

Sosyal Medya Uygulamalarında Kullanıcı Yorumlarının Metin Madenciliği ile Sınıflandırılması

Emre Dandıl^{1*}, ve Burhan Karakurt²

¹Bilgisayar Mühendisliği Bölümü, Bilecik Şeyh Edebali Üniversitesi, Türkiye
^{*}(emre.dandil@bilecik.edu.tr)

Özet –Sosyal medya kullanımı günümüzde insanlar için önemli bir günlük aktivite olmuştur. İnsan yaşamının odak noktalarından birisi haline gelen sosyal medya hem bir kişisel ihtiyaç olarak kullanılmakta hem de insanların eğlence gereksinimleri karşılamaktadır. Bu çalışmada, geliştirilen bir Android tabanlı mobil uygulamaya girilen kullanıcı yorumlarının gerçek zamanlı olarak Bayes algoritması ile sınıflandırılması gerçekleştirilmiştir. Böylece, sosyal medya kullanımını daha kaliteli ve bilinçli hale getirilmiş olmaktadır. Çalışmada eğitim için önceden hazırlanmış olumlu ve olumsuz kullanıcı yorumlarının bulunduğu bir veriseti kullanılmıştır. Eğitim aşamasından sonra mobil uygulamaya girilen kullanıcı yorumlarının Bayes algoritması ile pozitif veya negatif olduğunun sınıflandırılması sağlanmıştır. Mobil uygulama üzerinden gerçekleştirilen test işlemlerinde %88 doğruluk ile kullanıcı yorumlarının başarılı olarak sınıflandırıldığı görülmüştür.

Anahtar Kelimeler –Sosyal medya, mobil uygulama, Android, sınıflandırma, Bayes algoritması.

Classification of User Comments for Social Media Applications using Text Mining

Abstract –The use of social media has become an important daily activity for people today. Social media, which has become one of the focal points of human life, is used both as a personal need and meets the entertainment needs of people. In this study, Android based mobile application is developed and user comments entered into the application are classified in real time using Bayesian algorithm. Thus, the use of social media is made more quality and conscious. In the study, a pre-prepared dataset with positive and negative user comments is used for the training phase. After the training phase, the user comments entered into the mobile application are classified as positive or negative using Bayesian algorithm. It is seen that user comments are classified as successful with 88% accuracy in the test operations performed on mobile application.

Keywords –Social media, mobile application, Android, classification, Bayesian algorithm

I. GİRİŞ

Son yıllarda, akıllı telefon ve tablet gibi mobil cihazların özellikle günlük hayatta kullanımı önemli ölçüde artmıştır. İnsanlar artık mobil cihazlara yüklenen uygulamalarla birçok faaliyetini kolaylıkla yürütebilmektedir. Özellikle sosyal medya uygulamaları, kullanıcıların mobil cihazlarda gün içinde en çok kullandıkları servislerin başında gelmektedir[1]. Yaklaşık 82 milyon nüfusa sahip ülkemiz için internet kullanım ve sosyal medya istatistikleri incelendiğinde 59.36 milyon internet kullanıcısı, 52 milyon aktif sosyal medya kullanıcısı, 44 milyon aktif mobil sosyal medya kullanıcısı olduğu görülmektedir. İnsanların internette günde ortalama olarak 7 saat geçirdiği tahmin edilmektedir. Bu 7 saatin ortalama 2 saat 46 dakikasını ise sosyal medya platformlarında geçirdikleri öngörülmektedir[2].

İnternet ve sosyal ağlar ile insanlar farklı iletişim fırsatları oluşturmaktadırlar ve kullanılmaktadırlar. Sosyal ağlar insanlar için düşüncelerini paylaşacakları, paylaşım ve tartışma yapabilecekleri bir medya ortamı sunmaktadır. Sosyal medya ortamları, aynı zamanda kullanıcı tabanlı olmasının yanında kitleleri ve insanları bir araya getirmesi ve aralarındaki etkileşimi arttırması bakımından da önemlidir[3]. Kullanıcılar

sosyal medya uygulamalarını kullanmanın yanında, kendi yorumlarını da bu uygulamalar sayesinde, hem çevresi ile paylaşabilmekte hem de geniş kitlelere ulaştırabilmektedir. Bunun yanında bulut bilişim, sosyal medya ve mobil cihazların hem iş amaçlı hem de kişisel kullanımlar için yaygın olarak kullanılmaktadır.

