Product and Business Performance Evaluation Based on Customer Reviews

Abstract

Understanding consumer sentiment through online evaluations is crucial for developing business strategy in a digitally altered environment. Using machine learning models and sentiment analysis approaches, the research attempts to find out how customer reviews affect consumer behaviour. To guarantee quality, the dataset—which was obtained from Amazon and eBay—went through extensive investigation and cleaning procedures. Using naive Bayes, random forest, logistic regression, and support vector machines, the study classified reviews' sentiments with a high degree of accuracy (about 94%). Remarkably, the models demonstrated high recall rates, albeit a relatively small dataset was utilized. The results highlight how crucial it is for companies to use customer evaluations as a strategic tool to improve their products and reputation.

1. Introduction and Background

The availability of many types of data has led to enormous advancements in hardware and software technologies, which have also seen tremendous advancements in the field of data mining. This is especially true for text data, where the quick generation of vast libraries of various types of data has been made possible by the advancement of online and social network hardware and software platforms (Liu & Zhang, 2012).

In today's interconnected and rapidly evolving global marketplace, businesses face unprecedented challenges and opportunities. One of the defining features of this modern landscape is the proliferation of online platforms and e-commerce, which have transformed the way consumers research, evaluate, and purchase products and services. Simultaneously, the rise of the internet and social media has given consumers a powerful voice in the form of customer reviews (Majumder, et al., 2022).

The internet has made it easier to collect large amounts of product reviews from various sources. Product reviews is one of the most common sources of data for Sentiment Analysis and Sentiment Analysis has become increasingly important for businesses that want to monitor customer feedback and improve their products and services based on that feedback (Majumder, et al., 2022)

Even though natural language processing (NLP) and linguistics have a long history, until the year 2000, not much research had been done on people's attitudes and opinions. Since then, there has been a lot of activity in the field. This is due to a number of factors. First of all, it may be used in nearly any field with a broad range of applications. The growth of business applications has also led to a flourishing sector surrounding sentiment analysis. This offers a compelling reason to conduct research. Secondly, it presents numerous research topics that have never been examined previously (Liu B., 2022).

Sentiment analysis has become one of the most active fields of natural language processing research since the year 2000. It is also extensively researched in the fields of text, Web, and data mining. Because of its significance to business and society at large, it has actually expanded beyond computer science into the management sciences and social sciences. The industrial side of sentiment research has also flourished in recent years. A multitude of start-ups have surfaced. Numerous big businesses have developed their own internal resources. Systems for sentiment analysis have found use in nearly every industry and social setting (Liu B., 2022).

People and organizations are increasingly using the content found in social media—such as reviews, forum discussions, blogs, microblogs, Twitter, comments, and postings on social network sites—when making decisions due to the rapid rise of these online platforms. These days, there are a lot of user reviews and discussions about consumer products on the Web, so asking friends and relatives for advice is no longer the only option when looking to purchase a product. Because there is a wealth of publicly available information, an organization may no longer need to conduct surveys, opinion polls, and focus groups in order to acquire public viewpoints (Liu B., 2022).

There are various representational levels at which text data can be examined. Text data, for instance, is readily represented as a string of words or as a bag-of-words. To enable more insightful analysis and mining, it would be ideal to represent text data semantically in the majority of applications. Instead of describing text as a collection of words, for instance, storing text data at the level of named entities—such as individuals, groups, and places—as well as their relationships can make it possible to find more intriguing patterns (Liu & Zhang, 2012).

It is important to remember that sentiment analysis is an NLP issue. It covers every facet of NLP, including word sense disambiguation, negation handling, and coreference resolution, which creates additional challenges because they are unresolved issues in NLP. It is helpful to recognize that sentiment analysis is a very limited natural language processing (NLP) problem, though, as the system only needs to comprehend certain parts of each sentence or document, such as the positive or negative sentiments and their target entities or topics. Sentiment analysis, therefore, provides a fantastic platform for NLP researchers to make real advancements across the board, with the potential to have a significant practical impact (Liu B., 2022).

The primary objective of this research is to investigate the role of customer reviews in the evaluation of product and business performance. Specific research objectives include:

1. To analyse the impact of customer reviews on consumer purchasing decisions.

2. To examine the methodologies employed in analysing customer reviews, including sentiment analysis and natural language processing techniques.

3. To explore the strategic importance of leveraging customer reviews for enhancing product quality and optimizing business operations.

Customer reviews have emerged as a pivotal aspect of product and business performance evaluation. These reviews, found on Amazon and eBay, provide consumers with valuable insights into the quality, features, and overall satisfaction associated with a product or service. Furthermore, customer reviews influence purchasing decisions, brand reputation, and the overall success of businesses.

1.1 SENTIMENT ANALYSIS:

Sentiment analysis, often known as opinion mining, is a self-contained text analysis and summary method for online evaluations. The goal of opinion mining is to identify the feelings represented in the reviews, categorise them as positive or negative, and summarise them in a way that is easy for consumers to understand (Ganeshbhai & Shah B.K., 2015).

The goal of opinion spam detection is to find three characteristics—review content, review metadata, and actual product knowledge—that are indicative of a phoney review (Ravi. & Ravi. V., 2015). Machine learning algorithms are frequently used to analyse review content in order to detect dishonesty. Star ratings, IP addresses, geolocation, user IDs, and other information are examples of metadata; yet, in many situations, it is not available for analysis. The third approach makes use of practical experience. For example, reviews from that period may be questionable if a well-known product suddenly has a higher rating than the lesser product during that period (Ravi. & Ravi. V., 2015).

The survey paper by (Medhat, et al., 2014) discusses various algorithms for Sentiment Analysis, including machine learning approaches such as Naive Bayes, Support Vector Machines, Decision Trees, and Random Forests. Other algorithms mentioned in the paper include lexicon-based approaches and hybrid approaches that combine both machine learning and lexicon-based methods. Transfer learning and emotion detection are two related fields to Sentiment Analysis that are discussed in the survey paper . Transfer learning involves analyzing data from one domain and then using the results in a target domain, which can be useful for sentiment analysis tasks that involve different domains . Emotion detection, on the other hand, aims to extract and analyze emotions from text, which can be useful for sentiment analysis tasks that require a more nuanced understanding of the emotional content of text (Medhat, et al., 2014).

