

Review analyser using NLP

Justine Jiby Varghese

*Dept.of Computer Science and Engineering
St.Joseph's College of Engineering and Technology
Palai,Kottayam,Kerala
justinejibyvarghese2023@cs.sjcetpalai.ac.in*

Martin T.V

*Dept.of Computer Science and Engineering
St.Joseph's College of Engineering and Technology
Palai,Kottayam,Kerala
martinvarghesetv@gmail.com*

Prof. Mereen Thomas

*Assistant Professor, Department of CSE
St.Joseph's College of Engineering and Technology
Palai,Kottayam,Kerala
Mereen.Thomas@sjcetpalai.ac.in*

Melvin J Thomas

*Dept.of Computer Science and Engineering
St.Joseph's College of Engineering and Technology
Palai,Kottayam,Kerala
melvinjthomas2023@cs.sjcetpalai.ac.in*

Raghul Surendran

*Dept.of Computer Science and Engineering
St.Joseph's College of Engineering and Technology
Palai,Kottayam,Kerala
raghulsurendran2023@cs.sjcetpalai.ac.in*

Abstract—Online product reviews are a great source of information for consumers. From the sellers' point of view, online reviews can be used to gauge the consumers' feedback on the products or services they are selling. However, since these online reviews are quite often overwhelming in terms of numbers and information, an intelligent system, capable of finding the general sentiment from these reviews, will be of great help for both the consumers and the sellers. This system will serve two purposes: One is to Enable consumers to quickly understand the opinions of the reviews without having to go through all of them and the other is to help the sellers/retailers get the general consumer feedback, which will lead to better decisions and satisfied minds. The NLP's ultimate goal is to learn, interpret, comprehend and understand the human languages in a meaningful manner.

I. INTRODUCTION

Online product reviews give companies the ability to conduct an in-depth and detailed consumer environment study economically and expediently. In recent years, researchers and practitioners have been gradually drawing attention to these market data (Laroche et al., 2005; Gopal et al., 2006). Though online voices of the customer (VOC) are free text, they have proven to represent the fundamental market characteristics (Campbell et al. 2011; Godes and Mayzlin 2004; Duan et al. 2008; Shao 2012) and to use improvements in conventional marketing operations (Onishi and Manchanda 2012). Nevertheless, an barrier exists that prevents online product reviews from fulfilling their best potential. The online content generated by users is enormous and qualitative, making it difficult to quantify the data and achieve substantive information. (Godes et al. 2005). Because of the lack of effective methods for extracting key features from these texts, companies did not gain valuable information for constructing a map of the market structure. [1] However, recognizing these

points of view and market structure is essential for product growth, pricing, promotion / campaign and brand placement. Thus, businesses use quality scores as substitutes for product reviews.

For example, in Chintagunta et al. (2010), the relationship between customer feedback and movie industry sales is examined by product ratings, while Chevalier and Mayzlin (2006) use them for book industry research. Market structure reflects a partnership between brands based on different methods including set-up (Urban et al., 1984), brand-switching data (Cooper and Inoue, 1996) and brand associative networks (John et al, 2006). Researchers have begun to obtain standardised and quantitative market knowledge from online product reviews through the development of texts mining techniques based on natural language processing (NLP) (Feldman et al., 2007). Lee and Bradlow (2011), for example, have developed a text mining algorithm for online product reviews. [2] And Netzer et al. (2012) proposed a hybrid market structure surveillance text mining and semantic network analysis platform. However, these approaches require human intervention to discern related commodity properties, on the basis of the "bag of words" premise, and can not carry out the task of generating market structures entirely automatically.

For example, Lee and Bradlow (2011) use manual reading to classify 39 different product clusters from 99 K-means clusters. Their studies inspire us to use text mining techniques for the online product reviews. There is also a large amount of content online, such as web sites, newsgroup articles and online news sources. Automatic sentimental text analysis has been researched extensively in the form of sentiment classifiers, effect analysis, automatic testing, opinion extraction or recommendation systems. [4] Such techniques generally aim

to retrieve the general feeling exposed in a positive, negative or anywhere between the documents. Two complicated aspects of sentiment analysis are: first, while the general views on a subject are valuable, they are just a part of the interesting facts. The classification of the sentiment of the document level can not detect feelings about every aspect of the issue. In fact, for example, if one can be pleased with his car generally, he might be unhappy with the engine noise.

