ANALYSING CUSTOMER SENTIMENTS IN PRODUCT REVIEWS USING MACHINE LEARNING TECHNIQUES

[1] M.Raghavendra Rao, M.Tech, Assistant Professor, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[2] Sanapala. Mohan Murali Satyanarayana, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[3] Yerramsetti. Abhitha, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[4] Nethala. Bharath, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[5] Mohammad Asif, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

Abstract: Digital reviews play a crucial role in shaping consumer perceptions and influencing purchasing decisions globally. Major e-commerce platforms like Amazon and Flipkart serve as hubs for consumers to share their firsthand experiences, offering invaluable insights to prospective buyers. Extracting meaningful information from this vast repository of reviews necessitates categorizing them into positive and negative sentiments. Sentiment Analysis, a computational technique, facilitates this process by extracting subjective information from text data. In our research, we employed Sentiment Analysis to classify over 4, 000, 00 reviews into positive and negative sentiments. We utilized Naïve Bayes, Support Vector Machine (SVM), and Decision Tree models for classification, evaluating their performance through 10 Fold Cross Validation.

1. INTRODUCTION

With an ever increasing demand of smart phones, the mobile phone market is expanding at an exponential pace. With such a boom in the smart-phone industry, there is a need to realize the holistic review of the brand and the model of phone. There are numerous brands present in the market, out of which some are dominant and occupy quite a big part of the industry. For instance, Samsung, Apple, etc. are names associated with brands which are famous throughout the world. Electronic commerce plays a vital role in increasing the sales of the mobile phones and influencing consumer buying patterns. Reviews available on such e-commerce platforms act as a guiding tool for the consumers to make informed decisions. Retail websites like Amazon.com offer different options to the reviewers for writing their reviews. For instance, the consumer can provide numerical rating from 1 to 5 or write comments about the product. As there are innumerable products manufactured by many different brands, so providing relevant reviews to the consumers is the need of hour. Number of reviews associated with a product or a brand is increasing at an alarming rate, which is no less than handling the big data. Classifying the reviews on the basis of sentiment of customers into positive and negative sentiment provides sentiment orientation of the review, hence results in better judgement. Segregation of reviews on the basis of their sentiment can help future buyers to evaluate positive and negative feedback constructively and reach at better decisions as per their requirements. This evaluation acts as a testimony to the users who are looking to know the details and specifications of the smartphones; thereby increasing user credibility. In this research, unstructured data of Mobile Phone Reviews have been extracted from Amazon.com. It has been filtered to remove noisy data and has been pre-processed to evaluate sentiment of the reviews using supervised learning. The reviews have been classified using machine learning classification models like Naïve Bayes, Support Vector Machine (SVM) and Decision Tree and have been cross validated to find the best classifier for this purpose.

2. LITERATURE SURVEY

2.1 Semantic orientation applied to unsupervised classification of reviews.

Semantic orientation applied to unsupervised classification of reviews" introduces a methodological approach to categorize reviews based on their semantic orientation without the need for labeled training data. The paper explores techniques for determining the polarity of opinions expressed in reviews, focusing on the use of semantic orientation to discern positive and negative sentiments. Through a detailed examination of the methodology, the abstract underscores the significance of unsupervised classification in sentiment analysis tasks, particularly in scenarios where labeled data is limited or unavailable. The paper contributes to advancing the field of natural language processing by offering insights into innovative approaches for sentiment classification in textual data.

2.2 Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. textual data. Mining the peanut gallery: Opinion extraction and semantic

classification of product reviews" delves into the computational methods for extracting opinions and performing semantic classification of product reviews. The paper outlines techniques such as feature extraction, sentiment lexicons, and machine learning algorithms employed to categorize reviews based on sentiments and semantic meaning. Through a detailed exploration of methodologies, the abstract highlights the significance of opinion mining in understanding consumer sentiments and **preferences. The paper contributes to advancing the field of** natural language processing by providing insights into the challenges and advancements in extracting opinions from textual data, particularly in the context of product reviews.

