SENTIMENT ANALYSIS FOR TOP ENGINEERING COLLEGES IN TAMIL NADU

A MINI-PROJECT REPORT

Submitted by

GANESH K I 221701015

HARRSHAVARDHANN S 221701018

in partial fulfilment for the course

CD19651 Mini Project

for the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND DESIGN

RAJALAKSHMI ENGINEERING COLLEGE

RAJALAKSHMI NAGAR

THANDALAM

CHENNAI - 602 105

APRIL 2025

RAJALAKSHMI ENGINEERING COLLEGE

CHENNAI – 602105

BONAFIDE CERTIFICATE

Certified that this project report **"SENTIMENT ANALYSIS REVIEW FOR TOP ENGINEERING COLLEGES IN TAMIL NADU"** is the bonafide work of **"GANESH K I (221701015)**, **HARRSHAVARDHANN S(221701018)"** who carried out the project work for the subject CD19651 – Mini Project under my supervision.

SIGNATURE	**SIGNATURE**
Prof. Uma Maheshwar Rao	**Mr. Pradeep Kumar S ,**
Head of the Department	**Supervisor**
Professor and Head Computer Science and Design Rajalakshmi Engineering College Chennai - 602105	Assistant Professor Computer Science and Design Rajalakshmi Engineering College Chennai - 602105

Submitted to Project and Viva Voce Examination for the subject

CD19651 – Mini Project held on_____.

Internal Examiner External Examiner

ABSTRACT

This project presents a sentiment analysis system designed to evaluate public opinions about top engineering colleges in Tamil Nadu, India. By analyzing user-generated content such as reviews and social media posts, the system aims to provide insights into the perception and reputation of these institutions. The sentiment analysis is conducted using a Random Forest model and Natural Language Processing (NLP) techniques to classify sentiments as positive, negative, or neutral. Data is collected, preprocessed, and stored using MySQL. A Flask-based web application facilitates real-time review submission and sentiment prediction. The study offers visual representations of sentiment trends, aiding stakeholders in decision-making processes. The proposed solution demonstrates high accuracy in sentiment classification and can be further expanded to analyze opinions on a broader range of educational institutions.

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr.S.Meganathan, B.E, F.I.E.,** our Vice Chairman **Mr. Abhay Shankar Meganathan, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) Thangam Meganathan, Ph.D.,** for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N.Murugesan, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to our **Prof. Uma Maheshwar Rao** Professor and Head of the Department of Computer Science and Design for his guidance and encouragement throughout the project work. We convey our sincere thanks to our internal guide and Project Coordinator, **Mr. Pradeep Kumar S.** Department of Computer Science and Design, Rajalakshmi Engineering College for his valuable guidance throughout the course of the project.

GANESH K I(221701015)

HARRSHAVARDHANN S (221701018)

TABLE OF CONTENTS

S.NO	TITLE	PAGE.NO
1	Introduction	7
2	Literature Review	8
3	Software Methodology	10
4	Present Technology	13
5	Proposed Technology	17
6	Output	21
7	Conclusion	23

LIST OF FIGURES

S.NO	TITLE	PAGE.NO
1	Website Page	12
2	Login Page	21
3	Entering Mail and Password	21
4	Selecting College and Entering Review	22
5	Review Result	22

CHAPTER 1

INTRODUCTION

Sentiment analysis, also known as opinion mining, is a field of Natural Language Processing (NLP) that involves analyzing and interpreting textual data to determine the sentiment expressed within it. It is widely used in various industries to understand public opinion, customer satisfaction, and market trends. In the context of higher education, sentiment analysis can provide valuable insights into how students, alumni, and the general public perceive academic institutions.

This project focuses on conducting sentiment analysis for top engineering colleges in Tamil Nadu, India. By collecting and analyzing reviews and feedback from various platforms, the system aims to assess the general sentiment toward these institutions. Using Random Forest, a machine learning algorithm, along with advanced NLP techniques, the project classifies sentiments into categories such as positive, negative, or neutral.

The results of this analysis can be beneficial for prospective students in making informed decisions about their education. Additionally, educational institutions can leverage this feedback to improve their academic offerings, facilities, and student experiences. A Flask-based web application has been developed to collect user reviews and display sentiment analysis results in real time. The application integrates MySQL for data storage and ensures seamless communication between the frontend and backend.

Through this study, stakeholders gain actionable insights into the strengths and areas of improvement of each college, contributing to overall educational quality enhancement.

CHAPTER 2

LITERATURE REVIEW

The application of **sentiment analysis** in the field of education has gained significant traction over the past decade. Researchers have explored various techniques and models to analyze student opinions, feedback, and reviews to derive insights for institutional improvement. This section provides an overview of previous studies and their findings in sentiment analysis, particularly in the context of educational institutions.

