



Analyzing customer sentiments on E-Tail applications using NLP

¹T. Justin Roy, ²P. Sai Teja, ³V. Uday Kiran Reddy, ⁴B. Harshith, ⁵Dr. Y. Ambica

^{1,2,3,4} Student, Department of CSE (Artificial Intelligence & Machine Learning), CMR College Of Engineering & Technology, Hyderabad, Telangana, India

⁵ Associate Professor, Department of CSE (Artificial Intelligence & Machine Learning), CMR College Of Engineering & Technology, Hyderabad, Telangana, India

Abstract : With increasing interest in e-commerce and online shopping recently, purchasing items online has grown to be more and more fashionable outstanding in options like lower in prices, and best quality products with high positive reviews, therefore customers are seeking to shop online. User reviews are more useful in deciding product quality. Customer reviews are one of the most important elements that determine customer satisfaction with the products. Moreover, it gives a better picture of products to business proprietors. Hence, this paper aims to conduct a sentimental analysis approach on a group of customer reviews collected from Flipkart. As well, classify each review into one of these classes: positive review or negative review by using Natural Language Processing (NLP). Online reviews have enough potential to provide a conclusion to the buyers about the product and its quality, performance, and recommendations, in this manner providing a detailed picture of the product to the end buyers. These online reviews are not only useful for customers but also used for manufacturers to realize customer requirements. Both positive reviews and negative reviews of customers play a major role in determining the requirements of customers and extracting customer feedback about the product. Sentiment analysis is an approach that helps to extract useful information, like opinions, attitudes, emotions, etc, from text data. It consists of different approaches, including Extraction, Tokenization, Lemmatization, and Classification. Extracting reviews from the website using the Web scraping package Beautiful Soup, we can easily fetch the brand name, reviews, ratings, and other related things for a product, using NLTK library that helps in classifying the reviews as positive review and negative review, and finally, they will come to know which product having more number of positive reviews and best reviews.

IndexTerms- Customer reviews, NLTK, sentimental analysis, web scraping, beautiful soup, Flipkart.

I. INTRODUCTION

Shopping can take many different shapes and forms such as purchasing items and services from online sellers via various browsers and applications. Shoppers can select their desired product in online shops by looking around a wide variety of sites of different sellers who are selling same and getting the product, its accessibility and its cost. In the price comparison between one seller and another, a differentiation could be seen in both the determination of item and the accessibility of the sellers. Both clients and sellers focus on the prior audits of other customers and also pay more attention to a handful of and speak with various manufacturers. One of the most crucial means making customer able to get hold of the quality of the product that he has purchased is through reading or seeing customer reviews. Many a times, the platforms and venues like Amazon, flipkart encourage consumers to rate or make comments about the products they are buying. It should be noted that the customers' reviews serve not just due to the benefit of clients but for marketers as well. Combining and demonstrating shop owners with customer surveys helped them to understand top-selling product, the out-rated item, the most-rated product, the one who received most positive comments and the one which had most negative feedback from customers. In reality, retail sites like Amazon, flipkart provide numerous options for the analysts about writing their surveys. For the sake of occurrence, the client may provide rating in numeric form (it is typically ranging from one to five stars). Such form may serve as an alternative to open-ended comments from the client upon the item. The online surveys situated on the website are believed to increase the customer loyalty, enhance the rate of visitor visits, increase the site hits and time spent on the site. Responsive surveys of clients help to build the brand among the beginners. By doing so, the business can be given new life and grow to bring more customers onboard. Positive and negative audits can be considered of assistance by consumers or the manufacturers. Producers can receive the constructive feedback and know the limitations from the area they can move on with to improve their product or company's benefit.

II. RELATED WORK

1. Lexicon-Based Sentiment Analysis:

Sujata L. Sonawane; Pallavi Kulkarni.[4] One of the basic techniques used in Natural Language Processing (NLP) is Lexicon-based Sentiment Analysis. It uses pre-existing dictionaries or lexicons to determine sentiments expressed in a given text. Lexicon-based sentiment analysis based on a dictionary simply implies that words are scored according to sentiment

from pre-defined sentiment lexicons and such scoring can be employed for concluding the overall mood of a particular passage. This method calculates the score of each word and adds them up for whole text sentiment calculation.