2.3 Opinion Mining and Sentiment Analysis

"Opinion mining and sentiment analysis" explores the computational study of extracting subjective information from text. This field has significant implications for understanding consumer sentiments, product reviews, and social media discourse. Pang and Lee (2008) provide a comprehensive overview of methodologies and techniques, including lexicon-based approaches and machine learning algorithms, for classifying opinions and sentiments expressed in textual data. The abstract encapsulates the essence of sentiment analysis, emphasizing its importance in extracting valuable insights from large scale textual datasets across various domains.

2.4 Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologie

Sentiment analysis and opinion mining," as presented in the Synthesis Lectures on Human Language Technologies, provides a comprehensive exploration of techniques for extracting subjective information from textual data. This seminal work encompasses various methodologies, including lexicon-based approaches and machine learning algorithms, to classify sentiments and opinions expressed in diverse contexts. The abstract encapsulates the fundamental role of sentiment analysis in understanding consumer behavior, product reviews, and social media sentiments. Through a synthesis of key findings, the paper contributes to advancing the field of natural language processing by shedding light on methodologies and applications in sentiment analysis and opinion mining.

2.5 Serena Identifying the semantic orientation of terms using s-hal for sentiment analysis,

"Identifying the semantic orientation of terms using S-HAL for sentiment analysis" proposes a novel approach for determining the semantic orientation of terms in sentiment analysis tasks. The abstract outlines the utilization of S-HAL (Supervised Hierarchical Associative Lexicon) to infer the sentiment polarity of individual terms within a given context. Through detailed experimentation and analysis, the paper demonstrates the effectiveness of S-HAL in accurately identifying the sentiment orientation of terms, thus enhancing the overall sentiment analysis process. By synthesizing key findings, the abstract highlights the significance of incorporating advanced lexical resources like S-HAL in sentiment analysis frameworks to improve sentiment classification accuracy and robustness.abstract underscores the significance of SVM in effectively categorizing sentiments in textual data, contributing to advancements in sentiment analysis techniques.

2.6 Big data consumer analytics and the transformation of marketing

"Big data consumer analytics and the transformation of marketing" explores the profound impact of big data analytics on reshaping marketing strategies in the digital age. The abstract delves into the utilization of massive datasets generated from consumer interactions to derive actionable insights for targeted marketing campaigns. Through a comprehensive analysis, the paper highlights the transformative potential of big data analytics in understanding consumer behavior, preferences, and 5 trends. By synthesizing key findings, the abstract underscores the importance of leveraging advanced analytics techniques to optimize marketing efforts and enhance customer engagement in an increasingly data-driven landscape.

3. EXISTING SYSTEM

Number of reviews associated with a product or a brand is increasing at an alarming rate, which is no less than handling the big data. Classifying the reviews on the basis of sentiment of customers into positive and negative sentiment provides sentiment orientation of the review, hence results in better judgment.

DISADVANTAGES:

Limited Insight: Binary classification overlooks valuable feedback nuances, like specific suggestions or cultural context.

Subjectivity and Ambiguity: Sentiment analysis can vary among individuals and struggle with ambiguous language or sarcasm.

Neglect of Neutral Sentiment: Ignoring neutral feedback can lead to incomplete analysis and overlook valuable insights.
Scalability Challenges: Handling increasing review volumes may strain system efficiency and require more resources.
Lack of Contextual Understanding: Algorithms may struggle to grasp the full context of reviews, leading to misclassification and less accurate insights.

4. PROPOSED SYSTEM

Sentiment analysis is not only confined to the English language but has been implemented for various languages. Sentiment analysis of Chinese text by implementing four feature selection methods and five classifiers viz. Centroid classifier, K-nearest neighbor, Window classifier, Naive Bayes and SVM has been done Through this learning paradigm it was concluded that SVM outperforms all the other 9 learning methods in terms of sentiment classification. Sentiment analysis on travel reviews using three machine learning models namely, Naïve Bayes, SVM and character based N-gram model has been performed in which SVM and N-gram approaches have better performance than Naïve Bayes . It has been observed that in maximum number of cases SVM showcases best performance in comparison to other classification models.

