

Symbiosis Institute of Computer Studies & Research

**User Sentiment Analysis**

|  |
| --- |
| Submitted By:  Nachiket Chaudhari (22030142004) |

|  |
| --- |
| Submitted To:    Mrs. Anuja Bokhare |

**ABSTARCT**

The Internet brought revolution. People started using the Internet, and some companies took the technology as an opportunity and started selling their products over the Internet. After gaining popularity over the Internet, everybody started selling their product over the Internet and with heavy profit. As competition increases, many people start selling the same product with better quality and price. So, this indirectly increases the competition over the Internet. So, getting customer feedback becomes more critical to survive over the Internet. So, here sentiment analysis provides the company with a guide on what exactly the user wants and what are their expectation and reviews about their product, so the company will constantly work on their product to get a better version of what it is. So, by doing sentiment analysis company gets a strong position in the market. Sentiment analysis can also help the company grow and gain its customers' trust. Sentiment analysis divides the sentiment into three parts – positive, neutral, and hostile. With these three reactions, companies can improve their products and gain a solid market position. Though Sentiment analysis is a small part of development, it can play a crucial role. Sentiment analysis can be done for things like products review, movie reviews, company reviews, Etc.

**INTRODUCTION**

In the modern digital era, businesses and researchers are inundated with massive amounts of unstructured data, a significant portion of which is text containing invaluable customer feedback about products and services. The sheer volume of this data necessitates using automated methods to derive insights. One such method is Product Sentiment Analysis (PSA), a specialized form of sentiment analysis that focuses on discerning and interpreting consumer emotions, attitudes, and opinions toward specific products. With its roots in natural language processing (NLP), machine learning (ML), and computational linguistics, PSA endeavors to convert qualitative text data into quantitative sentiment scores that can be systematically analyzed. This process enables businesses to understand their customers better, informing product development, marketing strategies, and customer service initiatives. The present study will explore innovative techniques and applications of PSA, offering new insights into how technology can amplify our ability to understand and respond to consumer sentiment effectively.

While traditional methods for sentiment analysis have yielded considerable success, the dynamic nature of human language and its multifaceted context-related characteristics present persistent challenges. As we continue to evolve in a global, digital landscape, dialectical variations, linguistic nuances, and cultural factors have grown increasingly prevalent and difficult to ignore in the analysis of sentiments. Hence, newer, more sophisticated techniques that can handle such complexities are of great interest to academia and industry. The burgeoning field of deep learning and the subsequent advent of transformer-based models have initiated a significant shift in how we approach sentiment analysis. These models, equipped with their understanding of contextual relations between words, offer promising solutions to the challenges that traditional machine learning models face.

In this study, we aim to delve into the world of sentiment analysis, focusing on applying cutting-edge deep learning models in extracting and interpreting sentiment from product reviews. We further explore these models' capability in handling the complexities of real-world language and providing a more granular understanding of consumer sentiments. The findings from this study will contribute to the academic discourse in the realm of sentiment analysis and provide practical implications for businesses striving to understand their customers better.

PREVIOUS STUDY –

1] **Pang, B., Lee, L., and Vaithyanathan, S. (2002)** pioneered the application of machine learning techniques for the categorization of sentiment in movie reviews in their work, "Thumbs up? Sentiment Classification using Machine Learning Techniques". While they found that machine learning techniques were effective for the task, they also acknowledged the complexity of distinguishing between content that was pertinent and that which was not.

2] **Peter D (2002).** Turney introduced an unsupervised learning algorithm for the classification of reviews as positive or negative in his paper, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". His contributions have been a cornerstone in the subsequent developments in the field of sentiment analysis.

3] **Bing Liu's "Sentiment Analysis and Opinion Mining" (2012)** is a complete survey of sentiment analysis techniques and research. It encompasses a wide range of topics including sentiment categorization, summarizing opinions, and the detection of fraudulent reviews or opinions.

4] **Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013)** introduced a novel way of representing words and phrases and their associated meanings through neural networks, known as word2vec, in their paper "Distributed Representations of Words and Phrases and their Compositionality". These word embeddings, which can capture semantic relationships, have been key in enhancing sentiment analysis by providing a better comprehension of context.

5] **Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2019)** significantly improved the effectiveness of sentiment analysis with their introduction of BERT ("Bidirectional Encoder Representations from Transformers") in their paper "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". BERT models allow for a bidirectional understanding of context in words, providing a significant advancement in sentiment analysis performance.

