceği çok yönlü bir iletişime dönüşmüştür [1]. Güncel verilere göre dünyada yaklaşık 5 milyar sosyal medya kullanıcısı bulunmaktadır. En fazla kullanılan sosyal medya platformlarının başında sırasıyla; Facebook, YouTube, WhatsApp, Instagram gelmektedir [2]. Sosyal medya platformlarının günden güne popülerleşmesi ile sosyal medya pazarlama faaliyetleri pazarlama çalışmalarında önemli bir hale gelmiştir.

Günümüzde kullanılan birçok sosyal medya platformu içerikleri oluşturma ve paylaşma yoluyla kullanıcıların birbirleriyle iletişim ve etkileşim kurmasını sağlar. Bu platformların başında Twitter, YouTube, Wikipedia, WhatsApp gibi sosyal medya platformları gelir. Bu platformların ortak özelliği, içeriklerin kullanıcılar tarafından oluşturulmasıdır. Kullanıcılar tarafından oluşturulan içerik, platformda yer alan kullanıcıların içerik üretmesi olarak adlandırılır. YouTube’da yer alan videolar, Twitter’da yer alan gönderiler, Instagram’da paylaşılan resimler kullanıcılar tarafından üretilir. Sosyal medyaların en önemli özelliklerinin başında kullanıcı tarafından üretilen içeriklerin yer alması gelir [3]. İnternet ortamında büyük verinin (big data) oluşmasında kullanıcıların ürettiği içeriklerin payı oldukça büyüktür. Aşağıdaki Tablo 1’de en sık kullanılan içeriklerin kullanıcılar tarafından üretildiği platformlar yer almaktadır [3].

Tablo 1. İçeriklerin Kullanıcılar Tarafından Üretildiği Platformlar

Kullanıcı Tarafından Üretilen İçerik Türü	Platformlar
Görsel	Instagram, Pinterest, Snapchat, Flickr
Kişisel Paylaşım ve Networking	Twitter, Foursquare, Facebook, LinkedIn
Ürün ve Hizmet Yorumları	Yelp, Rotten, Tomatoes, ZocDoc, Amazon
Ansiklopedi ve Bilgi	Wikipedia, Wikia
Video	YouTube, Vine
Haber	NY Times, Online, WSJ Online
Kitlelesel Fonlama	Crowdrise, Kickstarter, IndieGoGo
Paylaşım Platformları	Uber, Airbnb, Couchsurfing
Sosyal Ödeme	Venmo, Square
Forum	Reddit, Quora
Blog	Wordpress, Tumblr

Markaların hedef pazarlarında yer alan tüketicilerini anlamak için tüketicilerin marka hakkında ürettiği içerikleri takip etmeleri gerekir. Tüketicilerin markalar hakkında sosyal medya platformlarında ürettiği içerikler markaların hem kendilerini değerlendirebilmek hem de rakiplerini değerlendirmek için önemli bir fırsat sunar. Bu fırsatların başında ürün ve hizmet geliştirmeleri, itibar yönetim, yeni pazar fırsatlarının ortaya çıkması, müşteri ile marka arasındaki iletişimin güçlendirilmesi gibi önemli konular yer almaktadır. En önemlisi, tüketicilerin marka hakkında sosyal medya platformlarında ürettiği içerikler markanın gelecek pazarlama planları için iç görü oluşturmaktadır [4]. Sosyal medya platformlarında paylaşılan içerikler, markaların hedef pazarlarını ve tüketicilerini daha iyi anlamasında için kullanılabilir en önemli kaynakların başında gelmektedir.

Kullanıcılar tarafından sosyal medyada içerikler markalar ve ürünleri ile ilgili değerlendirmeleri ve yorumları içerir. Pazarlamacılar için değerlendirildiğinde kullanıcılar tarafından sosyal medyada üretilen içerikler tüketicilerin marka algılarını anlamada ve bunları belirlemede büyük bir potansiyele sahiptir. Pazarlamacıların ana amacı, sosyal medya farklılıklarını dikkate alarak sosyal medya platformlarında kullanıcılar tarafından yaratılan içeriği anlamak ve bu potanseli markalar için değerlendirmek olmalıdır [5]. Pazarlamacılar sosyal medyada kullanıcılar tarafından oluşturulan içerikleri markaların stratejilerine yön verecek şekilde analiz etmelidir.

