

Duygu Analizi

1st Öğrenci Adı: Emirhan GÜLER
Bölüm Adı: Bilişim Sistemleri Mühendisliği
Üniversite Adı: Kocaeli Üniversitesi
Öğrenci No: 221307086
Grup No: 107
Email: eguler_73@hotmail.com

2nd Öğrenci Adı: Alperen GÖRMEZ
Bölüm Adı: Bilişim Sistemleri Mühendisliği
Üniversite Adı: Kocaeli Üniversitesi
Öğrenci No: 221307104
Grup No: 107
Email: alperen.41gs@hotmail.com

Abstract—This project was developed to collect, process, and analyze data on product reviews and ratings from Trendyol, aiming to perform sentiment analysis based on user-specified text. Web scraping techniques were employed to extract reviews across various product categories, with each review and rating compiled into a dataset alongside user-provided scores. These data were subsequently prepared for use in a sentiment analysis model. Python programming language was utilized in this project, employing libraries such as Requests, Selenium, and BeautifulSoup for web scraping tasks, while Pandas was used for organizing and saving the data in CSV format. This study demonstrates how customer feedback on e-commerce platforms can be transformed into sentiment analysis, facilitating the understanding of the emotional undertones of user-generated text.

Index Terms—Sentiment analysis, web scraping, e-commerce, data collection, Trendyol, natural language processing, Python, Selenium, BeautifulSoup

I. GİRİŞ

E-ticaret platformları, kullanıcıların ürünler hakkında geri bildirim sağlayarak ürünler hakkında kendi deneyimlerine göre ürün hakkında yorumlarda bulunur ve diğer potansiyel alıcıların bilinçli kararlar almasına katkıda bulunurlar. Özellikle Trendyol gibi geniş bir kullanıcı kitlesine sahip platformlarda, ürünlere yapılan yorumlar ve verilen puanlar, müşteri memnuniyeti, memnuniyetsizliği ve genel tüketici eğilimlerinin anlaşılmasında önemli bir rol oynamaktadır. Bu yorumlar, doğal dil işleme (NLP) teknikleriyle analiz edilerek, duygu analizine dönüştürülebilir ve belirli bir ürün veya hizmet hakkındaki kullanıcıların duygusal geri bildirimleri daha net bir şekilde ortaya çıkarılabilir.

Bu proje kapsamında, Trendyol platformundaki ürünlere ait kullanıcı yorumları ve puanları toplanarak duygu analizine tabi tutulmak istenmektedir. Kullanıcıların bir metin girişi üzerinden belirttikleri ifadelerin duygusal durumunu analiz eden bir yapay zeka modeli geliştirilmek istenmektedir. Bu model, müşterilerin ürünlere yönelik memnuniyet, şikayet veya önerilerini analiz ederek, kullanıcıların herhangi bir metindeki duygusal ifadeleri anlamasına olanak tanımaktadır.

Projenin amacı, Trendyol üzerindeki müşteri yorumlarının duygu analizine dönüştürülmesi ve elde edilen verilerin tüketici eğilimlerini anlamaya yönelik bir altyapı sağlamasıdır.

Identify applicable funding agency here. If none, delete this.

Bu çalışma, insanların yazdıkları yorumlar ve verdikleri puanlar aracılığıyla o an yazıyı yazarken nasıl bir duygu durumu içinde olduklarının anlaşılmasını sağlamayı hedeflemektedir.

II. VERİ TOPLAMA SÜRECİ

Bu çalışmada veri toplama süreci, Trendyol platformundaki ürünlere yapılan müşteri yorumları ve verilen puanları elde etmek için web scraping tekniklerini kullanarak gerçekleştirilmiştir. Python programlama dili ile Request, Selenium ve BeautifulSoup kütüphaneleri kullanılarak, belirli ürün kategorilerine ait yorumlar otomatik olarak çekilmiştir.

Veri toplama sürecinde, öncelikle Trendyol ana sayfasında yer alan ürün kategorileri listelenmiştir. Her kategoriye erişim sağlandıktan sonra, kategori sayfaları üzerinden belirli ürünler seçilmiş ve bu ürünlere ait yorum sayfalarına gidilmiştir. Yorum sayfalarından, müşterilerin ürünlere verdiği puanlar ve yazdıkları yorumlar alınarak bir veri seti oluşturulmuştur. Her bir yorum ve puan, analiz sürecine uygun olacak şekilde yapılandırılması için CSV formatında kaydedilmiştir.

