ÖZET

İnternet, öğrenmek, fikir üretmek ve ürünlerin, hizmetlerin, insanların ve yerlerin hakkındaki kullanıcı yorumlarını bulmak için vazgeçilmez bir kaynak haline gelmiştir. Her gün milyonlarca çevrimiçi inceleme ürünler, insanlar ve yerler hakkında yayınlanmaktadır. Müşteri incelemeleri, ürünlerini geliştirmek isteyen işletmeler için önemli bir performans göstergesidir ve gelecekteki müşterilerin geçmiş deneyimleri anlamaları için de hayati önem taşır. Bu nedenle, duygu analizi, hızla büyüyen bir araştırma alanıdır ve müşteri incelemelerini analiz etmek için doğal dil işleme, metin madenciliği ve istatistik gibi ilgili alanlar büyük bir rol oynamaktadır. Bu çalışmada, otomatik (sözlük tabanlı) yöntemlerin sınırlamalarını ve manuel yöntemlerin etkinliğini vurgulayarak duygu analizine katkıda bulunulmaktadır. R Studio yazılımı, önerilen çalışmanın uygulanmasını kolaylaştırır. İlk olarak, web kazıma teknikleri ve veri ön işleme yoluyla Amazon'dan ürün incelemeleri ve derecelendirmeleri alırız. Ardından, üç duygu sözlüğü (Bing, NRC, Afinn) ve manuel bir yöntem kullanarak duygu analizi gerçekleştirilir. İncelemeler daha sonra olumlu veya olumsuz olarak kategorize edilir ve ardından otomatik (Bing sözlük tabanlı) ve manuel yöntemlerin sonuçlarını analiz etmek için çoklu regresyon modeli geliştirilir. Sonuçlara bakıldığında, manuel yöntemin varyansın yaklaşık %45,44'ünü açıklaması, otomatik yöntemin ise yalnızca yaklaşık %3,26'sını açıklaması görülmüştür ve iki model karşılaştırıldığında otomatik (Bing sözlüğü) yöntemine kıyasla manuel yöntem sonucu oluşan model anlamlı çıkmıştır.

ABSTRACT

The internet has become an essential resource for learning, generating ideas, and finding reviews for products and services. Daily, millions of reviews are posted online about products, people, and places. Customer reviews of products are a critical performance indicator for businesses looking to improve their contributions. They’re also important for future customers to understand previous customers’ experiences. Therefore, sentiment analysis is a rapidly growing research field where natural language processing, text mining, and statistics plays a major role to analyze customer reviews. In this study, the aim is to **study contributes to sentiment analysis by highlighting the limitations of automatic (lexicon based)methods and the effectiveness of manual methods.** R Studio software is used to implement the proposed study. First, we extract a product reviews and ratings from Amazon through web scraping techniques and preprocessing the data. Then, we conduct sentiment analysis using three lexicons(Bing, NRC, Afinn) and manual method. Subsequently, the reviews are categorized into positive and negative senntiments, followed by building a multiple regression model to discuss the outcomes for automatic(bing lexicon based) and manual methods. The results show that the manual method yields significant results, explaining approximately 45.44% of the variance, whereas the automatic (bing-lexicon based) method does not, explaining only approximately 3.26%.

INTRODUCTION

In the digital age, the internet has become a necessary resource for consumers seeking information, generating ideas, and finding reviews of products and services. Every day, millions of reviews are posted online, providing valuable insights into user experiences and imfluencing purchasing decisions. For businesses, customer reviews are essential for understanding how satisfied their customers are and spotting areas where they can improve. Potential customers, on the other hand, depend on reviews to understand the advantages and disadvantages of a product prior to making a purchase. This highlights the growing importance of sentiment analysis, a field that utilizes natural language processing (NLP), text mining and statistical techniques to extract and analyze the emotional tone(positive,negative or neutral) from customer reviews.

Researches continuosly work to improve to accuracy of sentiment analysis methods, leading to the development of numerous packages in software programmes such as R and Phyton. Given the limitations of lexicon-based sentiment analysis methods, this study aims to investigate the effectiveness of manual sentiment analysis compared to automatic methods. Lexicon-based methods rely on predefined word lists (e.g, NRC’s eight emotions, Bing’s positive/negative, AFINN’s sentiment scores) to categorize sentiment. However, these methods can struggle with nuances and context. To address this, we will use web scraping to collect Amazon reviews and apply NRC, Bing, AFINN lexicons along with manual sentiment analysis. We will then compare the explanatory power of both methods using regression modeling to assess the effectiveness of lexicon-based sentiment analysis compared to manual analysis.

Various methods have been employed to address this task, utilizing a range of different techniques. Some of them use lexicon based methods while others utilize machine learning methods. There are also few simple yet effective methods which give satisfactory results.

In a study, the researches exploring sentiment analysis and word cloud techniques in text mining. Their research on Amazon earphone reviews shows how these methods can be used to understand customer sentiment, improve customer experience enhancement, and inform customer-centric decision-making. Their work highlights the potential of these techniques for extracting insights from big data readily available online. [1] (Karim et. al., 2020)

Another study explores a web-based text analytics approach using R to analyze and summarize review data. Their research focuses on health product reviews from Shaklee’s platform. This approach utilizes R’s data processing capabilities and integrates various Natural Language Processing (NLP) features into a user-friendly menu. The menu displays all NLP features in a step-by-step manner with clear labels, simplifying the text summarization process. While the study acknowledges the proposed approach's effectiveness compared to a baseline model, further investigation into specific performance metrics would strengthen the findings. [2] (Kadir and Aliman, 2020)

In another research, they investigates how to improve understanding customer sentiment on Amazon.com reviews. By analyzing a dataset of 1000 reviews (balanced between positive and negative), the study compares different classification methods for sentiment analysis. The key finding is that K-Nearest Neighbor combined with a lexicon technique achieved the highest accuracy (92.67%) in identifying positive or negative sentiment, followed by SVM with lexicon (91.33%). This suggests lexicon techniques can significantly improve sentiment analysis for e-commerce reviews. (Prakaso et. al., 2018)

[1] Kabir, A., Ahmed, K., & Karim, R. (2020). Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language. Informatica Economica.

[2] Nasibah Husna Mohd Kadir, & Sharifah Aliman. (2020). Text analysis on health product reviews using r approach. Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 18(3), 1303-1310.

[3] A. Aryo Prakoso, B. Winantesa Yananta, A. Fitra Setyawan and Muljono, "A Lexicon-Based Sentiment Analysis for Amazon Web Review," 2018 International Seminar on Application for Technology of Information and Communication, Semarang, Indonesia, 2018, pp. 503-508.