Sosyal medyada kullanıcılar tarafından üretilen içerik adından da anlaşılacağı gibi kullanıcılar tarafından üretilir ve kullanıcılar tarafından yönlendirilir. Kullanıcılar herhangi bir türde içerik üreterek bunu platformlarda yer alan diğer kullanıcılar ile paylaşırlar. Bu içerikler kullanıcı tarafından oluşturulan içerik, işletmeler tarafından içeriklerin sayısı açısından değerlendirildiğinde artmaktadır. Bu durum dijital ortamda bilginin işletme merkezinden tüketici merkezine kaydığını göstermektedir. Kullanıcıların ürettiği içerik ile dijital medya işletme – ürün merkezli olmaktan çıkarak tüketici merkezli bir yapıya dönüşmüştür [6]. Modern pazarlama anlayışı tüketici merkezli bir bakış açısına sahiptir. Bu nedenle dijital medya modern pazarlama anlayışının en önemli uygulama alanlarından biridir.

Araştırmada kullanıcılar tarafından üretilen içerik spor taraftarları açısından değerlendirilmiştir. Yapılan araştırma kapsamında derbi müsabakasında twitter üzerinde belirlenen hastagh etiketi ile atılan tweetler incelenmiştir. Araştırmanın amacını spor pazarlaması açısından pazarlama iletişiminin taraftar açısından değerlendirilmesi ve spor taraftarlarının yapmış olduğu paylaşımları anlamlandırmak oluşturmaktadır.

Araştırmada kullanıcılar tarafından üretilen içerik spor tüketiminin önemli paydaşlarından biri olan spor taraftarları açısından değerlendirilmiştir. Yapılan araştırmada, Süper Lig resmi twitter hesabı tarafından belirlenen #GSvBJK etiketi ile paylaşılan gönderiler incelenmiştir. Yapılan araştırmada taraftarları anlamak adına; taraftarların hangi paylaşım yaptıkları, paylaşımlarında geçen kelimelerin sıklığı, kullandığı emojiler incelenmiştir. Yapılan çalışmanın sonuçları frekans analizleri tabloları ve kelime bulutu şekilleri ile desteklenmiştir.

2. Materyal ve Metot

Yapılan Araştırma kapsamında, verilerin toplanmasında R programlama dili [7] ve RStudio [8] isimli yazılım ile academictwitterR [9], isimli kod paketi kullanılmıştır. Twitter API [10] isimli hizmet üzerinden akademik bir lisans [11] ile yüksek miktarda veri elde edilebilmektedir. Bu araştırmanın örneklem havuzu için Twitter üzerinde “#GSvBJK -is:retweet” şeklinde bir sorgu gönderilerek, Retweet biçiminde tekrar paylaşılan gönderiler hariç tutularak, derbi organizasyonunu etiketi üzerinden veriler elde edilmiştir.

Başlangıç tarihi olarak 4 Kasım 2022 - 10:00 girilen sorgunun, son tarihi 8 Kasım 2022 - 10:00 biçimindedir. Sorgu neticesinde 49258 adet gönderiden oluşan bir örneklem havuzu elde edilmiştir. Hazırlanan örneklem havuzu metin madenciliği yöntemi kullanılarak çeşitli açılardan ele alınmıştır. Bu aşamada dplyr [12], tm [13][14] stringr [15], tidyr [16], tidytext [17], tibble [18], emo [19] , wordcloud2 [20], webshot [21], htmlwidgets [22] , ggplot2 [23] , openxlsx [24] , xlsx [25] gibi çeşitli R kod dili paketlerinden faydalanılmıştır. Metin şeklindeki verilerin analiz için uygun hale getirilmesi amacıyla çeşitli ön-temizleme işlemleri uygulanmıştır. Örnek setindeki metin kısmı;

1. Küçük harflere dönüştürülmüş
2. Linkler temizlenmiş (http ve https)
3. Sayılar silinmiş
4. Mention ifadeleri (@kullanıcı adı biçiminde) temizlenmiş
5. Hashtag isimli etiketler (#) temizlenmiş
6. Noktalama işaretleri temizlenmiş
7. Emojiler temizlenmiş
8. Sık tekrar etmesi muhtemel ancak anlamsız kelimeler (stopwords) temizlenmiştir.
Kullanılan bu kelimeler: “ile” “için” için” şeklindedir.

