RESTAURANT REVIEW USING SENTIMENT ANALYSIS IN SOCIAL MEDIA

A. Prema[1], S. Devi Durga[2], S. Gayathri[3], S. Harani Shree[4], R. Siva Ranjani[5]

[1]Assistant Professor, PG & Research Department of Computer Science, Sri Meenakshi Govt Arts College for Women(A), Madurai.

[2,3,4,5] M.Sc., PG & Research Department of Computer Science, Sri Meenakshi Govt Arts College for Women(A), Madurai.

ABSTRACT Customer review sentiment research has a significant influence on a company's development plan. In the last ten years, the internet's development has generated vast amounts of data across all industries. There are presently a lot of reviews about the restaurant sector in online media, such as Twitter, Google reviews, etc., which have been growing daily due to customer actions. This industry needs to put more of an emphasis on customers by consistently improving customer service. Customer service begins with ensuring customer satisfaction. When classifying documents, sentiment analysis helps define whether they represent negative or positive opinions. Manually analysing the data from social media evaluations was nearly impossible. Different data mining algorithms are utilised for assessment and analysis. The Nave Bayes (NB), Random Forest, Support Vector Machine, and K-Nearest Neighbour (KNN) sentiment classification algorithms are used in this study to analyse the sentiment of users' tweets about restaurants. The suggested system has the ability to analyse tweet sentiment. For consumer reviews, a sample of actual data was therefore taken from Twitter. The proposed system has been implemented using the Python programming language. The findings are consistent with the idea that patrons have positive opinions of both restaurants and other people, particularly those who own them. Accuracy and error rate are two popular performance evaluation metrics used to gauge the results.

KEY WORDS: Customer satisfaction, Naïve Bayes, Sentiment analysis. Support Vector Machine, Random Forest Classifier, K-Nearest Neighbour.

1.INTRODUCTION

Sentiment analysis is the method for determining if a piece of text is partisan, neutral, positive, or negative. This text could be a tweet, a review of a book, movie, restaurant, or something else entirely. The sentiment analysis, commonly referred to as opinion mining, analyses people's opinions, assessments, feelings, and attitudes about a subject, person, or thing. The expressions can be categorised as either neutral, negative, or positive. An illustration. The phrase "I really liked the garlic noodles of your restaurant" is complimentary. The evaluations' overall sentiment polarity demonstrates a preference for service, which may encourage customers to "self-select" the meal they prefer. It is simple to acquire product reviews from a website and comprehend what customers are actually expressing as well as their sentiment with regard to a certain product thanks to natural language processing in artificial intelligence applications. Businesses that receive a lot of reviews can actually decipher them and utilise the information gathered to suggest new goods or services in accordance with client preferences. It is frequently used to determine how a customer or consumer feels about key components of a brand, such as tone, context, emotion, and so on. Reviews of this nature are crucial for improving the service and benefiting customers. From the consumer's point of view, hearing what other customers have to say about a service might help him form a general impression of the offering. Owners or service providers, on the other hand, utilise sentiment analysis to assess the acceptance of their products or to examine consumer feedback and satisfaction. However, as we can imagine, it takes a lot of time and effort to read through all of those reviews and personally analyse the tone of the contents. The general conclusion on the opinions and perspectives can be

quickly ascertained via sentiment analysis. It helps the business owners a better understanding of their customers as well as how they compare to the companies of their rivals. Then, three data mining classification algorithms—Support Vector Machine, Rule Based, and Random Forest—are employed in this study to classify the data and discover which one has the highest accuracy. Customer reviews are still written in text form; therefore restaurant reviews fall within the text mining area. The outcomes of this data will be categorised into two categories: positive or negative. Python was used to pre-process review data, such as removing stop words and punctuation.

2. LITERATURE SURVEY

[1] The potential issues with machine-based sentiment analysis are studied, recognised by their knowledge extraction tasks, and handled. [2] Machine learning techniques are used to derive emotions from social network data, revealing the ideas and experiences shared by substance users. [3] Sentiment analysis is carried out by employing quote phrases to apply an existing annotated corpus. [4] Sentiment analysis may be used in conjunction with supervised machine learning to detect phoney positive or negative reviews. The SVM algorithm is capable of false feedback in addition to passing text detection. [5] For such excellent research, this emotive analysis contrasts and discusses the favourite products that make consumers happy and examines unsupervised learning methods.

3. METHODOLOGY

Mentioned in the background section, we created a sentiment analysis system using the conventional machine learning methodology.

OBJECTIVE

This project focuses on determining the polarity of the emotion a sentence evokes using an input box. to put in place a system that automatically categorizes material as good, negative, or neutral. Sentiment analysis is used to evaluate whether the general public has a good, negative, or neutral view towards the topic at hand. A pie chart is used to depict it.

CONCERN STATEMENT

To create a method for classifying customer reviews based on their sentiment, which might be useful for analysing data where opinions are largely unstructured and either positive or negative.