(Chen, et al., 2016) created a hierarchical long short-term memory model (LSTM) that integrates user and product data through varying attention levels. Using the IMDB, Yelp2013, and Yelp2014 datasets, they demonstrated how their model outperforms models lacking user and product information by a substantial margin.

(Wang, et al., 2017) introduced an approach in 2017 that made it possible to analyse semantic content extraction systems in great detail. They put forth a novel method of webpage de-duplication using tf-idf and word vector distance.

Because it makes shopping more convenient and allows consumers to explore, browse, compare, and buy a wide range of products without being limited by time or location, this market is becoming more and more significant on a global scale. In order to assist customers, this study provides a decision support model for item comparison in e-commerce that makes use of online reviews and qualitative flexible multiple criteria techniques (Ji, et al., 2018).

(Atzori. & M. Atzeni, 2018) devised a technique in 2018 for converting commands and requests from natural language into computer code. To overcome this obstacle, academics in computer science used Semantic Web technologies to create CodeOntology, an open collaboration platform that enables open-source code to be treated as a first-class citizen on the Internet and linked to other resources. This function retrieves many methods and code examples by using Code Ontology. These are concatenated and graded to create Java source code from a natural language specification.

Deep learning is a promising, yet intricate, method that preserves word order and syntactic structures. Since the internet has become a ubiquitous part of life and false information spreads on the internet just as quickly as true information, opinion spam and fake review detection are major issues in sentiment analysis (Vosoughi, et al., 2018).

A method for classifying 300 transgender people's social media posts was presented by in 2020, and they serve as a representation. We use five different machine learning models to build sentiment analysis classifiers. To determine whether words are connected to effect classification, they cluster the terms using a logistic regression (Li W., 2020).

2020 saw the development of a method by (Li W., 2020) for deciphering the algorithm of natural language semantic understandings and a review suitable for preprocessing technology responses to robotic inquiries, the semantic examination of reputation, etc., comparative evaluations, and performance evaluation. Strong algorithms have been used to preprocess the text, showing how the words relate to one another. The device's flexibility and precision are limited, though.

A constant flow of ideas and viewpoints is made possible by the rapid expansion of social media, e-commerce, discussion boards, and websites that offer product reviews. Companies find it difficult to comprehend consumers' attitudes and opinions about items in general as a result of this expansion. With the use of tools like sentiment analysis and the proliferation of user-generated content, marketers can now obtain insights into how customers feel about their products (Cui, et al., 2023).

2. METHODS:

Sentiment Analysis techniques can be divided into machine learning approach, lexicon based approach and hybrid approach. The machine learning approach applies ML algorithms and uses linguistic features, while the lexicon-based approach relies on a sentiment lexicon, a collection of known and precompiled sentiment terms. The hybrid approach combines both approaches and is very common (Majumder, et al., 2022).

We used the machine learning approach.

The dataset was scraped from the websites of Amazon and eBay through the usage of Beautiful Soup, a python web-scraping library. The dataset has product reviews of earphones, keyboards and keypads, monitors, cameras, computers, screens, memory cards and video games.

Data quality is very important and shouldn’t neglected. We refer to data that has issues with data quality as "dirty data." The quality of the input data set and the accuracy of the results are correlated, and this relationship must be understood since a dirty data might compromise the accuracy of machine learning or data mining tasks (such as classification or clustering) (Qi, et al., 2018).

The dataset consisted of 1856 rows and 3 columns (review_text, review_star and product). Cleaning of the dataset was carried out. Duplicate rows and missing values were removed. Systematic data errors can make model training unreliable and ML problems are highly sensitive to dirty data, even when using robust techniques. Therefore, data cleaning is crucial for statistical analysis as it can significantly affect the accuracy and reliability of the result (Chu, et al., 2016). After cleaning, there was 1842 rows and 3 columns of data left.

It is important to recognize that eBay and Amazon share certain commonalities when comparing the two companies. For instance, both of them allow clients to post comments. It's beneficial since satisfied consumers' positive reviews will boost sales. If you have positive feedback scores, you will probably get purchases from future clients. Naturally, clients have the option to post unfavorable evaluations. Negative reviews on Amazon cannot be edited. If there is an issue, you will need to delete it entirely. You can edit comments on eBay that violate their regulations. Sellers on eBay have the ability to ask customers to modify their reviews. The buyer has complete control over whether or not to accept the seller's suggestion. Additionally, for misbehaving, buyers' risk having their accounts suspended. eBay will take down any unfavorable reviews they may have left if this occurs within ninety days of their transaction with you. (Harrison, 2022).

The ratings were already based on a scale of 1 to 5 for Amazon reviews. The eBay reviews were also converted.

To compare the sentiment across different products at a glance, the average sentiment per product was calculated and the plot is shown in figure 1. The percentage of ratings is also shown in figure 2, and 5-star rating has the highest percentage of 82.1%.