Bu veri toplama sürecinde karşılaşılan zorluklar arasında pop-up pencereler, sayfaların dinamik yapısı ve kategori sayfalarının farklı yüklenme süreleri yer almaktadır. Bu sorunları çözmek amacıyla Selenium'un bekleme fonksiyonları kullanılarak sayfa yüklenmeleri dinamik olarak yönetilmiştir. Ayrıca, veri toplama işlemleri sırasında sunucuya fazla istek gönderilmesini engellemek için rasgele sürelerle bekleme süreleri eklenmiştir.

Toplanan bu veri seti, Trendyol kullanıcılarının belirli bir ürün hakkındaki memnuniyet veya memnuniyetsizlik düzeylerini anlamak ve bu yorumların duygu analizine tabi tutulması için kullanılacak şekilde hazırlanmıştır.

III. VERİ ÖN İŞLEME

Bu projede toplanan yorumlar ve puanlar, duygu analizi modeline uygun hale getirilebilmesi için bir dizi ön işleme adımına tabi tutulmuştur. Yorumların metin analizi için optimize edilmesi ve model performansının artırılması amacıyla Python ve Zemberek kütüphanesi kullanılmıştır. Ayrıca, doğal dil işleme (NLP) tekniklerinden yararlanılmıştır.

A. Stop Words Temizliği

Türkçe diline özgü stop words (anlam taşımayan sık kullanılan kelimeler) temizlenmiştir. Bu işlem için NLTK kütüphanesi kullanılmış ve stopwords listesinden yararlanılmıştır. Yorumlardaki stop words kelimeler, metin analizinin daha anlamlı sonuçlar vermesi için çıkarılmıştır. Örneğin, "ve", "bir", "de" gibi kelimeler temizlenmiştir.

B. Tokenizasyon ve Küçük Harfe Dönüştürme

Yorumlar, kelime bazında bölümlere ayrılmış (tokenize edilmiş) ve tüm kelimeler küçük harfe dönüştürülmüştür. Bu işlem, analiz sırasında büyük/küçük harf farkının yaratabileceği sorunların önüne geçmek için uygulanmıştır.

C. Zemberek ile Kök ve Tür Analizi

Bu projede kullanılan Zemberek kütüphanesi, Türkçe diline özgü bir doğal dil işleme aracıdır ve Java tabanlı bir kütüphane olarak geliştirilmiştir. Python'da Zemberek'i kullanabilmek için, Zemberek'in JAR dosyası [12] numaralı kaynaktan yer alan Google Drive bağlantısından indirilmiştir. Python entegrasyonu için [10] ve [11] numaralı kaynaklarda sunulan Zemberek Python örneklerinden faydalanılmıştır.

Zemberek, her kelimenin kök halini (*lemma*) ve türünü (*primaryPos*) çıkararak Türkçe diline özgü dil bilgisi kurallarını dikkate almaktadır. Bu çalışmada, Zemberek'in kelime analizi özellikleri kullanılarak, yorumlar içerisindeki kelimeler köklerine indirilmiş ve türlerine göre sınıflandırılmıştır. Örneğin, "yazıyorum" kelimesi köküne ("yaz") indirgenmiş ve türü (fil) olarak belirlenmiştir.

Zemberek kütüphanesinin Java tabanlı olması nedeniyle, JPyype kütüphanesi ile Python'da entegrasyonu sağlanmıştır. JPyype, Python'dan Java kodlarını çalıştırmayı mümkün kılan bir araçtır ve bu proje kapsamında Zemberek'in Python ile kullanımı için kullanılmıştır.

D. İşlenmiş Yorumların Yapılandırılması

Yorumların her biri, kök ve tür bilgileriyle birlikte yeniden yapılandırılmıştır. Bu yapılandırılmış veri, daha sonra duygu analizi modeline giriş olarak kullanılacak şekilde düzenlenmiştir.

E. Verilerin Kaydedilmesi

İşlenmiş yorumlar ve analiz sonuçları bir pandas DataFrame'ine dönüştürülmüş ve `simplified_comments.csv` dosyası olarak kaydedilmiştir. Bu dosya, duygu analizi modeli için temiz ve tutarlı bir veri kaynağı oluşturmuştur.

Bu süreçler sonucunda, yorumlar ve puanlar, duygu analizinde kullanılabilecek temiz ve yapılandırılmış bir formata dönüştürülmüştür. Zemberek kütüphanesi, Türkçe diline özel dil bilgisi kurallarını analiz ederek veri ön işleme sürecinde önemli bir rol oynamıştır.