These individual weaknesses and strengths are equally important for manufacturers, or even more valuable than customers' overall customer satisfaction. [5] No e-commerce website is complete these days without good review. Though company reviews are fine, individual product reviews can have a huge effect on the customer travel and conversion rates. While product requirements and price are significant (no buyer would be willing to purchase a TV that is too big for his / her home), it is the influence of other people's opinions that allows the decision to buy. Reviews will include things like size, reliability, stability, suitability and more. Not just your customers benefit from all the lovely content produced by the consumer. [6] A. Feedback of Products Product feedback will provide a wonderful insight into the existing offerings, so you can continually adjust and develop the services. Usually, two phases of aspect- or finely known sentiment analysis address the challenge. The first phase aims to identify object characteristics, and the second phase classifies and summarizes each feature. In this paper we are concentrating on developing an initial process model: consumer satisfaction factor identification. Current aspects can be divided widely into two main approaches: supervised and unattended.

Controlled approaches to aspect detection involve a variety of pre-labeling training data and while supervised approaches can be successful, adequate labeled data are often costly and require a lot of human work. Since labeled data are not typically available, it is beneficial to create a model that works with unlabeled data. Thematic modeling is a method that allows automatic recognition of subjects in a text object and the derivation of hidden patterns shown in a text corpus. Review mining, or the examination of customer reviews and comments, is a notion that has been around for a while. There is now more client feedback accessible for firms to assess because to the growth of e-commerce and internet reviews. As a result, several tools and methods for sentiment analysis and review mining have been created. Rule-based methodologies were employed in the early stages of review mining and sentiment analysis to categorize reviews as good, negative, or neutral in terms of their sentiment.

These techniques, which depended on specific words or phrases appearing in the text, frequently fell short of adequately capturing the tone of a review. More complex methods for review mining and sentiment analysis have been created as a result of the development of natural language processing (NLP) and machine learning. These techniques are better capable of correctly categorizing the sentiment of a review since they employ algorithms to learn from a collection of labelled training data. Businesses across a wide range of sectors

now employ review mining and sentiment analysis to gather information and raise customer satisfaction. Researchers and analysts utilize it as well to examine consumer behaviours and spot patterns and trends in client feedback.

II. OBJECTIVE AND SCOPE

The following are some of the goals of our proposed paradigm.: To create a portable application for review mining and sentiment analysis of the products on online shopping websites, and To provide a cheap and user-friendly software. This application combines contemporary technologies such as NLP and spaCy. The system's scope is global because it can function with a smartphone and a stable internet connection.

III. OTHER METHOD'S

There are many other NLP methods as well which are used for analyzing and understanding online reviews. Some of them are listed below:

Text Summarization: Summarize the reviews into a paragraph or a few bullet points.

Entity Recognition: Extract entities from the reviews and identify which products are most popular (or unpopular) among the consumers.

Identify Emerging Trends: Based on the timestamp of the reviews, new and emerging topics or entities can be identified. It would enable us to figure out which products are becoming popular and which are losing their grip on the market.

Sentiment Analysis: For retailers, understanding the sentiment of the reviews can be helpful in improving their products and services.

IV. LITERATURE SURVEY

Yingwei et al. [1] presented a technique to use geotagged media platforms data to analyse public emotion and opinions, to evaluate how well tourist sites have recovered after an earthquake. They used the platform Twitter to find the related posts. In this the posts shared by the news channels are excluded using a simple keyword approach, machine produced posts are not identified for mining the massive number of Tweets regarding

Lombok and Bali in this work, a mixed techniques approach including sentiment analysis and LDA topic modelling is created. The variety and variance of people's views and viewpoints following the earthquakes have been revealed, in this they mainly talk about the places Lombok and Bali in which after so many earthquakes that happened in the two locations in August 2018. Natural and man-made disasters frequently cause damage to and even complete destruction of tourist infrastructures, attractions, and visitors' positive perceptions of destinations, raising concerns about stability and safety and reducing travelers' willingness to visit the areas impacted.