1.Sentiment Analysis in Education

Several studies have focused on analyzing student reviews to measure satisfaction levels. Researchers often apply Natural Language Processing (NLP) techniques to extract and classify sentiments.For instance, sentiment analysis has been used to evaluate course reviews, instructor feedback, and overall university experiences. The results provide actionable insights for improving educational quality.

2. Machine Learning Algorithms for Sentiment Analysis

Commonly used algorithms for sentiment classification include Naive Bayes, Support Vector Machines (SVM), and Random Forest. Among these, Random Forest is often preferred due to its accuracy and ability to handle large datasets.Recent studies have highlighted the effectiveness of ensemble models like Random Forest in minimizing classification errors and enhancing sentiment prediction accuracy.

3.Data Collection and Preprocessing

Data for sentiment analysis is generally gathered from online sources such as social media platforms, college review websites, and surveys. Preprocessing steps like tokenization, stopword removal, and lemmatization are applied to clean and prepare the text data.Researchers have also implemented TF-IDF (Term Frequency-Inverse Document Frequency) and word embeddings for feature extraction.

4. Web-Based Sentiment Analysis Systems

Many sentiment analysis models have been integrated into web-based platforms to allow real-time analysis. Using frameworks like Flask or Django, researchers have built applications where users can submit reviews, and the system predicts sentiments instantly.Additionally, MySQL and other relational databases are often employed for efficient data storage and management.

5. Gaps Identified in Previous Research

While prior studies have demonstrated effective sentiment classification, many lack focus on regional or domain-specific analysis.Few studies have specifically targeted the sentiment analysis of engineering colleges in Tamil Nadu, leaving a research gap that this project aims to address.Additionally, most existing systems do not provide real-time sentiment visualization for stakeholders, which this project intends to implement.

CHAPTER 3

SOFTWARE METHODOLOGY

This project employs a range of software tools and technologies to facilitate the development, deployment, and analysis of sentiment data. Each software component serves a specific purpose, contributing to the seamless functioning of the sentiment analysis system.

Python: Python serves as the core programming language for implementing the sentiment analysis model. Its extensive libraries, including scikit-learn for machine learning and NLTK for Natural Language Processing (NLP), are used for data preprocessing, model training, and sentiment classification.

Flask: A lightweight web framework in Python, Flask is used to build the backend of the application. It handles HTTP requests, connects to the sentiment analysis model, and provides API endpoints for submitting reviews and displaying results.

MySQL: The project uses MySQL as the database management system to store user reviews, sentiment predictions, and analysis results. Its structured query language ensures efficient data storage, retrieval, and management.

HTML, CSS, and Bootstrap: The frontend of the web application is designed using HTML for structure, CSS for styling, and Bootstrap for responsive design. This ensures an intuitive and user-friendly interface for submitting reviews and viewing sentiment analysis results.

Pandas and NumPy: For data manipulation and numerical operations, Pandas and NumPy are utilized. These libraries assist in organizing data, handling missing values, and performing statistical analysis during the preprocessing and model evaluation stage.

Matplotlib and Seaborn: To visualize the sentiment analysis results, Matplotlib and Seaborn are used. These libraries generate graphs and charts representing sentiment distributions, model performance metrics, and other insights.

Jupyter Notebook: During the model development phase, Jupyter Notebook is used for writing, testing, and debugging Python code. Its interactive environment allows for step-by-step implementation and visualization of the analysis.

Render or Heroku: For deployment purposes, Render or Heroku is used to host the web application, ensuring accessibility for users across different devices and platforms.

Together, these software tools provide a robust environment for conducting sentiment analysis, storing and managing data, building an interactive user interface, and deploying the application effectively.

Outcome and Impact:

The sentiment analysis project successfully delivers a functional system that classifies public opinions about the top engineering colleges in Tamil Nadu as positive, negative, or neutral using a Random Forest model and Natural Language Processing (NLP) techniques. The implementation of a Flask-based web application enables real-time sentiment analysis, allowing users to submit reviews and receive instant feedback on sentiment classification. Additionally, the use of MySQL ensures efficient data management, storing user reviews and sentiment results for further analysis. Visual representations of sentiment trends through graphs and charts offer stakeholders actionable insights.

The impact of the project extends to various stakeholders. Prospective students can make informed decisions based on authentic sentiment analysis, while educational institutions can leverage the feedback to improve their academic programs, faculty performance, and student satisfaction. Policymakers and administrators can use the aggregated sentiment data to monitor institutional performance and implement necessary improvements in the education sector. Furthermore, the project serves as a valuable resource for researchers exploring advancements in sentiment analysis and its applications in education. By promoting transparency and accountability, this sentiment analysis system contributes to enhancing the quality of higher education in Tamil Nadu, fostering continuous improvement and informed decision-making.