IV. PROJE UYGULAMA DETAYLARI

Bu proje kapsamında, Trendyol platformundan veri çekme işlemleri için Python programlama dili ve BeautifulSoup, Selenium, pandas, requests ve random kütüphaneleri kullanılarak bir kod geliştirilmiştir. Veri çekme işlemleri, Selenium ve BeautifulSoup kütüphaneleri yardımıyla gerçekleştirilmiştir. Bu araçlar, dinamik web sayfalarından veri toplama sürecinde rol almıştır.

Veri çekme işlemi iki ana adımda yapılmıştır: kategori sayfalarından alt kategori ve ürün bağlantılarının toplanması ve ürün sayfalarındaki kullanıcı yorumlarının alınması.

A. Kategori ve Ürün Bağlantılarının Toplanması

Veri çekme işlemi, Trendyol ana sayfasındaki ana kategorilerin ve alt kategorilerin tespiti ile başlamıştır. `fetch_categories` fonksiyonu, Trendyol ana sayfasına erişim sağladıktan sonra, sayfa üzerinde çıkan pop-up pencereleri otomatik olarak kapatmak için Selenium'un `WebDriverWait` fonksiyonunu kullanır. Kategoriler listelendikten sonra, her bir kategoriye tıklanır ve alt kategoriler BeautifulSoup ile analiz edilerek elde edilir. Kategori ve alt kategorilerde gezinme işlemleri sırasında sayfa yüklenme süreleri dinamik olarak yönetilmiştir.

B. Ürün Yorumlarının Toplanması

Alt kategorilere ulaşıldıktan sonra, her bir kategori sayfasındaki ürünler ziyaret edilmiştir. Ürünlerin bağlantıları alındıktan sonra, ürün yorumlarının olduğu sayfaya gidilerek kullanıcıların ürünlere verdiği puanlar ve yorumlar toplanmıştır. `scarape_comments` fonksiyonu, her ürün sayfasında yer alan yorumları `comment-text` ve `comment-rating` sınıflarına göre seçer. `rating_count` fonksiyonu, yorumlardaki yıldız puanlarını belirlemek için kullanılır ve `div` elemanlarından alınan yüzde oranlarını yorumlayarak doğru puan değerini hesaplar.

C. Veri Yapılandırma ve Kaydetme

Toplanan yorumlar ve puanlar, analiz sürecinde kullanılacak olan `comments_data` listesine eklenmiştir. Her yorumun puanı ve metni düzenlenmiş olarak bu listeye kaydedilmiştir. Son olarak, tüm yorumlar ve puanlar bir pandas DataFrame'ine dönüştürülmüş ve `comments2.csv` dosyası olarak kaydedilmiştir.

Bu süreçte, veri çekme işlemi sırasında sunucuya aşırı yüklenmeyi önlemek ve veri çekme işlemleri daha sağlıklı yönetebilmek için istekler arasında rastgele bekleme süreleri kullanılmıştır. Projenin sonunda, Trendyol kullanıcı yorumları ve puanları CSV formatında düzenlenmiştir.

V. GİTHUB VE VERİ PAYLAŞIMI

Bu proje kapsamında geliştirilen kodlar ve toplanan veriler, GitHub üzerinden paylaşılmıştır. Proje dosyaları, veri toplama ve duygu analizi sürecinde kullanılan tüm betikleri ve belgeleri içermektedir. Bu paylaşım, diğer grup üyelerinin veya geliştiricilerin projeyi incelemesine, yeniden kullanmasına ve geliştirmesine olanak tanımaktadır.

GitHub deposunda, projenin nasıl çalıştırılacağını ve kurulum için gereken adımları açıklayan bir 'README' dosyası bulunmaktadır. Bu dosyada, proje özeti, kullanılan teknolojiler, geliştirme ortamı ve projeyi çalıştırmak için izlenmesi gereken adımlar ayrıntılı bir şekilde belirtilmiştir. Ayrıca, projede kullanılan ana Python dosyaları ve yardımcı kodlar hakkında kısa açıklamalar sağlanmıştır.

Toplanan veri seti, GitHub deposunun boyut sınırlamaları nedeniyle doğrudan yüklenememiştir. Bunun yerine, veri seti Google Drive üzerinde depolanmış ve GitHub 'README' dosyasına bir indirme bağlantısı eklenmiştir. Bu bağlantı aracılığıyla, proje ile ilgili veri setine kolayca erişim sağlanmaktadır.