Jonah Zeenia et al. [2] presented a technology to Analyzing consumer product reviews statistically and emotionally. An extensive collection of online evaluations for mobile phones is analyzed for data in this study. Over 400,000 evaluations for

over 4500 mobile phones make up the data collection. The data collection is statistically analyzed to look at the relationships between the various attributes. The usefulness of reviews may be estimated to provide the designer the most relevant information, allowing him to improve the product or launch a new product by fully satisfying client demands. The industry will greatly benefit from the research's extensibility since it aims to reduce the time required to collect requirements and the costs associated with employing surveys, questionnaires, interviews, market research, and trends. High priced items received greater ratings than low-priced ones, resulting in increased levels of consumer satisfaction and better product quality.

Narayan and Twinkle [3] done research on product sentiment analysis using the random forests algorithm based on user reviews. The analysis of customer evaluations from the e-commerce industry is the main emphasis of this article. This project only extracts data from the site Flipkart. The reviews are genuine since they are taken directly from this site using machine learning approach and SO approach. Which uses a Random Forest Classifier for classification consisting of many decisions tree. This paper includes a standard sentiment analysis methodology that consists of three fundamental processes, including data preparation, review analysis, and sentiment categorization, as well as other supporting procedures. This method addresses the problem of difference between ratings and reviews. It overcomes the issues stated previously by providing Boolean reviews based on reviews rather than ratings. Since the website flipkart only used the review collection is only limited to single site .

AlZu'bi et al. [4] suggested a system for analyzing Amazon reviews in order to categorize reviews of products that are useful to users. The system's dataset is obtained from the Stanford Network Analysis Platform. The reviews are in JSON format, and the resulting data is standardized before further processing. The system selects reviews and votes from the received checks, then categorizes the reviews as helpful or not based on the upvote percentage. Deep learning technologies such as RNN were utilized by the system.

Abhinaya et al. [5] suggested a system for analyzing user evaluations in a web application. Users can log in to the system and post videos; other users can see the movies and leave feedback. Reviews that are uploaded will be saved in the database. After pre-processing the stored reviews, the system moves on to the sentiment analysis phase. The sentiment graph is obtained by the system and displayed to the user after processing.

Cheng et al. [6] devised a technique that mines user reviews to assess user reviews in a digital banking application. They claimed that the purpose of the paper is to collect vital knowledge that was previously unknown in online reviews. The system will apply LDA to the pre-processed text to extract only the relevant information based on their count frequency and other characteristics. The gathered results are then classified into two categories: "positive" and "negative." The system's major goal is to answer what the users are most worried

about and to identify the qualities that are evaluated positively and negatively. The dataset for the system is digital banking software from the Philippines, and the reviews were gathered using Python from the Google Play store. To perform the necessary processing, the Natural Language Toolkit Platform was employed along with the Apriori Algorithm to do the review analysis.

Guerreiro and Rita. [7] suggested a technique that scans tourist evaluations on social media and websites to discover explicit thoughts and attitudes. The algorithm made use of the Academic Yelp Dataset, which included reviews from individuals from a variety of industries. A text-mining approach based on Natural Language Processing was used to examine the reviews (NLP). For the sentiment analysis, IBM SPSS Modeler Text Analytics was used. For the analysis, multiple algorithms were explored, and the CHAID model was chosen and used due to its interpretability. The system had a test dataset accuracy of 66.05 percent and a training dataset accuracy of 61.28 percent

Kim and Chun. [8] suggested a system that analyses user reviews from a car review website and determines each automobile's qualities. The R programming language is used to create the system. The approaches utilized to get the desired output include text mining and association rules. Initially, the reviews are gathered via web scraping from the website "Edmunds.com" using the program "Parse Hub," and then the data is pre-processed into a CSV file containing all of the significant keywords. The frequency of each word is calculated using an R program utility called "Term Cloud," which displays the frequency of each word's occurrence in a word cloud format. The association rule is then used to establish the characteristics of the autos.