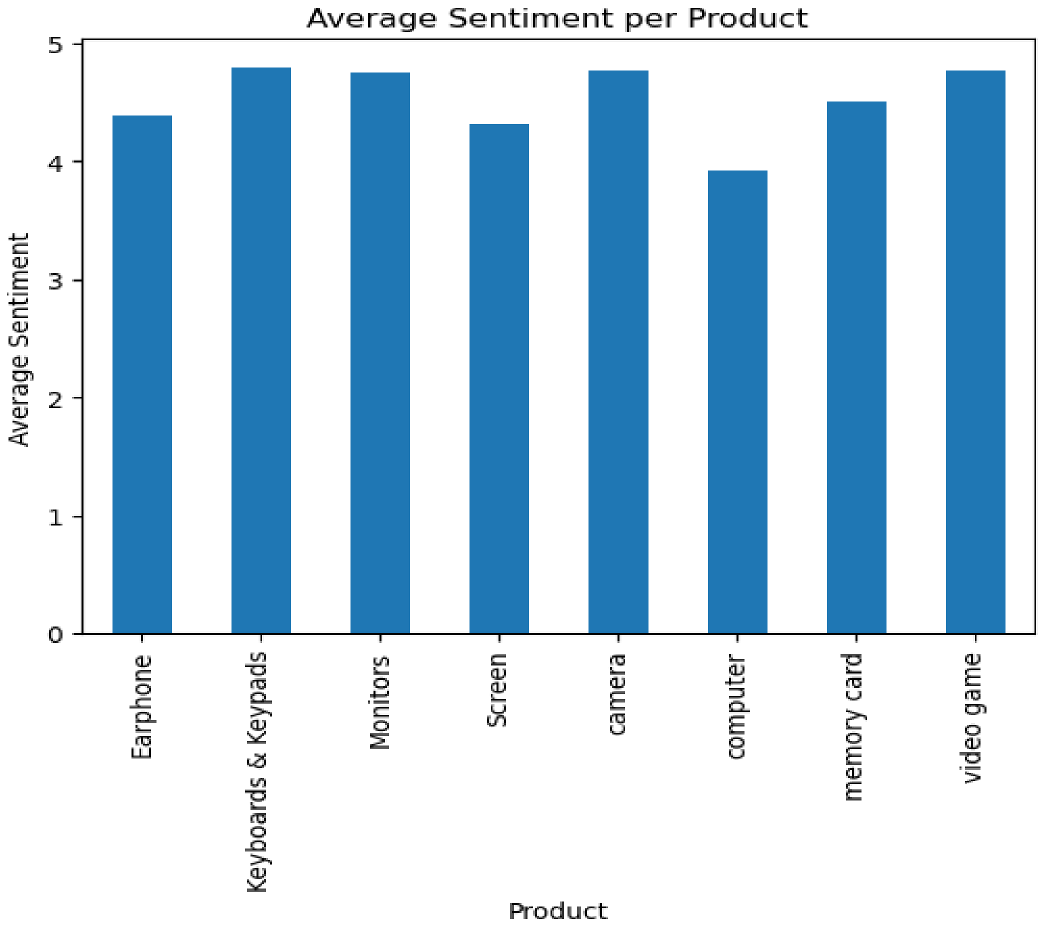

Figure 1 Average sentiment per product

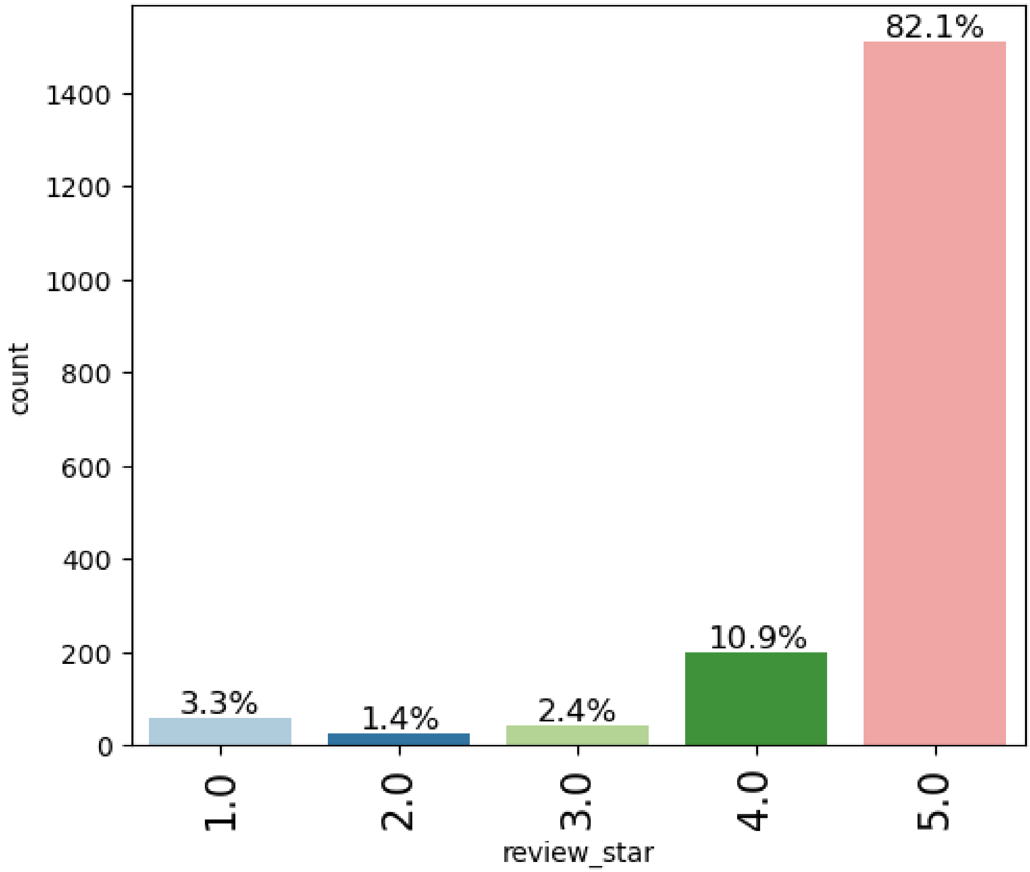

Figure 2 Count and percentage of each rating

Mapping was applied on the ratings. Ratings 1,2 and 3 were tagged “0” and “NEGATIVE” while ratings 4 and 5 were tagged “1” and “POSITIVE”. Figure 3 shows a chart of the percentage of both the negative and positive reviews. The classes are imbalanced; “POSITIVE” class has a count of 1712 while “NEGATIVE” class has just 130.

The wordcloud plot of the most used positive words and negative words used in the reviews is in figure 4 and 5 respectively while that of most used negative and positive words is in figure 6.