ADVANTAGES

Multilingual Support: By implementing sentiment analysis for various languages, the proposed system caters to a diverse user base and allows businesses to analyze sentiment across different linguistic contexts, expanding its applicability and reach.
Feature Selection: The use of four feature selection methods enhances the accuracy and efficiency of sentiment analysis by identifying the most relevant features for classification, thereby improving the overall performance of the system.
Multiple Classification Models: The utilization of multiple classifiers such as Centroid classifier, K nearest neighbor, Window classifier, Naive Bayes, and SVM provides flexibility and allows for comparison to determine the most effective model for sentiment classification in different scenarios or datasets.
Performance Comparison: Through performance evaluation of different machine learning models, the proposed system enables businesses to identify the most suitable approach for sentiment analysis, ensuring optimal results and better decision-making based on the specific requirements and characteristics of the data.
Optimized Performance: The observation that SVM often outperforms other classification models suggests that the proposed system can achieve high accuracy and reliability in sentiment classification tasks, leading to more accurate insights and actionable recommendations for businesses.

5. UML DIAGRAMS

1. CLASS DIAGRAM

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application. Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of object oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. It is also known as a structural diagram. Class diagram contains • Classes • Interfaces • Dependency, generalization and association.

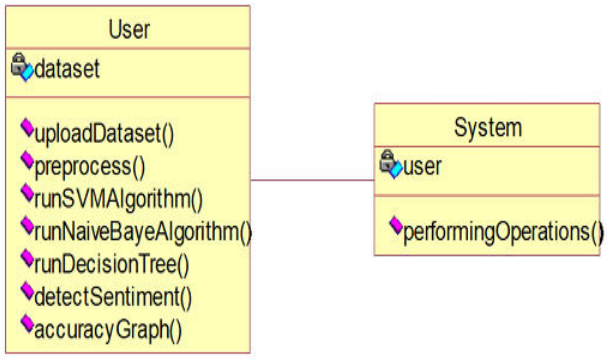

Fig 5.1 shows the class diagram of the project

2. USECASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

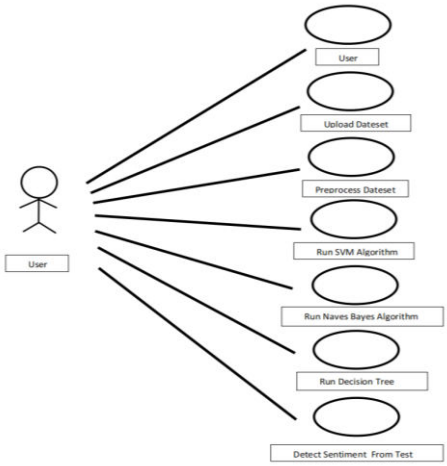

Fig 5.2 shows the Use case Diagram

3. SEQUENCE DIAGRAM:

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function. Sequence diagrams are used to formalize the behavior of the system and to visualize the communication among objects. These are useful for identifying additional objects that participate in the use cases. These diagrams are widely used by businessmen and software developers to document and understand requirements for new and existing systems.

