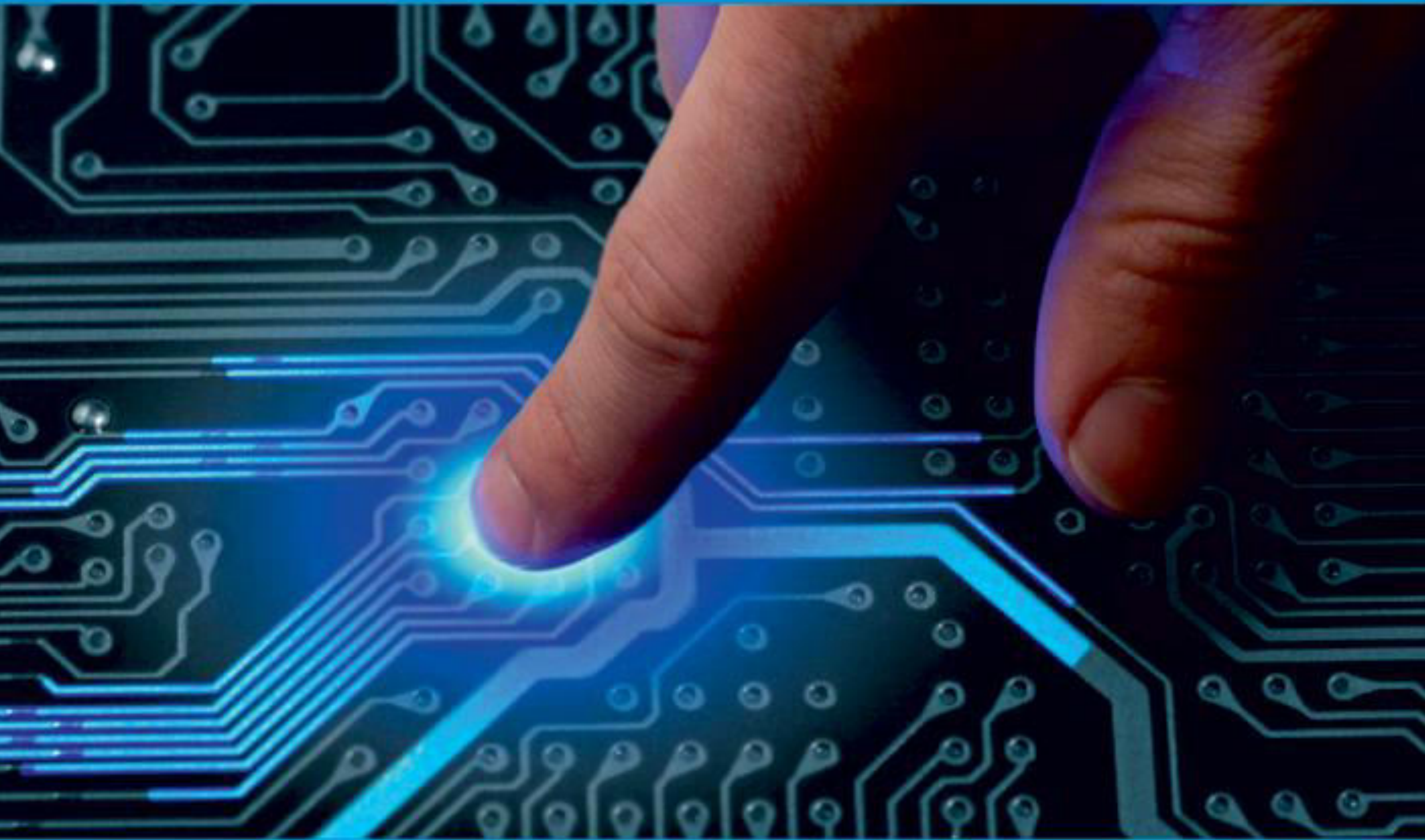
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204012 |

Sentiment Analysis for Social Media Marketing in E-Commerce Website

Dr.J.Jeyaboopathiraja[1], P.Selvadharsan[2]

[1]Assistant Professor, PG & Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, India