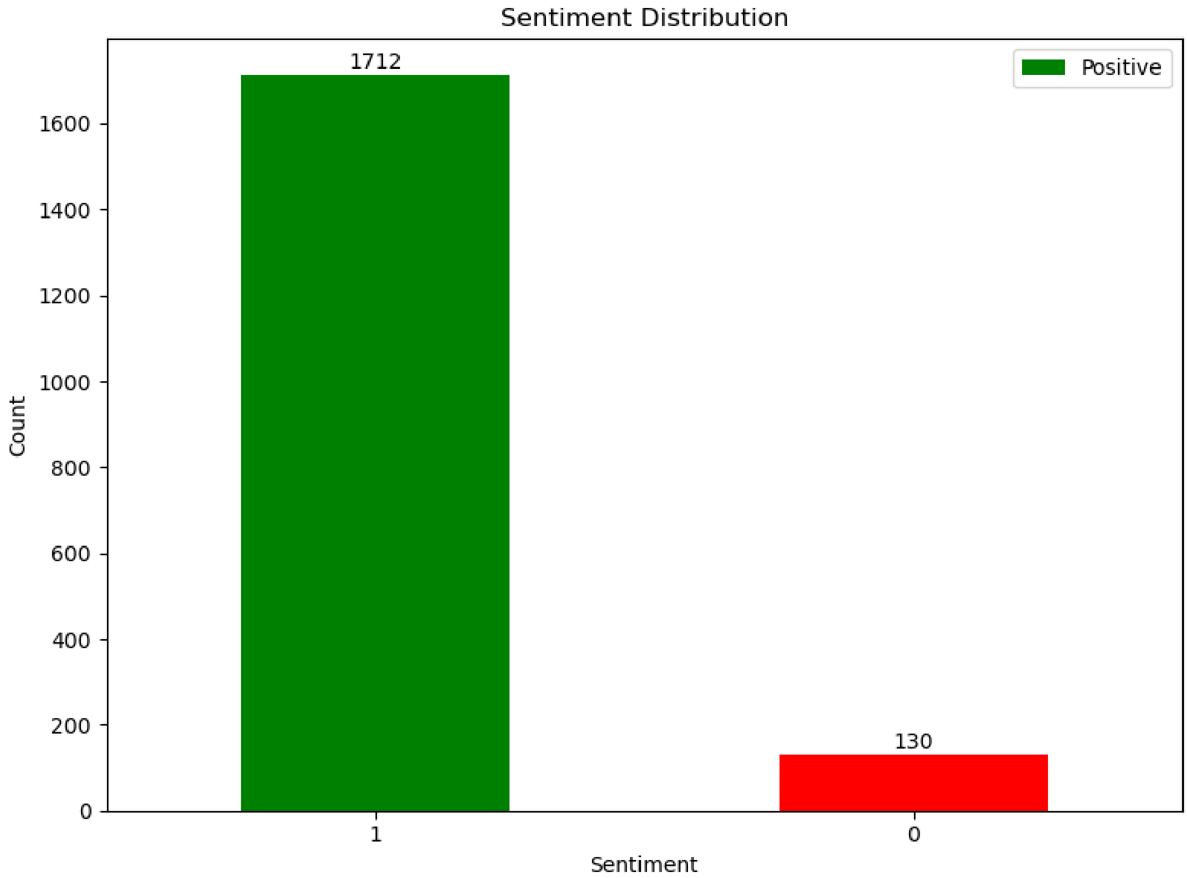

Figure 3 Bar chart of the positive and negative reviews

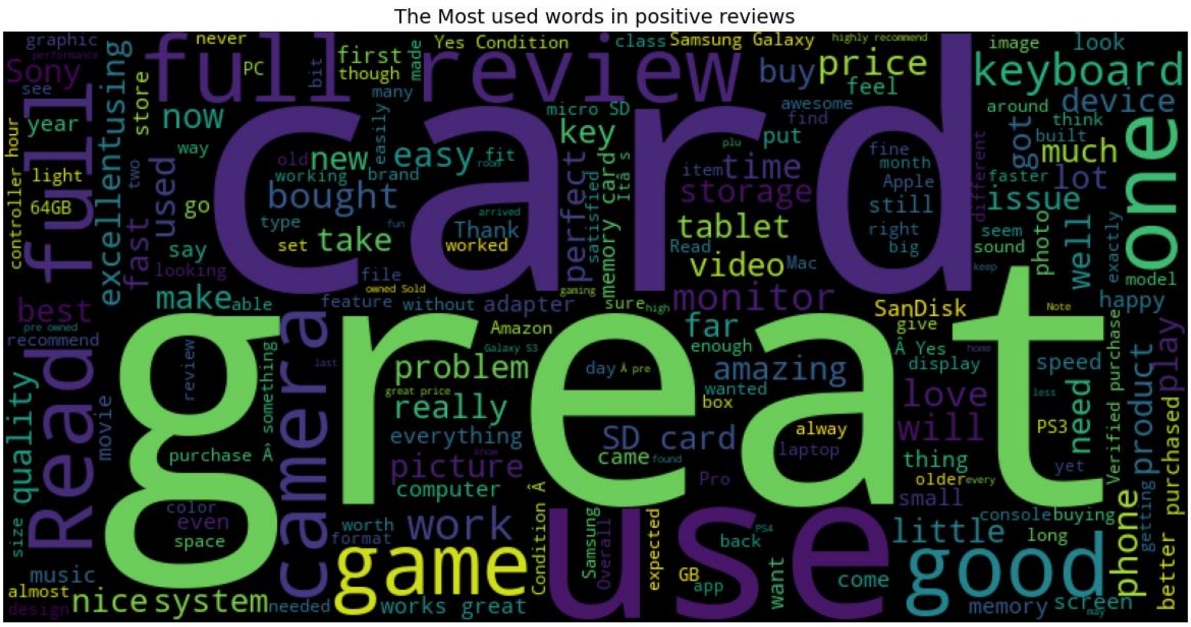

Figure 4 Wordcloud plot of Most used words in positive reviews

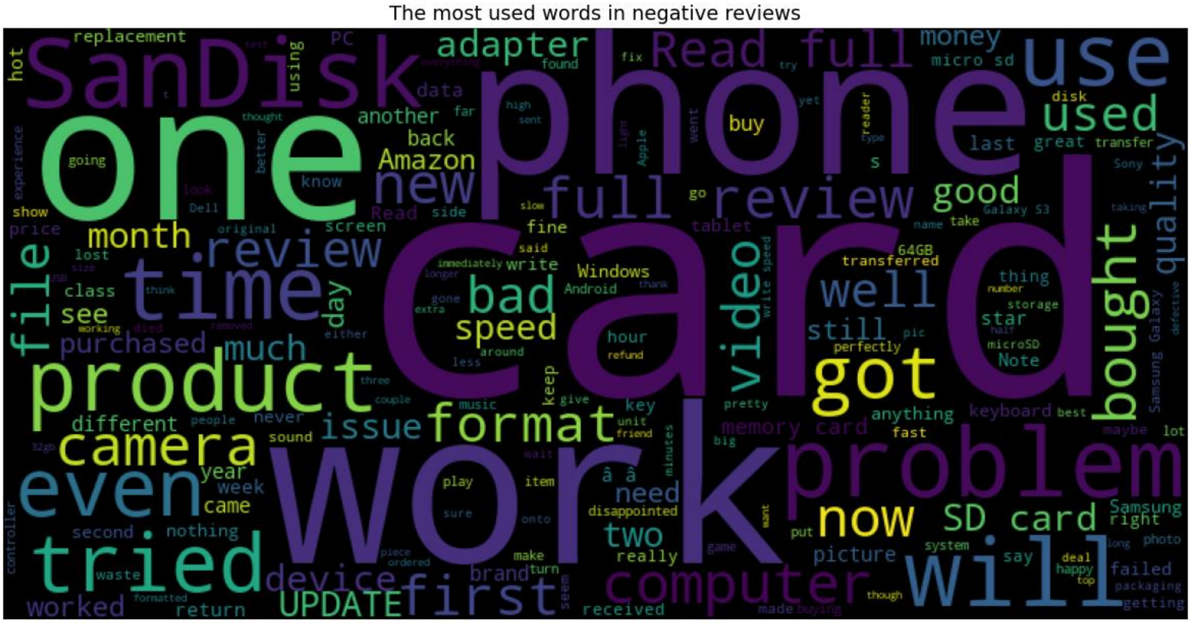

Figure 5 Wordcloud plot of Most used words in negative reviews

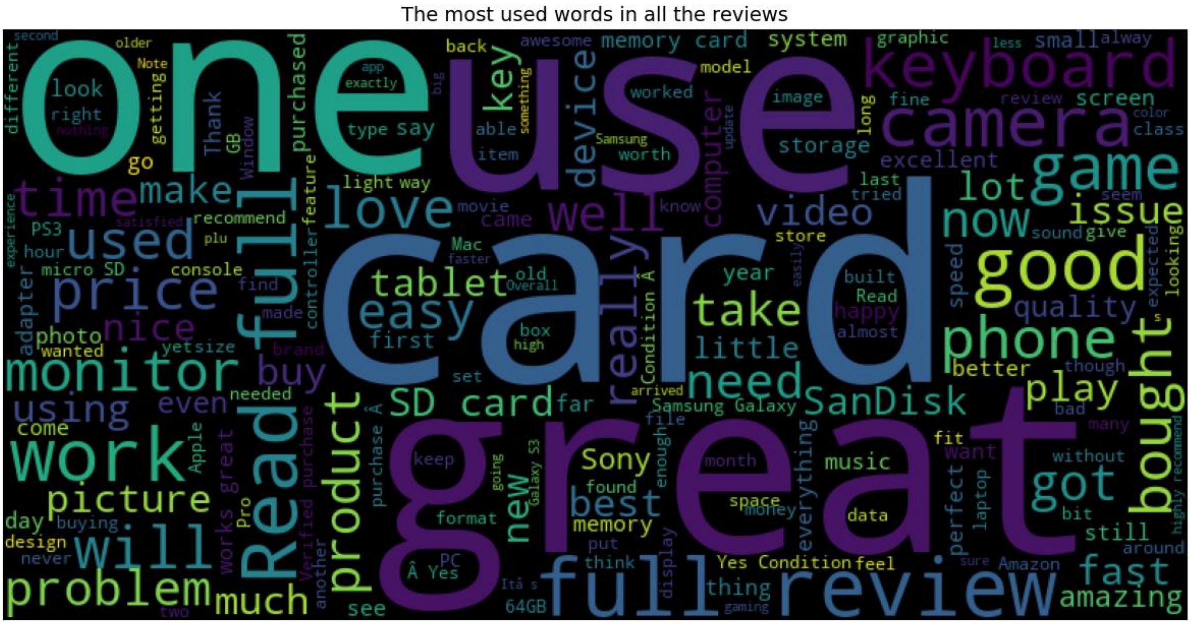

Figure 6 Wordcloud plot of most used words in all the reviews

It is suggested by (Medhat, et al., 2014) that text cleaning is an important pre-processing step in Sentiment Analysis, as it helps to remove noise and irrelevant information from the text data. Text cleaning techniques can include removing stop words, stemming, and removing punctuation and special characters. Some studies have used more advanced techniques, such as part-of-speech tagging and named entity recognition, to improve the accuracy of sentiment analysis.

We used text cleaning to remove punctuations, perform tokenization and stemming of the reviews.

For our feature selection, TF-IDF algorithm was used.

(Medhat, et al., 2014) state that TF-IDF stands for term frequency-inverse document frequency and is used to measure the importance of a term in a document. TF-IDF assigns a weight to each term based on its frequency in the document and its rarity in the corpus. Some studies have used TF-IDF as a feature selection method in combination with other techniques, such as PCA and LDA, to improve the accuracy of sentiment analysis.

TF-IDF algorithm is one of the many algorithms used to enhance the accuracy of machine learning in sentiment analysis. It is used in various studies to extract features from text data and has been shown to be effective in improving the performance of sentiment analysis models (Cui, et al., 2023).

