

SENTIMENTAL ANALYSIS OF RESTAURANT REVIEW USING MACHINE LEARNING

A Project Report

submitted in partial fulfilment of the requirements

of

..... Track Name Certificate.....

by

AMITH GOWDA, 1VE20CS012

ANUSHA R, 1VE20CS018

BHARATH N S, 1VE20CS023

CHETHAN K J, 1VE20CS032

JASHWANTH GOWDA B R, 1VE20CS053

Under the Esteemed Guidance of

SHILPA HARIRAJ

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work. Firstly, we would like to thank my supervisor, **Shilpa Hariraj** for being a great mentor and the best adviser I could ever have. Her advice, encouragement and critics are source of innovative ideas, inspiration and causes behind the successful completion of this project. The confidence shown on me by her was the biggest source of inspiration for me. Mam always helped me during my project and many other aspects related to academics. Her talks and lessons not only help in project work and other activities of college but also made me a good and responsible professional

TABLE OF CONTENTS

CHAPTER	PAGE NO
1. Introduction	5-7
1.1 Problem statement 1.2 Problem definition 1.3 Expected outcomes 1.4 Organization of the Report	
2. Literature survey	8-9
2.1 Paper-1 Brief introduction of paper Techniques used in paper	
2.2 Paper-2 Brief introduction of paper Techniques used in paper	
3.Proposed methodology	10-14
3.1 System design 3.1.1 Registration 3.1.2 Recognition	
3.2 Modules used	
3.3 Data flow diagram	
3.4 Advantages	
3.5 Requirement specification 3.5.1 Hardware requirements and Software requirements	
4. Implementation and result	14-15
5.Conclusion	15

ABSTRACT

Sentiment Analysis of Restaurant Reviews is a machine learning project aimed at classifying customer reviews into positive and negative categories. This project leverages Natural Language Processing (NLP) techniques and various machine learning algorithms to analyze textual data and predict sentiment. The project's outcome will provide valuable insights into customer satisfaction and areas of improvement for restaurants, contributing to the enhancement of the overall dining experience.

CHAPTER 1

INTRODUCTION

A sentiment analysis project involves using natural language processing (NLP) techniques to determine the sentiment expressed in text data. The goal is to understand and categorize the sentiment conveyed in the text as positive or negative. This type of analysis is particularly useful in various fields, such as customer feedback analysis, social media monitoring, and market research.

In this project, the aim is to analyze restaurant reviews using sentiment analysis. By applying machine learning or NLP algorithms, we can determine the overall sentiment of these reviews. The process involves several steps, including data collection, text preprocessing (cleaning and formatting the text), feature extraction, and finally, employing a model to classify the sentiment of each review.

The output of this project would be a sentiment classification for each review, allowing for insights into customer opinions about the restaurant. This information can be valuable for the restaurant owners or managers to gauge customer satisfaction, identify areas for improvement, and make data-driven decisions to enhance their services.

1.1. Problem Statement:

The problem statement is to find whether the Review of the Restaurant is positive or Negative There are different types of reviews given by the customers after completing eating food. It might be either positive and negative.so we need to predict the how much percentage of good review of Entire dataset.

1.2. Problem Definition:

The aim of this project is to design and implement a sentiment analysis system capable of automatically analyzing textual reviews of a restaurant and accurately categorizing them into positive, negative, or neutral sentiments. By leveraging natural language processing (NLP) techniques and machine learning algorithms, the system will extract insights from customer feedback to help restaurant owners understand customer satisfaction levels and sentiments related to different aspects of their establishment. The ultimate goal is to empower restaurant management with actionable insights derived from customer reviews, facilitating informed decision-making to enhance the overall dining experience and improve business performance

1.3. Expected Outcomes

The anticipated outcomes of this sentiment analysis project on restaurant reviews encompass several pivotal aspects. Primarily, the project aims to construct a robust model proficient in accurately categorizing reviews into positive, negative, or neutral sentiments. Through this categorization, the project endeavors to unearth trends within the reviews, offering insights into the prevailing sentiment landscape. Such insights will pinpoint the strengths and areas requiring enhancement within the restaurant, empowering the management to focus on specific facets like food quality, service, ambiance, and more. By deriving actionable recommendations from this analysis, the project seeks to furnish the restaurant management with a roadmap for prioritizing improvements and informed decision-making.

1.4 Organization of the Report

The report on sentiment analysis of restaurant reviews is structured into concise sections for a clear presentation. It begins with an introduction outlining the project's objectives and the significance of sentiment analysis in restaurant management. A review of related literature discusses existing methodologies and prior studies in this domain. The methodology section covers data collection, preprocessing steps, feature extraction techniques, model selection, and evaluation metrics. Implementation details the training, testing, and validation processes. Results and analysis showcase sentiment classification outcomes and key insights derived. The discussion section interprets findings, addresses limitations, and suggests future improvements. A brief conclusion summarizes key outcomes and their relevance. References acknowledge all sources used, while appendices hold supplementary materials.