[2]UG Student, PG & Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, India

ABSTRACT: Sentiment analysis serves as a cornerstone for online product companies, enabling them to delve into the sentiments expressed within user reviews, thereby facilitating better decision-making processes and product recommendations. This analytical approach involves the systematic examination of collections of reviews to extract valuable insights that can guide user preferences and enhance their overall experience. The industry widely relies on sentiment analysis to align product offerings with user sentiments gleaned from reviews, thereby optimizing customer satisfaction and driving sales. In a comprehensive comparative study focusing on sentiment analysis within the Electronics category on Amazon, a range of machine learning models were meticulously evaluated. These models, including Support Vector Machines, Naive Bayes Classifier, Random Forest Classifier, BERT Model, Stochastic Gradient Descent (SVM Linear), and Linear SVC models, were deployed to process and dissect the content of reviews. Their objective was to gain a profound understanding of user sentiments towards various products offered on the platform. The study endeavours to augment the comprehension of how diverse machine learning algorithms perform sentiment analysis on Amazon product reviews. By meticulously comparing the effectiveness of these models, companies can make well-informed decisions grounded in user feedback. This, in turn, facilitates the enhancement of product recommendations, thereby fostering improved customer satisfaction and loyalty within the competitive landscape of the online marketplace. The significance of sentiment analysis cannot be overstated in the realm of online product companies. It not only aids in deciphering user sentiments but also serves as a catalyst for understanding consumer behavior and preferences. Through sentiment analysis, companies can uncover valuable insights into the strengths and weaknesses of their products, enabling them to refine their offerings and stay ahead of market trends.

The deployment of machine learning models such as Support Vector Machines, Naive Bayes Classifier, and Random Forest Classifier underscores the commitment of online product companies to harness the power of data-driven insights. These models, when employed effectively, can sift through vast volumes of user-generated content, distilling meaningful patterns and sentiments that inform strategic decision-making.

KEYWORDS: filtering, sentimental analysis, SVM, Naïve bayes , Random forest algorithm

I. INTRODUCTION

Social media plays a pivotal role in modern society, providing individuals with a platform to express their opinions and sentiments about products available on e-commerce websites. Customer reviews and opinions play a crucial role in shaping consumer decisions, as they offer insights into the quality, features, and overall performance of products. In the fast-paced world of e-commerce, where new products flood the market regularly, customers heavily rely on these reviews to guide their purchasing choices. By analysing customer reviews, businesses can identify trends, strengths, weaknesses, and areas for improvement in their products. This data-driven approach allows companies to tailor their offerings to meet customer expectations, enhance product performance, and ultimately boost customer satisfaction and loyalty. Leveraging sentiment analysis tools and techniques, businesses can extract actionable insights from reviews, enabling them to refine their products, address customer concerns, and stay competitive in the dynamic e-commerce landscape. In essence, the analysis of customer reviews is a powerful tool that empowers businesses to understand consumer sentiments, improve product offerings, and enhance the overall shopping experience for customers. By delving into the nuances of customer feedback, businesses can adapt to market demands, optimize their strategies, and foster long-term relationships with their clientele.

Customer reviews or ratings aim to define the attitude of the writer towards the product. It may be positive, negative, or neutral. Some people give a product four or five star and express their final satisfaction with it, and others give product one or two stars and express their final dissatisfaction with it. This does not present any difficulty in sentiment analysis. However, other people give three stars, although obviously expressing their final satisfaction with it. This leads to confusing other customers, as well as companies, who want to know their actual opinion.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204012 |

Consequently, customers and companies face difficulty with respect to analyzing reviews and understanding consumer satisfaction. So, the three-star rating doesn't actually represent a neutral sentiment, because in practice people who assign a 3-star rating to a product or service don't necessarily mean that they're absolutely balanced in their opinion between positive and negative. Based on this argument, this research proposes a sentiment analysis to predict the polarity of Amazon mobile phone dataset reviews. We will leave the 3-star rating as is and consider it to represent a neutral sentiment. This is done with the purpose of increasing the challenge and difficulty of this study and to measure the efficiency of state-of-the-art NLP models, like BERT in solving difficult classification problems. Furthermore, four machine-learning models with different feature extraction approaches will be used in this research: Logistic Regression, Naïve Bayes, Random Forest, and Bi-LSTM. Then, we analyse the best performance model in order to investigate its sentiment classification. At the end of the study, we will take the best performing model and re-train it on the dataset with the neutral class removed, effectively recasting the problem as a binary-classification problem. We'd like to measure how much this recasting of the problem will affect model performance.

