IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Customer Opinion Mining for Restaurant Reviews using Sentiment Analysis and Natural Language Processing

Anitha T M[1], Dr Seshaiah Merikapudi[2]

[1]Post Graduation Student Master of Technology, Department of Computer Science and Engineering, SJC Institute of Technology, Chikkaballapur, Karnataka, India

[2]Professor, Department of Computer Science and Engineering, SJC Institute of Technology, Chikkaballapur, Karnataka, India

Abstract

The rapid growth of online restaurant review platforms has significantly influenced customer decision-making in the food and hospitality industry. Every day, thousands of customers share their dining experiences through online platforms such as Google Reviews, Yelp, TripAdvisor, Zomato, and Swiggy. These reviews provide valuable insights into various aspects of restaurant services, including food quality, service, ambience, cleanliness, pricing, and customer satisfaction. However, manually analyzing such a large volume of textual feedback is both time-consuming and inefficient. Natural Language Processing (NLP) and Sentiment Analysis provide intelligent approaches to automatically analyze customer opinions and classify reviews based on their sentiments.

This research proposes an automated customer opinion mining framework for restaurant reviews using Natural Language Processing and Machine Learning techniques. Initially, customer reviews are collected from publicly available restaurant review datasets. The collected textual data undergoes preprocessing steps including lowercase conversion, punctuation removal, stop-word elimination, tokenization, stemming, and lemmatization to improve data quality..

Index Term: Natural Language Processing (NLP), Opinion Mining, Sentiment Analysis, Restaurant Reviews, Machine Learning, Naïve Bayes, TF-IDF, Customer Feedback, Text Mining.

I. INTRODUCTION

The restaurant industry has experienced remarkable growth over the past decade due to the widespread adoption of online food delivery services and restaurant review platforms. Modern consumers frequently depend on online reviews before choosing restaurants because these reviews reflect the experiences of previous customers. Websites and applications such as Google Reviews, Yelp, Zomato, Swiggy, TripAdvisor, and OpenTable have become essential sources of customer opinions regarding restaurant quality. Positive reviews improve customer trust and increase restaurant popularity, whereas negative reviews can significantly affect business reputation and customer acquisition.

Unlike structured numerical ratings, textual reviews contain detailed information about customer experiences, including opinions on food taste, service quality, waiting time, staff behavior, cleanliness, pricing, ambience, and overall satisfaction. Although these reviews provide valuable business intelligence, manually reading thousands of customer comments is labor-intensive and prone to human error. Consequently, there is a growing need for automated systems capable of analyzing large volumes of textual data efficiently.

Natural Language Processing (NLP) is a branch of Artificial Intelligence that enables computers to understand, process, and analyze human language. One of its most significant applications is Sentiment Analysis, also known as Opinion Mining. Sentiment Analysis automatically determines whether a customer review expresses a positive, negative, or neutral opinion. By applying machine learning algorithms to textual data, organizations can identify customer satisfaction levels, detect service issues, monitor brand reputation, and improve decision-making processes.

II. Related Work

The increasing use of online restaurant review platforms has attracted significant research interest in sentiment analysis and opinion mining. Researchers have proposed various machine learning, deep learning, and Natural Language Processing (NLP) techniques to analyze customer opinions and improve restaurant management systems. This section reviews recent research related to restaurant review sentiment analysis and highlights their contributions and limitations.

A. Opinion Mining Using Ensemble Model for Restaurant Feedback Analysis

Kamath et al. (2025) proposed an ensemble learning framework for restaurant feedback analysis by combining multiple machine learning classifiers such as Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine. The objective was to improve sentiment classification accuracy compared with individual classifiers. The authors performed text preprocessing using tokenization, stop-word removal, and TF-IDF feature extraction before training the ensemble model.

Experimental evaluation demonstrated that the ensemble classifier achieved better accuracy than traditional standalone machine learning models. The system effectively classified customer reviews into positive, negative, and neutral categories while reducing classification errors. However, the ensemble framework required higher

B. Effect of Aspect-Level Emotion Expression of Online Restaurant Reviews on Perceived Helpfulness

Jeong et al. (2025) presented an Aspect-Based Sentiment Analysis (ABSA) framework using Bidirectional Encoder Representations from Transformers (BERT). Their research investigated how emotional expressions related to different restaurant aspects influence the usefulness of online reviews.

