**Sentiment Analysis of Online Product Review: A Comparison of Natural Language Processing Techniques**

|  |  |  |
| --- | --- | --- |
| |  | | --- | | **Chinedu Okonkwo, Abdelrhman Alroudi, Karla De La Cruz, Mihn Le, Ryan Gill**  The University of Texas at San Antonio | |  | |

**Introduction/Methods**

Sentiment analysis refers to the examination of people’s opinions and feelings towards an entity or topic (Medhat et al., 2014). For a business to improve the quality of their services, customers’ reviews on their products and services are vital to properly address the needs of their customers. With the advent of the internet, millions of reviews are readily available about a product or service, which makes it increasingly difficult to manually examine the opinions of consumers expressed through online reviews (Shivaprasad & Shetty, 2017). Natural Language Processing (NLP) is a computerized method of text analysis that is founded on a set of theories as well as a set of technology (Liddy, 2001). Different techniques of NLP have been utilized in the past for sentiment analysis of online product review such as TextBlob, Naïve Bayes classifier, lexicon-based method, supervised classifier, Named entity based method, Bidirectional Encoder Representations from Transformers (BERT) (Bade Shrestha & Bal, 2020; Duwairi & Qarqaz, 2014; Laksono et al., 2019; Taboada et al., 2011), with different levels of accuracy on different datasets. The aim of this study is to compare the various techniques on the same dataset to determine which technique is most suitable for sentiment analysis of online product reviews.

**Problem Statement**

Different NLP techniques have been applied to different datasets for sentiment analysis of news, online product reviews, etc. These techniques have shown different levels of accuracy on different datasets. Using different datasets, however, makes it difficult to juxtapose these techniques. This presents a need to compare these techniques on the same dataset to better identify which technique is better for the sentiment analysis of online reviews. In this study, we will compare five NLP techniques for sentiment analysis. The five techniques include TextBlob, Naïve Bayes classifier, lexicon-based method, supervised classifier, and Named entity-based method. The result of this study is expected to help businesses identify the most suitable NLP technique to better understand how their customers feel about their products and the services they render.

**Methodology**

To achieve the aim of this study, online reviews from amazon.com will be used as data source. Amazon is one of the largest e-commerce platforms with a wide variety of products and customer reviews, as well as other useful information like the credibility of the reviewer and rating of the product, which could be useful in selecting the most appropriate dataset for the study. Beautiful Soup, a python library will be used to extract raw data from amazon.com. the data will be preprocessed with WordNet and Parts of Speech Tagging (POST) before implementing the five NLP techniques for sentiment analysis. The output of the techniques will be compared to determine which of the five techniques performs best in analyzing online product reviews.

**Experimental Setup**

Either using Beautiful Soup, as described above, or already generated csv files from a website that has already compiled the data like data.world: <https://data.world/datafiniti/consumer-reviews-of-amazon-products> . The Keras library within tensorflow is more than capable of handling POST and implementations of the other methods and WordNet is accessible through NLTK. Once our implementations are complete, we can compare the different techniques on individual reviews and the whole dataset to determine if there are any notable findings.

**References**

Bade Shrestha, B., & Bal, B. K. (2020). Named-Entity Based Sentiment Analysis of Nepali News Media Texts. *Proceedings of the 6th Workshop on Natural Language Processing Techniques for Educational Applications*, 114–120. https://aclanthology.org/2020.nlptea-1.16

Duwairi, R. M., & Qarqaz, I. (2014). Arabic Sentiment Analysis Using Supervised Classification. *2014 International Conference on Future Internet of Things and Cloud*, 579–583. https://doi.org/10.1109/FiCloud.2014.100

Laksono, R. A., Sungkono, K. R., Sarno, R., & Wahyuni, C. S. (2019). Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes. *2019 12th International Conference on Information Communication Technology and System (ICTS)*, 49–54. https://doi.org/10.1109/ICTS.2019.8850982

Liddy, E. D. (2001). *Natural Language Processing*. 15.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Shivaprasad, T. K., & Shetty, J. (2017). Sentiment analysis of product reviews: A review. *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 298–301. https://doi.org/10.1109/ICICCT.2017.7975207

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267–307. https://doi.org/10.1162/COLI\_a\_00049