The usage of multiple algorithms can prove advantageous for sentiment analysis by enhancing the precision and resilience of sentiment identification (Tang, et al., 2009).

We used four machine learning algorithms for the sentiment analysis, including Random Forest, Naive Bayes, Support Vector Machines (SVM), and Logistic Regression.

3. RESULTS

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Logistic Regression	0.94	0.94	100	0.97
Support vector Machine (SVM)	0.93	0.94	99	0.96
Random Forest	0.94	0.94	99	0.97
Naïve Bayes	0.94	0.94	100	0.97

Logistic Regression, Random Forest, and Naïve Bayes have very similar performance across Accuracy, Precision, Recall, and F1-score, scoring around 94% in accuracy. These models demonstrate a high level of consistency in their predictive capabilities for the reviews.

Support Vector Machine (SVM), while slightly lower in accuracy at 93%, still maintains strong Precision, Recall, and F1-score in the range of 94-99%. This suggests that the SVM model performs slightly less accurately overall but has a good balance between Precision and Recall.

The models, except for SVM, show high Recall percentages (around 99-100%), indicating their ability to capture a vast majority of the positive instances of the reviews in the sentiment analysis).

Precision scores are also high (around 94%), implying that when these models predict a positive review, they are correct most of the time.

F1-scores for all models are also high, indicating a good balance between Precision and Recall. Figures 7,8,9 and 10 show the confusion matrices for each of the classifiers used.