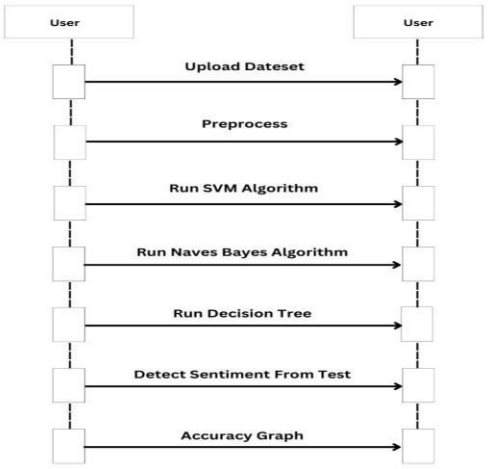

Fig 5.3 Shows the Sequence Diagram

6. RESULTS

6.1 Output Screens

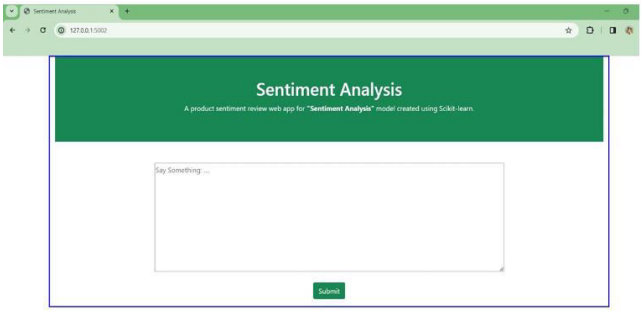

Fig 6.1 Home Page

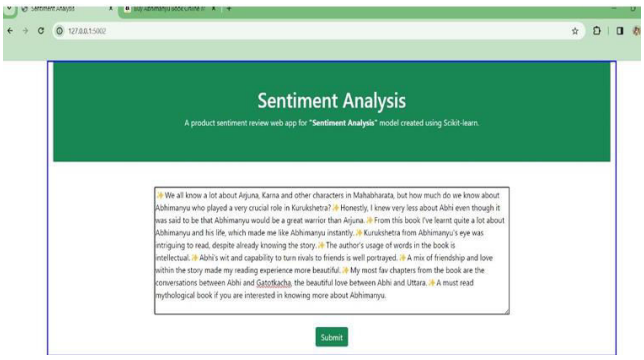

Fig 6.2 enter the text message

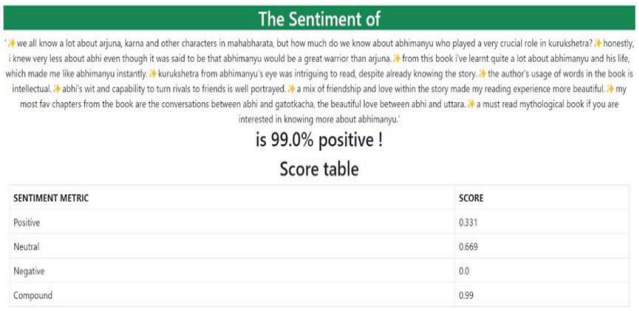

Fig 6.3 percentage of rating given from the customers

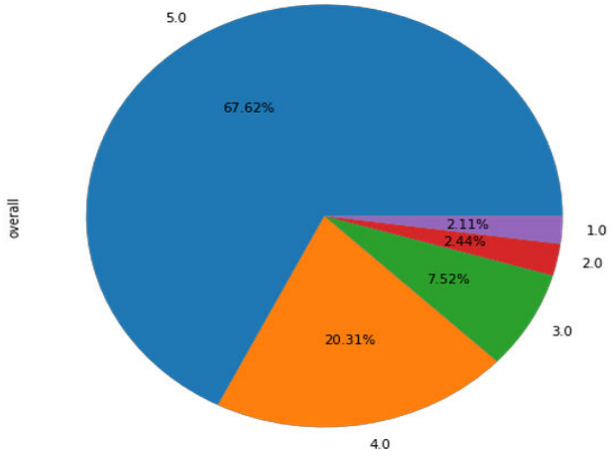

Fig 6.4 sentiment analysis in pie chart

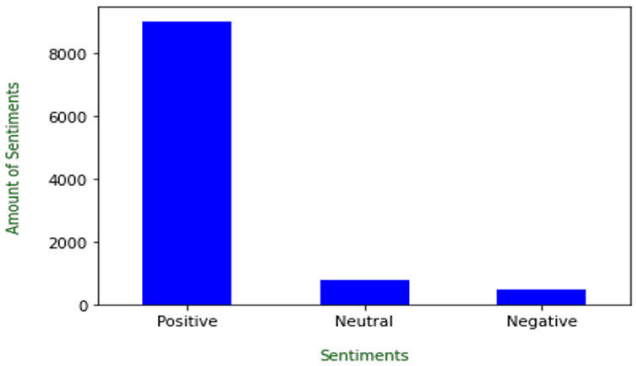

Fig 6.5 amounts of each sentiments based rating

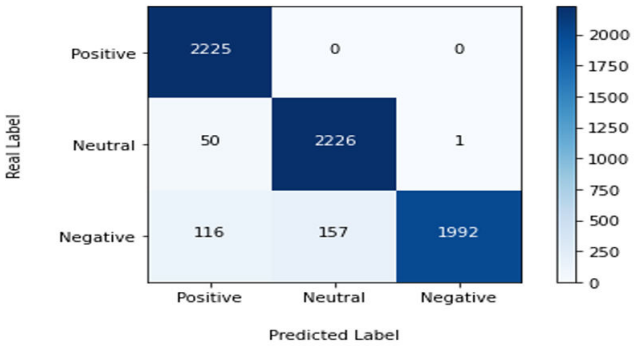

Fig 6.6 confusion matrix for sentiment analysis.

7. CONCLUSION

In conclusion, the surveyed literature spanning from 2016 to 2022 provides a comprehensive overview of sentiment analysis in product reviews using machine learning techniques. The studies highlight the evolution of sentiment analysis methodologies, from traditional machine learning algorithms to deep learning architectures. Key themes include the application of machine learning models for sentiment classification, aspect-based sentiment analysis, and the challenges associated with analyzing customer sentiments in product reviews. The surveyed papers underscore the importance of sentiment analysis in understanding customer feedback and guiding business decisions. They showcase the advancements in machine learning techniques and the growing interest in deep learning approaches for sentiment analysis tasks. Additionally, the papers identify key challenges such as noisy data, domain adaptation, and aspect-level sentiment analysis, suggesting avenues for future research. Overall, the literature survey provides valuable insights into the state-of-the-art techniques and methodologies for analyzing customer sentiments in product reviews. It serves as a foundation for further research in sentiment analysis and offers guidance for practitioners and researchers in the field.

FUTURE SCOPE