Following tokenization, individual token's (words) sentiments are calculated by summing up scores of words that have been mapped from the sentiment lexicon. This approach is good where moods related to just single words contribute towards contextualizing moods in larger texts like customer feedback, social media posts, or reviews made online. Again, there might be an issue with more sophisticated things like spelling mistakes, grammar errors, and so on. Moreover, they may not accurately determine feelings when sentences become complex due to complicated diction.

2. Rule-Based Sentiment Analysis:

Isanka Rajapaksha; Chanika Ruchini Mudalige; Dilini Karunarathna; Nisansa de Silva; Gathika Rathnayaka; Amal Shehan Perera.[5] Rule-based sentimental analysis is a fundamental method in natural language processing (NLP), in which there are predetermined sets of rules those rules serve as a guide for sentiment classification. This method works especially well when it is possible to specify clear rules for sentiment interpretation. This allows the rule-based sentimental analysis to be applied in a variety of situations, including social media material, customer evaluations, and financial news analysis. This method has the advantage of being flexible enough to meet the needs of different domains, allowing for the creation of explicit rules that may be used to capture complex sentiment expressions. To perform well Rule-based systems when there is a precise set of rules that are defined for sentiment classification; this makes them ideal for jobs like sentiment tracking on social media or customer feedback research. Rule-based systems may find it difficult to accommodate linguistic expression variations or sentiments that vary depending on the situation, sometimes defining the set of rules is complicated, Which limits their ability to manage more complex linguistic nuances.

3. Manual Annotation and Labeling:

Manual annotation and labeling in the sentimental analysis involves human annotators, as the title itself says that it is mostly done manually. Human annotators carefully read and interpret the textual data to classify them as positive, negative, or neutral. This is a hands-on practical method. A labeled dataset is created by annotators, this dataset consists of marking each text with its corresponding sentiment category. So, that dataset works as a training dataset for sentimental analysis models that enable the models to learn from human annotators. This dataset also works as an evaluation set. This manual annotation and labeling is crucial for tasks where context and intricate linguistic nuances play a significant role such as research studies, qualitative analysis, etc. Because they offer a human-centric method for classifying attitudes in textual data, manual annotation, and labeling are essential to sentiment analysis. For sentiment analysis models to be trained and be able to generalize and comprehend the nuances of sentiment expression in many situations, high-quality labeled datasets must be manually annotated.

III. METHODOLOGY

Our Methodology consists of 5 stages. Each stage describes each task to accomplish the objective.

1. Data Collection:

Extracting the data is the primary task in NLP to gather the data we use Web scraping techniques like Beautiful Soup library to get customer reviews from the E-Kart application. In addition to reviews, we can extract product names, review text, ratings, and other metadata.

2. Preprocessing of Data:

Any model can work efficiently on well-defined data, so Data preprocessing was used to preprocess the data like removing special characters, HTML elements, punctuations, and removal of stopwords. Later tokenize the text so that the whole text is converted into words or tokens. To ensure consistency in word representation, we use Lemmatization and stemming so that words are reduced to their base form.

3. Sentiment Analysis with NLTK:

Sentimental Analysis with NLTK is the main stage in our approach this is the heart of the model. In this stage, we apply sentimental analysis on the preprocessed data which we obtained in the previous stage. We use methods to represent the textual data in a mathematical manner such as bag-of-words also known as (BoW). To train a sentiment classifier we can also use built-in NLTK classifiers.

4. Model Evaluation and Validation:

Test how well the model differentiates between positive reviews and negative reviews and classifies the attitudes of different customers by analyzing their reviews of the product. If necessary, adjust the model's parameters or investigate different algorithms to maximize performance.

5. Visualization and Interpretation:

The final part of our paradigm is the output, which is the visualization and interpretation stage. In our case, we visualize the outcome of the analysis using the word cloud to mean particular trends or changes in consumer experience preferences and behaviours over time or with varying products. Synthesizing these results, you can get more information about users' attitudes, preferences, and experiences with e-commerce applications. Also, you can synthesize valid findings and recommendations to e-commerce players drawn from the sentiment analysis.