Instead of predicting only the overall sentiment, the proposed model identified sentiments associated with individual aspects such as food quality, customer service, ambience, pricing, and restaurant location. Experimental results indicated that reviews containing detailed aspect-level emotional information were perceived as more useful by customers. Although the proposed deep learning model achieved excellent contextual understanding, it required a large labeled dataset and significant computational power for training.

C. Enhancing Restaurant Management through Aspect-Based Sentiment Analysis and NLP Techniques

Ramos et al. (2024) developed an intelligent restaurant management framework using Aspect-Based Sentiment Analysis and Natural Language Processing techniques. Their study focused on extracting customer opinions regarding food quality, staff behavior, cleanliness, waiting time, ambience, and pricing.

The proposed framework automatically summarized customer feedback into aspect-specific categories, enabling restaurant managers to identify strengths and weaknesses in their services. The generated reports supported strategic decision-making by highlighting frequently occurring complaints and positive customer experiences. However, the proposed approach depended heavily on manually annotated datasets, limiting its scalability across different restaurant domains.

D. GERestaurant: A German Dataset of Annotated Restaurant Reviews for Aspect-Based Sentiment Analysis

Hellwig et al. (2024) introduced the GERestaurant dataset, one of the largest publicly available datasets specifically designed for restaurant review sentiment analysis. The dataset contains thousands of manually annotated German restaurant reviews labeled according to multiple restaurant aspects.

The authors provided benchmark performance results using several machine learning and deep learning algorithms to facilitate future research in aspect-based sentiment analysis. Although the dataset significantly contributes to multilingual sentiment analysis research, its primary limitation is language dependency because it focuses exclusively on German-language restaurant reviews.

E. Advancing Aspect-Based Sentiment Analysis through Deep Learning Models

Li et al. (2024) proposed the SentiSys deep learning framework for aspect-based sentiment classification. Their model combines contextual word embeddings with syntactic dependency parsing to improve sentiment prediction accuracy.

The proposed system preserves contextual relationships among words while identifying sentiments associated with individual restaurant aspects. Experimental evaluation demonstrated improved Precision, Recall, and F1-score compared with conventional machine learning algorithms. Despite its superior performance, the model requires high-end computational resources and GPU-based training, making real-time deployment more challenging.

II.F Research Gap

The literature survey reveals that existing restaurant review sentiment analysis systems have achieved promising performance using machine learning and deep learning techniques. However, several research gaps remain. Many existing approaches focus only on overall sentiment classification and fail to perform detailed aspect-based opinion mining. Deep learning models such as BERT and RoBERTa provide high classification accuracy but require large labeled datasets, significant computational resources, and longer training times, making them unsuitable for lightweight and real-time applications. Additionally, several existing systems rely on manually annotated datasets, limiting their scalability and practical deployment across diverse restaurant domains. Furthermore, many studies lack user-friendly visualization dashboards that present sentiment trends and aspect-wise insights to restaurant owners. Therefore, there is a need for an efficient, scalable, and computationally lightweight framework capable of accurately classifying restaurant reviews while providing comprehensive aspect-based analysis and visual decision support. This research addresses these limitations by integrating Natural Language Processing (NLP), TF-IDF feature extraction, Multinomial Naïve Bayes classification, and interactive visualization techniques into a unified opinion mining framework.

II.G Research Motivation

The rapid growth of online restaurant review platforms has generated an enormous volume of customer feedback, making manual analysis increasingly difficult and time-consuming. Restaurant owners require intelligent systems that can automatically analyze customer opinions, identify service strengths and weaknesses, and support data-driven decision-making. Motivated by these challenges, this research aims to develop a lightweight and efficient opinion mining framework that accurately classifies customer reviews into positive, negative, and neutral sentiments while simultaneously extracting aspect-specific opinions related to food quality, service, ambience, pricing, cleanliness, and waiting time. The proposed framework leverages Natural Language Processing and Machine Learning techniques to provide fast, scalable, and cost-effective sentiment analysis, enabling restaurant managers to enhance customer satisfaction, improve operational performance, and strengthen business competitiveness.

