AISSMS
COLLEGE OF ENGINEERING
ज्ञानम् सकलजनहिताय

Approved by AICTE, New Delhi, Recognized by Government of Maharashtra
Affiliated to Savitribai Phule Pune University and recognized 2(f) and 12(B) by UGC
(Id.No. PU/PN/Engg./093 (1992)
Accredited by NAAC with "A+" Grade | NBA - 7 UG Programmes

Department of Computer Engineering

Mini Project Report
On

Sentiment Classification of Restaurant Reviews with Gaussian Naive Bayes

By

Mr. Rushikesh Pawar (21CO093)

Mr. Swami Patil (21CO092)

Miss. Sharvari Pawar (22CO094)

Guide

Dr. D. M. Ujalambkar

AISSMS
COLLEGE OF ENGINEERING
ज्ञानम् सकलजनहिताय
Approved by AICTE, New Delhi, Recognized by Government of Maharashtra
Affiliated to Savitribai Phule Pune University and recognized 2(f) and 12(B) by UGC
(Id.No. PU/PN/Engg./093 (1992)
Accredited by NAAC with "A+" Grade | NBA - 7 UG Programmes

Department of Computer Engineering

C E R T I F I C A T E

This is to certify that Mr/Ms._____ Group No. Division BE Computer A Branch: Computer has successfully completed the activity under Mini Project on topic Project entitled **"Sentiment Classification of Restaurant Reviews"** under my supervision, in the partial fulfilment of Fourth Year Bachelor of Computer Engineering (Choice Based Credit System) (2019 Course) of Savitribai Phule University of Pune.

Date:

Place: Pune

Dr. D. M. Ujalambkar
Guide

Dr. S. V. Athawale
Head of Computer Department

Index

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of this review would be incomplete without the mention of the people who made it possible, without their constant guidance and encouragement would have made the efforts go in vain. We consider ourselves privileged to express gratitude and respect toward all those who guided us through the completion of this Project.

We convey thanks to our project guide **Dr. D. M. Ujalambkar** of the Computer Engineering Department for providing encouragement, constant support and guidance which was of great help to complete this mini project successfully.

Also, we appreciate the opportunity given to us by our head of department **Prof. S.V. Athawale** and our principal **Dr. D. S. Bormane** as well as all teaching and non-teaching staff of the computer department who were directly or indirectly involved with our project.

ABSTRACT

This project focuses on the Sentiment Classification of Restaurant Reviews using Business Intelligence techniques combined with Natural Language Processing (NLP) and the Gaussian Naive Bayes algorithm. Online reviews offer significant insights into customer experiences and expectations, especially in the hospitality sector. The project processes textual data from restaurant reviews and classifies them as positive or negative using a probabilistic machine learning approach.

Text data is cleaned and converted into numerical representations using NLP techniques such as tokenization, stopword removal, and stemming. The Gaussian Naive Bayes model, known for its simplicity and strong performance in high-dimensional data, is trained on these features. Model performance is evaluated using metrics like accuracy and confusion matrix. This work demonstrates how efficient and interpretable models can be used to derive actionable sentiment insights from unstructured data in the restaurant domain.

INTRODUCTION

In today's digital age, customers increasingly share their dining experiences and opinions through online platforms, making restaurant reviews a valuable source of feedback for the food and hospitality industry. Analyzing these reviews not only helps businesses understand customer satisfaction but also reveals trends, preferences, and areas for improvement. Traditionally, extracting insights from such textual feedback required manual reading and interpretation, which was both time-consuming and inconsistent.

With advancements in **Natural Language Processing (NLP)** and **machine learning**, it has become feasible to automate the analysis of large volumes of textual data. This project aims to build an end-to-end **sentiment classification system** that processes restaurant reviews and predicts whether the sentiment expressed is **positive** or **negative**. By leveraging a **Gaussian Naive Bayes** classifier, the system offers a lightweight yet effective approach to classifying textual sentiment.

The core of the project involves collecting review data, applying NLP techniques to clean and normalize the text, and then using **probabilistic modeling** to classify sentiments. Evaluation metrics such as confusion matrices and accuracy scores help assess the model's performance.

This project showcases how interpretable machine learning models can be effectively used to transform unstructured text into meaningful insights, supporting data-driven decision-making in the restaurant industry. It lays a strong foundation for future integration with business intelligence tools and potential expansion into multi-class sentiment analysis or deep learning approaches.