6] **Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., and Potts, C. (2013)** introduced a model that comprehends phrases and longer texts through the construction of a parse tree. The model operates by synthesizing the sentiments of child nodes to gain insight into the parent node.

7] **Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014)** put forward a technique for learning sentiment-specific word embeddings in their research, "Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification". This approach drastically enhanced the performance of sentiment classification.

8] **Severyn, A., & Moschitti, A. (2015).** In the paper "Twitter Sentiment Analysis with Deep Convolutional Neural Networks", the authors proved the efficacy of Convolutional Neural Networks (CNNs) for sentiment analysis, particularly with Twitter data.

9] **Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011).** In the paper "Learning Word Vectors for Sentiment Analysis", they introduced a novel dataset composed of movie reviews. This dataset has since been extensively utilized for testing sentiment analysis algorithms.

10] **Hutto, C.J., Gilbert, E.E. (2014).** Introduced VADER (Valence Aware Dictionary and Sentiment Reasoner) in their work "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text". VADER is a model employed for textual sentiment analysis, adept at identifying both polarity (positive/negative) and intensity (strength) of emotion. It's particularly effective when applied to social media text.

**PROBLEM FORMULATION**

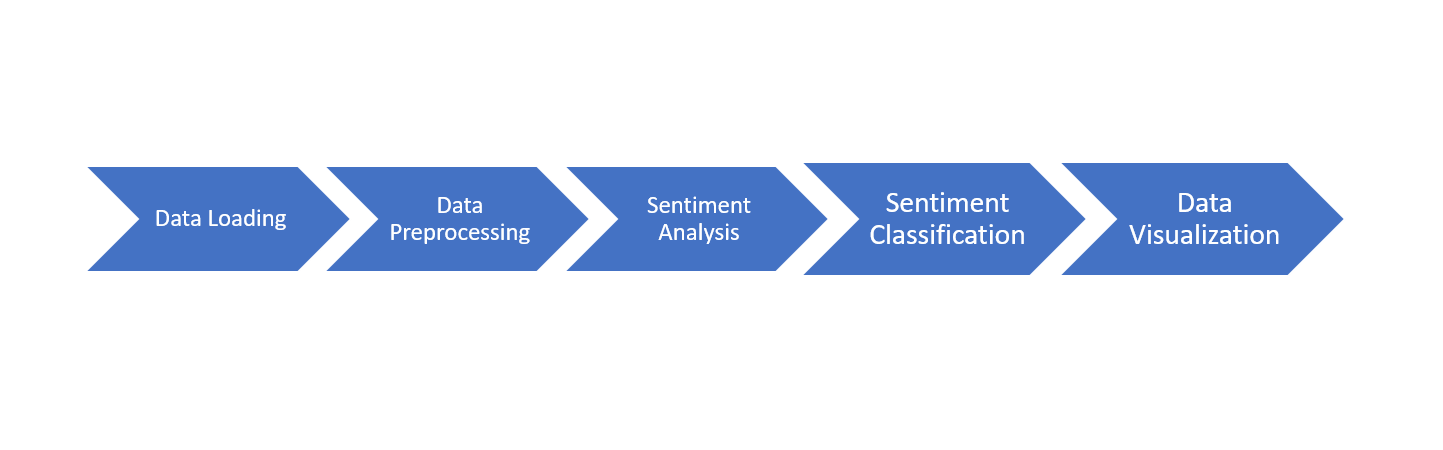
User Sentiment Analysis, a subset of sentiment analysis, primarily focuses on extracting, identifying, and understanding the sentiment expressed by users in text format towards a particular subject or context. While significant advancements have been made in this field, several challenges persist that need to be addressed to improve the efficiency and effectiveness of existing User Sentiment Analysis.

The core problem we aim to address in this study involves developing an advanced user sentiment analysis system capable of capturing the nuances and complexities inherent in user-generated content. User language often involves slang, abbreviations, emoticons, and code-mixed language, which poses a significant challenge for standard natural language processing algorithms. Moreover, linguistic features such as sarcasm, irony, and implicit sentiments are difficult to detect and interpret accurately.

Another aspect of the problem lies in the classification granularity of sentiments. Traditional sentiment analysis models are usually confined to binary (positive/negative) or ternary (positive/negative/neutral) classification. However, real-world sentiments are often more complex and can encompass a range of emotions beyond just positive, negative, or neutral.

Furthermore, the scalability and adaptability of the sentiment analysis system across various domains, languages, and cultures represent a substantial part of the problem. Sentiments can be expressed differently across diverse cultural, geographical, and linguistic contexts, necessitating a model capable of understanding these variations.