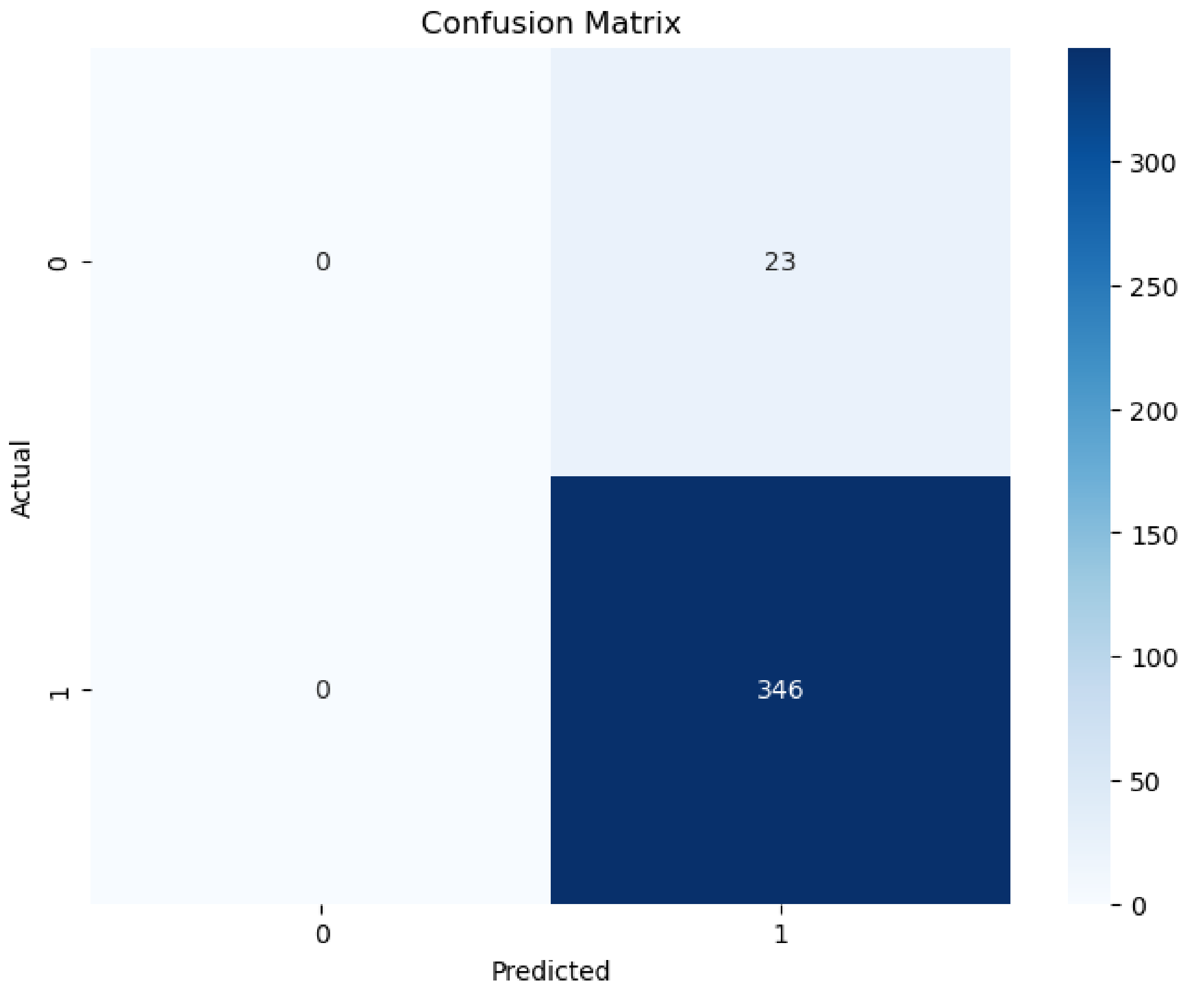

Figure 7 Confusion matrix for Logistic Regression

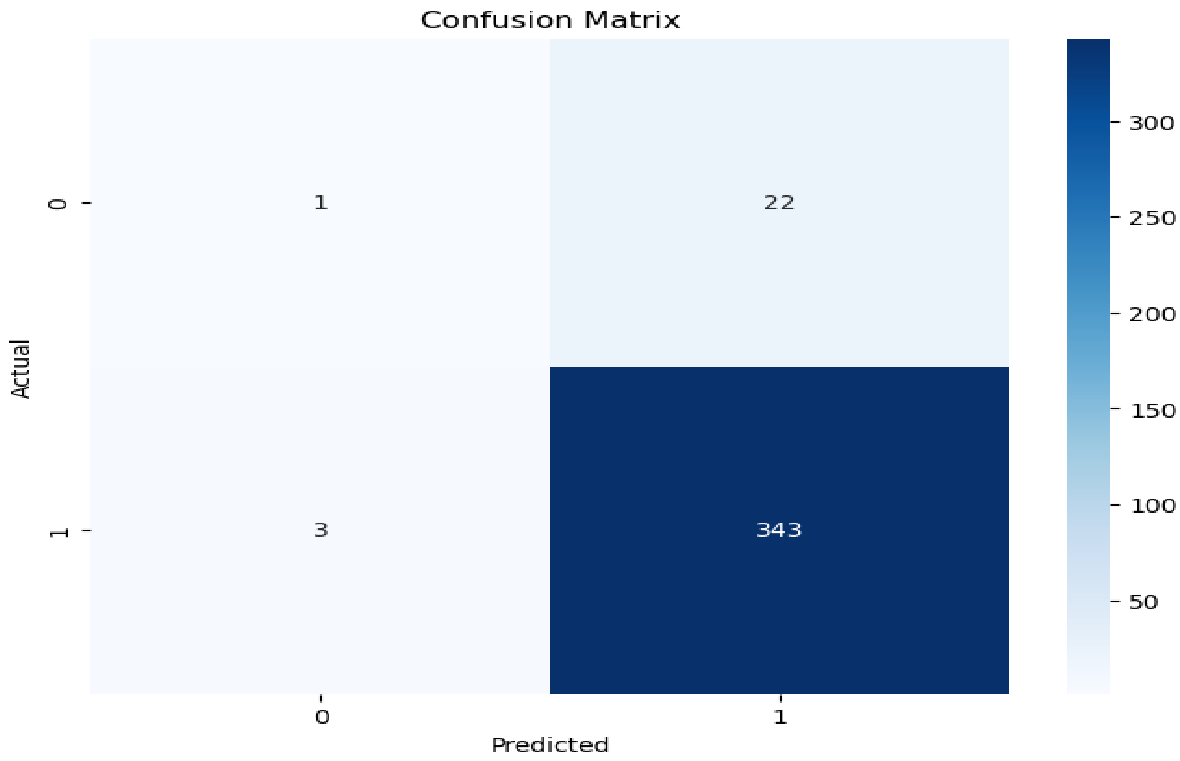

Figure 8 Confusion Matrix for Random Forest

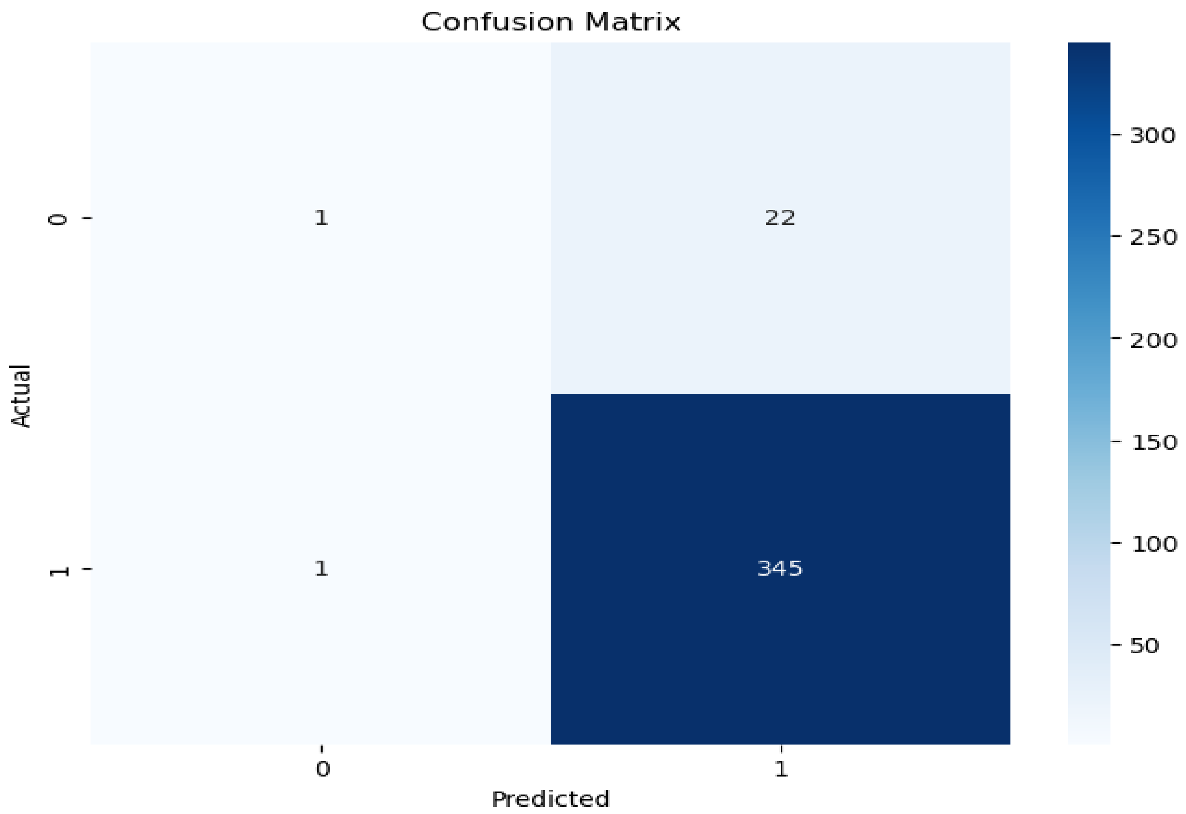

Figure 9 Confusion matrix for SVM

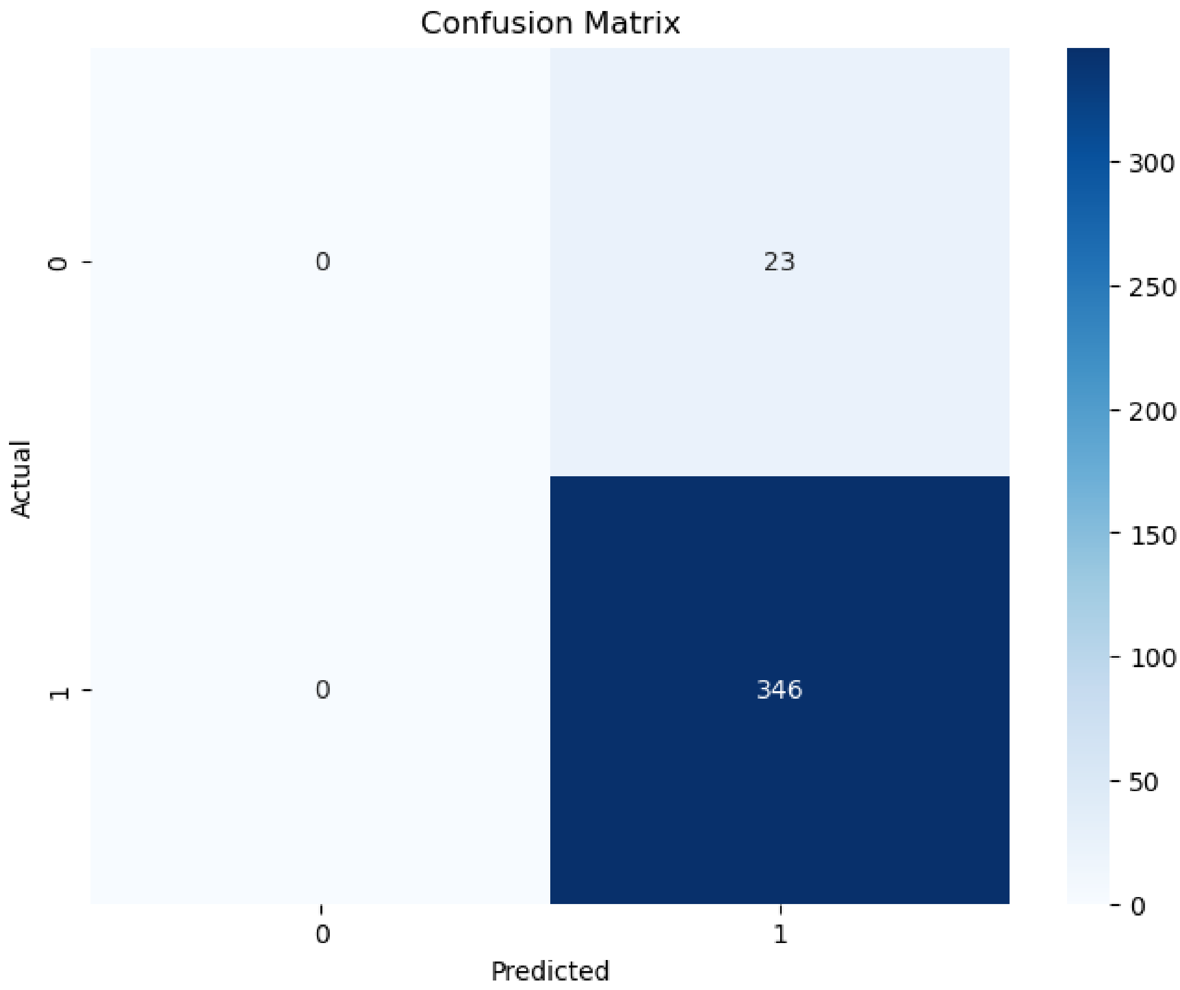

Figure 10 Confusion Matrix for Naive Bayes

4. CONCLUSION AND RECCOMMENDATION

In the rapidly evolving landscape of online commerce, this study has underscored the profound impact of customer reviews on product evaluation and business performance. The analysis has delved deep into the multifaceted realm of sentiment analysis, highlighting its pivotal role in shaping consumer decisions and brand reputation. Through an extensive examination of sentiment analysis techniques and dataset exploration, the study established robust machine learning models, notably Logistic Regression, Support Vector Machine (SVM), Random Forest, and Naïve Bayes, showcasing remarkable accuracy and precision around 94%. These models demonstrated high recall percentages, signifying their ability to effectively capture positive sentiments in reviews.

The confusion matrices unveiled the models' performance in classifying sentiments, emphasizing their accuracy in discerning positive and negative sentiments within customer reviews. Precision, recall, and F1-score metrics further substantiated the models' proficiency in accurately predicting sentiments, thereby aiding businesses in better understanding consumer perceptions.

We used a relatively small dataset for our sentiment analysis. Our future work will involve the usage of a much larger dataset.

The following recommendations are made:

1. Investment in more sophisticated sentiment analysis techniques, such as deep learning or hybrid models, to further refine the accuracy of sentiment analysis. Exploring cutting-edge algorithms could enhance the ability to extract nuanced sentiments from customer reviews.