Choudary et al. [9] I proposed a technique for collecting smartphone reviews from websites and analyzing their mood. The system's dataset is obtained from the website "Twitter." The reviews collected from Twitter are saved locally in a CSV format and used for additional data processing. The frequency of the words is then analyzed using a word cloud. The sentiment analysis phase is conducted using the "syuzhet" R programming tool. Using the NRC sentiment dictionary in the syuzhet package, the estimated sentiment of the text is classified into several sentiments.

Kshirsagar and Deshkar [10] suggested a technique for evaluating the polarity of reviews from multiple online shopping sites using WEKA classifiers. A web crawler is used to identify online shopping websites using the required product URL as its search phrase. Once the crawler finds the review, it mines all of the user reviews from the website and returns them to the system. The mining reviews are sorted, and sentiment analysis is performed. The WEKA classifier, which employs the Naive Bayes classifier technique, is then used to classify the reviews. Once the classification is complete, the review's features are extracted and shown.

Sheikh [11] proposed a system to Legitimate and Spurious Reviews using Opinion Mining. Additionally, service providers, distributors, or manufacturers may utilise these

reviews to learn what the general public thinks as restrictions on their goods or services. But the underlying reality of internet reviews reveals something else. These reviews may be fictitious, posted or written with ulterior motives; they frequently include a positive or supportive opinion in an effort to enhance, promote, and publicise their goods and services, or they may have a pessimistic intention in order to harm the reputation and business of their competitors. Therefore, it is important to explore and investigate these reviews before using opinion mining. In this paper, a methodology is proposed that entails the acquisition of reviews about mobile devices of various makes using the Tag path clustering approach, along with metadata from Flipkart, and the detection of fake reviews based on identical and nearly identical reviews using semantic similarity with review length.

Abinaya [12] proposed a system of Automatic Sentiment Analysis of User Reviews. The process of extracting information from raw data is known as data mining. Data mining is primarily used to collect the needed data, to draw out information that is usable from the data, and to analyse the data. The current method uses Dual Sentiment Analysis and the Bag of Words model to categorise evaluations as positive, negative, or neutral. Due to certain basic shortcomings in how it addresses the polarity shift problem, Bag of Words' performance can occasionally still be considered restricted. The suggested approach properly classifies reviews as favourable, negative, or neutral using a dictionary-based categorization. The Support Vector Machine technique is used to improve the categorization of neutral reviews. The sentiment graph, which is created based on the reviews for each of the product videos, may be used by both the product owner and the consumer to determine the quality of the product. In order to boost the effectiveness of visual representation, a comparison study of sentiment graphs is conducted.

Rangkuti [13] proposed a system for Sentiment Analysis on Movie Reviews Using Ensemble Features and Pearson Correlation Based Feature Selection. Microblogging has emerged as a very popular form of media among internet users. As a result, microblogging emerged as a rich source of information for thoughts and evaluations, particularly of films. The sentiment analysis of movie reviews employing ensemble features, Bag of Words, and selection is what we suggested. Features Using Pearson's correlation, one may obtain the best feature combinations by reducing the dimension of the feature. By lowering the dimension of the feature and obtaining the best feature combinations, feature selection is used to enhance classification performance. The process of classifying data using several Naive Bayes models, including the Bernoulli Naive Bayes model for binary data, the Gaussian Naive Bayes model for continuous data, and the Multinomial Naive Bayes model for numeric data. According to the study's findings, accuracy, precision, memory, and f-measure were all improved when nonstandard words were included in tweet evaluations. These improvements were accuracy (82percent), recall (86percent), recall (82.69percent), and f-measure (82.69percent). The evaluation's accuracy rose by 8percent following manual

word normalisation, rising to 90percent accuracy, 92percent precision, 88.46percent recall, and 90.19percent f-measure utilising 85percent feature selection. Based on these findings, it is possible to draw the conclusion that word standardisation can enhance classification and feature selection performance while decreasing the overall amount of dimension features.