**PROPOSED MODEL**

****

**Data Loading**: Here, the data has been provided to the model.

**Text Preprocessing**: Clean the text data by removing unnecessary noise, including URLs, numbers, and special characters. Convert the text into lowercase and apply stemming to reduce words to their root form.

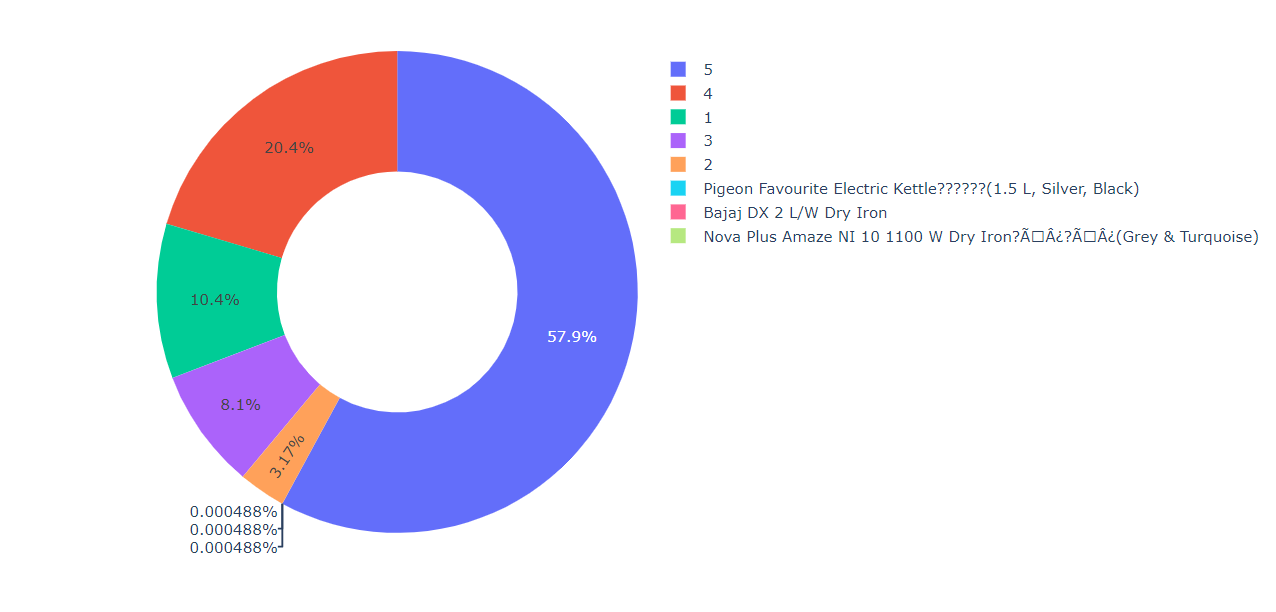
**Sentiment Analysis**: Use VADER from NLTK to analyse the sentiments of the text data.

**Sentiment Classification**: Classify sentiments as positive, negative, or neutral based on compound score.

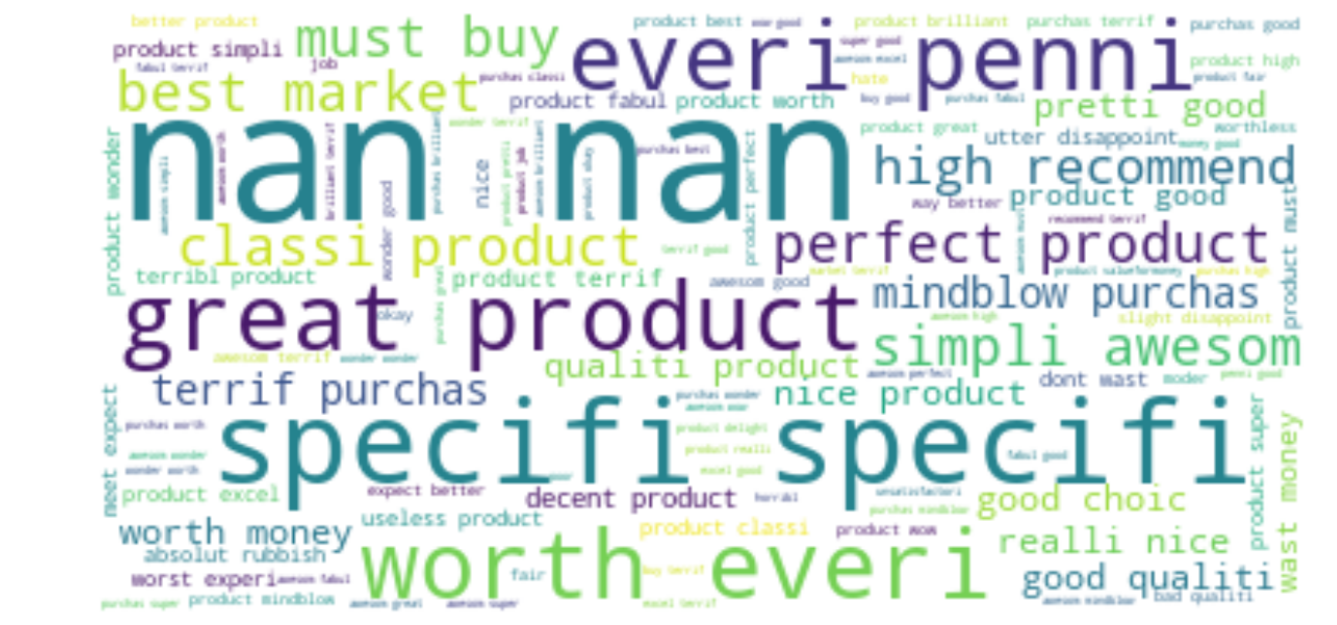
**Data Visualization**: Show your result using various graphs.