Web 2.0'ın gelişmesiyle internet üzerinden sosyal medya siteleri, bloglar, wikiler, multimedya paylaşımları ve zengin web uygulamaları ile kullanıcı ile etkileşim kurulması ön plana çıkmıştır[4]. Aynı zamanda, bu ortamlar üzerinde ticari şirketler büyük veriler üzerinde, müşterilerinin favorilerini ve gelecek için neler talep edebileceklerinden yola çıkarak birçok analiz yapmaktadır. Özellikle sosyal medya uygulamaları üzerinde kullanıcıların metinsel içeriklerinden yararlanarak, müşterilerin duygusal eğilimlerinin ortaya çıkarılması daha fazla önem kazanmıştır.

Kullanıcıların sosyal medya ortamlarında, mobil uygulamalar ile paylaştıkları içeriklerden duygusal çıkarımların yapılması bir sınıflandırma problemi olarak karşımıza çıkmaktadır. Literatürde bu şekilde önerilen birçok çalışma olduğu görülmektedir. Turney ve Peter [5] araba kullanıcılarının yorumlarından kullanıcı eğilimlerinin belirlenmesi için konuşmaların parçalanarak etiketlenmesini

$$C_{MAP} = \operatorname{argmax}_{c_j \in C} (P(a_1, a_2, \dots, a_n | c_j) P(c_j)) \quad (5)$$

Sade Bayes sınıflandırıcıda, hedef değerleri olasılıksal olarak birbirinden bağımsız olduğu için Eşitlik (5), daha sade olarak Eşitlik (6)'daki gibi verilebilir. Eşitlik (6)'da C_{NB} Bayes sınıflandırıcının çıkış değerini göstermektedir.

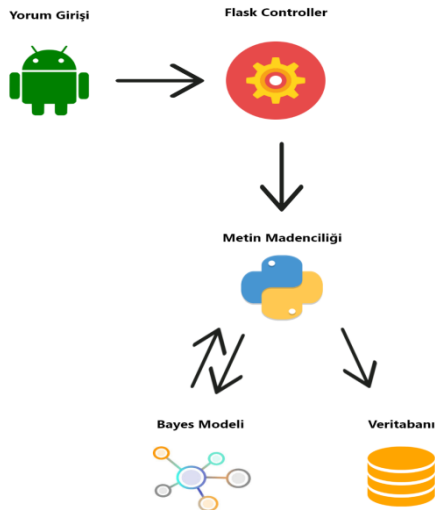
$$C_{NB} = \operatorname{argmax}_{c_j \in C} (P(c_j) \prod_i P(a_i | c_j)) \quad (6)$$

Sade Bayes sınıflandırıcıda, eğer öznitelik değerleri sayısal değerlerden oluşuyorsa, sonrasal olasılık, Eşitlik (7)'de sunulan ve normal dağılımı gösteren Gauss denklemi ile hesaplanır. Burada, μ aritmetik ortalamayı, σ ise varyansı temsil etmektedir.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (7)$$

III. TASARLANAN MOBİL UYGULAMA

Çalışmada, kullanıcı yorumlarının sınıflandırılması için Java dili ile Android tabanlı bir mobil uygulama tasarlanmıştır. Tasarlanan mobil uygulamanın özetçe mimari yapısı Şekil 3'te sunulmuştur. Mobil uygulamada Android' e sağlanan veriler Python ile yazılan mikroservisler ile işlenmektedir. Uygulama geliştirilirken, Android altyapısında MVC mimarisinden yararlanılmıştır.



Şekil 3. Geliştirilen mobiluygulamanın mimari organizasyonu

Tablo 1. Geliştirilen Naive Bayes algoritması için oluşturulan fonksiyonlar ve görevleri

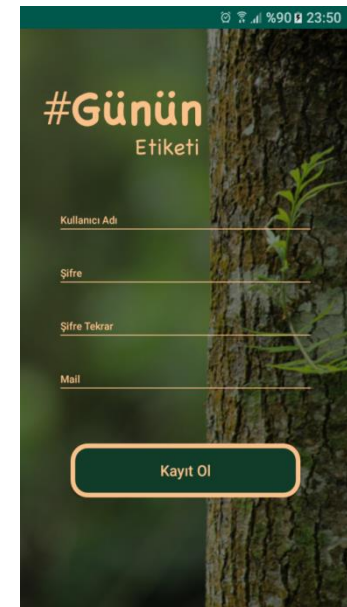
Fonksiyon Adı	Görevi
Remove_stopwords()	Verisetindeki cümleler İçerisindeki Nötr kelimeleri ayıklar.
Train_test_split()	Tüm cümleleri kelimelere ayırarak CountVectorizer nesnesini oluşturur.
Metrics.accuracy_score()	Oluşturulan test verisinin doğruluk oranını hesaplar.
NB.feature_count()	Kelimelerine ayrılan cümleyi parametreler halinde verisetinde arayarak pozitif – negatif oranlarını bulur.