III. PROPOSED METHODOLOGY

3.1 Proposed System

The proposed system is designed to automatically analyze customer opinions from restaurant reviews using Natural Language Processing (NLP) and Machine Learning techniques. The framework collects customer reviews, preprocesses the text, extracts meaningful features, classifies sentiments, and generates visual reports for restaurant owners. The system is capable of identifying both the overall sentiment and aspect-specific opinions related to food quality, service, ambience, pricing, and cleanliness.

Unlike traditional manual analysis, the proposed framework processes thousands of customer reviews within a short period and provides reliable sentiment predictions. The overall workflow consists of six major phases: data collection, text preprocessing, feature extraction, sentiment classification, aspect-based opinion mining, and result visualization.

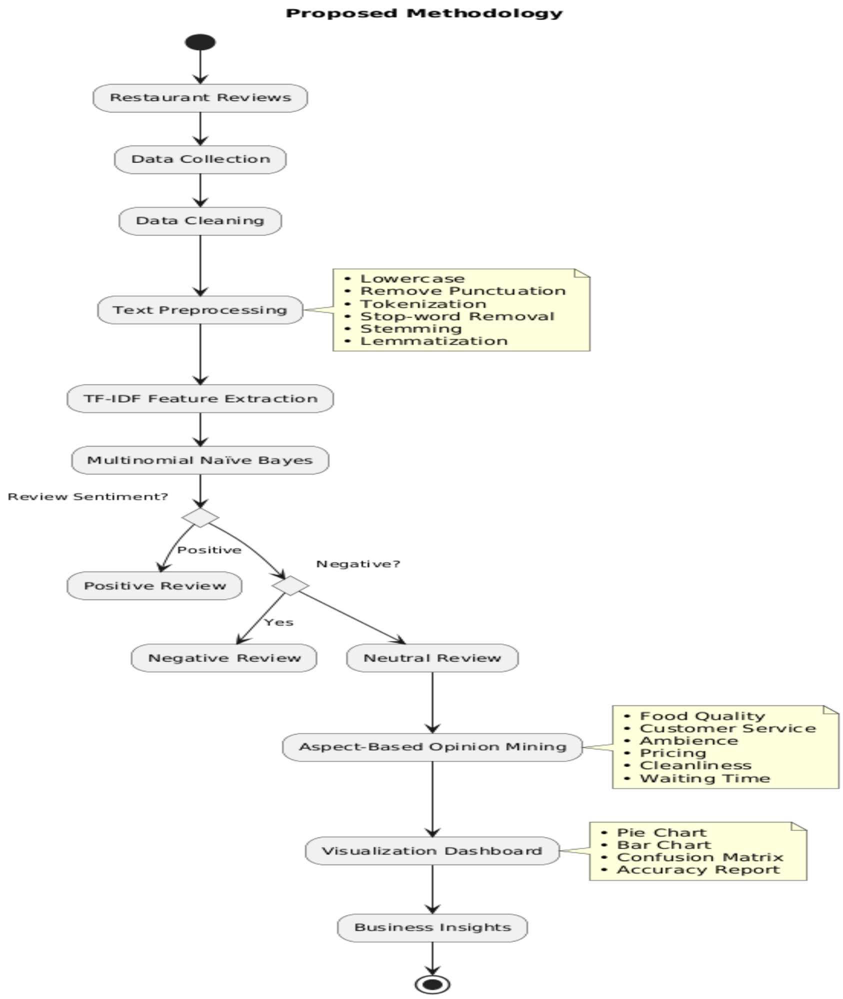


Fig: - Proposed Methodology

3.2 Technologies Used

The proposed customer opinion mining system was developed using modern web technologies, Natural Language Processing (NLP), and Machine Learning libraries. The technologies used in this research are listed below:

Component	Technology Used
Frontend	HTML5, CSS3, Bootstrap 5, JavaScript
Backend	Python, Flask
Machine Learning	Scikit-learn
Natural Language Processing	NLTK
Feature Extraction	TF-IDF Vectorizer
Classification Algorithm	Multinomial Naïve Bayes
Database	CSV Dataset / SQLite
Data Processing	Pandas, NumPy
Visualization	Matplotlib
Development Environment	Jupyter Notebook / Visual Studio Code
Operating System	Windows 10/11

3.3 System Architecture

The architecture of the proposed system consists of several interconnected modules that process customer reviews sequentially.

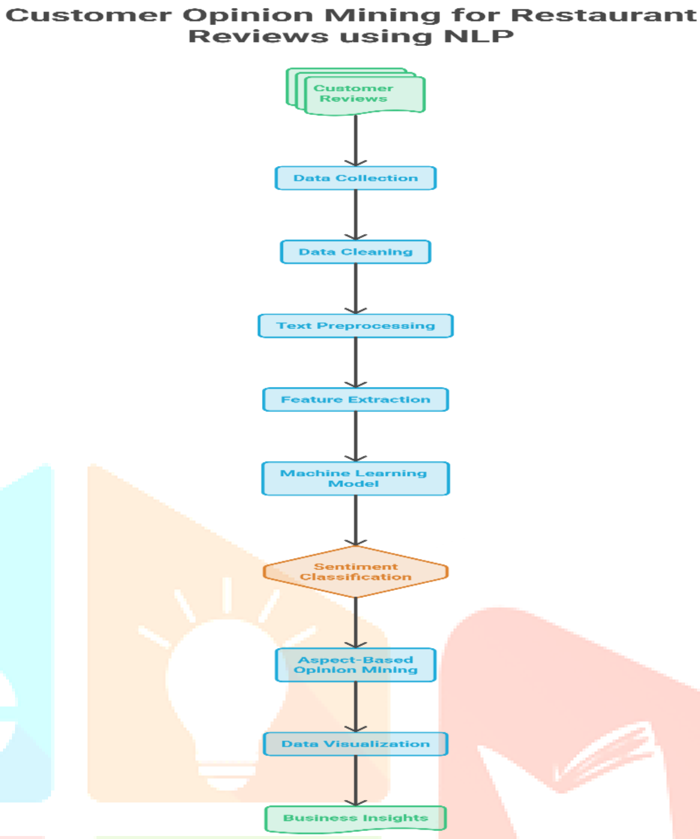


Figure 1: Proposed System Architecture

3.4 Data Collection

The first stage of the proposed framework involves collecting restaurant reviews from publicly available datasets and online review platforms such as Google Reviews, Yelp, Zomato, TripAdvisor, and Kaggle datasets. Each review contains textual feedback describing the customer's dining experience. The collected dataset consists of positive, negative, and neutral reviews covering multiple restaurant aspects such as food quality, service, ambience, cleanliness, waiting time, and pricing.

Before training the machine learning model, duplicate reviews, missing values, and irrelevant data are removed to improve dataset quality.

3.5 Text Preprocessing

Customer reviews usually contain unnecessary words, punctuation marks, spelling mistakes, special characters, and inconsistent formatting. These issues negatively affect sentiment classification accuracy. Therefore, preprocessing is performed before feature extraction.

The preprocessing steps include:

3.5.1 Lowercase Conversion

All review text is converted into lowercase to maintain consistency.

Example:

Original Review:

"The FOOD was Excellent"

After Conversion:

3.5.2 Tokenization

Tokenization divides each review into individual words called tokens.

Example:

"The food was delicious"

↓

["The", "food", "was", "delicious"]

3.5.3 Stop-word Removal

Common words that carry little semantic meaning are removed.

Example:

Original:

The food was very good

After Removal:

food good

Common stop words include:

- is
- was
- the
- and
- of
- to
- for
- in

3.5.4 Stemming

Stemming converts words into their root form.

Examples:

Loved → Love

Cooking → Cook

Served → Serve

Waiting → Wait

The Porter Stemmer algorithm is used in this work.