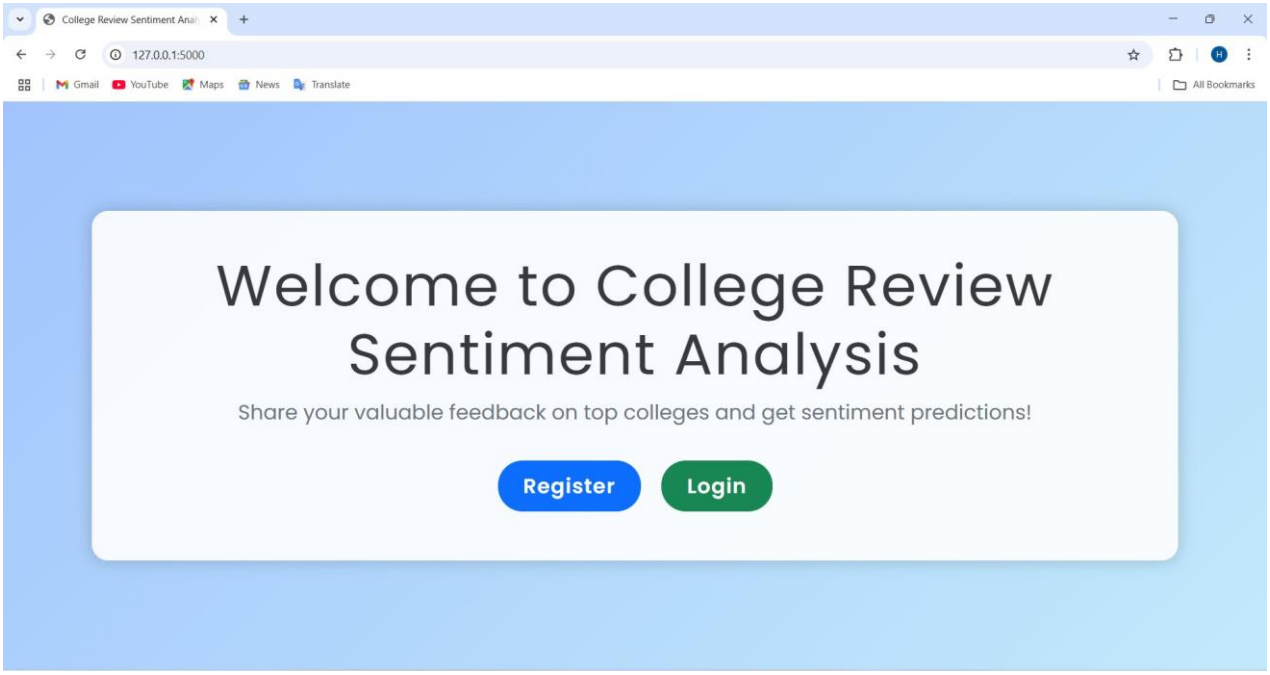


Fig 1: Website page for login.

CHAPTER 4

PRESENT TECHNOLOGY

The sentiment analysis project follows a three-tier architecture consisting of the Frontend, Backend, and Database layers, ensuring efficient data processing, real-time analysis, and an intuitive user experience. Each layer uses modern technologies to provide seamless functionality.

Frontend Layer

The Frontend is designed using HTML for the structural framework, CSS for styling, and Bootstrap for responsive design. These technologies ensure a clean, mobile-friendly user interface.Users interact with the application by submitting reviews and viewing sentiment analysis results.JavaScript is used for dynamic updates and improved interactivity on the client side.

Backend Layer

The Backend is powered by Flask, a Python-based micro web framework that handles API requests, processes user input, and serves responses.The sentiment analysis model is implemented using Python with Scikit-Learn and NLTK libraries for Natural Language Processing (NLP).The backend performs sentiment classification using the Random Forest algorithm, ensuring accurate predictions.REST APIs facilitate communication between the frontend and backend for seamless data exchange.

Database Layer

MySQL serves as the relational database management system, storing user reviews, sentiment analysis results, and other application data.The database is optimized using structured tables to ensure efficient data retrieval and management.SQL queries are used for storing, fetching, and analyzing data.

Machine Learning Model

The sentiment analysis model is trained using historical review data and built using the Random Forest classifier.Pandas and NumPy are used for data preprocessing, handling missing values, and extracting features using techniques like TF-IDF (Term Frequency-Inverse Document Frequency).The model is evaluated using accuracy metrics and fine-tuned for optimal performance.