**DATASET AND EXPERIMENT DISCUSSION**

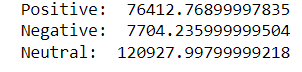
The Flipkart sentiment analysis dataset taken from Kaggle is a collection of customer reviews from Flipkart. It includes both the text of the study and the rating given by the customer. The dataset contains six columns with 205053 records. This dataset consists of the main features of any product, like product\_name, price, rate, review, summary, and sentiment.



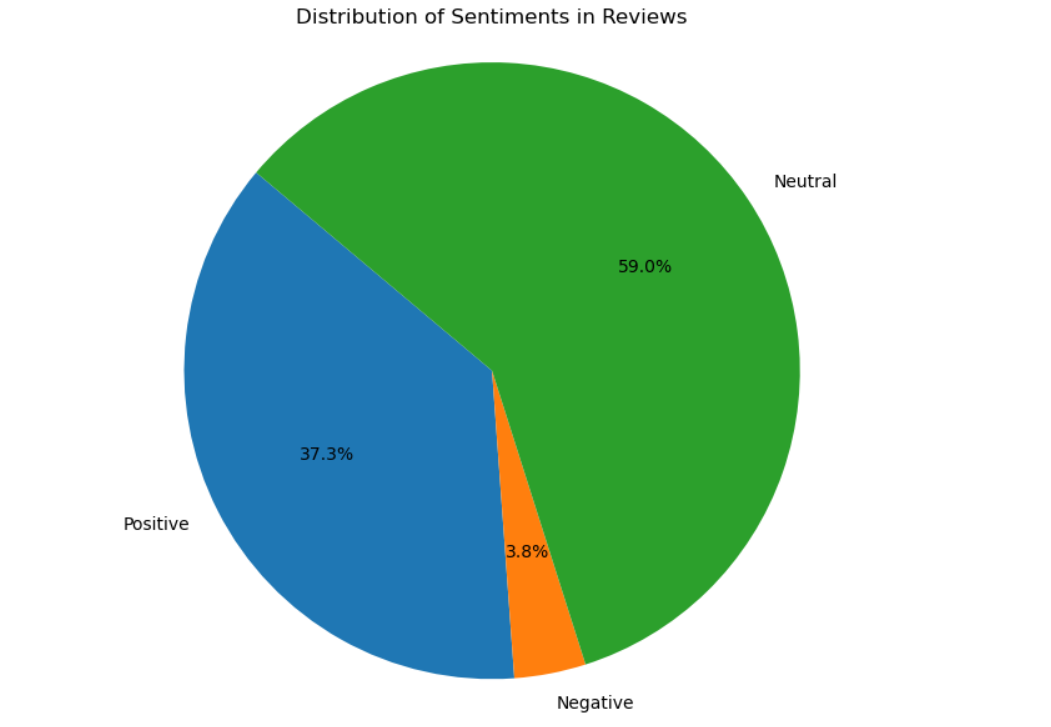
The figure shows How much Rating is given by people to Flipkart products.