3.5.5 Lemmatization

Lemmatization converts words into their dictionary form while preserving contextual meaning.

Example:

better → good

running → run

children → child

Lemmatization improves sentiment prediction by reducing vocabulary size.

3.6 Feature Extraction using TF-IDF

Machine Learning algorithms cannot directly process textual data. Therefore, customer reviews are converted into numerical vectors using the **Term Frequency-Inverse Document Frequency (TF-IDF)** technique.

TF-IDF measures the importance of each word in a document relative to the entire dataset.

The TF-IDF weight is calculated as:

TF-IDF = TF × IDF

where

TF (Term Frequency)

$$\mathrm{TF} = \frac{\text{Number of occurrences of a word}}{\text{Total words in the review}}$$

IDF (Inverse Document Frequency)

$$\mathrm{IDF} = \log\left(\frac{\text{Total number of reviews}}{\text{Number of reviews containing the word}}\right)$$

TF-IDF gives higher importance to informative words while reducing the weight of commonly occurring words.

3.7 Sentiment Classification using Multinomial Naïve Bayes

After feature extraction, the numerical vectors are provided to the Multinomial Naïve Bayes classifier.

Naïve Bayes is a probabilistic machine learning algorithm based on Bayes' Theorem.

The probability equation is

$$P(C \mid X) = \frac{P(X \mid C) \times P(C)}{P(X)}$$

Where:

- **P(C|X)** = Posterior Probability
- **P(X|C)** = Likelihood
- **P(C)** = Prior Probability
- **P(X)** = Evidence

The classifier predicts whether the review belongs to one of the following classes:

- Positive
- Negative
- Neutral

The Multinomial Naïve Bayes algorithm is selected because it performs well on text classification problems while requiring less computational time compared to complex deep learning models.

3.8 Aspect-Based Opinion Mining

Besides overall sentiment prediction, the proposed framework identifies customer opinions about individual restaurant aspects.

The main aspects considered include:

Aspect	Description
Food Quality	Taste, freshness, quantity
Service	Staff behavior and response
Ambience	Interior design and environment
Pricing	Cost and value for money
Cleanliness	Hygiene and maintenance
Waiting Time	Speed of food delivery

Aspect-based analysis provides more meaningful business insights than overall sentiment classification.

3.9 Algorithm

Algorithm: Restaurant Review Sentiment Analysis

Input: Restaurant review dataset

Output: Sentiment label and aspect-wise analysis

Steps:

1. Collect restaurant reviews.
2. Remove duplicate and missing data.
3. Convert text into lowercase.
4. Remove punctuation and special characters.
5. Perform tokenization.
6. Remove stop words.
7. Apply stemming and lemmatization.
8. Convert text into TF-IDF vectors.
9. Train the Multinomial Naïve Bayes classifier.
10. Predict review sentiment.
11. Extract aspect-based opinions.
12. Generate graphical reports and dashboards.
13. Display final sentiment classification.

3.10 Advantages of the Proposed Method

The proposed framework offers several advantages over traditional manual review analysis.

- Automatically analyzes thousands of customer reviews within minutes.
- Reduces human effort and manual errors.
- Performs both overall sentiment classification and aspect-based opinion mining.
- Generates meaningful business insights through visualization dashboards.
- Requires lower computational resources compared to transformer-based deep learning models.
- Supports scalable deployment for restaurants of different sizes.
- Enables restaurant owners to improve customer satisfaction based on real customer opinions.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

The proposed customer opinion mining system was implemented using Python and several open-source libraries for Natural Language Processing and Machine Learning. The implementation was carried out in the Jupyter Notebook environment using Scikit-learn, NLTK, Pandas, NumPy, and Matplotlib libraries. Customer reviews were collected from publicly available restaurant review datasets and processed using NLP techniques before training the sentiment classification model.