Visualization and Reporting

Matplotlib and Seaborn are employed for data visualization, providing clear representations of sentiment trends and classification results.Visual dashboards and charts offer stakeholders actionable insights.

Deployment

The application is deployed using cloud platforms such as Render or Heroku, ensuring scalability and accessibility for users.API endpoints are hosted on the cloud, providing real-time sentiment predictions.

3.1 LIMITATIONS:

While the sentiment analysis project offers valuable insights into public opinions about top engineering colleges in Tamil Nadu, it also has certain limitations that may affect its accuracy and performance.

Data Quality and Bias: The accuracy of the sentiment analysis heavily depends on the quality of the input data. Reviews containing slang, sarcasm, or ambiguous language can lead to incorrect sentiment classification. Additionally, if the dataset has inherent biases, the model may reflect those biases in its predictions.

Limited Context Understanding: Despite using Natural Language Processing (NLP), the model may struggle with understanding the deeper context of certain reviews. Complex sentiments that involve mixed emotions or comparative statements might be misclassified.

Dependence on Predefined Data: The model's performance is limited to the training data it has been exposed to. It may not accurately analyze sentiments from emerging language patterns or new slang without regular updates and retraining.

Inadequate Handling of Multilingual Data: The current implementation primarily supports sentiment analysis in English. Reviews written in regional languages or using a combination of languages (code-switching) may result in lower accuracy.

Lack of Domain-Specific Understanding: The model lacks domain-specific knowledge, which can affect the interpretation of technical or academic terms frequently used in engineering college reviews.

Real-Time Performance Limitations: While the model is efficient, real-time analysis of a large volume of data may lead to delays in prediction results, especially if the server or database resources are limited.

False Positives and Negatives: Despite using a Random Forest classifier, the model may still produce false positives or negatives due to limitations in capturing nuanced sentiments.

No Emotion Detection: The system only classifies reviews into positive, negative, or neutral categories. It does not detect specific emotions like joy, anger, or sadness, which could provide deeper insights.

Limited Visualization Capabilities: While visualizations using Matplotlib and Seaborn are helpful, they may not be sufficient for analyzing large datasets or identifying hidden patterns without further customization.

Scalability Constraints: The current deployment using platforms like Render or Heroku may face scalability challenges if there is a sudden surge in user reviews, requiring infrastructure adjustments.

CHAPTER 5

PROPOSED TECHNOLOGY

To overcome the limitations of the current sentiment analysis system and enhance its accuracy, performance, and usability, the following redesign proposals are suggested:

Improved Data Quality and Preprocessing:
Implement advanced text preprocessing techniques using BERT (Bidirectional Encoder Representations from Transformers) or RoBERTa for better understanding of complex sentences.Apply context-aware filtering to reduce noise from irrelevant or misleading reviews.Incorporate data augmentation techniques to expand the training dataset and improve model robustness.

Contextual and Sentiment Understanding:
Upgrade the sentiment analysis model to a transformer-based model like BERT or GPT for better contextual understanding and improved accuracy.
Implement a multi-label classification system to detect nuanced sentiments, including emotions like joy, anger, or satisfaction, instead of limiting to positive, negative, or neutral categories.

Multilingual Support:
Integrate Multilingual BERT (mBERT) or XLM-R models to handle reviews in regional languages, such as Tamil and Hindi, enabling sentiment analysis for a wider audience.Provide a language detection module using LangDetect or FastText to automatically identify and process multilingual reviews.

Enhanced Model Training and Evaluation:
Employ Transfer Learning by using pre-trained models on larger datasets for more accurate sentiment predictions.Implement an active learning framework to continuously improve model performance based on user feedback and real-time predictions.

Real-Time Performance Optimization:
Utilize asynchronous processing and microservices architecture to improve real-time response rates.Integrate load balancers and scale using platforms like AWS or Azure to manage high traffic during peak usage.

Emotion and Aspect-Based Analysis:
Implement Aspect-Based Sentiment Analysis (ABSA) to extract specific opinions on aspects like faculty, infrastructure, or placement opportunities.Integrate emotion detection to provide detailed sentiment reports based on user emotions.