EXISTING SYSTEM

The number of user-generated material on social media platforms like Facebook, Twitter, review sites, etc. is rapidly increasing. These viewpoints may be collected and used as corporate intelligence for a range of purposes, including marketing and forecasting, among others. In most cases, sentiment analysis is used to assess an author's knowledge on a certain subject. Sentiment analysis, however, is not yet used in our social networking platforms. The sentiment analysis in certain surveys is based on a static sent word dataset. But in order to determine the polarity of the microblogs, a suitable solution must be found.

PROPOSED SYSTEM

Through the text field, we will capture the unstructured data. The data is handled as follows after being converted to lower case.

Pre-processing

The review text goes through a pre-processing stage where the following actions are conducted on the review text before the feature extractor can utilize the reviews to construct feature vectors. With more information provided, these stages transform the review's plain language into parts that may be processed and used by feature extractors. Third-party technologies that were designed to handle the particulars of review text were employed for each of these processes. An integrated development environment (IDE) for computer programming is called PyCharm. It was made by the Czech company JetBrains. Code analysis, a graphical debugger, and an integrated unit testing are all included.

Step 1: Tokenization

Tokenization is the act of breaking down text as a string into smaller, more manageable pieces known as tokens. These components can be words, emojis, URL links, hashtags, or punctuation when used in a review. " An insanely amazing.... Text was divided into "an", "insanely", and "awsum". Frequently, a space is used to separate these components. However, as a term used as a hash tag could have a different sentimental meaning than a word used often in the text, hash tags with the character "#" before the tag must be preserved.

Step 2: Tags for Parts of Speech The features of a word in a sentence based on grammatical categories of words are known as parts of speech (POS) tags. This data is necessary for sentiment analysis since the sentiment value of words might vary based on their POS tag. For instance, the word "good" as an adjective has a positive mood but the word "good" as a noun does not. A POS is given to each token retrieved in the last stage.

Step 3: Dependency Parsing, Extracting the relationship between words in a phrase is what dependency parsing serves for in this context. This is helpful for determining the link between "not" and "good" in expressions like "not really good," because the relationship is not necessarily with the adjacent word.

Table1. Class and Reviews

Fig.1 Sample Reviews for sentiment class

CLASS	REVIEWS
Positive	A Wonderful Experience so long ago. incredible experience servicing the elite
Negative	The worst hospitality, the slowest service, the worst experience...
Neutral	My initial visit was fantastic, but with time it evolved to be just okay. Quantity reduction and, to some extent, quality reduction

4. CLASSIFICATION ALGORITHM

In this work, four classifiers namely random forest, support vector machine, k-nearest neighbour and are applied for the pre-processed data set. Among these classifiers, random forest works better for our data set.

4.1 RANDOM FOREST

The preferred algorithm for machine learning A component of the supervised learning approach is the random forest. It may be applied to classification and regression issues in machine learning. It is based on the idea of ensemble learning, which is the practise of integrating many classifiers to address a complicated issue and enhance the model's performance. In order to enhance the displayed accuracy of the dataset, the random forest classifier averages the results from many decision trees applied to various subsets of the input dataset.

4.2 MULTINOMINAL NAVIE BAYES

A technique to Bayesian learning that is well-liked in Natural Language Processing (NLP) is the Multinomial Naive Bayes algorithm. Using the Bayes principle, the programme makes an educated prediction about the tag of a text, such as an email or news article. It determines the likelihood of each tag for a particular sample and outputs the tag with the highest likelihood. Each feature being classified by the Naive Bayes classifier is distinct from every other feature, which unites the various algorithms that make up the classifier. The inclusion or removal of one characteristic does not depend on the presence or absence of another feature.

4.3 SUPPORT VECTOR MACHINE

One of the most popular supervised learning algorithms is called the Support Vector Machine, or SVM, and it may be applied to both classification and regression issues. However, it is mainly used in machine learning to solve categorization issues. The SVM algorithm's objective is to establish the optimal line or decision boundary that can divide n-dimensional space into classes, allowing us to quickly classify fresh data points in the future. The name of this best-choice boundary is a hyperplane. SVM selects the extreme vectors or points that aid in the creation of the hyperplane. The method is recognised as a support vector machine because of the term "support vectors," which refers to these extreme situations.

4.4 K-NEAREST NEIGHBOUR

One of the simplest machine learning algorithms based on the supervised learning method is K-Nearest Neighbour. The K-NN method places the new instance into the category that is most similar to the available categories based on the assumption that the new case or data and the available cases are comparable. When a new data point is added, the K-NN algorithm classifies it based on how similar the existing data is to it and saves it all. This implies that using the K-NN method, fresh data may be quickly and accurately categorised into a suitable category. Although the K-NN technique may be applied to classification and regression problems, it is most frequently used for classification issues. When the K-NN technique is non-parametric, it makes no assumptions about the underlying data.