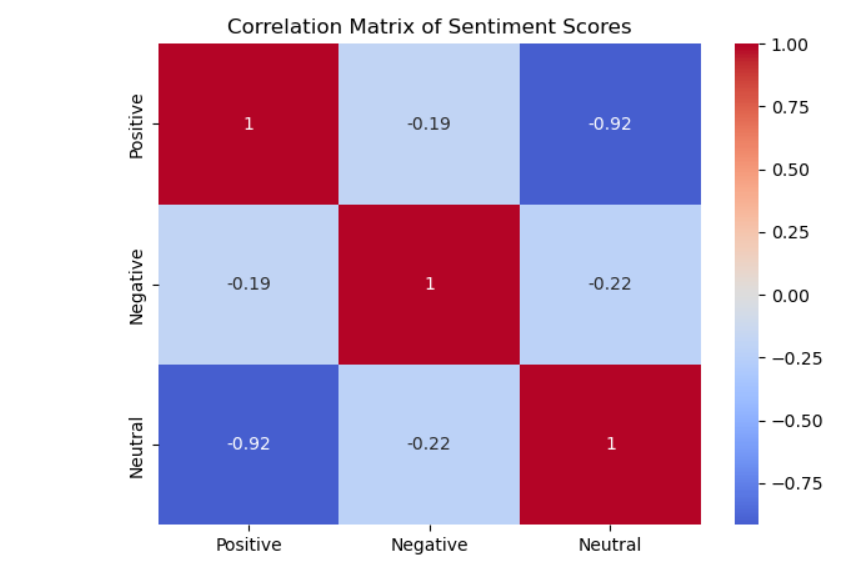
The presented image is a word cloud that graphically illustrates the User Sentiment Analysis found in the input text. Each word's prominence and scale within the cloud correlate with its occurrence rate in the text. User Sentiment Analysis, like great products, busy buys, and more, are depicted in the word cloud, providing a straightforward visualization of the most frequently occurring User sentiment analysis instances.



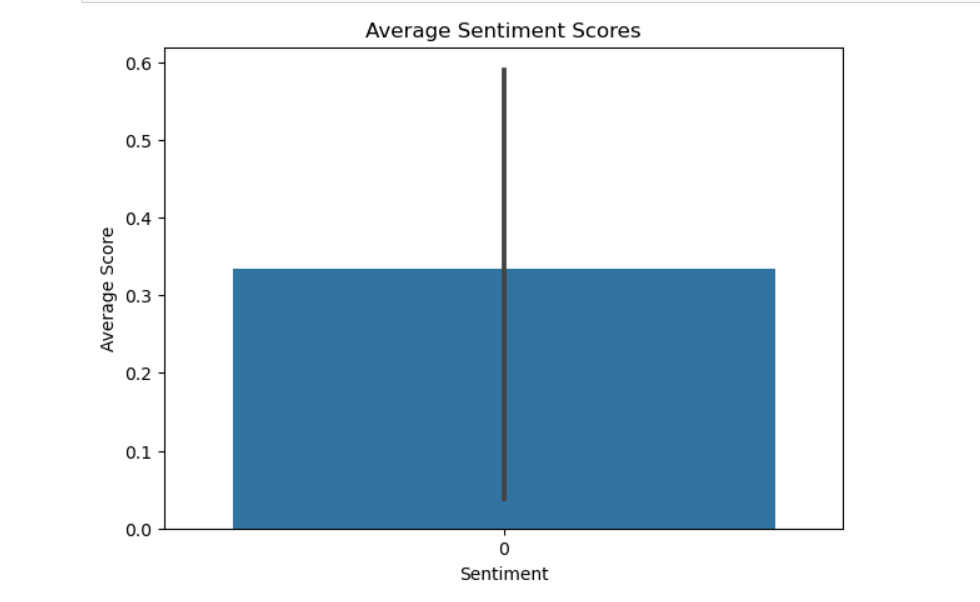
The sentiment analysis from the user feedback unveils some compelling data. A significant chunk of the input, close to 120,927, is categorized as 'Neutral,' suggesting these users exhibit a fair or indifferent view towards the product or service. On the positive side, a substantial number of sentiments, approximately 76,412, reflect a favorable impression or positive response. Conversely, the 'Negative' sentiments, around 7,704, while fewer, must be addressed as they represent a set of users with less-than-positive experiences. This information highlights the crucial need for ongoing enhancements to products or services with the goal of converting neutral and negative impressions into positive feedback.