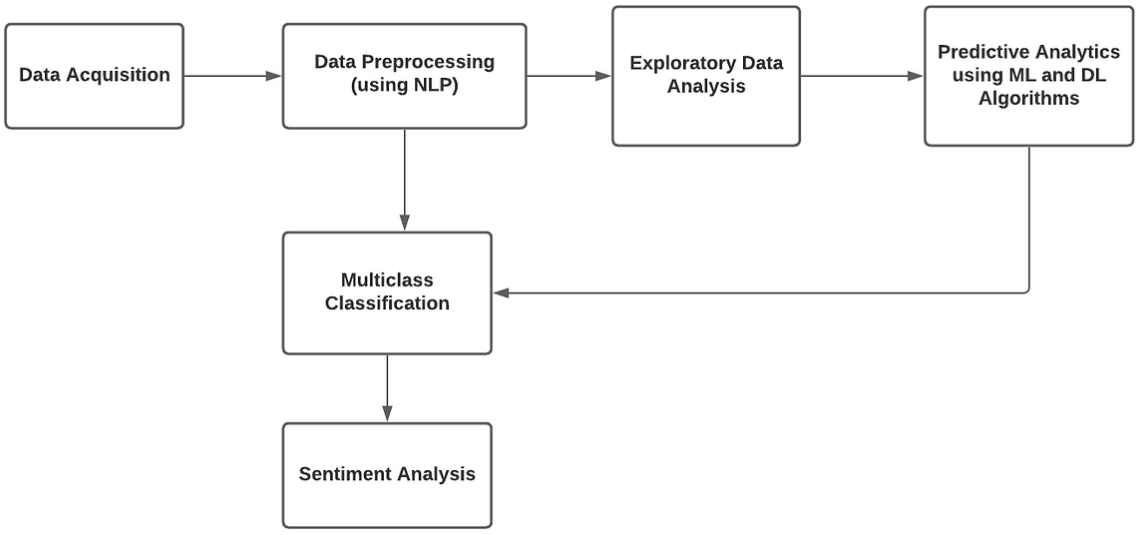
Improved Visualization and Reporting:
Develop interactive dashboards using tools like Plotly or Power BI to provide detailed reports with customizable views.Implement time-series analysis to observe sentiment trends over time and predict future sentiment changes.

Enhanced User Experience:
Provide a review summary generator using NLP to present key highlights from large volumes of reviews.Enable sentiment comparison between different colleges using visual charts for users to make informed decisions.

Scalability and Deployment:
Migrate the application to a containerized environment using Docker and Kubernetes for scalable deployments.Implement CD/CI pipelines for efficient updates and version management.

Feedback Integration:
Provide a user feedback system to allow users to rate the accuracy of sentiment predictions, which can be used to further refine the model.

5.1 USER FLOW DIAGRAM:



5.2 ADVANTAGES:

Implementing sentiment analysis for a website offers numerous benefits, particularly in gaining valuable insights from user feedback. Here are some key advantages:

Enhanced Decision-Making:

Sentiment analysis provides actionable insights by identifying the overall sentiment of user reviews.Stakeholders can make informed decisions regarding service improvements, product modifications, or content adjustments based on user opinions.

Real-Time Feedback Monitoring:

The system allows for real-time analysis of user reviews, enabling prompt responses to negative feedback and immediate recognition of positive sentiment.This is particularly useful for websites offering customer support or managing brand reputation.

Improved User Experience:

By understanding user sentiments, website owners can tailor the user experience to meet customer expectations. Personalized content or targeted promotions can be delivered based on user sentiment patterns.

Competitive Analysis:

Sentiment analysis can be extended to monitor competitor reviews and assess market perception.This helps businesses understand how they compare to competitors and identify areas for improvement.

Efficient Data Management:

Using advanced Natural Language Processing (NLP) and Machine Learning (ML) algorithms, large volumes of data can be processed quickly and accurately.Websites with thousands of user reviews can efficiently analyze data for meaningful insights.

Brand Reputation Management:

By continuously monitoring user sentiments, companies can detect early signs of negative public perception and take corrective measures.Positive feedback can also be highlighted for brand promotion and marketing campaigns.

Increased Customer Engagement:

Providing users with transparent sentiment reports or visualizing public opinion through dashboards can enhance engagement.Websites can also create opportunities for users to interact by submitting reviews and viewing sentiment trends.

Cost-Effective Analysis:

Automating sentiment analysis using AI models reduces the need for manual data analysis, saving time and resources.It offers a scalable and efficient solution for analyzing large datasets without extensive human intervention.

Product and Service Enhancement:

Websites offering products or services can identify areas for improvement based on customer feedback.Features that users appreciate can be enhanced further, while pain points can be addressed promptly.

Predictive Insights:

By analyzing sentiment trends over time, businesses can predict future user behavior and preferences.This helps in strategic planning and enhancing customer satisfaction.