In future, the work can be extended to perform multiclass classification of reviews which will provide delineated nature of review to the consumer, hence better judgment of the product. It can also be used to predict rating of a product from the review. This will provide users with reliable rating because sometimes the rating received by the product and the sentiment of the review do not provide justice to each other. The proposed extension of work will be very beneficial for the e-commerce industry as it will augment user satisfaction and trust.

8. REFERENCES

1) Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 417-424.

2) Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. Proceedings of the 12th International Conference on World Wide Web (WWW), 519-528.

3) Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1-2), 1-135. 4) Q. Peng, and Y. Cheng, “Identifying the semantic orientation of terms using s-hal for sentiment analysis,” KnowledgeBased Systems, vol. 35, pp. 279–289, 2012.

5) Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies,5(1),1-167.

6) S. Erevelles, N. Fukawa, and L. Swayne, “Big data consumer analytics and the transformation of marketing,” Journal of Business Research, vol. 69, no. 2, pp. 897 904, 2016.

7) D. N. Devi, C. K. Kumar, and S. Prasad, “A feature based approach for sentiment analysis by using support vector machine,” in Advanced Computing (IACC), 2016 IEEE 6th Interna-tional Conference on. IEEE, 2016, pp. 3–8.

8) Anshumaan Chauhan (september 2021)Analysis of Machine Learning Algorithms and Feature Extraction Methods 10.1109/ICSES52305.2021.9633882. for Sentiment analysis doi 37

9) S. Sabba, N. Chekired, H. Katab, N.Chekkai and M. Chalbi, "Sentiment analysis for IMDb reviews using deep learning classifier," in 7th International Conference on Image and Signal Processing and their applications,2022,pp. 1-6,doi:10.1109/ISPA54004.2022.9786284.

10) 10) Sahithya Soundearajah(October 2022) Sentiment Analysis of ASOS Product Reviews Using Machine Learning Algorithms by Comparing Several Models. 2022 International Research Conference on Smart Computing and Systems Engineering (SCSE)