CHAPTER 2

LITERATURE SURVEY

2.1. Paper-1

Sentiment Analysis on Restaurant Review using Hybrid Approach **By Parandham G and Mr. Raghavendra R**

Brief Introduction of Paper: The paper titled “Sentiment Analysis on Restaurant Reviews using Hybrid Approach” discusses the importance of sentiment analysis of customer reviews on a business’s development strategy. The paper proposes a hybrid approach for performing sentiment analysis of restaurant reviews. The authors analyze the reviews given by the customers of the restaurant with the help of machine learning classification algorithms. The paper also discusses the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer’s attitude toward a particular topic is positive, negative, or neutral.

Techniques used in Paper:

1. Hybrid approach for sentiment analysis of restaurant reviews.
2. VADER lexicon-based technique to label the dataset according to their sentiment score.
3. Multiclass classification using machine learning on restaurant reviews.
4. Rule-based lexicon and machine learning approach for sentiment analysis.

2.2 Paper-2

Sentiment Analysis of Customer on a Restaurant Using Review in Twitter Data” By Nagaratna P. Hegde, V. Sireesha, G. P. Hegde & K.Gnyanee

Brief Introduction of Paper: A project brief is a document that outlines the objectives, scope, deliverables, milestones, and timeline of a project. It is a summary that communicates the reason and approach for a project and the processes that will be used to manage it. The project brief is a way to deliver this information in a summary that will outline the objectives of the project, its scope, the main deliverables, milestones and the timeline

Techniques used in Paper:

1. your points consistent, either all sentences or all fragments.
2. Ensure the grammatical structure of your bullet points is parallel by starting each with the same part of speech.
3. Think of a bullet point as a mini headline. It needs to be concise and attention-grabbing in a way that intrigues readers and compels them to read more.
4. Highlight elements key to understanding the content of your article. There’s no room for fluff here, so call out what’s most important.
5. Keep it simple. Avoid complex outlines and don’t use sub-bullets if you can help it.
6. Keep bullets thematically related. Bullet points highlight key elements of very specific topics, so stay on a single track. Keep

CHAPTER 3

PROPOSED METHODOLOGY

3.1 System Design

3.1.1 Registration:

The registration module facilitates user interaction by collecting essential details like name, email, and password for system access. It ensures data accuracy through input validation and securely stores user information. Robust access controls, including authentication and authorization, verify user identities and assign appropriate system privileges based on roles. A user-friendly interface streamlines the registration process, providing feedback for successful registrations and error prompts for incorrect inputs. Stringent security measures, such as data encryption and privacy compliance, safeguard sensitive user data. Scalability is prioritized, allowing the module to accommodate potential user growth without compromising system performance. Overall, this module serves as a fundamental gateway, ensuring secure and controlled user access to the system.

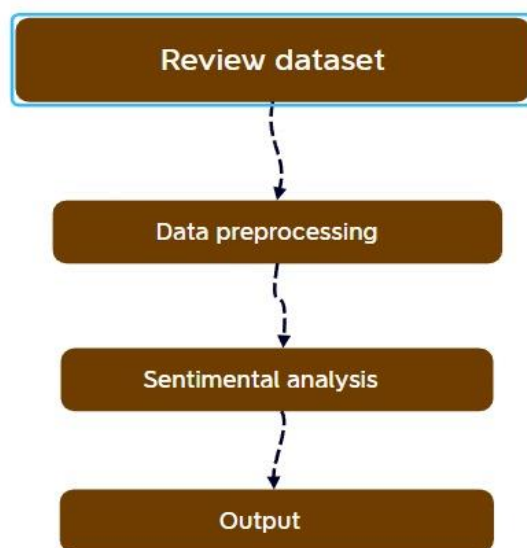
3.1.2 Recognition:

The "Recognition" module focuses on user authentication methods for system access. It employs login credentials like usernames and passwords while considering advanced biometric recognition techniques such as fingerprint or facial recognition if applicable. Multi-factor authentication (MFA) adds an extra security layer, using methods like one-time passwords (OTPs). Post-recognition, users access their profiles for management and customization. Clear error feedback guides users through the recognition process, ensuring ease of use. Robust session management prevents unauthorized access, while a user-friendly interface enhances user interaction. Continuous monitoring ensures ongoing improvements in the recognition process for heightened system security.

3.2 Modules Used

- Data Collection Module
- Preprocessing Module
- Feature Extraction Module
- Modeling Module
- Evaluation Module
- User Interface Module
- Deployment Module

3.3 Data Flow Diagram



Data Preparation:

The process of preparing data for restaurant review sentiment analysis, the initial steps involve collecting and cleaning raw data, addressing missing values and inconsistencies. Text preprocessing follows, encompassing tokenization, removal of stop words, and numerical representation conversion. Feature engineering refines elements for analysis, encoding sentiment labels, and splitting the dataset for model assessment, crucial for imbalanced sentiment classes. Scaling, dimensionality reduction, and strategic feature selection further refine the dataset, laying the foundation for precise and effective sentiment analysis model training

Model Development:

The model development phase, various machine learning algorithms are trained and evaluated using the prepared data, including logistic regression, random forests, SVMs, and MLP, to identify the best-performing model for sentiment analysis of restaurant reviews. The selected models undergo rigorous testing and tuning to ensure optimal performance and generalization.