Rahamathulla [14] proposed a Feature Based Approach for Sentiment Analysis using SVM and Coreference Resolution. In the modern era of technology, online shopping is one of the most convenient methods of purchasing. People routinely purchase things online and give reviews of the goods they have utilised. The user's perspective will be expressed through the tweets or product evaluations they submit on an ecommerce website. These reviews will be crucial in determining how well-regarded the items are among consumers. These reviews will also assist the makers in enhancing the features of the product as necessary, however it is highly challenging to manually read the evaluations and determine their mood. This issue may be resolved by developing an automated system that analyses user evaluations and extracts the users' perceptions of a certain feature. In this study, we used a classifier called the Support Vector Machine to construct a process for "feature-based sentiment analysis".

Adak [15] proposed a system of Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review. Customers' demand for having food delivered to their doorsteps during the COVID-19 crisis has fueled the expansion of meal delivery businesses (FDSs). Since all restaurants have gone online and joined FDSs like UberEATS, Menulog, or Deliveroo, customer ratings on internet review sites have become a valuable resource for learning about the success of the business. In order to improve customer satisfaction, FDS organisations seek to compile complaints from consumer feedback and use the data efficiently to identify areas that need improvement. In order to forecast client attitudes in the FDS sector, this work reviewed machine learning (ML), deep learning (DL), and explainable artificial intelligence (XAI) methodologies. A survey of the literature found that lexicon-based and ML approaches are often used to forecast consumer attitudes from FDS evaluations. Due to the lack of model interpretability and decision explanationability, there are few research using DL approaches. The following are the main conclusions of this systematic review: Organizations can make a case for the system's explainability and trustworthiness despite the fact that 77percent of the models are inherently unintelligible. Although DL models in other domains perform well in terms of accuracy, they don't have the same level of explainability, which XAI implementation may provide.

Jansher [16] For this project, data that includes product reviews for 8000 different products was scraped from the Amazon website. He uses Naive Bayes and Support Vector Machine classifiers, together with machine learning methods, to categorise the buyer reviews (SVM). As a sentiment analysis algorithm, Naive Bayes classifier (NBC) is employed; it bases its operations on the likelihood of condition. SVM is

used in conjunction with the Python module scikit-learn as a supervised learning technique for classification. The lovely Soup library in Python scrapes reviews from Amazon; for each review's uniqueness, an Amazon Standard identifying number (ASIN) is provided. Following data cleaning and input into classifier algorithms, the Naive classifier divides the reviews into positive and negative ones. Which was able to give an accuracy of 84.74percent. The SVM classifier did a little better than the NBM with an accuracy of 86.59percent. Hence the SVM classifier performs better, although both algorithms are performing well. Ms. Budhwar et al. [17] examined an Amazon product data set with approximately 35,000 product reviews. Using the classifying data, this was utilised to train the model using methods such as Naive Bayes, K-nearest Neighbor, Linear Support Vector Machine, and Long Short Term Memory. The model was a sentimentally hybrid classifier in which all of the aforementioned algorithms were used and the reviews were classified into three categories: positive, negative, and neutral reviews. The proposed model was created to provide precise solutions and address a problem with earlier sentimental analysis techniques.

Shrestha et al. [18] created a model with recurrent neural networks (RNN) and gated recurrent units (GRU) based on 3.5 million product reviews scraped from the amazon.com website. The product database is then cleaned, and the data is fed into RNN with GRU to capture information that exists between reviews that belong to a specific product sequence. The sentiment is then classified using the Support vector machine, which distinguishes between positive, negative, and neutral review labels. They also created a web service to avoid inconsistent review and rating pairs when using SVM. The model's accuracy is 81.82percent.

Mrs. Harika et al. [19] suggested a system to anticipate positive and negative attitudes using machine learning algorithms based on datasets gathered from amazon.com. The algorithms used for sentimental classification include K-Nearest Neighbor, Random Forest, Decision Tree, and Support Vector Machine. The data set was collected from Amazon using a web crawler they created that pulled product reviews from the site page by page. The raw data is then preprocessed and tokenized, then NLP is used to simplify the data set before sentiment and the emotion each token represents are examined. It uses the NRC Emotion Lexicon to associate words with positive and negative sentiments as well as the eight basic emotions.

