International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.48047/INTJECSE/V15I4.83 ISSN: 1308-5581 Vol 15, Issue 04 2023

SENTIMENT ANALYSIS OF CUSTOMER PRODUCT REVIEW

1 Mr K KRISHNA REDDY,
ASSISTANT PROFFESSOR, DEPARTMENT OF CSE,
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY,TELANGANA,INDIA,
Mail id :krishnareddy @sreyas.ac.in
2 BEJJANKI PRAGATHI,
DEPARTMENT OF CSE,
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY,TELANGANA,INDIA,
Mail id : bejjankipragathi@gmail.com
3 KANUGU SIDDHARTHA,
DEPARTMENT OF CSE,
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY, TELANGANA,INDIA,
Mail id: Luckysiddartha25@gmail.com
4 ALUKARAJAU PRAVEEN,
DEPARTMENT OF CSE,
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY, TELANGANA,INDIA,
Mail id : praveenalukaraju@gmail.com
5 BANDA MANIKANTA,
DEPARTMENT OF CSE,
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY, TELANGANA,INDIA,
Mail id : manikanta18511@gmail.com
Bandlaguda, Beside Indu Aranya, Nagole, Hyderabad-500068, Ranga Reddy Dist.

ABSTRACT

The importance of internet reviews in modern consumer communication and purchasing decisions is undeniable. E-commerce behemoths like Amazon, Flipkart, etc. provide customers a place to voice their opinions and give prospective purchasers honest feedback on how well a product works. This thing may be represented in the form of an occasion, a single blog entry, or a product experience.

In this study, we used an Amazon.com dataset that includes customer ratings and comments on digital cameras, computers, smart phones, tablets, televisions, videos, and surveillance systems. Classifying reviews into positive and negative sentiment is necessary for gaining useful insights from a big number of evaluations. Computing research into extracting subjective information from text is known as "sentiment analysis." This research finds that the Product Reviews may be effectively categorized using Machine Learning Techniques. In the study suggested, Sentiment Analysis is used to categorize over 4,000 evaluations into positive and negative categories. Naive Bayes, Support Vector Machine (SVM), and Decision Tree are just some of the categorization methods that have been used to organize reviews. 10Fold Cross Validation is used for the assessment of models.

INTRODUCTION

Sentiment analysis is a great tool for organizations to analyze how people socially connect with their brand, product, or service. It's a way for companies to gauge and interpret how consumers feel about their wares. To classify data, Naive Bayes is a supervised learning method. The use of natural language processing (NLP) and machine learning methods for analyzing the tone of product evaluations written by actual customers is quite useful. Sentiment analysis is the process of deducing the underlying emotional tone represented in customer evaluations and comments. Companies and organizations may use the results of this research to better understand their consumers and make more informed decisions about their products and marketing tactics. Sentiment analysis, often known as opinion mining, is a branch of natural language processing that seeks to mechanically identify the positive, negative, or neutral emotional tone of a piece of text. The goal of sentiment analysis in the context of consumer product evaluations is to identify whether the reviews are favorable, negative, or neutral. Sentiment analysis, often known as opinion mining, is a method for analyzing online comments to identify the overall tone of the comments. Performing sentiment analysis on textual data is a common practice for firms looking to track consumer feelings about a product or service, as well as better understand what those customers want. These days, it's also used to summarize user evaluations to assist people find what they're looking for while they're purchasing or watching a movie online. Its

primary use is in high-dimensional training datasets for text categorization. It makes predictions based on the object's likelihood since it is a probabilistic classifier. Naïve Bayes method is based on Bayes theory and used for solving classification issues. It is named Naïve because it believes that the existence of a specific trait is independent of the occurrence of other features. For instance, an apple may be recognized as such because it is red, has a round form, and is tasty. To better comprehend consumer feedback and make adjustments to goods and services, companies may use sentiment analysis of customer product evaluations. It makes use of natural language processing and machine learning to glean insights from large volumes of unstructured text data.