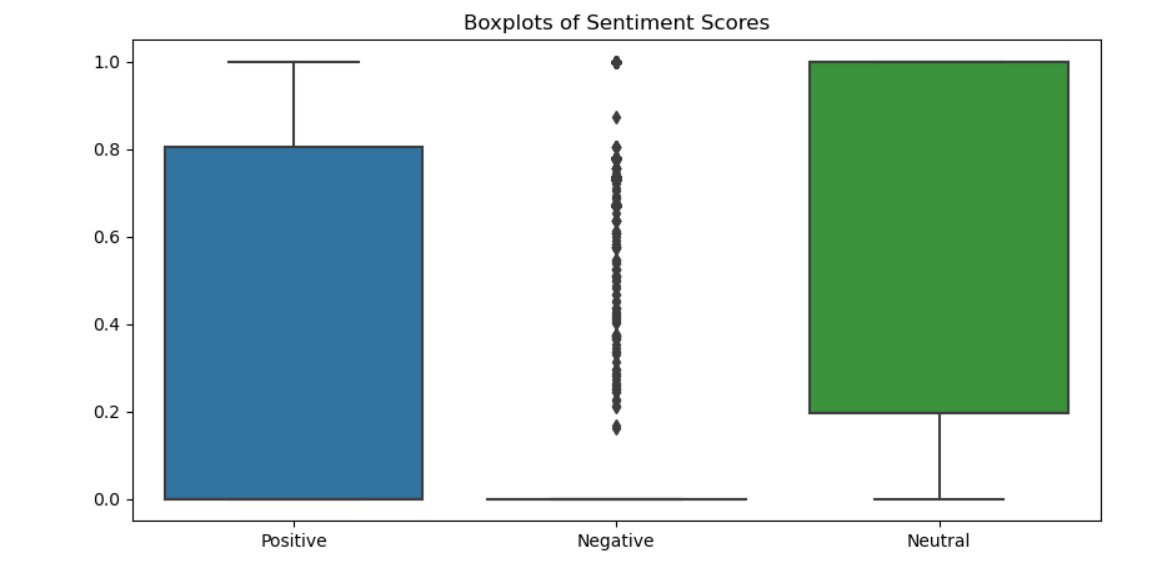
The Figure shows the distribution of Positive, Negative, and Neutral reviews given by users.



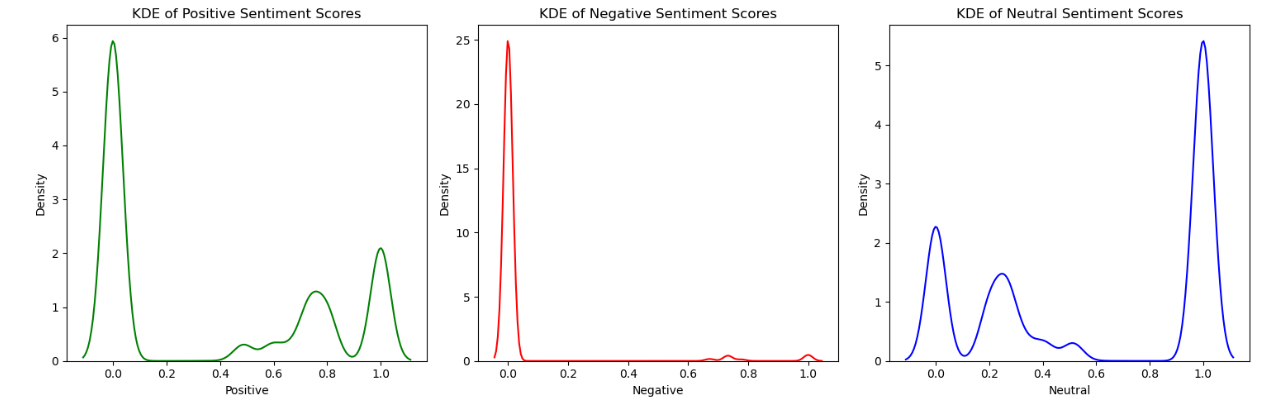
The heatmap of the correlation matrix represents the correlation coefficients between pairs of variables (sentiment scores). Correlation coefficients range from -1 to 1. Positive values denote positive linear correlation; negative values denote negative linear correlation; zero value suggests no linear correlation.