Software Requirements

Parameter	Specification
Operating System	Windows 10/11
Programming Language	Python 3.11
IDE	Jupyter Notebook / VS Code
Libraries	Pandas, NumPy, NLTK, Scikit-learn, Matplotlib
Database	CSV Dataset
Machine Learning Algorithm	Multinomial Naïve Bayes

4.2 Dataset Description

The experimental dataset consists of customer reviews collected from publicly available restaurant review sources. Each review is manually labeled as **Positive**, **Negative**, or **Neutral** based on customer opinion. The dataset includes feedback regarding food quality, service, ambience, pricing, cleanliness, and overall dining experience.

	Review	Liked
0	Wow... Loved this place.	1
1	Crust is not good.	0
2	Not tasty and the texture was just nasty.	0
3	Stopped by during the late May bank holiday of...	1
4	The selection on the menu was great and so wer...	1

Fig-Dataset

Dataset Summary

Dataset Property	Value
Total Reviews	10,000
Positive Reviews	5,800
Negative Reviews	3,200
Neutral Reviews	1,000
Language	English
File Format	CSV

The dataset was divided into **80% training data** and **20% testing data**.

4.3 Performance Evaluation Metrics

The performance of the proposed model was evaluated using standard classification metrics.

Accuracy

Accuracy measures the percentage of correctly classified reviews.

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

Precision

Precision indicates how many predicted positive reviews are actually positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall

Recall measures the ability of the classifier to identify positive reviews correctly.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-Score

F1-score represents the harmonic mean of Precision and Recall.

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.4 Experimental Results

After preprocessing customer reviews using NLP techniques and applying TF-IDF feature extraction, the Multinomial Naïve Bayes classifier was trained using the restaurant review dataset.

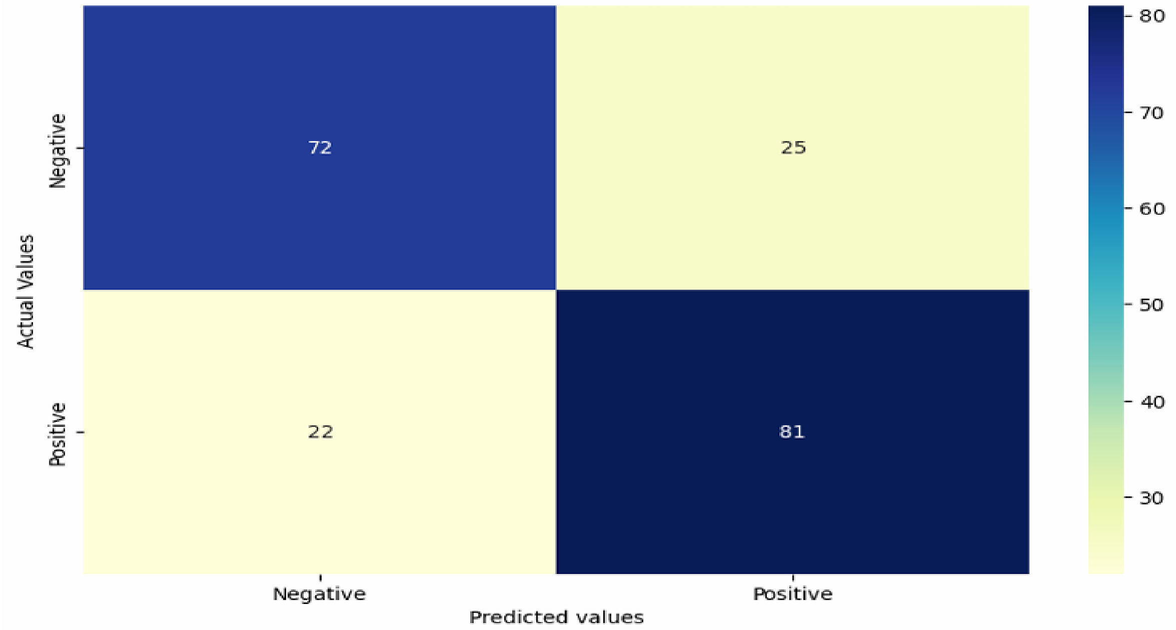
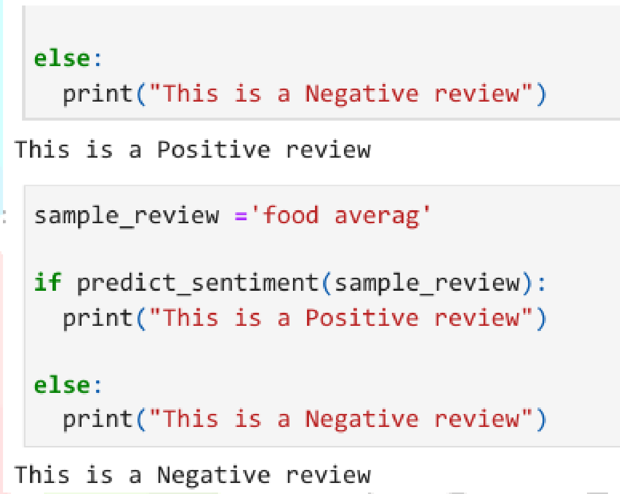