A supervised learning method, Support Vector Machine (SVM) is put to use in classification and regression analysis. Classification applications benefit more from it, although regression may also make good use of it. In order to divide the information into distinct categories, SVM locates a hyper-plane. In 2-dimensional space, this hyper-plane is nothing but a line. The N-dimensional space used in SVM corresponds to the number of features or characteristics in the dataset. After that, we'll determine which hyper plane best divides the information. As a result, you should now realize that, by design, SVM can only do binary classification.

Related Work

The reviews of movies have been classified according to their tone using a hybrid categorization method. Integration of diverse feature sets and classification methods such as Naïve Bayes, Genetic algorithm has been carried out to examine performance on the basis of accuracy. Compared to the basic classifier, the hybrid NB-GA is more effective and efficient, and when comparing NB and GA, GA is the more effective of the two. Document polarity is also a crucial factor in text mining. The use of tree kernels in the engineering of the future has been considered. When compared to other methods, this one produces superior results. The paper's author defines two types of categorization schemes—a two-way scheme and a three-way scheme. In 2-way classification, sentiments are categorized into either positive or negative while in 3-way classification, sentiments are classified into positive, negative or natural. In the tree kernel technique, the author explores a tree-based representation of tweets. The most accurate and best-performing model was a tree kernel-based one. The experimental method improved performance by 4% compared to the unigram model. Cascading classification may be performed using a hierarchical method to sentiment analysis. The author used a hierarchical structure, with three levels: objective, subjective, polar, and non-polar.

LITERATURE SURVEY

An important aspect of any research effort, including those involving the sentiment analysis of consumer product evaluations, is a literature survey, often known as a literature review. Academic and professional literature is surveyed and summarized for new ideas, knowledge gaps, and an appraisal of the present status of the area.

- In the context of product reviews written by customers, please explain the significance of sentiment analysis.
- Provide an overview of your literature review's goals and scope.
- Please detail the resources and databases you utilized to conduct your search.
- Describe the terms and phrases you used to find the research and articles you needed.
- Talk about the criteria you used to decide which papers to include and which to leave out.

Please detail the resources and databases you utilized to conduct your search. Describe the progress of sentiment analysis across time, from rule-based methods to current-day machine learning methodologies.

Focus on significant achievements and seminal works.

Explain the theoretical foundations of sentiment analysis, including polarity detection, subjectivity analysis, and feature-based sentiment analysis. Describe the techniques and strategies used for classifying sentiments in sentiment analysis. Understanding the current research, methodology, obstacles, and breakthroughs in the field of sentiment analysis of consumer product evaluations is the primary goal of a literature assessment. The purpose of this literature review is to offer a thorough overview of the current research environment, approaches, problems, and developments in the subject of sentiment analysis of consumer product evaluations. This survey may be used by researchers, practitioners, and corporations to learn more about the field and get ideas for future projects. The use of opinion and emotion-describing adjectives and adverbs in a sentence might seem like a good indicator of whether or not it has a positive or negative tone, but there are many other factors to consider, such as the phrase's or word's context, which can give different meanings using similar words.

There are many different techniques and models that may be utilized with various Algorithms for feature extraction, training, and classification. In this chapter we quickly examine some of the related work which employed these algorithms and experimented on various forms of training data, examining the strengths and weaknesses of each

DOI:10.48047/INTJECSE/V15I4.83 ISSN: 1308-5581 Vol 15, Issue 04 2023

word. In order to guarantee that your study is founded on the most up-to-date discoveries and best practices in the area of sentiment analysis, doing a well-structured literature review is essential.

EXISTING SYSTEM

With the use of cutting-edge machine learning and NLP methods, the current system for analyzing consumer sentiment in product evaluations has grown more complex. It helps organizations better comprehend and react to consumer feedback and, in turn, enhance their goods and services.

You may avoid starting from scratch when it comes to creating a system for sentiment analysis by using already tools like VADER (Value Aware Dictionary and Sentiment Reasoned). Choose one of these options and incorporate it into your client product review platform based on the needs of your project and the degree of customization you desire.