II. RELATED WORKS

Due to the proliferation of online reviews, Sentiment analysis has gained much attention in recent years. Therefore, many studies have been devoted to this research area. In this section, some of the most related research works to this thesis are presented. Joachims (1998) experimented SVM for text classification and showed that SVM performed well in all experiments with lower error levels than other classification methods. Pang, Lee and Vaithyanathan (2002) tried supervised learning for classifying movie reviews into two classes, positive and negative with the help of SVM and Na¨ıve Bayes and maximum entropy classification. In terms of accuracy all three techniques showed quite good results. In this study they tried various features and it turned out that the machine learning algorithms performed better when bag of words was used as features in those classifiers.

In a recent survey that was conducted by Ye et al. (2009), three supervised machine learning algorithms, Naive Bayes, SVM and N-gram model have been attempted on online reviews about different travel destinations in the world. In this study, they found that in terms of accuracy, well trained machine learning algorithms performs very well for classification of travel destinations reviews. In addition, they have demonstrated that the SVM and N-gram model achieved better results than the Naive Bayes method. However, the difference among the algorithms reduced significantly by increasing the number of training data set. Chaovalit and Zhou (2005) compared the supervised machine learning algorithm with Semantic orientation which is an unsupervised approach to movie review and found that the supervised approach provided was more reliable than the unsupervised method.

Agarwal, Y.et al. [3] opinion mining is increasingly influencing the market. The goal of this research was to give a review and update on the sentimental analysis that is often conducted via social networking. Numerous publications and studies in this sector have been published and completed.

Arti, Dubey, K. P., & Agrawal, S. [5] Authorized users may use Twitter's API to gather data and insights from tweets to analyze public opinion for sporting events such as the 2016 Indian Premier League. The result reflects the population's positive and negative opinions. This kind of opinion analysis might give valuable information to the organization and assist them in detecting an unfavourable change in Twitter audience understanding. Early detection of unfavourable trends allows businesses to make educated choices about focusing certain parts of their services and goods on boosting consumer satisfaction.

Da'u, A.et al. [7] An R.S. model based on the ABOM approach is presented in this study. Initially, these authors demonstrated how user ratings might be derived using a deep learning technique. To further improve the recommendation system, the recovered features were used to generate aspect-based scores, which were then put into a tensor factorization engine. ABOM and rating prediction make up the majority of the proposed method. Itutilised latent dirichlet allocation (LDA) to combine the extracted aspects with the underlying opinion terms to form latent aspects. Next, for each feature of the review, a score was calculated based on a linguistic technique. To further include the user's perspective on many elements and overall ratings, a three-dimensional T.F. approach is used to determine the representation of the underlying components.

Hasan, K. M. A.et al. [9] these authors gathered information from reliable sources. The data were preprocessed to remove noise and improve usability.

Kumar, P.et al. [12] indicated that a comprehensive examination of any issue might be undertaken by gathering a representative sample of Twitter user thoughts. Such an investigation might give helpful information to organizations or creators of movies, television shows, and businesses and warn them of an alarming change in the public's impression of their area. Identifying unfavourable tendencies early on may enable businesses to make more informed judgments about enhancing consumer satisfaction by concentrating on specific product characteristics. This study shows that the machine learning classifier utilized significantly impacts the inquiry's overall accuracy.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204012 |

An area of natural language processing (NLP) called opinion mining, commonly referred to as sentiment analysis, focuses on the computer analysis of people's verbally stated opinions, feelings, and attitudes [1]. Opinion mining has grown more significant in a range of areas, including marketing, politics, healthcare, and customer service [2] because to the internet's exponential increase in user-generated information. Opinion mining aims to identify and extract subjectivity from text, such as reviews of products, comments on social media, and news articles, and to categorise it as good, negative, or neutral [3]. This might offer insightful information on public opinion, assist businesses in improving their goods and services, and have an impact on legislative choices [4].