5. PERFORMANCE EVALUATION

To analyse the performance of our algorithms the following two metrics are used:

- **Accuracy**

- **Error rate**

Accuracy

Simply put, accuracy measures how often the classifier makes accurate predictions. By dividing the total number of forecasts by the percentage of accurate predictions, accuracy is calculated.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Error Rate

The term "error rate" refers to a measurement of the magnitude of a model's prediction error with regard to the actual model. Accuracy and error rate are directly connected.

Error rate = 1.0 – accuracy

Table 2 shows the accuracy and error rate achieved using the proposed algorithm. The Multinominal Nave Bayes algorithm achieves higher accuracy of 90%, and lower error rate of 10%.

ALGORITHM	ACCURACY	ERROR RATE
RANDOM FOREST	88.25%	11.75
MULTINOMINAL NB	90%	10
SVM	75.51%	24.49
KNN	81.31%	18.69

Table2. Analysis of Accuracy and Error Rate for Classifiers
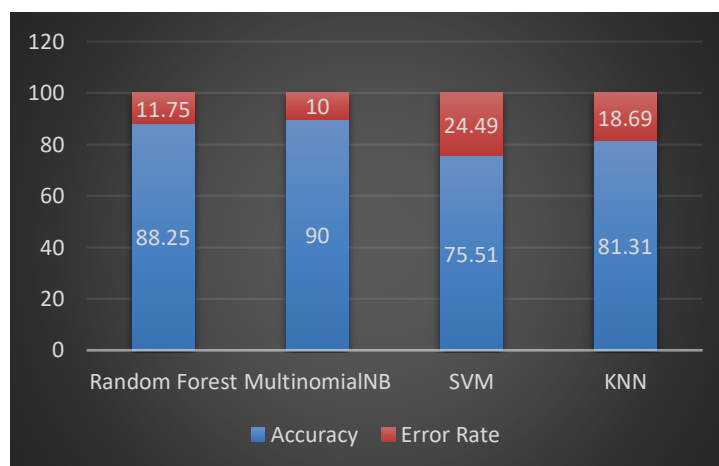

Fig.2 Accuracy and Error Rate

6. CONCLUSION

The paper examines crucial methods for determining reviews' emotional content. In this research, many techniques are used to determine the review text's sentiment analysis. We compare all of the currently used procedures and cover the most ground possible. Tokenization, stop word and punctuation removal, and stemming are all done using NLP. Additionally, use the Naive multinomial, Support Vector Machine, Random Forest and K-nearest neighbour method for classification, which is a supervised machine learning approach.

7. REFERENCE

[1] Aydoğan, Ebru, and M. Ali Akcayol. "A comprehensive survey for sentiment analysis tasks using machine learning techniques." In *2016 international symposium on innovations in intelligent systems and applications (INISTA)*, pp. 1-7. IEEE, 2016.

[2] Raza, Hassan, M. Faizan, Ahsan Hamza, Mushtaq Ahmed, and Naeem Akhtar. "Scientific text sentiment analysis using machine learning techniques." *International Journal of Advanced Computer Science and Applications* 10, no. 12 (2019).

[3] Isah, Haruna, Daniel Neagu, and Paul Trundle. "Bipartite network model for inferring hidden ties in crime data." In *Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015*, pp. 994-1001. 2015.

[4] Naithani, Kanchan, and Yadav Prasad Raiwani. "Realization of natural language processing and machine learning approaches for text-based sentiment analysis." *Expert Systems* (2022): e13114.

[5] Yang, Christopher C., Xuning Tang, Y. C. Wong, and Chih-Ping Wei. "Understanding online consumer review opinions with sentiment analysis using machine learning." *Pacific Asia Journal of the Association for Information Systems* 2, no. 3 (2010): 7.

[6] M. Govindarajan, "Sentiment Analysis of Restaurant Reviews Using Hybrid Classification Model," IRF International Conference, 2014.

[7] Y. Woldemariam, "Sentiment analysis in a cross-media analysis framework," Hangzhou, 2016.

[8] A.B. Pawar, M.A. Jawale and D.N. Kyatanavar, "Fundamentals of Sentiment Analysis: Concepts and Methodology", An Environment of Computation Intelligence, Springer International Publishing Switzerland 2016, pp. 25-35.

[9] Sharif, Omar, Mohammed Moshiul Hoque, and Eftekhar Hossain. "Sentiment analysis of Bengali texts on online restaurant reviews using multinomial Naïve Bayes." *2019 1st* international conference on advances in science, engineering and robotics technology (ICASERT). IEEE, 2019.

[10] Krishna, Akshay, et al. "Sentiment analysis of restaurant reviews using machine learning techniques." Emerging Research in Electronics, Computer Science and Technology: Proceedings of International Conference, *ICERECT 2018*. Springer Singapore, 2019.

[11] Zahoor, Kanwal, Narmeen Zakaria Bawany, and Soomaiya Hamid. "Sentiment analysis and classification of restaurant reviews using machine learning." 2020 21st International Arab Conference on Information Technology *(ACIT)*. IEEE, 2020.