IV. FLOWCHART

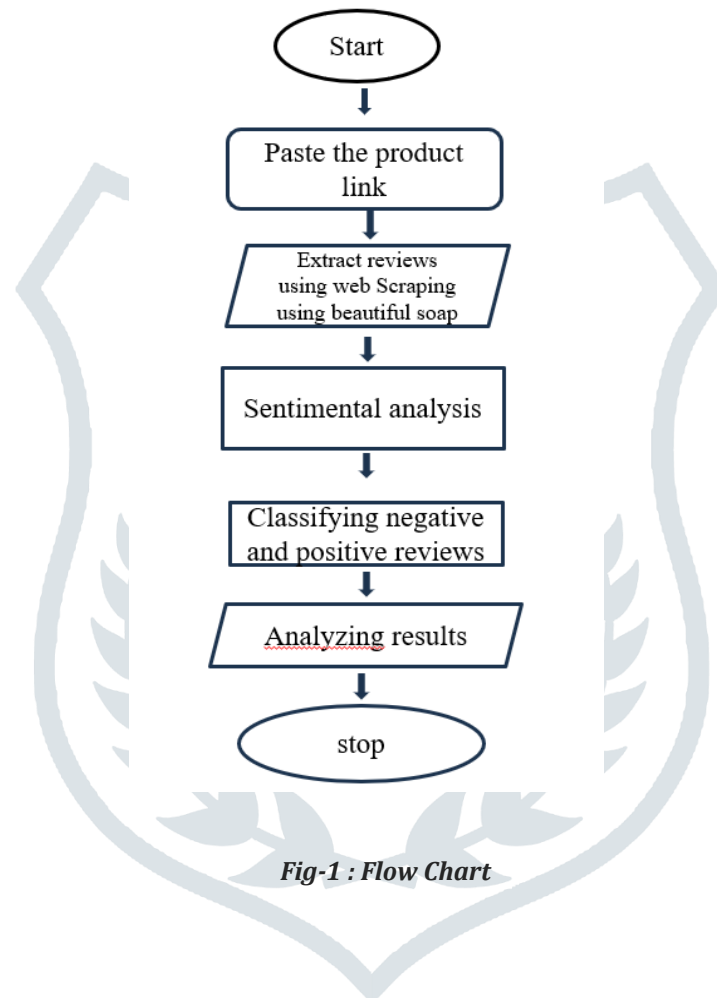


Fig-1 : Flow Chart

V. RESULT AND DISCUSSION

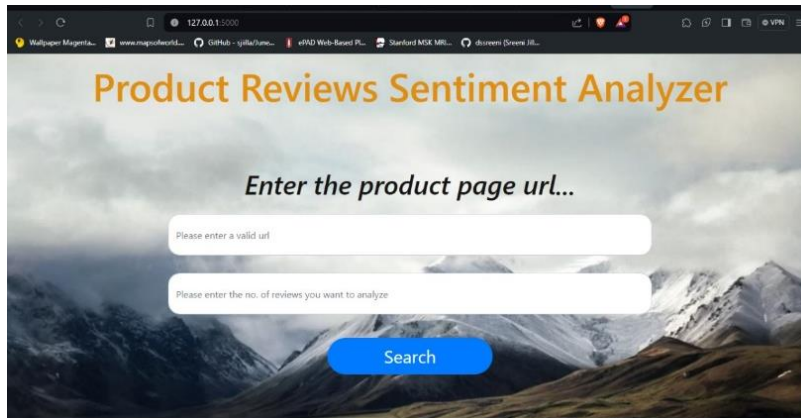


Fig-2: Homepage

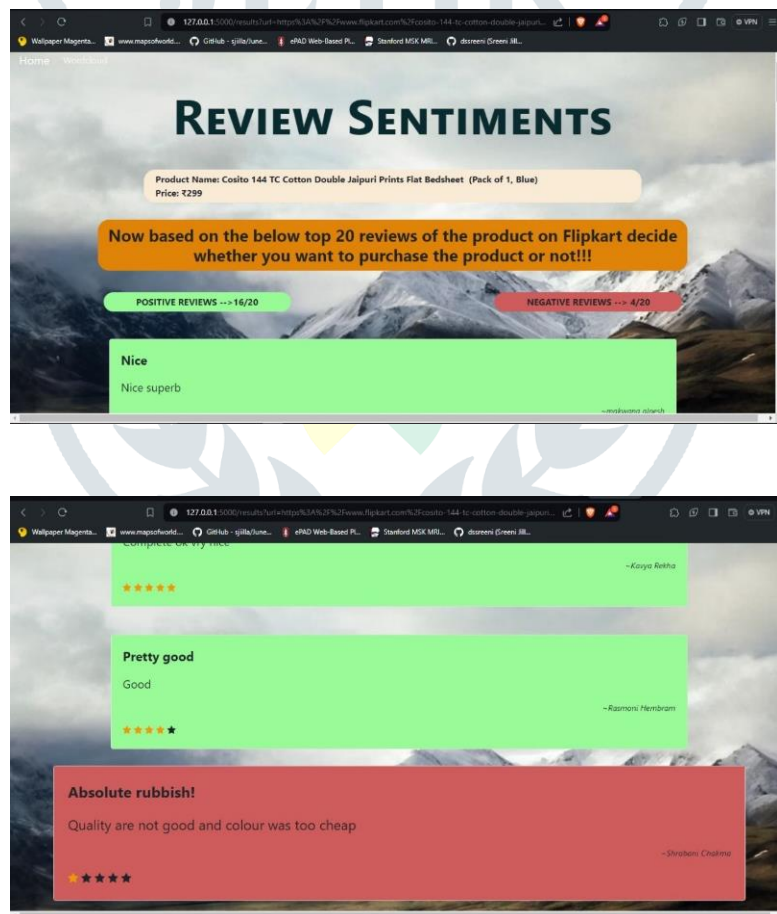


Fig-3,4: Displaying the reviews given by customers on a product (negative reviews are displayed in red an positive reviews in green)

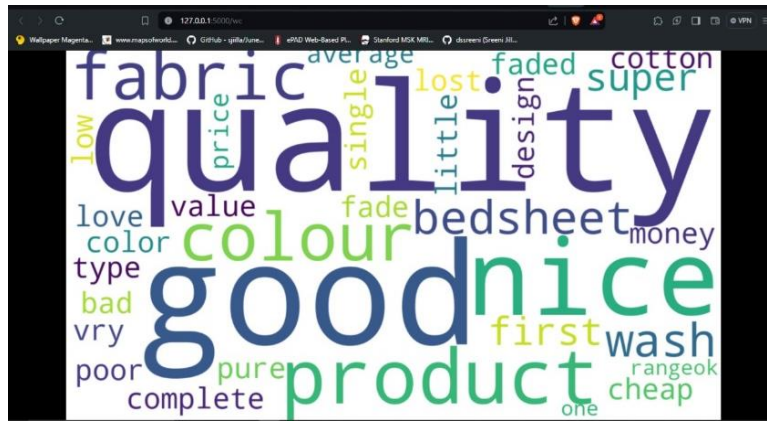


Fig-5: wordcloud

VI. CONCLUSION

With the improvement of online and e-commerce, more and more e-commerce platforms have the requests of the target clients' emotions. However, there are a few related sorts of inquiries about within the e-commerce reviews space and traditional sentiment analysis strategies can not prepare enormous sums of data at a high speed. In this paper, the sentimental Analysis of product reviews of the Flipkart site is accomplished with NLP and web scraping. The product points of interest are extracted by utilizing Web scraping. This work classifies positive and negative words from surveys and it calculates the rate of positive and negative words. In this manner, the result examination of the review rate makes a difference in the client's conclusion based on the positive survey rate of the product. Future work can be concentrated on mining surveys from different sites and numerous products etc. The same work can be boosted to consolidate many more classification algorithms which can offer assistance to us to choose or to select the most excellent classifier for opinion mining and sentiment analysis.