Kshirsagar [20] has presented a system in which analysis is completed on over 1,000 Facebook postings on newscasts, and in which the attitude of Rai, the Italian public broadcasting service, toward the rapidly expanding and extremely dynamic private organisation la7 is examined. This paper can become aware of the system obtaining information of in-text statistics, to get the direct association between good and bad ratings, in Amazon products, using the statistical analysis of the NPS score. The sentimental algorithms Naive Bayes classifier (NBC), support vector machine (SVM), and contextual linguistics seek are the technology employed here

(CSS).

V. PROPOSED SYSTEM

The presented system consists of a Web application that receives the product URL from the user and collects the reviews from the web page using the web scraping approach. Python Requests and BeautifulSoup libraries are used to achieve this functionality. The request library gets the raw HTML data from the webpage and the BeautifulSoup library parses the HTML file and extracts the contents. The retrieved reviews are then evaluated using NLTK and Spacy to determine their sentiment. The assessed result is then delivered to the customer, along with the option to seek for the product on a separate online purchasing website. Fig 3.1 depicts the block diagram of the proposed system.

A web application for an NLP (Natural Language Processing) review analyzer is a software application that can be accessed through a web browser, allowing users to analyze customer reviews, feedback, and opinions about a product or service. The main purpose of this web app is to provide a convenient and user-friendly interface for users to access the NLP review analyzer and its various features, such as sentiment analysis, topic modeling, keyword extraction, and more. The web app typically includes a user interface that allows users to input the text data they want to analyze, such as customer reviews, product specifications, and so on. The app then processes the data using NLP techniques and algorithms, and presents the results in a visually appealing and easy-to-understand format, such as graphs, charts, and tables. When the app is launched, the user is presented with a start button. When the user clicks the start button, a search bar with a URL entry field appears. Once the user has input the URL and continues, the system will perform the necessary processing and provide the product information to the user, along with the opportunity to examine the reviews and the nature of the product.

A backend, often known as a "server-side" or "server-based" application, refers to the component of an application that operates on a server rather than on a user's device. The backend is responsible for storing and maintaining data, processing business logic, and performing other operations that are not immediately connected to the display of the program to the user. The backend of the application is implemented using the Python Flask framework. Python Flask is a simple micro-web framework used to develop web applications.

The frontend, also known as the "client-side" or "client-based" application, refers to the portion of an application that runs on a user's device and is responsible for presenting data to the user and handling user interactions. The front end of the application is done using HTML, CSS and JavaScript. The frontend of a web application for an NLP (Natural Language Processing) review analyzer refers to

the client-side components that provide the user interface and visual elements for the application. The frontend is responsible for presenting the data and results generated by the NLP review analyzer to the user, and allowing the user to interact with the application. The frontend typically includes the following components:

User Interface: To present the data and results generated by the NLP review analyzer to the user, such as graphs, charts, tables, and so on.

Input Forms: To allow the user to input the text data they want to analyze, such as customer reviews, product specifications, and so on.

Navigation: To provide a clear and intuitive navigation structure, allowing the user to easily access the various features and functions of the application.

Visualizations: To present the results of the NLP analysis in a visually appealing and easy-to-understand format, such as pie charts, bar graphs, and word clouds.

Web scraping can be used to gather data for NLP review analysis. The process involves sending HTTP requests to a website's server to retrieve its HTML code, then parsing that HTML to extract relevant data, such as product reviews. This data can then be cleaned and preprocessed for NLP techniques, such as sentiment analysis, to gain insights on customer opinions and feedback. However, it is important to respect websites' terms of use and to avoid scraping them too frequently, as this can put a strain on their servers and potentially result in legal consequences. Python is used for web scraping. It's done with the `python-requests` and `pythonBeautifulSoup` libraries. The `requests` library sends HTTP request to the webpage and fetches the webpage as raw HTML data. The raw data is then sent to `python BeautifulSoup` library. The `BeautifulSoup` library parses the HTML file and extracts the data from the HTML according to the needs. The data is obtained from the HTML file using the `findall` method in the `BeautifulSoup` library.