III. METHODOLOGY

The proposed system in the project sentiment analysis for social media marketing in e-commerce websites aims to analyze the sentiments expressed in customer reviews on e-commerce platforms. The main objectives of this system include:

•Understanding Customer Sentiments: The system is designed to analyse the opinions, attitudes, and emotions of customers towards products and services offered by e-commerce businesses. This helps in gauging the overall sentiment towards a particular product or brand.

•Improving Marketing Strategies: By analysing customer sentiments, businesses can identify emerging trends, evaluate the effectiveness of marketing campaigns, and monitor brand reputation in real-time. This allows them to make data-driven decisions to improve their marketing efforts.

•Enhancing Customer Engagement: The system can help businesses tailor their communication strategies to better engage with customers, address concerns promptly, and build stronger relationships.

•Optimizing Product Development: Sentiment analysis can provide valuable insights into customer preferences and pain points, enabling businesses to refine their product offerings and enhance features to meet customer expectations.

It is an estimator that fits randomised decision trees to distinct subsamples of a dataset to boost accuracy and control over data fitting. To produce random splits for each of the maximum characteristics before dividing the node's data in half rather than predetermining which split is ideal. With 75 trees used in its design and 50 trees used in its construction, the Extra Trees Classifier had an accuracy rate for the test data of 90.27 percent and 90.73 percent, respectively. This model is not further developed since adding more trees does not significantly increase accuracy.

$$\varphi_i^* = \underset{\varphi i \varepsilon \rho}{argmin}\{\textstyle\sum_{j=1}^{m} l(yj, yj)\} \text{ ------ (1)}$$

$$= argmin\{\textstyle\sum_{j=1}^{m} l(flightGBM(\varphi i, xj), yj\} \text{------- (2)}$$

In Eq. (7), ℓ (•) is the loss function, yˆj is the prediction yˆj=flightGBM(ψi, xj), and yj is the given actual class.

Let k be a scalar of n bits, P and Q two points of an ellipticallycurvedefined on a finite field F by the equation

$$E: Y^2 + a1XY += X^3 + a_2X^2 + a_{4x}X + a_6 \text{ ------- (3)}$$

where a1; a2; a3; a4 2 F. In this work, to consider the finite prime fields Fp. Eq. (9) then becomes:

$$E: Y^2 = X^3 + aX + b \text{ ------ (4)}$$

The system leverages natural language processing (NLP) and machine learning algorithms to classify text into positive, negative, or neutral sentiments. This allows businesses to quantify customer sentiments and track changes over time, enabling them to make informed decisions based on customer feedback.

In our approach we have developed the application in two phases in the first phase we have collected the sentiment data from the e-commerce website and separated it into words using Parts of Speech Tagging (POST). The second phase of our proposed system model gives the real-time analytics of the data gathered from an e-commerce site. The collected data is fed to the trained model which determines the sentiment in real-time. For this purpose, we have developed a Restful web service approach which is implemented using FLASK and python8 . Scalability, fault tolerance, and availability are the challenging aspects of sentiment analysis. To overcome these challenges the Spark NLP and SparkML based PySpark is used. It has an inbuilt Natural Language Toolkit (NLTK) library to perform sentiment analysis.

Many reviews are collected in the form of structured and unstructured forms. We need in-memory distributed data processing for doing sentiment analysis. The Apache Spark framework gain importance due to in-memory distributed processing. It has in-built SparkML library to accomplish certain natural language processing tasks but doesn't provide fully fledged solution for natural language processing. To overcome this issue the John Snow Labs contributed in the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204012 |

development of Spark NLP. It is built to does NLP tasks completely. The data collection and data modeling process of sentiment analysis is shown in the proposed system architecture as hown in Figure 1.