VADER (Valence Aware Dictionary and sentiment Reasoned) is an application developed for assessing the emotional tone of textual material using a combination of a vocabulary and rule-based sentiment analysis. Researchers C.J. Hutto and Eric Gilbert created VADER to efficiently handle sentiment analysis tasks, and it has since found widespread usage in natural language processing and sentiment analysis applications. VADER employs a sentiment lexicon that rates the positive or negative emotional tone of each English word. Words may be given ratings ranging from 0 (very negative) to 100 (extremely positive) to show how they make the reader feel. The current study also discovers textual polarity. A positive, negative, or neutral polarity will be shown. Word cloud was used to find novel words that were then given polarities in this study. Each word is compared to a polarity word file, and the polarity of the text file is tallied, much as in sentiment analysis. Sentiment package allows for the examination of six distinct emotions: anger, contempt, fear, pleasure, sorrow, and surprise. By employing word cloud frequently recurring words were recorded. These common expressions were given an emotional overtone. The updated sentiment file includes these newly-added phrases and emotions. At the moment, we're using the bayes algorithm. Sentiment analysis method checks each word with terms in sentiment file and assigns count for each sentiment.

PROPOSED SYSTEM

The suggested solution integrates state-of-the-art natural language processing (NLP) methods with real-time analysis, user-friendly interfaces, and the rest of the company's operations to provide actionable insights into consumer attitudes and facilitate data-driven decision making. Outlining the architecture and components of the system to fulfill your unique aims and needs is an important part of designing a suggested system for sentiment analysis of consumer product evaluations. A suggested system for sentiment analysis of consumer product evaluations intends to expand the current methodologies and technology, making the process more efficient, accurate, and insightful. Connect too many data sets, such as user profiles, online stores, and review sites. Gather data in real time using web scraping and application programming interfaces. Extra information may be gleaned by collecting metadata like user demographics and review context.

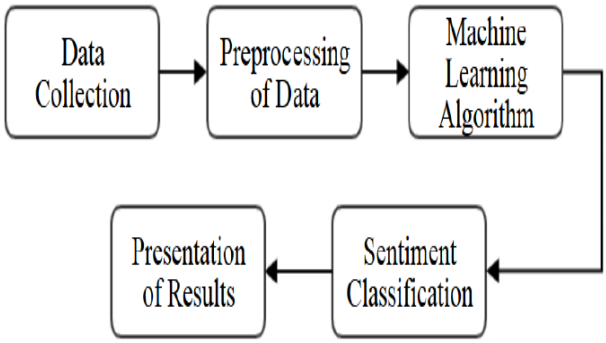

Fig. 1 Proposed system working

Advantages

Machine learning models can handle massive amounts of customer evaluations fast and effectively, making it feasible to examine a tremendous quantity of data in a short period of time. For companies who get a lot of client feedback, this is invaluable. Algorithms based on machine learning are more reliable than human reviewers since they provide consistent findings and don't introduce bias into the process. This guarantees that all reviews are reviewed in a consistent way. Using machine learning models, organizations can monitor client feedback in real

time and act swiftly in response to any concerns that may arise. The result may be happier and more loyal customers.

Customer feedback from different locations and marketplaces may be analyzed with the use of machine learning models that can be taught to do sentiment analysis in numerous languages.

Machine learning can analyze client feedback over time to spot patterns and trends that might inform company choices like product iterations and advertising campaigns.

SYSTEM DESIGN

IMPORTANCE OF DESIGN

In order to solve the issue detailed in the requirements document, the design phase is carried out. Software architecture is the process of defining your code's methods, functions, objects, and general structure and interaction so that the final functionality may meet the needs of your users. The proper abstraction can help you better grasp the requirements and provide the desired results. This eliminates unnecessary steps and maximizes reusability. The transition from the issue domain to the solution domain begins at this stage. In other words, design leads us to how to meet requirements once we start with what is required. Perhaps more than any other single stage, the design of a system has a profound effect on the final product's quality (through testing and into the maintenance phase). The design document is the product of this stage. This paper serves as a guide for the solution's installation, testing, and upkeep phases. System Design and Detailed Design are two common sub-phases of the overall design process.