Bu veri paylaşımı yöntemi, veri setinin büyük boyutuna rağmen grup üyeleri tarafından erişilebilir olmasını sağlamaktadır. Ayrıca, bu yöntemin kullanımı, veri paylaşımında sürdürülebilirliği sağlamak ve veri bütünlüğünü korumak açısından önemlidir.

VI. SONUÇLAR VE GELECEK ÇALIŞMALAR

Bu proje kapsamında, Trendyol platformundan toplanan müşteri yorumları ve puanları kullanılarak duygu analizi yapılmıştır. Elde edilen sonuçlar, müşterilerin ürünlere yönelik olumlu veya olumsuz görüşlerini daha iyi anlamamıza olanak sağlamıştır. Web scraping ve doğal dil işleme tekniklerinin birleşimiyle, kullanıcı geri bildirimlerinin duygu analizi modeline uygun hale getirilmesi sağlanmış ve model doğruluğunu artırmak için gerekli veri ön işleme adımları başarıyla uygulanmıştır.

Projede kullanılan yöntemler, e-ticaret platformlarındaki müşteri memnuniyeti analizine yönelik değerli içgörüler sunmaktadır. Yapılan analizler sonucunda, kullanıcıların ürünler hakkında belirttikleri duygusal ifadelerin sınıflandırılması mümkün olmuş ve bu sınıflandırmalar doğrultusunda olumlu ve olumsuz eğilimler ortaya konmuştur. Modelin doğruluğu, elde edilen verilerin kalitesine ve ön işleme adımlarının etkinliğine bağlı olarak olumlu bir performans sergilemiştir.

Gelecek Çalışmalar: Bu çalışmanın devamı olarak, model doğruluğunu artırmak ve daha fazla veri türünü analiz sürecine dahil etmek amacıyla ek geliştirmeler yapılabilir. Özellikle, farklı e-ticaret platformlarından veri toplanarak daha geniş bir veri seti oluşturulabilir. Ayrıca, yorumların sadece olumlu veya olumsuz olarak sınıflandırılmasının ötesine geçerek, "memnuniyet", "öneri", "şikayet" gibi daha ince duygusal kategorilerle sınıflandırma yapılabilir. Derin öğrenme algoritmalarının kullanımı ile duygu analizi modelinin doğruluğu daha da artırılabilir.

Bu önerilen geliştirmeler, duygu analizi sürecini daha kapsamlı hale getirerek müşteri davranışlarının analiz edilmesine daha derin bir bakış sağlayabilir.

KAYNAKÇA

REFERENCES

- [1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [2] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130–137, 1980.
- [3] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, 2009.
- [4] J. Brownlee, "A Gentle Introduction to Sentiment Analysis," *Machine Learning Mastery*, 2020. [Online]. Available: <https://machinelearningmastery.com/sentiment-analysis/>.
- [5] A. Balahur, R. Steinberger, M. Kabadjov, V. Zavarella, E. Van Der Goot, M. Halkia, B. Pouliquen, and J. Belyaeva, "Sentiment analysis in the news," *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pp. 2216–2220, 2010.
- [6] J. Ramos, "Using tf-idf to determine word relevance in document queries," *Proceedings of the First Instructional Conference on Machine Learning*, 2003.
- [7] G. Pang, J. Cao, L. Chen, and L. Liu, "Text classification algorithms: A survey," *IEEE Access*, vol. 8, pp. 166890–166906, 2020.
- [8] Selenium Documentation, "WebDriver API," [Online]. Available: <https://www.selenium.dev/documentation/en/webdriver/>. [Accessed: Nov. 10, 2024].
- [9] BeautifulSoup Documentation, "Beautiful Soup Documentation," [Online]. Available: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>. [Accessed: Nov. 10, 2024].
- [10] Berkay Öztürk, "Zemberek Python Examples," GitHub Repository. [Online]. Available: <https://github.com/ozturkberkay/Zemberek-Python-Examples>. [Accessed: Nov. 10, 2024].
- [11] Ahmet Akın, "Zemberek NLP Library," GitHub Repository. [Online]. Available: <https://github.com/ahmetaa/zemberek-nlp>. [Accessed: Nov. 10, 2024].
- [12] Zemberek JAR Dosyası, Google Drive. [Online]. Available: <https://drive.google.com/drive/u/0/folders/1FN80VbqesnqU21us4c4Pvgv2VqUsSf2z>. [Accessed: Nov. 10, 2024].