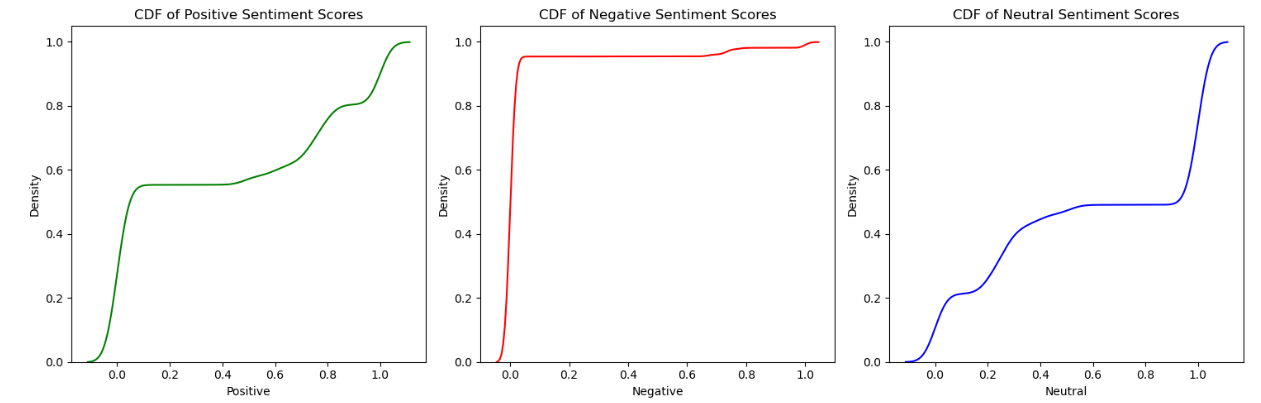
A bar chart gives you a comparative view of the average positive, negative, and neutral scores.



Boxplots have given a good idea of how the positive, negative, and neutral scores are spread across the data and can also help you identify any potential outliers.



The given figure shows Kernel Density Estimate (KDE) plots for the positive, negative, and neutral sentiment scores contained within the dataset.



The figure shows Cumulative Distribution Function (CDF) plots for the positive, negative, and neutral sentiment scores found within the dataset. Each plot visually represents the cumulative probability for the respective sentiment scores.

**FUTURE SCOPE**

The future of user sentiment analysis holds immense potential, with several transformative developments on the horizon. Promising areas include:

* Emotion AI for more nuanced sentiment interpretation.
* Multilingual and cross-cultural models for global applicability.
* Real-time sentiment analysis for immediate response during live interactions.

Moreover, integrating sentiment analysis with IoT devices could lead to more responsive and intuitive systems. At the same time, advancements in NLP could improve the detection of complex language constructs like sarcasm and irony. Also, merging textual data with non-textual elements such as voice tonality or facial expressions could offer a more comprehensive understanding of user sentiment. Despite these exciting prospects, overcoming challenges such as data scarcity, privacy concerns, model interpretability, and bias remains crucial.

**CONCLUSION**

User sentiment analysis is an essential tool in today's digital era, with far-reaching applications across diverse domains, from business and marketing to public opinion research and mental health assessment. Extracting and interpreting user emotions from textual data offers valuable insights that can drive strategic decision-making, enhance customer experience, and promote a better understanding of public sentiment on various issues. Although current models have made significant strides, future developments hold immense promise, such as real-time and emotion-specific sentiment analysis, integration with IoT devices, and improved handling of linguistic complexities like sarcasm and irony. However, these advancements also underscore the necessity of addressing challenges around data privacy, algorithmic bias, and model interpretability. By tackling these hurdles, the field of sentiment analysis can continue to evolve and play a pivotal role in shaping our digital interactions and understanding of human sentiment.

Through sentiment analysis, businesses can enhance customer relationship management, develop more effective marketing strategies, and design better products catering to customers' sentiments. Similarly, sentiment analysis can aid in gauging public opinion on policies or social issues in the public sector, helping to create more citizen-centric initiatives. In the realm of social media, it can monitor user reactions and trends, enabling platforms to tailor user experiences. Moreover, as technological advances like AI and machine learning continue to revolutionize this field, the accuracy and effectiveness of sentiment analysis are set to soar, allowing for more nuanced understanding and applications. However, it's critical that as we continue to innovate, we also diligently address the ethical implications concerning data privacy and fairness to ensure the responsible use of this powerful tool.

**REFERENCE**

1] Hasan, Md Rakibul, Maisha Maliha, and M. Arifuzzaman. "Sentiment analysis with NLP on Twitter data." *2019 international conference on computer, communication, chemical, materials and electronic engineering (IC4ME2)*. IEEE, 2019.

2] Dang, Nhan Cach, María N. Moreno-García, and Fernando De la Prieta. "Sentiment analysis based on deep learning: A comparative study." *Electronics* 9.3 (2020): 483.