System Design, often known as top-level design, is the process of determining what components should make up a system, what those components should be, and how they should work together to achieve the intended outcomes. The system's modules' internal logic and requirements are finalized during the detailed design phase. At this stage, specifications for the data are often written in a high-level design description language that is not tied to any particular programming language. While module identification is a primary concern throughout system design, detailed design is where each module's logic is developed. Developers fill in the blanks between the requirements specification (the result of requirements elicitation and analysis) and the final product (the delivered system) during system design. Product design, graphic design, web design, architecture, and many more areas rely heavily on the skills of designers. Several factors highlight why design is so crucial:

Appeal to One's Sense of Beauty is a Common Goal in Design. Products, environments, websites, and visuals that have been thoughtfully created provide a good first impression because they are appealing to the eye.

Customer Service: User interface and experience design has a major effect on how people utilize technology in the digital world. Intuitive and user-friendly design may improve productivity and satisfaction for every given user.

Usability: Design isn't just about looks; it's about functionality and issue solving. Products and systems with a functional design are more likely to be effective, user-friendly, and useful. Design is essential in creating and maintaining a distinct identity for a brand. Logos, colors, and typefaces that are used consistently across all brand materials help build consumer recognition and loyalty.

Design as a Means of Expressive Communication. Whether it's via the design of a logo, a piece of art, or an info graphic, design may transmit messages, emotions, and information in a visually attractive manner.

Architecture of Systems

A system architecture diagram would be used to depict the interconnections between the various parts of the system. In their most basic form, they serve to depict the interplay between components of a system, both physical and virtual. On the other hand, it may be made for use in online apps as well. Developing a system architectural diagram for sentiment analysis is not always a simple undertaking, and the final product might look quite different depending on the resources at hand and the desired outcomes.

Complexity, new parts, and interdependencies among modules are all possible in a real-world setting. This schematic depicts the overall framework for sentiment analysis.

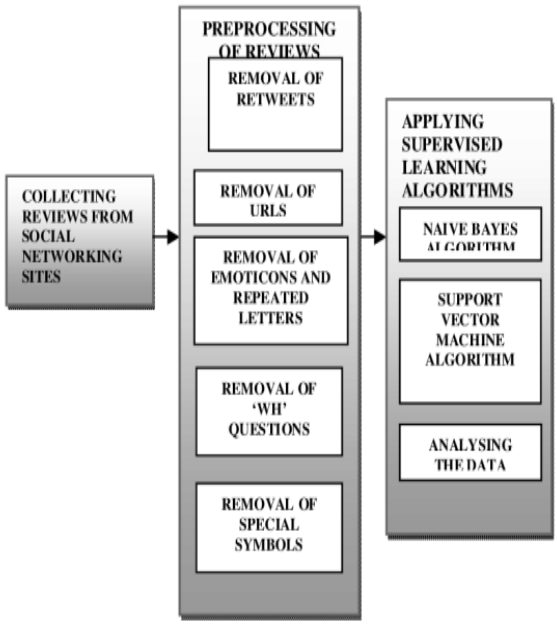

Fig.2 Architecture for Sentiment Analysis

FUNCTIONAL REQUIREMENTS

Hardware Requirements

- Processor (CPU) – Multi-Core processor
- Memory (RAM) –minimum 8 GB
- Hard Disk – 40 GB

Software Requirements

- Programming Language – Python
- Operating System – Windows/Linux

TESTING

IMPORTANCE

Errors in software can only be found via testing. It's the most important metric for gauging worker quality throughout the coding process. The purpose of testing is to ensure that the software behaves as intended by running it with a collection of test cases and analyzing its results. The software testing process is the last look at the specifications, designs, and codes that make up the program. The rising profile of software as a system component and the consequential monetary penalties of a software failure are driving forces behind our deliberate, methodical approach to testing. The goal of testing is to uncover bugs in software by running it repeatedly. Software and other engineered product testing may be just as difficult to develop as the original product design.

TYPES OF TESTING

The software development process employs many tiers of testing procedures to ensure that the final product is free of bugs.