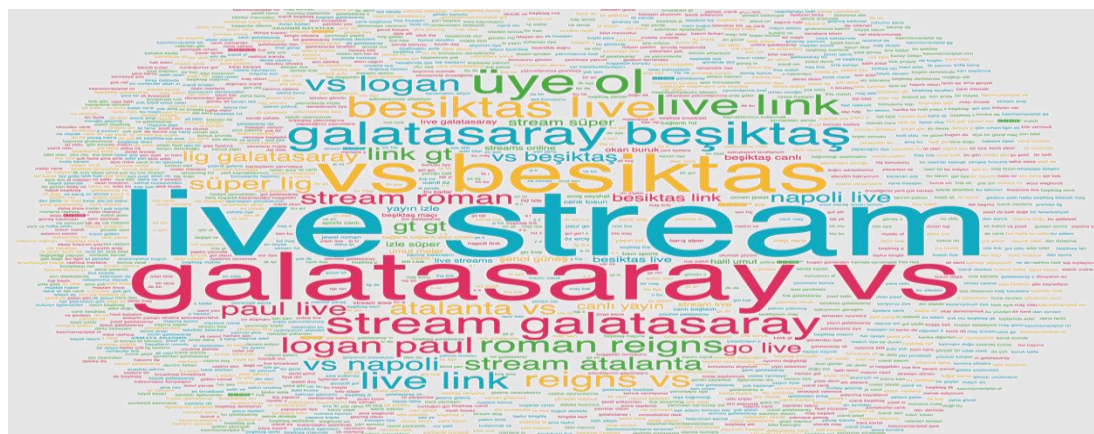
Hashtag (etiket) ve emoji verilerinde ise ilgili kısımlara yönelik analiz yapılmıştır. Ngram adı verilen birbirleriyle bir arada tekrar eden kelime grupları için 2’li ngram ve 3’lü ngramlara çalışmada yer verilmiştir.

Ngram adı verilen birbirleriyle bir arada tekrar eden kelime grupları için 2'li ngram ve 3'lü ngramlara çalışmada yer verilmiştir. Tablo 5'te Ngram adı verilen birbirleriyle bir arada tekrar eden kelime grupları için 2'li ngram tablosu yer almaktadır.

Tablo 5. Gönderilerde En Fazla Kullanılan 2'li Ngram Tablosu

Sıralama	2'li Ngram	Kullanım Sıklığı	Sıralama	2'li Ngram	Kullanım Sıklığı
1	live stream	10324	26	besiktas link	1141
2	galatasaray vs	7418	27	canlı yayın	1080
3	vs besiktas	5989	28	stream süper	1036
4	üye ol	3767	29	besiktas live	878
5	galatasaray beşiktaş	3576	30	beşiktaş canlı	721
6	stream galatasaray	3337	31	şenol güneş	713
7	besiktas live	3189	32	izle süper	693
8	live link	3179	33	yayın izle	693
9	live link	2635	34	live galatasaray	663
10	logan paul	2409	35	stream live	659
11	reigns vs	2407	36	halil umut	655
12	roman reigns	2407	37	live streams	611
13	vs logan	2407	38	okan buruk	606
14	paul live	2024	39	umut meler	589
15	atalanta vs	2021	40	streams online	576
16	stream atalanta	2021	41	cenk tosun	534
17	vs napoli	2021	42	beşiktaş maçı	507
18	stream roman	1995	43	bağlantı bağlantı	461
19	süper lig	1875	44	bağlantı hd	461
20	link gt	1858	45	canlı bağlantı	461
21	napoli live	1585	46	barış alper	457
22	go live	1583	47	canlı izle	448
23	lig galatasaray	1581	48	erciş seyahat	440
24	gt gt	1513	49	öz erciş	440
25	vs beşiktaş	1433	50	napoli link	436

Tablo 5 incelendiğinde gönderilerde en fazla kullanılan 2'li Ngram'ların başında live stream, Galatasaray vs, vs besiktas, üye ol, Galatasaray Beşiktaş, stream galatasaray, besiktas live, live link gibi kelimelerin geldiği görülmektedir. Genel olarak tablo incelendiğinde kaçak yayına ait kelimelerin daha fazla kullanıldığı görülmektedir. Bunun yanında takım, futbolcu, teknik direktör ve hakem isimlerinin en fazla kullanılan 2'li Ngram'ların içerisinde olduğu söylenebilir. Şekil 5'de araştırma kapsamında gönderiler içerisinde en fazla kullanılan 2'li Ngramlara ait frekanslar hesaplanarak oluşturulan kelime bulutu yer almaktadır.