2. Implement a streamlined system for real-time monitoring and response to customer reviews. Quick and personalized responses to both positive and negative feedback can significantly impact brand perception and customer satisfaction.

3. Continuously refine machine learning models by incorporating new data and adjusting algorithms. This iterative process ensures that models remain relevant and adaptive to evolving consumer sentiments and language trends.

4. Integrate customer reviews strategically across marketing campaigns and product development processes. Leveraging positive reviews as testimonials and addressing negative feedback in product enhancements can strengthen brand credibility and trust.

References

1.Chen,H., Cunchao,T., Maosong,S., Yankai,.L. & Zhiyuan,L. (2016) *Neural sentiment classification with user and product attention.* Austin, Texas, Association for Computational Linguistics.

2. Chu,X., Ihab,F.I., Jiannan W., Krishnan, S. (2016) *Data Cleaning: Overview and Emerging Challenges*, SIGMOD Conference.

3. Cui,J., Cumbria,E. , Seng-Beng,H., & Zhaoxia,W. (2023) *Survey on sentiment analysis:* Evolution of research methods and topics. Springer, p. 8469–8510.

4. FreeCode Camp (2022) Machine Learning For Everybody - Full Course. Available Online: https://www.youtube.com/watch?v=i_LwzRVP7bg

5. FreeCode Camp (2020) *Web Scraping with Python - Beautiful Soup Crash Course.* Available Online: https://www.youtube.com/watch?v=XVv6mJpFOb0

6. Harrison, A., 2022. *Sagemailer.* [Online]
Available Online: https://sagemailer.com/blog/ebay-vs-amazon/
[Accessed 13 December 2023].

7. Ji, P., Wang, J.Q, & Zhang, H.Y. (2018) *A fuzzy decision support model with sentiment analysis for items comparison in e-Commerce:* The case study of PConline.com. *IEEE*, October, pp. 1-12.

8. John,W.R. (2020) *How I Scrape Multiple Pages on Amazon with Python, Requests and BeautifulSoup* Available online: https://www.youtube.com/watch?v=4VfqVpTz4Q4