Mobil uygulamada, kullanıcı yorumlarının olduğu veriseti üzerinde, Python dili ile yazılan Scikit-learn kütüphanesi ile eğitilerek işlem yapılmaktadır. Eğitilen model mikroservise bağlanarak, Bayes sınıflandırıcı sayesinde gelen yorumun pozitif veya negatif olma durumuna göre içerik hakkında veritabanında işlem yapılmaktadır. Tasarlanan mobil uygulamanın mimari yapısında bulunan bazı temel fonksiyonlar Tablo 1'de sunulmuştur. Bu fonksiyonlar, Remove_stopwords(), Train_test_split(), Metrics.accuracy_score() ve NB.feature_count() şeklinde sıralanabilir. Bunların görevleri ve ne amaçla kullanıldıkları yine tabloda gösterilmiştir. Çalışmada geliştirilen mobil uygulamanın veritabanı MySQL ile tasarlanmıştır. Toplamda ü tane tablo kullanılmıştır. Bu tablolar Kullanıcılar, Etiketler ve Yorumlar şeklinde sıralanabilir.

Hazırlanan Android tabanlı uygulamanın giriş yap sayfası Şekil 4 ve Şekil 5'te sunulmuştur. Burada kullanıcı sisteme üye olduktan sonra kullanıcı bilgileri ile sisteme giriş yapabilmektedir.



Şekil 4. Geliştirilen mobil uygulama için giriş ekranı



Şekil 5. Geliştirilen mobil uygulama için yeni üye kayıt ekranı

Uygulamaya giriş yapıldıktan sonra kullanıcının karşısına gelen Şekil 5 ' deki ekranda o günün içeriğini yorumları ile birlikte görülebilmektedir. Bu ekranda içeriğin altına gelen yorumların pozitif veya negatif mi olduğu da kullanılan emojilerden anlaşılmaktadır. Uygulamanın ana ekranının sağ alt tarafındaki buton ile içeriğe kendi yorumunuzu ekleyebilmektedir. Ayrıca sol menüden tüm etiketler denilerek, istenilen herhangi bir etiket hakkında kullanıcıların pozitif veya negatif yorumları görülebilmektedir. Şekil 6 'da verisetindeki kullanıcı yorumlarının pozitif ve negatif kelimelerin analizi ve doğruluk puanı hesabının ekran çıktısı görülmektedir.

```
Naive Bayes
Doğruluk puanı: 88.14%
Karışıklık Matrisi:
[[2303 235]
 [ 358 2104]]
Analysis
Token Sayısı : 68275
Pozitif Token : 29815
Negatif Token 38460
```

Şekil 6. Verisetinin Doğruluk Puanı ve Analiz Sonuçları

Şekil 7 ve 8 'de Bayes 'e gönderilen "kötü filmdi" , "iyi filmdi" yorumlarının, ilk olarak parçalanması ve token dizisine atılması gösterilmiştir. Sonrasında kelimelerin, tek tek negatif ve pozitif puanlarını hesaplanarak kelimenin pozitif mi negatif mi olduğuna karar verilmiştir. Şekil 7 'de görüldüğü gibi "iyi" kelimesinin pozitif puanı 2277, negatif puanı ise 1440 'dır. Bu durumda "iyi" kelimesi pozitif olarak belirlenmektedir. Şekil 8'de ise "kötü" kelimesinin pozitif puanı 264, negatif puanı ise 1517 'dir. Bu durumda "kötü" kelimesi negatif olarak belirlenmektedir.

```
arama sonuçları token/s: ['iyi', 'filmdi']
Token Positive Negative
21042 filmdi 325.0 363.0
30444 iyi 2277.0 1440.0
```

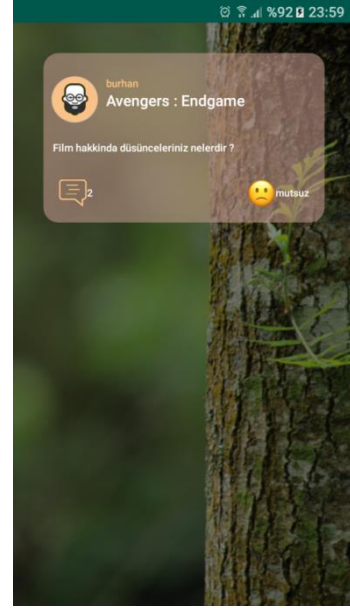
Şekil 7. Verisetine gönderilen olumlu cümle sonucu

```
arama sonuçları token/s: ['kötü', 'filmdi']
Token Positive Negative
21042 filmdi 325.0 363.0
38016 kötü 264.0 1517.0
```