The obtained reviews are passed into the NLTK library. In this stage the obtained reviews are cleaned and converted into tokens which then passed into the sentiment analysis model. The NLTK's `WordNetLemmatizer` database is used for the lemmatization and stemming process. Certain words and numbers are removed from the word list. The `TF-IDF Vectorizer` algorithm is used here. The algorithm is used for transforming text into meaningful representation of numbers which is used to fit the model. Data cleansing for NLP review analysis involves the following steps:// Removing duplicates: Duplicate reviews can skew the analysis results.// Removing irrelevant data: Any data that does not pertain to the product being analyzed should be removed.// Removing HTML tags, URLs and special characters: HTML tags, URLs and special characters can affect the analysis results, so they should be removed.

Removing stop words: Common words such as "a", "and", "the", etc., can be removed as they do not provide any

meaningful information.

Removing punctuation: Punctuation can also affect the analysis results, so it should be removed.

Stemming/Lemmatization: This involves reducing words to their root form to avoid having multiple variations of the same word.

Converting to lowercase: Converting the data to lowercase can help in reducing the size of the data and improve analysis results.

Removing biased language: If the data contains any biased language, it can affect the analysis results, so it should be removed.

The sentiment analysis is carried out using the NLTK and Spacy. Reviews are taken from over ten pages of reviews on the website and kept in raw form in a text document. The NLTK's `SentimentIntensityAnalyzer()` method is used to initially classify the words into three lists of positive, negative and neutral reviews hence the sentiment analysis is done. Once the nature of the reviews is analyzed the corresponding list is used to display the word cloud of the occurrences of different words and the intensity of each word in multiple reviews in order to identify the keywords which made the reviews positive or negative. Sentiment analysis and word cloud are two techniques used in NLP review analysis to determine the overall sentiment and the most frequently mentioned words respectively. **Sentiment analysis:** This involves determining the polarity (positive, negative, neutral) of the reviews and calculating the overall sentiment of the data. This can be done using various algorithms such as Naive Bayes, SVM, and Deep Learning. **Word Cloud:** This is a visual representation of the most frequently mentioned words in the reviews. The size of the word in the cloud represents its frequency. Word clouds are useful for quickly identifying the most frequently discussed topics in the data. By combining these two techniques, a comprehensive analysis of the reviews can be performed to understand the customer sentiment and their most common concerns or opinions.

VI. TECHNOLOGIES USED

A. Python IDE

An IDE (or Integrated Development Environment) is a program dedicated to software development. As the name implies, IDEs integrate several tools specifically designed for software development. An IDE (or Integrated Development Environment) is a program dedicated to software development. As the name implies, IDEs integrate several tools specifically designed for software development.

B. VS Code

Visual Studio Code, also commonly referred to as VS Code, is a source-code editor made by Microsoft with the Electron Framework, for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git.

C. spaCy

spaCy is an open-source software library for advanced natural language processing, written in the programming languages Python and Cython. The library is published under the MIT license and its main developers are Matthew Honnibal and Ines Montani, the founders of the software company Explosion.

CONCLUSION

We have used Python Flask web framework for the development of our project. As of now the system can collect the URL from the user and display the product information and also collect the user reviews from the webpage and perform sentiment analysis on these reviews to get the nature of the reviews. After the analysis the system can display the nature of the reviews to the user. The system can classify the nature of reviews to three different classes they are positive, negative and neutral. The system can also display the keywords which have made the reviews positive or negative and can display them in the form of word clouds according to their occurrence intensity. On experimental testing the system can perform up to an accuracy of greater than 85 percent in real time. At the current stage the system collects information from ten review pages of the products which will be around 200 reviews of the product.