CHAPTER 6

PROJECT LINK: http://127.0.0.1:5000/

OUTPUT

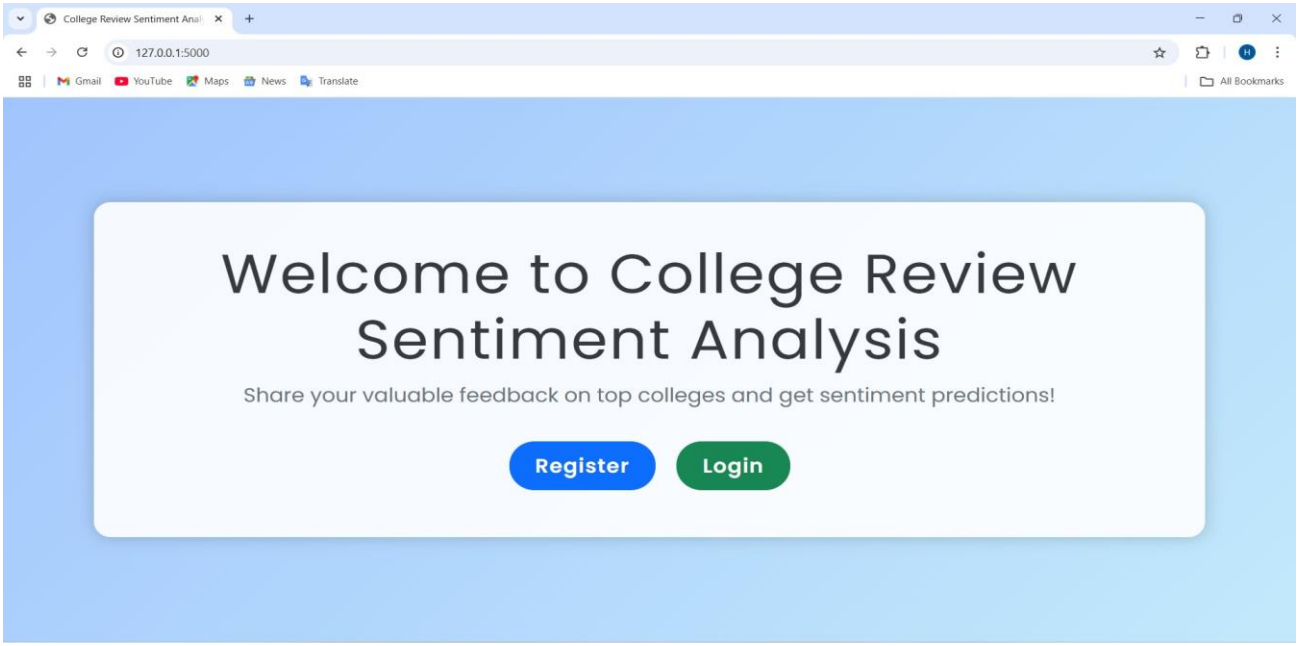


Fig 1: Website Login

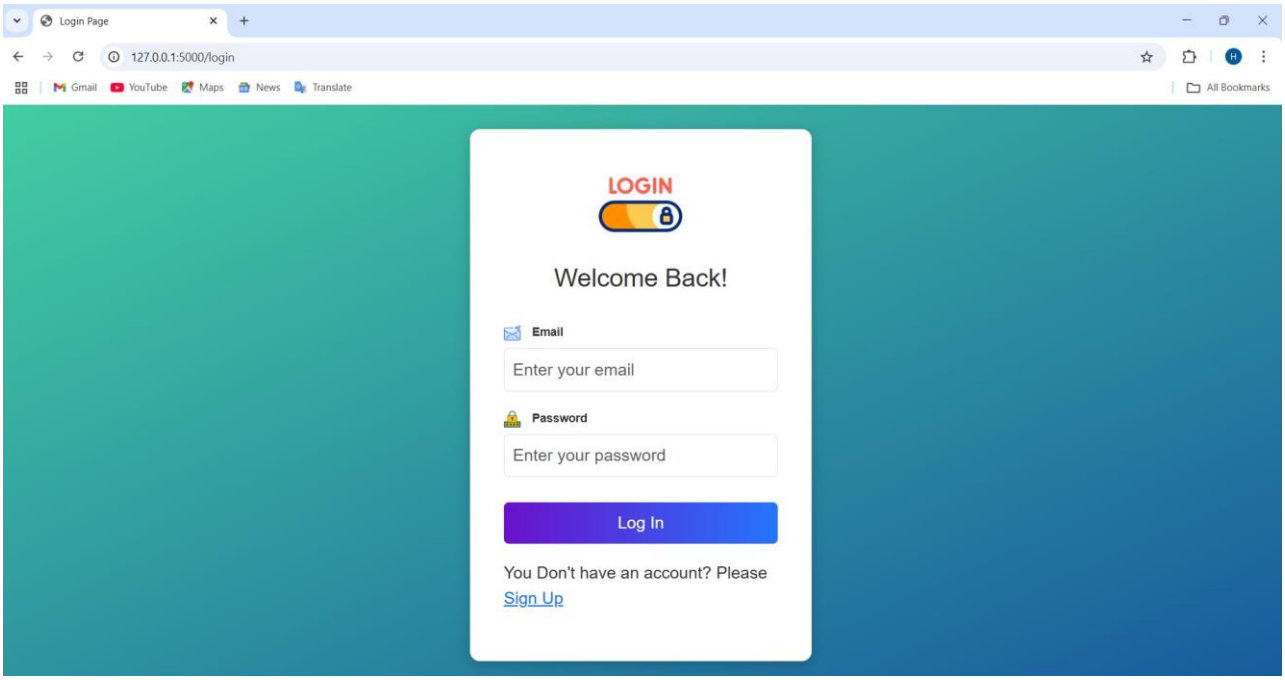


Fig 2: Enter through email and password

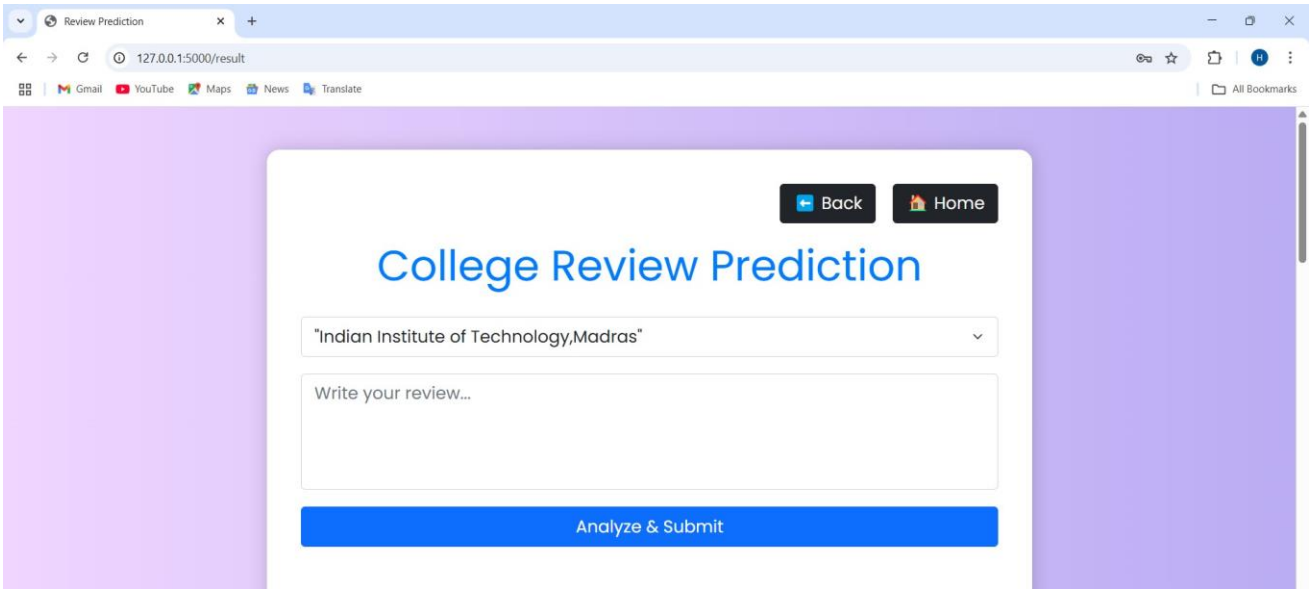


Fig 3: Selecting College and Entering Review

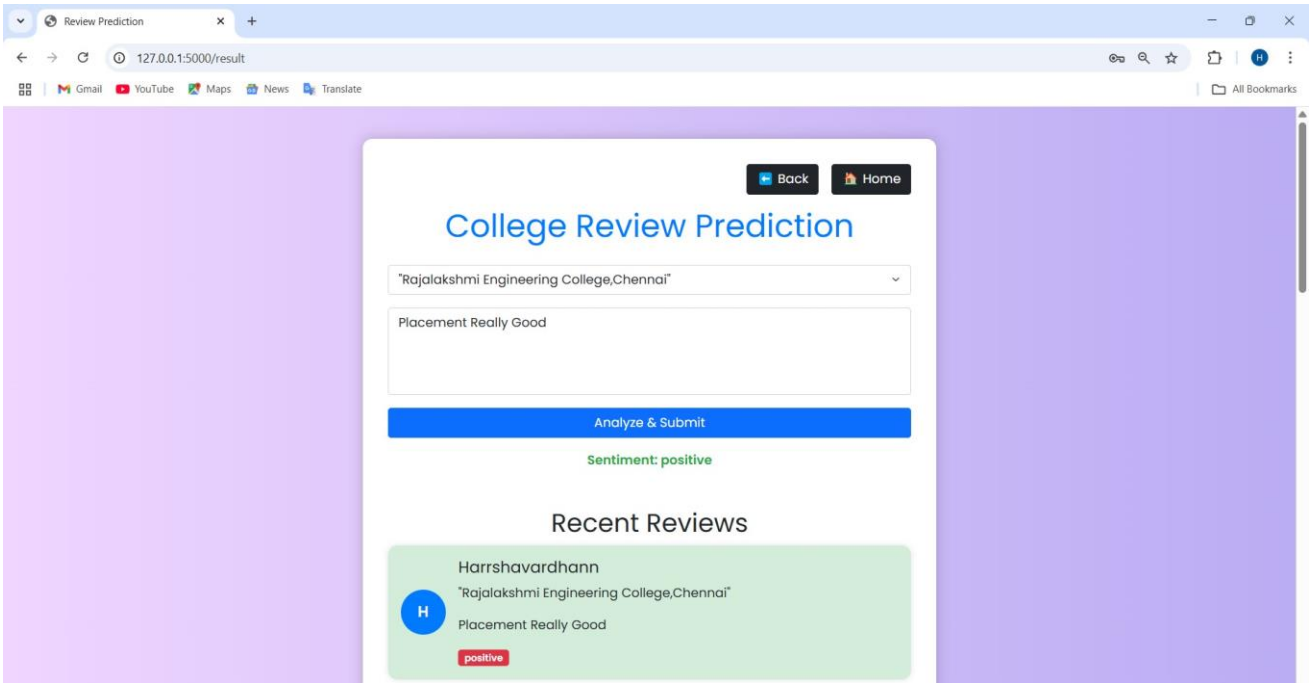


Fig 4:After Selecting College and Entering Review it gives whether the review is Positive,Negative,Neutral as shown above.

CHAPTER 7

CONCLUSION

The implementation of sentiment analysis for the website analyzing reviews of top engineering colleges in Tamil Nadu provides valuable insights into public perception. By leveraging Natural Language Processing (NLP) and Machine Learning (ML) techniques, particularly the Random Forest model, the system efficiently classifies user sentiments as positive, negative, or neutral. The use of a user-friendly Flask-based web application ensures seamless interaction, while the integration of MySQL enables secure storage and management of review data.