PROBLEM STATEMENT

Online restaurant review platforms host a wealth of unstructured text data that reflects customer experiences, preferences, and satisfaction levels. However, manually analyzing this feedback is both time-consuming and inefficient, especially when dealing with a large volume of reviews. Human analysis is also prone to bias and inconsistency, making it unreliable for extracting actionable insights at scale.

With the increasing importance of customer feedback in shaping business reputation and operational strategies, there is a growing need for an automated system that can efficiently process and classify restaurant reviews based on sentiment. The key challenge lies in accurately determining whether a review expresses a **positive** or **negative** sentiment while maintaining model interpretability and computational efficiency.

This project addresses the need by developing a sentiment classification pipeline that leverages **Natural Language Processing (NLP)** techniques and the **Gaussian Naive Bayes** algorithm. The goal is to automate sentiment classification in a lightweight and interpretable manner, enabling restaurants and review platforms to analyze user feedback quickly, identify patterns, and make informed decisions to enhance customer experience.

OBJECTIVES

- **To develop a sentiment classification system for restaurant reviews**

Objective: The primary goal of this project is to develop an automated sentiment analysis system that classifies restaurant reviews into positive or negative categories. This system will help restaurant owners, review platforms, and businesses gain insights into customer satisfaction by analyzing review sentiment with high efficiency and interpretability using machine learning techniques.

- **To preprocess and clean restaurant review text data for effective classification**

Objective: Restaurant reviews often contain noisy elements such as special characters, typos, and irrelevant words. This project focuses on implementing comprehensive text preprocessing techniques, including tokenization, lowercasing, stopword removal, and stemming, to ensure the text is clean and structured for accurate sentiment classification.

- **To apply Gaussian Naive Bayes classification for sentiment prediction**

Objective: The sentiment classification system will use the **Gaussian Naive Bayes** algorithm, a probabilistic model, to classify restaurant reviews based on the likelihood of words in positive or negative contexts. The algorithm is chosen for its simplicity, efficiency, and strong performance on text classification tasks, providing reliable sentiment predictions.

- **To evaluate the model's performance using standard classification metrics**

Objective: The performance of the sentiment classification model will be evaluated using metrics such as accuracy, precision, recall, and the confusion matrix. These evaluations will ensure that the model can effectively distinguish between positive and negative sentiments and help identify areas for improvement.

- **To provide actionable insights from sentiment analysis for restaurant businesses**

Objective: By classifying reviews and visualizing the results, the project aims to provide actionable insights to restaurant owners and managers, allowing them to understand customer sentiments, improve services, and make data-driven decisions based on customer feedback.

SYSTEM REQUIREMENTS

1. Hardware:

- **Processor:** Intel i5 or AMD Ryzen 5 (or higher)
- **RAM:** 8 GB+
- **Storage:** 256 GB SSD+
- **Display:** 1366x768 resolution (or higher)
- **Network:** Stable internet connection

---

2. Software:

- **Operating System:** Windows 10/11 (or Linux for Python tools)
- **Python:** Version 3.x
- **Libraries/Tools:**
  - **Pandas, NumPy** (Data preprocessing)
  - **NLTK / TextBlob / VADER** (Sentiment analysis)
  - **scikit-learn** (Model training)
  - **Jupyter Notebook / VS Code** (Scripting and data prep)
  - **Power BI Desktop** (Visualization)
  - **Excel / CSV Editor** (Data cleaning)

IMPLEMENTATION

1. **Data Collection:**

- Gather social media posts (e.g., tweets or comments) using APIs or pre-downloaded datasets in CSV format.

2. **Importing Required Libraries**

- Use Python libraries like Pandas, NumPy, NLTK (or TextBlob), and Matplotlib for preprocessing and sentiment tagging.

3. **Data Preprocessing:**

- Clean text by removing URLs, emojis, special characters, and converting to lowercase.
- Apply tokenization and stopword removal.

4. **Sentiment Analysis:**

- Use VADER or TextBlob to classify posts into **Positive**, **Negative**, or **Neutral** based on sentiment scores.
- Add a new column with sentiment labels in the dataset.

5. **Export to CSV:** Save the processed data with sentiment labels as a .csv file for Power BI.