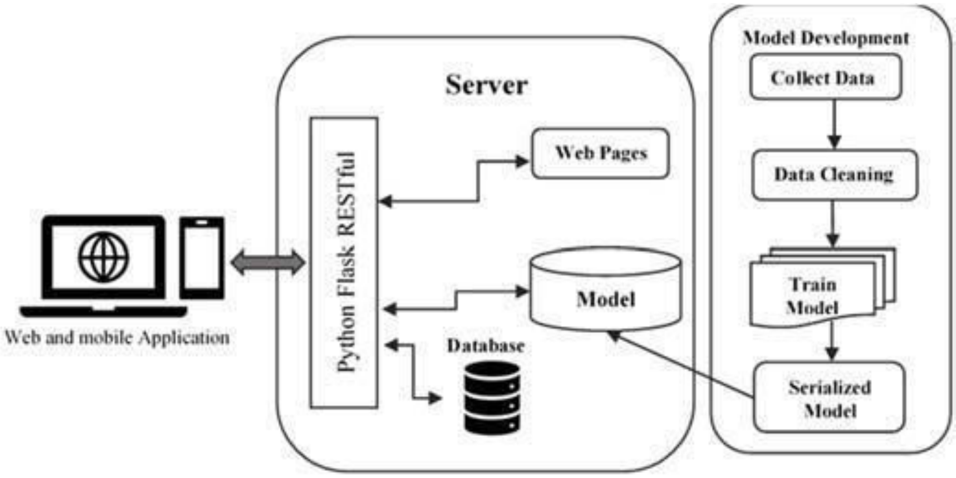


Figure-1 System Architecture

The Spark NLP is an open-source natural language processing library. It has annotators which are either transformers or estimators, which utilize machine learning, deep learning, and rule-based algorithm. The estimators are basically the learning algorithm whereas the transformers convert one dataframe from one format to another. The result of Spark NLP is an annotation. Some of the important annotators are Tokenizer, Normalizer, Stemmer, Lemmatizer etc. which does the natural language processing. It has wide ranges of pre-trained model also known as Annotator model such as NerDLModel, Deep Sentence Detector, Lemmatizer Model etc. This model helps in transforming on DataFrame into another.

IV. RESULTS AND DISCUSSION

The proposed method has implemented by using python programming language with spyder 3.8 framework.

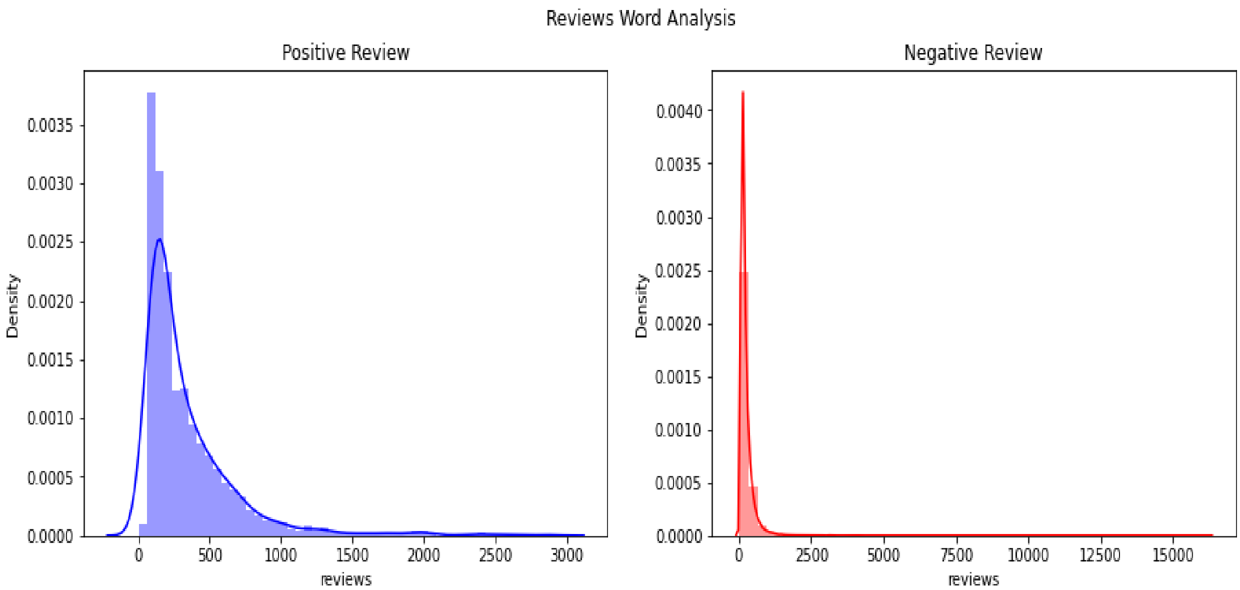


Figure 2: Reviews and analysis

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204012 |

Figure 2 is a graph that illustrates reviews and analysis instances, illustrating both positive and negative types of input. The reviews are shown on the x-axis, while the density of those reviews is shown on the y-axis. The graph resembles a histogram, with bars of varying heights standing in for the number of reviews falling into various densities. The testimonials in Figure 2 probably refer to written evaluations or remarks made by consumers or users of a certain item or service. The density metric might be a score or rating determined using natural language processing (NLP) methods, which would be a measure of the sentiment or tone of those evaluations.