VII. REFERENCES

- [1] R. Y. Kim, "Using Online Reviews for Customer Sentiment Analysis," in IEEE Engineering Management Review, vol. 49, no. 4, pp. 162-168, 1 Fourthquarter, Dec. 2021, doi: 10.1109/EMR.2021.3103835. <https://ieeexplore.ieee.org/document/9527429>
- [2] M. R. Danendra and Y. Sibaroni, "Sentiment Analysis on Beauty Product Reviews using LSTM Method," 2021 9th International Conference on Information and Communication Technology (ICoICT), Yogyakarta, Indonesia, 2021, pp. 616-620, doi: 10.1109/ICoICT52021.2021.9527429. <https://ieeexplore.ieee.org/document/8716454>
- [3] Alrehili and K. Albalawi, "Sentiment Analysis of Customer Reviews Using Ensemble Method," 2019 International Conference on Computer and Information Sciences (ICIS), Sakaka, Saudi Arabia, 2019, pp. 1-6, doi: 10.1109/ICISci.2019.8716454. <https://ieeexplore.ieee.org/document/9442190>
- [4] Lexicon-Based Sentiment Analysis: Extracting sentiments from reviews: A lexicon-based approach, Sujata L. Sonawane; Pallavi Kulkarni, 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM) <https://ieeexplore.ieee.org/document/8122144>
- [5] Rule-Based Sentiment Analysis: Rule Based Approach of sentimental analysis for Party in Legal Opinion Texts Isanka Rajapaksha; Chanika Ruchini Mudalige; Dilini Karunarathna; Nisansa de Silva; Gathika Rathnayaka; Amal Shehan Perera 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer) <https://ieeexplore.ieee.org/document/9325435>
- [6] Current Trends on Sentimental Analysis on Youtube Videos, Sangeetha J; V. Maria Anu, 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI) <https://ieeexplore.ieee.org/document/9777200>
- [7] P. V. Rajeev, V. S. Rekha, "Recommending Products to Customers using Opinion Mining of Online Product Reviews and Features," in Proc. of 2015 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, India, 2015.
- [8] Ahlam Alrehili Computer Science Department, Kholood Albalawi Computer Science Department Taibah University, Sentiment Analysis of Customer Reviews Using Ensemble Method ISBN Information: INSPEC Accession Number: 18674179 DOI: 10.1109/ICISci.2019.8716454 <https://ieeexplore.ieee.org/document/8716454>
- [9] Sentiment Analysis of Chinese E-commerce Reviews Based on BERT Song Xie; Jingjing Cao; Zhou Wu; Kai Liu; Xiaohui Tao; Haoran Xie 2020 IEEE 18th International Conference on Industrial Informatics (INDIN) <https://ieeexplore.ieee.org/document/9442190>

- [10] Z. Singla, S. Randhawa, and S Jain, "Sentiment analysis of customer product reviews using machine learning," in Proc. of 2017 international conference on intelligent computing and control (I2C2), Coimbatore, India, pp. 1-5, 2017.
- [11] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews," in Proc. of 2020 International Conference on Contemporary Computing and Applications (IC3A), Lucknow, India, pp. 217-220, 2020.
- [12] T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," in Proc. of 2018 IEEE international conference on innovative research and development (ICIRD), Bangkok, Thailand, pp. 1-6, 2018.
- [13] G. Kaur, and A. Singla, "Sentimental analysis of Flipkart reviews using Naïve Bayes and decision tree algorithm," International Journal of Advanced Research in Computer Engineering & Technology, vol. 5, no. 1, 2016.
- [14] P. V. Rajeev, V. S. Rekha, "Recommending Products to Customers using Opinion Mining of Online Product Reviews and Features," in Proc. of 2015 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, India, 2015.
- [15] S. Xie, J. Cao, Z. Wu, K. Liu, X. Tao and H. Xie, "Sentiment Analysis of Chinese E-commerce Reviews Based on BERT," 2020 IEEE 18th International Conference on Industrial Informatics (INDIN), Warwick, United Kingdom, 2020, pp. 713-718, doi: 10.1109/INDIN45582.2020.9442190.
- [16] R. Abinaya., P. Aishwaryaa, S. Baavana, and N. T. Selvi, "Automatic sentiment analysis of user reviews," in Proc. of 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), Chennai, India, pp. 158-162, 16 July 2016.
- [17] C. Rain, "Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning," M.S thesis, Department of Computer Science, Swarthmore College, Swarthmore, PA, USA, 2013.
- [18] S. S. Sikarwar, Dr. N. Tiwari, "Analysis The Sentiments Of Amazon Reviews Dataset By Using Linear SVC And Voting Classifier," International journal of science and technology research, vol. 9, no. 6, pp. 461-465, 2020.