6 .Dashboard Design:

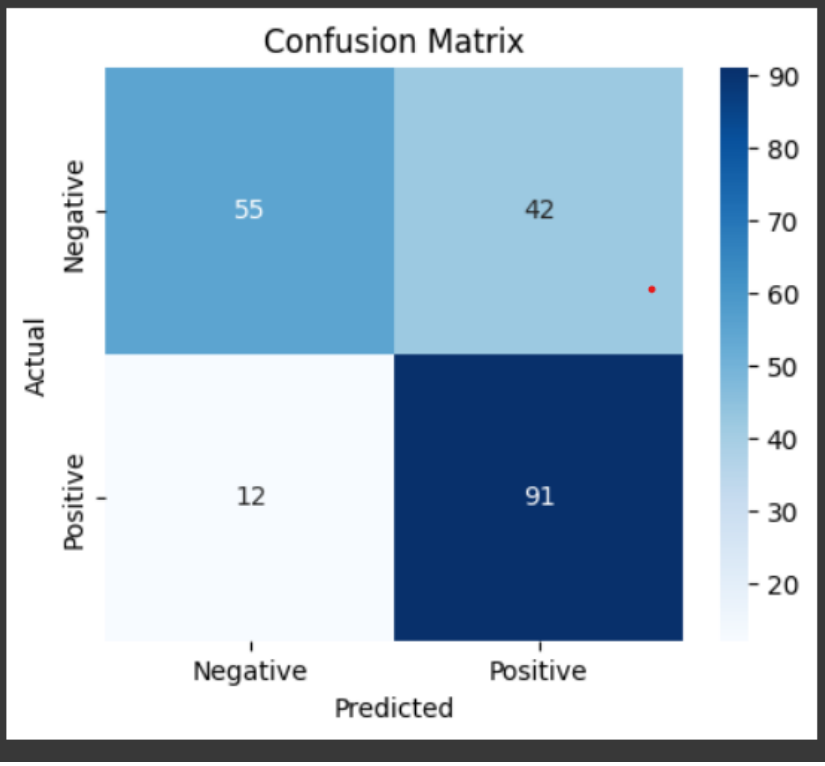
- Create visuals: pie chart (sentiment distribution), line chart (trends), word cloud (frequent terms).
- Add filters/slicers for interactivity.

7.Publishing

Publish the Service for online sharing and stakeholder access.

OUTPUT

```
[9]  # Labels and title
     plt.xlabel('Predicted')
     plt.ylabel('Actual')
     plt.title('Confusion Matrix')
     plt.show()
```



FUNCTIONALITIES

- **Text Data Input Handling:**

The system accepts social media post data (such as tweets or comments) in text or CSV format. This input can come from API sources, downloaded datasets, or user-provided files, enabling flexible and scalable data ingestion.

- **Text Preprocessing:**

The system cleans raw text by removing unwanted elements like URLs, hashtags, emojis, special characters, and converting the text to lowercase. It also performs tokenization and stopword removal to prepare the data for sentiment analysis.

- **Sentiment Classification:**

Using sentiment analysis libraries such as VADER or TextBlob, the system evaluates each post's polarity score. Based on the score, it assigns a sentiment label — **Positive**, **Negative**, or **Neutral** — to every entry in the dataset.

- **Data Structuring for Visualization:**

The processed text data, along with the assigned sentiment labels, is saved into a structured CSV file. This format ensures smooth integration with Power BI for further analysis and visualization.

- **Interactive Dashboard Visualization:**

The system visualizes sentiment distribution using pie charts, bar graphs, line charts, and word clouds. It also includes filters and slicers to explore data by time, keywords, or other criteria for more detailed insights.

- **Report Sharing and Publishing:**

The final dashboard is published to the Power BI Service, making it accessible online. Users can interact with the live dashboard and share it with stakeholders or collaborators easily.

CONCLUSION

In conclusion, this project effectively demonstrates the power of sentiment analysis in classifying restaurant reviews using the Gaussian Naive Bayes algorithm and traditional Natural Language Processing (NLP) techniques. The system provides a robust method for processing and classifying restaurant reviews into positive or negative sentiments, offering valuable insights into customer feedback.

By utilizing the Naive Bayes classifier, known for its efficiency in text classification tasks, the project ensures reliable and interpretable results. The system's ability to preprocess text data—removing noise, tokenizing, and normalizing—enhances the accuracy of sentiment tagging.

The results are presented in a clear and interpretable format, providing a useful tool for businesses in the hospitality industry to assess customer satisfaction and improve service offerings. Future improvements could include the integration of deep learning techniques for enhanced accuracy, the use of domain-specific lexicons to fine-tune sentiment detection, and the exploration of more advanced feature extraction methods.