Parts Checking

Individual modules are tested in isolation as computer code. It is limited solely by the specifications of the creator. The following methods are available for testing each module. 47

Contrast-Based Testing

In this method certain test cases are produced as input conditions that completely execute all functional requirements for the program. In order to identify problems in the following areas, testing has been applied:

Faulty or absent features

- Difficulties in data access or organization

- Mistakes in performance
- Errors during start-up and shut-down

IMPLEMENTATION

MODULE DESCRIPTION

The "CustomerReviewSentimentAnalyzer" component was developed to do sentiment analysis on product reviews written by customers using machine learning methods. Businesses may get significant insights into consumer satisfaction and product performance by automatically categorizing and determining the emotion represented in text-based product evaluations (positive, negative, or neutral). The implementation stage is essential since it is when abstract concepts are translated into usable products. To make sure the final product or solution is in line with the original vision and goals, it is common practice to plan, monitor, and perform quality assurance. Effective implementation is vital for attaining success in numerous sectors and contexts.

Naive Bayes is a popular and straightforward classification technique that uses Bayes' theorem, a cornerstone of probability theory, to make predictions. It's frequently used in machine learning for text classification tasks including spam detection, sentiment analysis, and document categorization. The term "naive" comes from the classification method's underlying assumption that the characteristics being utilized are conditionally independent, which is usually an oversimplification but nevertheless produces decent results in many real-world contexts.

SAMPLE CODE

1.Initializing the Libraries And Modules

```
import numpy as np

import pandas as pd

from sklearn.feature_extraction.text import
TfidfVectorizer

from sklearn.model_selection import train_test_split
```

From sklearn.naive_bayes import MultinomialNB

From sklearn.svm import SVC

RESULTS

Customer product reviews that have undergone sentiment analysis are sorted into three broad categories: positive, negative, and neutral. In most cases, the analysis will consist of:

Finding reviews that are good about the product or service and full of praise or praiseworthy experiences. The positives and advantages are often highlighted in these analyses.

Recognizing evaluations that express disappointment, criticism, or a poor experience. These may draw attention to problems with the product or suggest ways to enhance it.

Reviews with a neutral sentiment are those that don't lean much in either direction emotionally and instead focus on the facts.

Comparison:

Both the Naive Bayes and the support vector machine do well in sentiment analysis. However, performance might shift depending on the details of the dataset and the precision with which features are extracted.

Naive Bayes is a basic method that may get accurate results when working with tiny datasets and when the independence assumption is not severely broken. However, SVM can handle non-linear separations, despite its complexity and potential need for further adjustment.

Scalability: Naive Bayes may be quicker and needs less data for training compared to SVM. Still, SVM can handle high-dimensional data and may provide superior generalization.

Naive Bayes Model Accuracy: 78.11%

	precision	recall	f1-score	support
0	0.70	0.88	0.78	88
1	0.88	0.71	0.78	113
accuracy			0.78	201
macro avg	0.79	0.79	0.78	201
weighted avg	0.80	0.78	0.78	201

Fig.3. Classification Report of Naïve Bayes

SVM Model Accuracy: 81.59%

	precision	recall	f1-score	support
0	0.75	0.88	0.81	88
1	0.89	0.77	0.82	113
accuracy			0.82	201
macro avg	0.82	0.82	0.82	201
weighted avg	0.83	0.82	0.82	201

Fig.4 Classification Report of Support Vector Machine

```
[14] new_review = "This product is great! I highly recommend it."
     predicted_sentiment_nb = predict_sentiment(new_review, tfidf_vectorizer, nb_classifier)
     predicted_sentiment_svm = predict_sentiment(new_review, tfidf_vectorizer, svm_classifier)

     print("Naive Bayes Predicted Sentiment:", predicted_sentiment_nb)
     print("SVM Predicted Sentiment:", predicted_sentiment_svm)

     Naive Bayes Predicted Sentiment: 1
     SVM Predicted Sentiment: 1
```

Fig 5 Output of New Review Using Naïve Bayes and SVM

Since the Support Vector Machine (SVM) machine learning algorithm is more accurate than the Nave Bayes machine learning algorithm, we will not be employing both of them. Whether you choose with Naive Bayes or SVM for sentiment analysis may come down to the details of your dataset, the power of your computer, and the trade-off you're willing to make between precision and 56 computing complexity. The relative merits of the two approaches may often be seen by applying them to the provided dataset. Companies may immediately understand the distribution of positive, negative, and neutral attitudes when the data are displayed using graphs, charts, or summaries. Insights into the strengths and shortcomings of offers may be gained by analyzing patterns over time and comparing sentiment across various goods or services, which can then be used to inform decisions about changes, marketing tactics, and customer service enhancements.