9. Jonathan, Smith (2023) Web Scrapping: A beginner Project. Amazon store

10. Li, W. (2020) *Analysis of Semantic Comprehension Algorithms of Natural Language Based on Robot's Questions and Answers.* s.l., IEEE.

11. Mattia,A. & Maurizio,A. (2018) *Translating natural language to code:* An unsupervised ontology-based approach," in 2018 IEEE first international conference on artificial intelligence and knowledge engineering (AIKE). s.l., s.n.

12. Madhumit,G. M., Sangita,D.G. & Justin,P. (2022) *Perceived usefulness of online customer reviews:* A review mining approach using machine learning & exploratory data analysis. *Journal of Business Research,* 150(0148-2963), pp. 147-164.

13.Medhat,W., Ahmed,H. & Hoda,K. (2014) *Sentiment analysis algorithms and applications*: A survey. *Ain Shams engineering journal,* pp. 1093-1113.

14. Mine, D. & Ctinkaya-Rundel, M. (2020) *Web Scraping in the Statistics and Data Science*: Challenges and Opportunities. Journal of statistics and Data Science Education, pp.1-3.

15. Ravi,K. & Vadlamani,R. (2015) *A survey on opinion mining and sentiment analysis*: tasks, approaches and applications. *Science Direct,* pp. 14-46.

16. Simplilearn (2018) *Logistic Regression in Python* : Machine Learning Algorithms. Available online: https://www.youtube.com/watch?v=XnOAdxOWXWg

17. Su-Zhen,W. & Qing-Chuan,Z. (2017) *Natural language semantic corpus construction based on cloud service platform.* s.l., IEEE.

18. Shah,B.K. & Solanki, Y.G.(2015) *"Feature based opinion mining: A survey," 2015 IEEE International Advance Computing Conference (IACC).* Banglore, India, IADCC.

19. TileStats (2022) Support Vector Machines (SVM) : The Basics Simply Explained. Available Online: https://www.youtube.com/watch?v=gUzEN2TxnxE

20. Vosoughi,S., Roy,D. & Sinan,A. (2018) *The spread of true and false news online. Science, pp. 1146-1151.*