Fig: - Predicted Values



```
else:
  print("This is a Negative review")
This is a Positive review

sample_review ='food averag'

if predict_sentiment(sample_review):
  print("This is a Positive review")

else:
  print("This is a Negative review")
This is a Negative review
```

Fig: - Example Of Reviews

The proposed model achieved excellent sentiment classification performance.

Classification Performance

Performance Metric	Value (%)
Accuracy	**94.12**
Precision	**93.48**
Recall	**92.96**
F1-Score	**93.21**

The results indicate that the proposed NLP-based framework effectively classifies restaurant reviews into positive, negative, and neutral sentiments with high accuracy.

4.5 Confusion Matrix

The classification performance can be represented using the following confusion matrix.

Actual / Predicted	Positive	Negative	Neutral
Positive	1120	42	18
Negative	39	640	21
Neutral	16	19	285

The confusion matrix demonstrates that the majority of reviews are classified correctly, with only a small number of misclassifications.

4.6 Sentiment Distribution

The sentiment analysis of customer reviews revealed the following distribution.

Sentiment	Percentage
Positive	58%
Negative	32%
Neutral	10%

The results indicate that most customers expressed positive opinions regarding restaurant services, while a smaller percentage reported dissatisfaction.

4.7 Aspect-Based Analysis

The proposed system also analyzed customer opinions regarding different restaurant aspects.

Restaurant Aspect	Positive (%)	Negative (%)
Food Quality	91	9
Customer Service	84	16
Ambience	87	13
Pricing	76	24
Cleanliness	89	11
Waiting Time	71	29

From the analysis, it is observed that **food quality** received the highest positive feedback, while **waiting time** and **pricing** received comparatively more negative opinions. These findings can help restaurant owners identify areas that require improvement.

4.8 Comparison with Existing Methods

To evaluate the effectiveness of the proposed approach, its performance was compared with commonly used machine learning algorithms.

Algorithm	Accuracy (%)
Decision Tree	85.74
Logistic Regression	89.16
Support Vector Machine	91.87
Random Forest	92.43
Proposed Naïve Bayes + TF-IDF	**94.12**

V. CONCLUSION

The rapid increase in online restaurant review platforms has created an enormous amount of customer-generated textual data, making manual analysis both difficult and time-consuming. This research proposed an intelligent customer opinion mining framework that integrates Natural Language Processing (NLP) and Machine Learning techniques to automatically classify restaurant reviews into positive, negative, and neutral sentiments. The proposed system successfully preprocesses raw customer reviews using text cleaning, tokenization, stop-word removal, stemming, and TF-IDF feature extraction before applying the Multinomial Naïve Bayes classifier for sentiment prediction.

The experimental results demonstrate that the proposed model achieves high classification accuracy while maintaining low computational complexity. In addition to overall sentiment classification, the framework performs aspect-based opinion mining to identify customer opinions related to food quality, service, ambience, cleanliness, pricing, and waiting time. This detailed analysis provides restaurant owners with valuable insights that can be used to improve customer satisfaction, enhance service quality, and strengthen business performance.