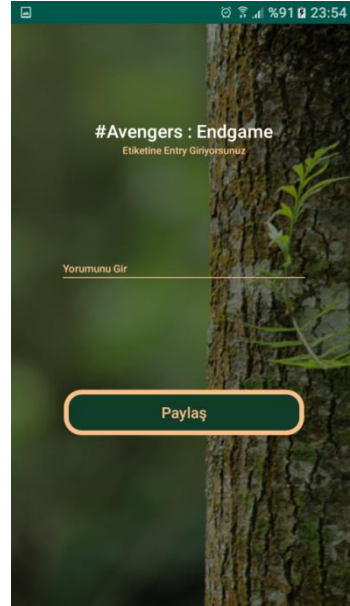
Şekil 8. Verisetine gönderilen olumsuz cümle sonucu

IV. DENEYSEL ÇALIŞMALAR

Geliştirilen mobil uygulama için Şekil 9 'da uygulama içerisindeki tüm içeriklerin listelendiği ekran görülmektedir. Bu ekranda kullanıcı herhangi bir içeriğe giderek yorumlarını okuyabilmektedir. Yorum sayısını ve içeriğin durumunu görebilmektedir. Şekil 10 ' da ise kullanıcının seçtiği içeriğe yorum ekleyebileceği form gösterilmiştir.

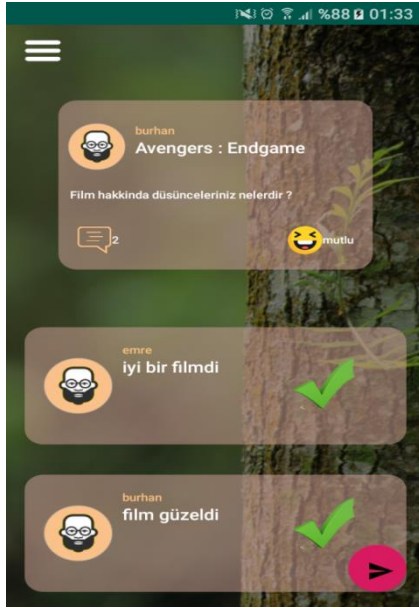


Şekil 9. Mobil uygulama üzerinde tüm içeriklerin listelenmesi

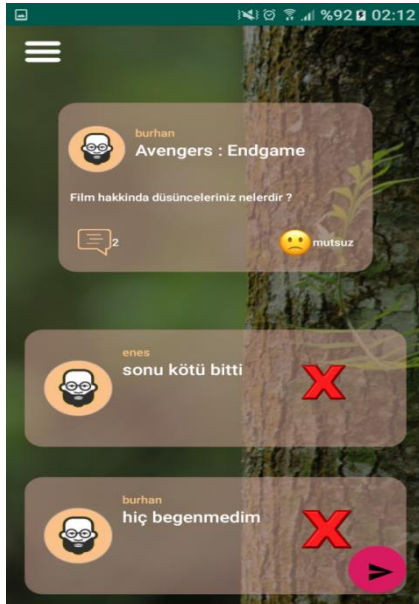


Şekil 10. Bir içeriğe kullanıcı tarafından yorum eklenmesi

Şekil 11'de tasarlanan mobil uygulama üzerinde, iki farklı kullanıcı tarafından girilen yorumların Olumlu, Şekil 12'de ise olumsuz olarak sınıflandırılması gösterilmiştir. Burada kullanıcı bir içeriğe yorum girdikten sonra, mobil uygulama üzerinde gerçek zamanlı prosedürler işletilmekte ve Bayes sınıflandırıcı ile kullanıcının Olumlu veya Olumsuz duygu analizi tespit edilmektedir.