REFERENCES

- [1] L. C. Cheng and L. R. Sharmayne, "Analysing Digital Banking Reviews Using Text Mining," 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2020, pp.
- [2] Guerreiro, J., Rita, P. (2019). How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *Journal of Hospitality and Tourism Management*.
- [3] Kim, E.-G., Chun, S.-H. (2019). Analyzing Online Car Reviews Using Text Mining. *Sustainability*, 11(6), 1611. doi:10.3390/su11061611.
- [4] M. Choudhary and P. K. Choudhary, "Sentiment Analysis of Text Reviewing Algorithm using Data Mining," 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT), 2018, pp. 532-538, doi: 10.1109/ICSSIT.2018.8748599
- [5] A. A. Kshirsagar and P. A. Deshkar, "Review analyzer analysis of product reviews on WEKA classifiers," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015, pp.
- [6] Yan, Y., Chen, J. and Wang, Z. (2020) "Mining public sentiments and perspectives from geotagged social media data for appraising the post-earthquake recovery of tourism destinations," *Applied Geography*, 123, p. 102306.
- [7] Singla, Z., Randhawa, S. and Jain, S. (2017) "Statistical and sentiment analysis of Consumer Product Reviews," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) [Preprint].
- [8] Singla, Z., Randhawa, S. and Jain, S. (2017) "Statistical and sentiment analysis of Consumer Product Reviews," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) [Preprint].
- [9] AlZu'bi, S. et al. (2019) "A brief analysis of Amazon Online Reviews," 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS) [Preprint].
- [10] Sun, Y., Wang, L. and Deng, Z. (2009) "Automatic sentiment analysis for web user reviews," 2009 First International Conference on Information Science and Engineering [Preprint]. 2019.
- [11] Aijaz Ahmad Sheikh;Tasleem Arif;Majid Bashir Malik;Suhail Iqbal Bhat; (2020). Opinion Mining: Legitimate vs Spurious Reviews . 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN).
- [12] Abinaya, R; Aishwaryaa, P; Baavana, S; Selvi, N.D. Thamarai (2016). [IEEE 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) - Chennai, India (2016.7.15-2016.7.16)] 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) - Automatic sentiment analysis of user reviews. , (), 158–162.
- [13] Saputra Rangkuti, Fachrul Rozy; Fauzi, M. Ali; Sari, Yuita Arum; Sari, Eka Dewi Lukmana (2018). [IEEE 2018 International Conference on Sustainable Information Engineering and Technology (SIET) - Malang, Indonesia (2018.11.10-2018.11.12)] 2018 International Conference on Sustainable Information Engineering and Technology (SIET) - Sentiment Analysis on Movie Reviews Using Ensemble Features and Pearson Correlation Based Feature Selection. , (), 88–91
- [14] Krishna, M. Hari; Rahamathulla, K.; Akbar, Ali (2017). [IEEE 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT) - Coimbatore, India (2017.3.10-2017.3.11)] 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT) - A feature based approach for sentiment analysis using SVM and coreference resolution. , (), 397–399
- [15] Adak A, Pradhan B, Shukla N. Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review. *Foods*. 2022 May 21;11(10):1500. doi: 10.3390/foods11101500. PMID: 35627070; PMCID: PMC9140678..
- [16] Jansher, Rabnawaz. (2020). Sentimental Analysis of Amazon Product Reviews Using Machine Learning Approach. 10.13140/RG.2.2.36392.80645.
- [17] Jyoti Budhwar, Sukhdip Singh, 2021, Sentiment Analysis based Method for Amazon Product Reviews, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH TECHNOLOGY (IJERT) ICACT – 2021 (Volume 09 – Issue 08).
- [18] Nishit Shrestha and Fatma Nasoz.(2019).Deep Learning Sentiment Analysis of Amazon.Com Reviews and Ratings.International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.8, No.1, February 2019.
- [19] Mrs.Kopparthi Harika, K. Mani Veera Venkata Ratna Kumari, M.Sai Anusha, M.Anila, S.Tejaswini, B.Tirupati Swamy(2020). Sentiment Analysis on Amazon product reviews. ISSN : 0950-0707.
- [20] Kumar, K Kshirsagar, Pravin. (2020). Sentiment Analysis of Amazon Product Reviews using Machine Learning.
- [21] Python IDEs and Code Editors (Guide) Real Python <https://realpython.com/python-ides-code-editors-guide/>
- [22] Visual Studio Code - Code Editing. Redefined Microsoft <https://code.visualstudio.com/>
- [23] spaCy · Industrial-strength Natural Language Processing in Python <https://spacy.io/>