3] Yaakub, Mohd Ridzwan, Muhammad Iqbal Abu Latiffi, and Liyana Safra Zaabar. "A review on sentiment analysis techniques and applications." *IOP conference series: materials science and engineering*. Vol. 551. No. 1. IOP Publishing, 2019.

4] Kanakaraj, Monisha, and Ram Mohana Reddy Guddeti. "Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques." *Proceedings of the 2015 IEEE 9th international conference on semantic computing (IEEE ICSC 2015)*. IEEE, 2015.

5] Jing, Nan, Zhao Wu, and Hefei Wang. "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction." *Expert Systems with Applications* 178 (2021): 115019.

6] Dang, Cach N., María N. Moreno-García, and Fernando De la Prieta. "Hybrid deep learning models for sentiment analysis." *Complexity* 2021 (2021): 1-16.

7] Akhtar, Md Shad, et al. "A hybrid deep learning architecture for sentiment analysis." *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2016.

8] Abd El-Jawad, Mohammed H., Rania Hodhod, and Yasser MK Omar. "Sentiment analysis of social media networks using machine learning." *2018 14th international computer engineering conference (ICENCO)*. IEEE, 2018.

9] Prabowo, Rudy, and Mike Thelwall. "Sentiment analysis: A combined approach." *Journal of Informetrics* 3.2 (2009): 143-157.

10] Birjali, Marouane, Mohammed Kasri, and Abderrahim Beni-Hssane. "A comprehensive survey on sentiment analysis: Approaches, challenges and trends." *Knowledge-Based Systems* 226 (2021): 107134.

11] Abirami, A. M., and V. Gayathri. "A survey on sentiment analysis methods and approach." *2016 Eighth International Conference on Advanced Computing (ICoAC)*. IEEE, 2017.

12] Li, Gang, and Fei Liu. "A clustering-based approach on sentiment analysis." *2010 IEEE international conference on intelligent systems and knowledge engineering*. IEEE, 2010.

13] Alessia, D., et al. "Approaches, tools and applications for sentiment analysis implementation." *International Journal of Computer Applications* 125.3 (2015).

14] Yu, Yang, Wenjing Duan, and Qing Cao. "The impact of social and conventional media on firm equity value: A sentiment analysis approach." *Decision support systems* 55.4 (2013): 919-926.

15] Agarwal, Basant, et al. "Machine learning approach for sentiment analysis." *Prominent feature extraction for sentiment analysis* (2016): 21-45.

16] Feldman, Ronen. "Techniques and applications for sentiment analysis." *Communications of the ACM* 56.4 (2013): 82-89.

17] Sharef, Nurfadhlina Mohd, Harnani Mat Zin, and Samaneh Nadali. "Overview and Future Opportunities of Sentiment Analysis Approaches for Big Data." *J. Comput. Sci.* 12.3 (2016): 153-168.

18] Tang, Duyu, Bing Qin, and Ting Liu. "Deep learning for sentiment analysis: successful approaches and future challenges." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 5.6 (2015): 292-303.

19] Tang, Duyu, Bing Qin, and Ting Liu. "Deep learning for sentiment analysis: successful approaches and future challenges." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 5.6 (2015): 292-303.

20] Shivaprasad, T. K., and Jyothi Shetty. "Sentiment analysis of product reviews: A review." *2017 International conference on inventive communication and computational technologies (ICICCT)*. IEEE, 2017.

21] Wankhade, Mayur, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. "A survey on sentiment analysis methods, applications, and challenges." *Artificial Intelligence Review* 55.7 (2022): 5731-5780.

22] Liang, Ting-Peng, et al. "What in consumer reviews affects the sales of mobile apps: A multifacet sentiment analysis approach." *International Journal of Electronic Commerce* 20.2 (2015): 236-260.

23] Liang, Ting-Peng, et al. "What in consumer reviews affects the sales of mobile apps: A multifacet sentiment analysis approach." *International Journal of Electronic Commerce* 20.2 (2015): 236-260.

24] Birjali, Marouane, Mohammed Kasri, and Abderrahim Beni-Hssane. "A comprehensive survey on sentiment analysis: Approaches, challenges and trends." *Knowledge-Based Systems* 226 (2021): 107134.

25] Liang, Ting-Peng, et al. "What in consumer reviews affects the sales of mobile apps: A multifacet sentiment analysis approach." *International Journal of Electronic Commerce* 20.2 (2015): 236-260.