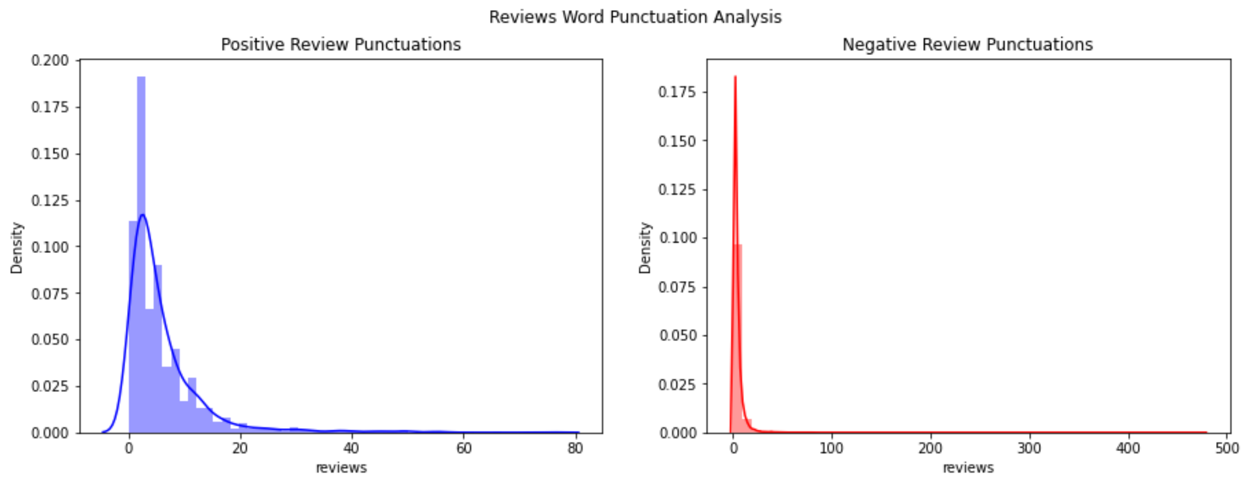


Figure 3:Reviews word punctuation analysis

The examination of punctuation for words is shown in figure 3. The graphic highlights both positive and negative evaluations of the product. This graph shows reviews along the x-axis, while density is shown along the y-axis.

Table 1: Training and validation accuracy and loss

Epoch	Training Loss	Validation Loss	Training Accuracy	Testing Accuracy
1	0.1544	0.0679	0.9547	0.9779
2	0.0537	0.0526	0.9837	0.9820
3	0.0352	0.0596	0.9890	0.9826
4	0.0243	0.0465	0.9922	0.9848