This sentiment analysis system offers numerous advantages, including real-time feedback analysis, enhanced decision-making, and brand reputation management. It empowers stakeholders, including students, college administrators, and policymakers, to make informed decisions based on accurate sentiment insights. The visual representation of sentiment trends further enhances clarity and understanding.

Although the system has limitations, such as challenges in understanding multilingual data and detecting complex emotions, proposed improvements like integrating multilingual support and adopting advanced deep learning models can enhance its effectiveness. Ultimately, this project serves as a robust tool for continuous feedback analysis, contributing to the growth and development of educational institutions by promoting transparency and data-driven decision-making.

REFERENCE

1. Pang, B., & Lee, L. (2008). "Opinion Mining and Sentiment Analysis." Foundations and Trends in Information Retrieval.
2. Liu, B. (2012). "Sentiment Analysis and Opinion Mining." Morgan & Claypool Publishers.
3. Kumar, A., & Sebastian, T. M. (2012). "Sentiment Analysis: A Perspective on its Past, Present, and Future." IJAET.
4. Medhat, W., Hassan, A., & Korashy, H. (2014). "Sentiment analysis algorithms and applications: A survey." Ain Shams Engineering Journal, 5(4), 1093-1113.
5. Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). "New avenues in opinion mining and sentiment analysis." IEEE Intelligent Systems, 28(2), 15-21.
6. Feldman, R. (2013). "Techniques and applications for sentiment analysis." Communications of the ACM, 56(4), 82-89.
7. Zhang, L., Wang, S., & Liu, B. (2018). "Deep learning for sentiment analysis: A survey." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.
8. Yadav, V., & Vishwakarma, D. K. (2020). "Sentiment analysis using deep learning architectures: a review." Artificial Intelligence Review, 53, 4335-4385.

Online Resources and Links

1. **NLTK Documentation** - https://www.nltk.org/
2. **Scikit-Learn Documentation** - https://scikit-learn.org/
3. **Flask Documentation** - https://flask.palletsprojects.com/
4. **MySQL Documentation** - https://dev.mysql.com/doc/
5. **Kaggle** - https://www.kaggle.com/
6. **Towards Data Science** - https://towardsdatascience.com/