Şekil 11. Mobil uygulama üzerinde kullanıcı yorumlarının olumlu olarak sınıflandırılması



Şekil 12. Mobil uygulama üzerinde kullanıcı yorumlarının olumsuz olarak sınıflandırılması

Veriseti üzerinde yapılan deneysel çalışmalarda, Şekil 13'te görüldüğü gibi sınıflandırma başarı oranının yaklaşık olarak %88 olarak hesaplandığı görülmektedir. Şekil 13'teki Naive Bayes 'in metrics kütüphanesinin accuracy_score metoduna, test verisi ve eğitilmiş veriyi parametre olarak gönderilmiştir. Bu fonksiyondan geriye, doğruluk puanı hesaplanarak döndürülmektedir.

```
NE = MultinomialNB()
NE.fit(X_train_dtm, y_train)
y_pred = NE.predict(X_test_dtm)
print('\nNaive Bayes')
print('Doğruluk puanı: ', metrics.accuracy_score(y_test, y_pred) * 100, '%', sep='')
print('Karışıklık Matrisi: ', metrics.confusion_matrix(y_test, y_pred), sep='\n')
```

Şekil 13. Sınıflandırma başarı oranı

Tablo 2'de ise veriseti hakkında toplam farklı kelime, pozitif kelime, negatif kelime analizleri bulunmaktadır. Bunların hesabı ise mobil uygulama üzerinde CountVectorizer ile yapılmaktadır.

Tablo 2. Veriseti üzerinde sınıflandırma başarı oranı

Token Sayısı	68275
Pozitif Token	29815
Negatif Token	38460
Doğruluk Puanı	%88.14

V. SONUÇLAR

Bu çalışmada, tasarlanan Android tabanlı bir mobil uygulama üzerinde sosyal medya yorumları Sade Bayes sınıflandırıcısı ile sınıflandırılarak kullanıcı yorumlarının Olumlu veya Olumsuz olarak tahmin edilmesi ile metin içerikleri üzerinde duygu analizi yapılmıştır. Çalışmada eğitim için önceden hazırlanmış olumlu ve olumsuz kullanıcı yorumlarının bulunduğu bir veriseti kullanılmıştır. Eğitim aşamasından sonra mobil uygulamaya girilen kullanıcı yorumlarının Bayes algoritması ile pozitif veya negatif olduğunun sınıflandırılması sağlanmıştır. Mobil uygulama üzerinden gerçekleştirilen test işlemlerinde %88 doğruluk ile kullanıcı yorumlarının başarılı olarak sınıflandırıldığı görülmüştür.

İleride, mevcut sistemin verisetini oluşturmak ve daha detaylı analizler yapabilmek için daha farklı içeriklerin yorumları verisetine eklenerek başarı oranı artırılabilir ve daha doğru tahminler yapılabilir.

KAYNAKLAR

- [1] L. Zhang, K. Hua, H. Wang, G. Qian, and L. Zhang, "Sentiment analysis on reviews of mobile users," *Procedia Computer Science*, vol. 34, pp. 458-465, 2014.
- [2] Dijilopedi. (09.06.2019). 2019 Türkiye İnternet Kullanım ve Sosyal Medya İstatistikleri. <https://dijilopedi.com/2019-turkiye-internet-kullanim-ve-sosyal-medya-istatistikleri/>, Erişim Tarihi: 28/07/2019
- [3] B. Solmaz, G. Tekin, Z. Herzem, and M. Demir, "İnternet ve Sosyal Medya Kullanımı Üzerine Bir Uygulama," *Selçuk üniversitesi iletişim fakültesi akademik dergisi*, vol. 7, pp. 23-32, 2013.
- [4] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, "Ontology-based sentiment analysis of twitter posts," *Expert systems with applications*, vol. 40, pp. 4065-4074, 2013.
- [5] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th annual meeting on association for computational linguistics*, 2002, pp. 417-424.
- [6] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, 2002, pp. 79-86.
- [7] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168-177.
- [8] K. Sailunaz and R. Alhaji, "Emotion and Sentiment Analysis from Twitter Text," *Journal of Computational Science*, 2019.
- [9] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, 2011, pp. 30-38.
- [10] L. Terán and J. Mancera, "Dynamic profiles using sentiment analysis and twitter data for voting advice applications," *Government Information Quarterly*, 2019.
- [11] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment analysis in tourism: capitalizing on big data," *Journal of Travel Research*, vol. 58, pp. 175-191, 2019.
- [12] M. Yapnaz. (27/0/2019). *Scikit-Learn kütüphanesi ile basit bir duygu analizi örneği*. <https://github.com/merveyapnaz/Sentiment-Analysis>, Erişim Tarihi: 28/07/2019.
- [13] D. Lewis, "Representation and learning in information retrieval," University of Massachusetts at Amherst, 1992.
- [14] X. Chai, L. Deng, Q. Yang, and C. X. Ling, "Test-cost sensitive naive bayes classification," in *Fourth IEEE International Conference on Data Mining (ICDM'04)*, 2004, pp. 51-58.
- [15] J. Vaidya, M. Kantarcioğlu, and C. Clifton, "Privacy-preserving naive bayes classification," *The VLDB Journal*, vol. 17, pp. 879-898, 2008.