Şekil 5. Gönderilerde En Fazla Kullanılan 2'li Ngram'ların Ait Kelime Bulutu

3. Sonuç ve Tartışma

Günümüz pazarlama dünyası büyük bir dönüşüm içerisinde. Pazarlamacıların en önemli amacı tüketicileri anlamak ve onların ihtiyaçlarına karşılayacak, rakiplerinden farklılaşarak bir değer oluşturmaktadır. İnternetin yaygınlaşması, akıllı telefonların günden güne hayatımızın ayrılması bir parçası haline gelmesi nedeniyle tüketicilerin yeni dijital davranışları ortaya çıkmaya başlamıştır. Dijital dünyanın en önemli pazarlama iletişimi mecralarının başında sosyal medya platformları gelmektedir. Sosyal medya platformuna sahip olmak, sosyal medya platformlarında tüketicilerin yapmış oldukları paylaşımları takip etmek işletmeler için bir zorunluluk haline gelmiştir.

Pazarlama araştırmaları, pazarlamacıların tüketicileri anlamak için başvurdukları en önemli kaynakların başında gelmektedir. Dijital dünya ile geleneksel pazar araştırma yöntemleri dijital yöntemler ile desteklenmelidir. Dijital dünyada sosyal medya platformlarının ortaya çıkması, içerik üretiminin kullanıcılara geçmesi ile muazzam bir datanın ortaya çıkmasına neden olmuştur. Pazarlama dünyası dijital dünyayı ve dijital dünyanın yeni tüketicisini anlamlandırmak adına kullanıcıların dijital platformlarda üretmiş oldukları içerikleri incelemelidir.

Yapılan araştırma kapsamında süper lig resmi twitter hesabı tarafından belirlenen #GSvBJK etiketi ile paylaşılan toplam 49258 gönderi incelenmiştir. Etiket kapsamında gönderilerde en fazla kullanılan kelimeler, etiketler, emojiler, 2'li ve 3'lü ngramlar incelenmiştir. İncelemeler sonucunda yapılan paylaşımların büyük çoğunluğunun kaçak yayın ve yasa dışı bahis ile ilgili olduğu görülmüştür. Maç ile ilgili etkileşimlerin düşük yoğunlukta olduğu gözlemlenmiştir. Analiz sonucunda özellikle taraftarların değil bot ve reklam hesaplarının daha fazla paylaşım yaptığı görülmektedir. Bu durum en popüler sosyal medya kanallarından biri olan Twitter'a olan güveni azaltmaktadır.

Gelecek çalışmalarda daha fazla etiket ve daha fazla gönderi datasına yer verilerek daha anlamlı sonuçlar alınabilir. Ayrıca araştırma birkaç spor branşı özelinde incelenerek karşılaştırmalar yoluyla geliştirilebilir. Bunun yanında diğer sosyal medya platformlarında benzer etiketle paylaşılan gönderiler yoluyla sonuçlar karşılaştırmalı olarak incelenebilir. Ayrıca spor taraftarını ve tüketicisini anlamlandırmak için sonuçlar veri işleme yönteminin yanı sıra, nitel ve nicel farklı araştırma metodlarıyla incelenecek sonuçlar daha fazla desteklenebilir.

Yazar Katkıları: Bu çalışmada giriş literatür taraması analizler ve sonuç bölümü yazar tarafından yapılmıştır.

Finansman: Bu çalışma için herhangi bir finansal kaynak sağlanmamıştır.

Çıkar Çatışması: Yazarlar herhangi bir çıkar çatışması olmadığını beyan etmemektedir.