Ultimately, the trade-offs between speed, interpretability, and accuracy that matter most for your application should guide your decision between Naive Bayes and SVM.

Fig.6 Classification of New Review Using SVM

If the dataset is large or complicated, or if the relationships between features (words) in the data are not well captured by the simple independence assumption of Naive Bayes, Support Vector Machines (SVM) often provide higher accuracy in sentiment analysis than Naive Bayes. By modeling non-linear correlations and adjusting to the specifics of the data, SVM is a more robust and versatile classification system. It is important to weigh the trade-offs between accuracy and computing complexity when selecting an algorithm for sentiment analysis. Both SVM and Naive Bayes may benefit greatly from fine-tuning of hyper parameters and feature extraction techniques.

CONCLUSION

The study of how individuals feel about things (their "sentiments") is known as "sentiment analysis" or "opinion mining." In this study, we address a fundamental challenge in sentiment analysis: how to properly classify the polarity of emotions. The data utilized in this analysis comes from customers' reviews posted on Amazon.com. Figure 2 depicts a suggested method for classifying sentiments according to their polarity, along with thorough explanations of each stage. Both sentence-level and review-level classification experiments have been conducted. Sentiment analysis models will play a crucial role in assisting organizations in making data-driven choices, enhancing product and service quality, and boosting customer happiness as the world becomes more interconnected and interdependent. With the use of sentiment analysis, companies will be able to better understand their customers' feelings and respond to their needs by tailoring their products and services. Businesses may get a lot of insight into their customers' thoughts and feelings by using sentiment analysis powered by machine learning. Companies' approaches to customer happiness, product development, and market strategies are evolving as a result of machine learning's improved ability to gauge sentiment from textual data. There will be much more progress and influence in the future of sentiment analysis using machine learning.

In conclusion, machine-learning-based sentiment analysis is a robust and rapidly developing technology with several present and future uses. Automation, scalability, real-time analytics, and the capacity to extract useful information from consumer feedback are just some of the benefits it provides to companies and organizations. It's anticipated that sentiment analysis will grow in breadth as it becomes more industry-specific, incorporates multimodal data sources, and improves its ability to comprehend emotions, context, and fine-grained analysis. Integrating sentiment analysis with other AI technologies like chat bots and virtual assistants will result in more tailored and efficient consumer interactions and the ethical issues around sentiment analysis will get greater attention. As technology evolves, it's possible that sentiment analysis will expand to incorporate the analysis of multimodal data and focus on increasingly niche, industry-specific applications. Businesses will be able to make better choices based on deeper insights into consumer sentiment made possible by emotion recognition, context comprehension, and fine-grained analysis.

REFERENCE

[1] Samaneh Moghaddam and Martin Ester, "Opinion Digger: An Unsupervised Opinion Miner from Unstructured Product Reviews", Proceedings of 19th ACM International Conference on Information and Knowledge Management, pp. 1825-1828, 2010.

[2] Aurangzeb Khan, Baharum Baharudin and Khairullah Khan, "Sentiment Classification from Online Customer Reviews using Lexical Contextual Sentence Structure", Proceedings of International Conference on Software Engineering and Computer Systems, pp. 317-331, 2011.

[3] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews", Proceedings of 10th ACM International on Knowledge Discovery and Data Mining, pp. 166-177, 2005.

[4] A.M. Popescu and O. Etzioni, "Extracting Product Features and Opinions from Reviews", Proceedings of International Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 339-346, 2005.

[5] G. Vinodhini and R.M. Chandrasekaram, "Sentiment Analysis and opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, No. 6, pp. 28-35, 2012.

[6] A. Collomb, C. Costea, D. Joyeux, O. Hasan and L. Brunie, "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation", Available at: https://liris.cnrs.fr/Documents/Liris-6508.pdf.