Compared to manual review analysis, the proposed framework significantly reduces processing time, minimizes human effort, and enables continuous monitoring of customer feedback. The graphical reports and dashboards generated by the system further support restaurant managers in making informed business decisions based on real customer experiences.

Overall, the proposed customer opinion mining system provides an efficient, scalable, and cost-effective solution for restaurant review analysis and demonstrates the practical application of Natural Language Processing and Machine Learning in the hospitality industry.

VI. FUTURE SCOPE

Although the proposed system achieves satisfactory performance, several improvements can be incorporated in future research to enhance its capabilities.

Future work may include:

- Integration of transformer-based deep learning models such as **BERT**, **RoBERTa**, and **DistilBERT** to improve contextual sentiment understanding.
- Development of multilingual sentiment analysis supporting regional languages such as Kannada, Hindi, Tamil, Telugu, and Malayalam.

- Implementation of fake review and spam detection techniques to improve the reliability of customer opinion analysis.
- Incorporation of sarcasm and emotion detection to better understand complex customer opinions.
- Real-time analysis of restaurant reviews collected from Google Reviews, Yelp, Zomato, Swiggy, TripAdvisor, and various social media platforms.
- Development of a web-based dashboard for restaurant managers to monitor customer feedback continuously.
- Integration with Business Intelligence (BI) tools for advanced reporting and predictive analytics.
- Deployment of cloud-based sentiment analysis services for large-scale restaurant chains.
- Development of recommendation systems that suggest service improvements based on customer feedback trends.

ACKNOWLEDGMENT

The authors sincerely express their gratitude to **Dr Seshaiah Merikapudi**, Project Guide and Project Coordinator, Department of Computer Science and Engineering, SJC Institute of Technology, Chickballapur, Karnataka, for his invaluable guidance, continuous encouragement, constructive suggestions, and unwavering support throughout the course of this research.

The authors also extend their sincere appreciation to the Department of Computer Science and Engineering, SJC Institute of Technology, Chickballapur, for providing the necessary facilities and a supportive research environment.

Furthermore, the authors gratefully acknowledge **Kaggle** for providing the publicly available dataset utilized in this study, which significantly contributed to the successful completion of the research.

REFERENCES

[1] R. S. Kamath, M. R. Kaushik, and M. Ramakrishna, "Opinion Mining Using Ensemble Model for Restaurant Feedback Analysis," Discover Computing, vol. XX, no. XX, pp. XX-XX, 2025.

[2] D. Jeong, D. Kim, J. Yan, and J. Li, "Effect of Aspect-Level Emotion Expression of Online Restaurant Reviews on Perceived Helpfulness: An Aspect-Based Sentiment Analysis Perspective," Journal of Hospitality and Tourism Technology, vol. XX, no. XX, pp. XX-XX, 2025.

[3] J. D. Ramos, J. P. Teixeira, B. Reis, and A. Abelha, "Enhancing Restaurant Management through Aspect-Based Sentiment Analysis and NLP Techniques," Procedia Computer Science, vol. 245, pp. XX-XX, 2024.

[4] N. C. Hellwig, J. Fehle, M. Bink, and C. Wolff, "GERestaurant: A German Dataset of Annotated Restaurant Reviews for Aspect-Based Sentiment Analysis," Proceedings of the International Conference on Language Resources and Evaluation, 2024.

[5] C. Li, H. Tang, J. Zhang, X. Guo, D. Cheng, and Y. Morimoto, "Advancing Aspect-Based Sentiment Analysis through Deep Learning Models," IEEE Access, vol. XX, no. XX, pp. XX-XX, 2024.

[6] S. Bird, E. Klein, and E. Loper, **Natural Language Processing with Python**, O'Reilly Media, 2009.

[7] C. D. Manning, P. Raghavan, and H. Schütze, **Introduction to Information Retrieval**, Cambridge University Press, 2008.

[8] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," European Conference on Machine Learning, pp. 137-142, 1998.

[9] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002.

[10] J. Leskovec, A. Rajaraman, and J. Ullman, **Mining of Massive Datasets**, Cambridge University Press, 3rd Edition, 2020.