Kaynaklar

- [1] C. Wyrwoll, "User-Generated Content", içinde *Social Media: Fundamentals, Models, and Ranking of User-Generated Content*, C. Wyrwoll, Ed. Wiesbaden: Springer Fachmedien, 2014, ss. 11-45. doi: 10.1007/978-3-658-06984-1_2.
- [2] hannahcurrey, "DIGITAL 2022: ANOTHER YEAR OF BUMPER GROWTH", *We Are Social UK*, 26 Ocak 2022. <https://wearesocial.com/uk/blog/2022/01/digital-2022-another-year-of-bumper-growth-2/> (erişim 20 Kasım 2022).
- [3] M. Luca, "Chapter 12 - User-Generated Content and Social Media", içinde *Handbook of Media Economics*, c. 1, S. P. Anderson, J. Waldfogel, ve D. Strömberg, Ed. North-Holland, 2015, ss. 563-592. doi: 10.1016/B978-0-444-63685-0.00012-7.
- [4] A. J. Kim ve K. K. P. Johnson, "Power of consumers using social media: Examining the influences of brand-related user-generated content on Facebook", *Computers in Human Behavior*, c. 58, ss. 98-108, May. 2016, doi: 10.1016/j.chb.2015.12.047.
- [5] A. N. Smith, E. Fischer, ve C. Yongjian, "How Does Brand-related User-generated Content Differ across YouTube, Facebook, and Twitter?", *Journal of Interactive Marketing*, c. 26, sy 2, ss. 102-113, May. 2012, doi: 10.1016/j.intmar.2012.01.002.
- [6] E. Moriuchi, *Social media marketing: Strategies in utilizing consumer-generated content*. Business Expert Press, 2019.
- [7] R. C. Team, "R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria", <http://www.R-project.org/>, 2013.
- [8] Rs. Team, "RStudio: integrated development for R. RStudio", *Inc., Boston, MA*, c. 700, s. 879, 2015.
- [9] C. Barrie ve J. C. Ho, "academictwitterR: an R package to access the Twitter Academic Research Product Track v2 API endpoint", *Journal of Open Source Software*, c. 6, sy 62, s. 3272, 2021.
- [10] "Twitter API Documentation". <https://developer.twitter.com/en/docs/twitter-api> (erişim 20 Kasım 2022).
- [11] "Twitter API for Academic Research | Products". <https://developer.twitter.com/en/products/twitter-api/academic-research> (erişim 20 Kasım 2022).
- [12] H. Wickham ve L. Henry, "tidyr: Tidy Messy Data. R package version 1.1.2", *CRAN. R-project.org/package=tidyr*, 2020.

-
- [13] I. Feinerer ve H. K. Tm, "Text Mining Package. R package version 0.7-7. 2019". 2020.
- [14] I. Feinerer, K. Hornik, ve D. Meyer, "Text Mining Infrastructure in R", *Journal of Statistical Software*, c. 25, ss. 1-54, Mar. 2008, doi: 10.18637/jss.v025.i05.
- [15] H. Wickham ve RStudio, "stringr: Simple, Consistent Wrappers for Common String Operations". 20 Ağustos 2022. Erişim: 20 Kasım 2022. [Çevrimiçi]. Erişim adresi: <https://CRAN.R-project.org/package=stringr>
- [16] H. Wickham, M. Girlich, ve RStudio, "tidyr: Tidy Messy Data". 08 Eylül 2022. Erişim: 20 Kasım 2022. [Çevrimiçi]. Erişim adresi: <https://CRAN.R-project.org/package=tidyr>
- [17] J. Silge ve D. Robinson, "Text mining and analysis using tidy data principles in r", *J. Soc. Struct*, sy 1.
- [18] K. Müller ve H. Wickham, "Tibble: Simple data frames". 2019.
- [19] H. Wickham, R. François, L. Henry, ve K. Müller, "dplyr: A Grammar of Data Manipulation. R package version 0.4. 3", *R Found. Stat. Comput., Vienna*. <https://CRAN.R-project.org/package=dplyr>, 2015.
- [20] D. Lang ve G. Chien, "wordcloud2: Create Word Cloud by 'htmlwidget'. R package version 0.2. 1". 2018.
- [21] W. Chang, "Webshot: Take screenshots of web pages", *R package version*, c. 5, 2017.
- [22] R. Vaidyanathan, Y. Xie, J. Allaire, J. Cheng, ve K. Russell, "htmlwidgets: HTML Widgets for R. R package version 1.5. 1". 2019.
- [23] H. Wickham, "Package 'ggplot2': elegant graphics for data analysis", *Springer-Verlag New York*. doi, c. 10, ss. 978-0, 2016.
- [24] P. Schauburger ve A. Walker, "openxlsx: read, write and edit xlsx files (R package version 4.2. 3)". 2020.
- [25] A. Dragulescu ve C. Arendt, "xlsx: Read, Write, Format Excel 2007 and Excel 97/20 0 0/XP/20 03 Files. R package version 0.6. 5". 2020.