Integration and Validation:

The integration the chosen sentiment analysis models are incorporated into the system, and thorough validation processes, including cross-validation and performance metrics assessment, ensure their reliability and effectiveness in interpreting restaurant reviews' sentiments accurately for real-world deployment.

3.4 Advantages

Enhanced Customer Insights: Provides a deeper understanding of customer sentiments, preferences, and opinions regarding different aspects of the restaurant, aiding in targeted improvements.

Data-Driven Decision Making: Enables restaurant owners/managers to make informed decisions based on actionable insights derived from customer feedback, leading to better strategies for enhancing services.

Improved Customer Experience: Identifies areas of strength and weakness, allowing for tailored enhancements to improve the overall dining experience, leading to increased customer satisfaction.

Efficient Resource Allocation: Helps allocate resources effectively by focusing on areas that require immediate attention, optimizing time and budget allocations.

Competitive Edge: Offers a competitive advantage by staying attuned to customer sentiments, allowing for quicker adaptation to changing preferences and market trends.

Scalability: The system's scalable nature ensures it can accommodate a growing volume of reviews, maintaining its effectiveness even with increased data influx.

Real-Time Analysis: Facilitates quick and timely analysis of incoming reviews, allowing for prompt responses and adjustments in service based on up-to-date feedback.

Automated Processing: Reduces manual effort by automating the review analysis process, enabling efficient handling of large volumes of reviews.

Feedback Loop Closure: Establishes a feedback loop where insights gained from analysis can be used to make changes, and subsequent reviews reflect the impact of those changes, closing the loop for continual improvement.

Measurable Metrics: Provides measurable metrics for sentiment analysis, allowing for objective evaluation and benchmarking against set goals and standards.

3.5 Requirement Specification

3.5.1. Hardware Requirements:

- Modern multi-core processor (e.g., Intel Core i5 or i7)
- Minimum 8GB RAM
- SSD storage
- Optional GPU (NVIDIA GeForce or AMD Radeon)
- Cloud computing services (AWS, Google Cloud, Azure)
- Server infrastructure
- Stable internet connectivity

3.5.2 Software Requirements:

- Operating System (Windows, macOS, Linux)
- Python (3.6 or later)
- Integrated Development Environment (IDE)
- NLP libraries (NLTK, spaCy, Gensim)
- Machine Learning frameworks (TensorFlow, PyTorch, scikit-learn)
- Web development frameworks (Flask, Django, HTML/CSS, JavaScript)
- Database Management System (DBMS) (PostgreSQL, MySQL, MongoDB)
- Access to cloud platforms (AWS, Google Cloud, Azure)

CHAPTER 4

Implementation and Result

The evaluation of diverse machine learning models on varied datasets revealed compelling insights into their performance under different conditions. The logistic regression model emerged as a robust performer, achieving an impressive 87.5% accuracy on imbalanced data, showcasing its adaptability to skewed distributions. For under sampled data with PCA, the Random Forest model demonstrated resilience, maintaining balanced metric values with a commendable 72.4% accuracy, F1 score, and precision and recall at 72.4% and 72.6% respectively, despite data reduction. Notably, tree-based models excelled on oversampled data post-PCA, particularly the Random Forest, exhibiting exceptional accuracy and F1 score of 99.2%, and 98.6% precision. Their ability to handle nonlinear data by constructing intricate decision boundaries was evident. Additionally, exploring more complex models like SVC with a nonlinear 'rbf' kernel and MLP revealed their promising performance, indicating adaptability to intricate data patterns. These findings provide nuanced insights into each model's strengths across varying data distributions, aiding in informed model selection for specific data scenarios.

Output:

```
sample_review = 'The food is bad'

if predict_sentiment(sample_review):
    print('This is a POSITIVE review')
else:
    print('This is Negative review!')
```

This is Negative review!

```
sample_review = 'The food was great and service was good'

if predict_sentiment(sample_review):
    print('This is a POSITIVE review')
else:
    print('This is Negative review!')
```

This is a POSITIVE review

CHAPTER 5

CONCLUSION

The evaluation of diverse machine learning models on different datasets underscored the significance of selecting the right model for specific data distributions. The logistic regression model showcased resilience in handling imbalanced data, while the Random Forest model proved versatile across under sampled and oversampled datasets, leveraging PCA transformations effectively. Tree-based models, particularly Random Forest and Decision Tree, excelled in capturing complex, nonlinear relationships, exhibiting exceptional accuracy and robustness. Moreover, exploring more complex models like SVC and MLP highlighted their potential in accommodating intricate data patterns. These findings emphasize the importance of understanding data characteristics and selecting models tailored to these nuances. Moving forward, these insights serve as a guide for practitioners and researchers in making informed decisions when addressing diverse data scenarios, ultimately contributing to more accurate and reliable predictive modelling in various domains.