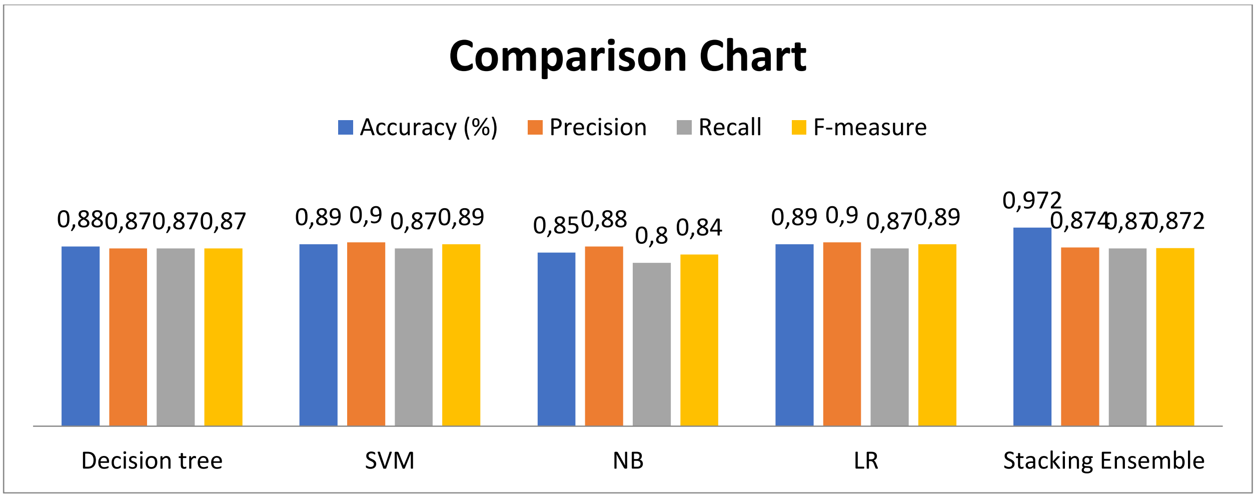


Figure 4: Comparison chart

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204012 |

The classification performance matrices for a variety of models, including Decision Tree, SVM, Naive Bayes (NB), Logistic Regression (LR), and Stacking Ensemble, are displayed in Table . Metrics including accuracy, precision, recall, and F-measure are included in the matrix. The Stacking Ensemble model exceeds the other models in terms of accuracy, scoring 0.972 as opposed to 0.85 to 0.89 for the other models. The SVM and LR models had the greatest accuracy ratings (0.90), closely followed by the NB model (0.88). The Decision Tree and Stacking Ensemble models are closely behind the SVM and LR models in terms of recall, scoring 0.87 and 0.870, respectively.

V. CONCLUSION

In conclusion, this paper on sentiment analysis for an e-commerce website has been a significant endeavor in leveraging natural language processing techniques to extract valuable insights from customer feedback. Through the analysis of sentiments expressed in reviews and comments, this project has shed light on the crucial role sentiment analysis plays in understanding customer satisfaction levels, product preferences, and areas for improvement within the e-commerce domain. The development and implementation of a sentiment analysis model have proven to be instrumental in accurately categorizing sentiments as positive, negative, or neutral, thereby empowering the e-commerce platform to address customer concerns promptly, enhance user experience, and tailor marketing strategies based on real-time feedback. The successful deployment of the sentiment analysis model underscores its potential to revolutionize how e-commerce websites engage with their customers, make data-driven decisions, and ultimately drive business growth through customer-centric strategies. By harnessing the power of sentiment analysis, the e-commerce website can not only identify customer sentiments but also predict trends, anticipate needs, and personalize interactions to create a more seamless and satisfying shopping experience for users.

REFERENCES

1. B. McMahan et al., "Communication-Efficient Learning of Deep Networks From Decentralized Data", *Artificial Intelligence and Statistics Proc. PMLR*,vol. 10, no. 1, pp. 1273-82, 2017.
2. Napoleon D. and Praneesh M. “Detection of Brain Tumor using Kernel Induced Possiblistic C-Means Clustering”, volume no.3, issue no.9, pp 436-438, 2013
3. Pimwadee Chaovalit and Lina Zhou. Movie review mining: A comparison between supervised and unsupervised classification approaches. In System Sciences, 2005. HICSS’05. Proceedings of the 38th Annual Hawaii International Conference on, pages 112c–112c. IEEE, 2005.
4. Nello Cristianini and John Shawe-Taylor. An introduction to support vector machines and other kernel-based learning methods. Cambridge university press, 2000.
5. P´adraig Cunningham, Matthieu Cord, and Sarah Jane Delany. Supervised learning. In Machine learning techniques for multimedia, pages 21–49. Springer, 2008.
6. Sajib Dasgupta and Vincent Ng. Topic-wise, sentiment-wise, or otherwise?: Identifying the hidden dimension for unsupervised text classification. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2, pages 580–589. Association for Computational Linguistics, 2009.
7. Xing Fang and Justin Zhan. Sentiment analysis using product review data. Journal of Big Data, 2(1):5, 2015.
8. Emma Haddi, Xiaohui Liu, and Yong Shi. The role of text pre-processing in sentiment analysis. Procedia Computer Science, 17:26–32, 2013.
9. Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 168–177. ACM, 2004.