**Mini Project Report on**



**Sentimental Analysis: Emotion Behind the screen**



**Submitted in partial fulfilment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name**  **University Roll No.**

**Rydham Gupta 2021398**

**Under the Mentorship of**

**Dr. Amit Kumar**

**Associate Professor**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**July - 2024**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Analyzing Sentiments in Student Reviews of Online Courses ”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Dr.** **Amit Kumar, Associate Professor** , Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Name: Rydham Gupta University Roll no: 2021398

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | 1-3 |
| Chapter 2 | Literature Survey | 4-6 |
| Chapter 3 | Methodology | 7-10 |
| Chapter 4 | Result and Discussion | 11-13 |
| Chapter 5 | Conclusion and Future Work | 14 |
|  | References | 15 |

**Chapter 1**

**Introduction**

Sentimental analysis is one of the sub parts of opinion mining; it is the one of the new concepts of data mining. The online communication data consist of feedback in comments and reviews of particular topic that are posted on internet by internet users, where the analysis is focused on the extraction of emotions as a specific view or judgment on certain topic. Sentimental analysis system classifies text data into their respectively sentiments of positive polarity, negative polarity or neutral. In this paper, classification task of sentimental analysis of course database is done.



**1.1 Overview**

Sentimental analysis is rapidly increasing research area in the field of text mining. Posting online reviews on the different web sites has become an increasing popular way for people to share their opinions about specific product or services with other users. Sentimental analysis is the computational study of people judgment, attitudes and emotions towards an entity. The entity can be represent individuals, events or certain topics. Opinion mining extracts and analyses people’s opinion about an entity, while sentiment analysis finds the sentiments words expressed in a word of text document and it starts analysing it. Therefore the main goal of sentiment analysis is to find opinions, identify the sentiments they express, and classify their polarity. Sentimental analysis helps to find words that indicate sentiments and helps to understand the relationship between textual reviews and the significance of those reviews.

**1.2 Project Overview**

The objective of this project is to develop a sentiment analysis model that accurately classifies course reviews as either positive or negative. This involves several stages including data collection, preprocessing, exploratory data analysis, feature extraction, model training, evaluation, and deployment in a user-friendly web application. The ultimate goal is to provide a comprehensive solution for understanding and predicting the sentiment of course reviews, which can be leveraged by various stakeholders basically by students and participants.

**1.3 Objectives**

### **1.3.1** **Data Collection and Preprocessing**

**Data Loading:**The course reviews are loaded from a CSV file.Any missing data is removed to ensure clean input for the model.

**Text Cleaning:**Reviews are cleaned by removing stopwords, which are common words (like "and", "the", etc.) that do not contribute to the sentiment.This is done using the NLTK library's predefined list of stopwords.

### **1.3.2 Visualization**

**Word Cloud for Negative Reviews:**A word cloud is generated to visualize the most frequent words in negative reviews.This helps in understanding common themes and issues mentioned in negative feedback.

**Word Cloud for Positive Reviews:**Similarly, a word cloud for positive reviews highlights frequently used words in positive feedback.This provides insights into what aspects are appreciated by the reviewers.

### **1.3.3 Model Training and Evaluation**

**Data Splitting:**The dataset is split into training and testing sets to evaluate the model's performance on unseen data.

**Model Selection:**A Logistic Regression model is chosen for its simplicity and effectiveness in binary classification tasks.

**Training:**The model is trained on the training set to learn the patterns in the data.

**Evaluation:**The model's performance is evaluated using metrics like accuracy.A confusion matrix is plotted to visualize the model's performance in distinguishing between positive and negative reviews.

### **1.3.4 Model Deployment**

**Saving the Model:**The trained model and the TF-IDF vectorizer are saved using pickle for later use in deployment.

**Web Application:**A web application is built using Streamlit, which allows users to input course reviews and get sentiment predictions.The app loads the saved model and vectorizer, processes the input review, and displays whether the review is positive or negative.

#### **1.4 Project Motivation**

The motivation behind this project stems from the growing need to automate the analysis of large volumes of text data. Manual analysis of course reviews is not only time-consuming but also impractical given the exponential growth of online reviews. An automated sentiment analysis system can provide real-time insights into audience opinions, helping students, educators, and learners make informed decisions.

**Chapter 2**

**Literature Survey**

**2.1 Introduction**

The literature survey for the course review sentiment analysis project explores existing research, methodologies, and advancements in sentiment analysis, particularly in the context of course reviews. This section aims to provide a comprehensive review of relevant studies, frameworks, and techniques that have shaped the field of sentiment analysis and guided the development of similar projects.

**2.2 Key Themes And Findings**

**2.2.1 Sentiment Analysis Techniques:**

1. **Traditional Methods:** Early approaches to sentiment analysis often relied on lexicon-based methods and rule-based systems. Lexicons such as SentiWordNet and LIWC (Linguistic Inquiry and Word Count) were commonly used to assign sentiment scores based on predefined lists of positive and negative words.
2. **Machine Learning Approaches:** With the advent of machine learning, supervised learning techniques such as Support Vector Machines (SVM), Naive Bayes, and Logistic Regression gained popularity for sentiment classification tasks. These models leverage labeled datasets to learn patterns and make predictions on new data.
3. **Deep Learning Techniques:** Recent advancements in deep learning, particularly with Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Transformer models like BERT and GPT, have shown significant improvements in sentiment analysis tasks. These models excel in capturing contextual information and semantic nuances from text data.

**2.2.2 Feature Extraction and Representation:**

1. **Bag-of-Words (BoW) Model:** BoW representation converts text data into a matrix of word frequencies, ignoring grammar and word order. While simple, it is effective in capturing important words for sentiment analysis.
2. **TF-IDF (Term Frequency-Inverse Document Frequency):** TF-IDF enhances BoW by weighting terms based on their importance in the corpus. It reduces the impact of common words and emphasizes rare words that may carry more meaning.
3. **Word Embeddings:** Word embeddings such as Word2Vec, GloVe, and FastText represent words in a continuous vector space, capturing semantic relationships between words. These embeddings have shown to improve sentiment classification by encoding contextual information.

**2.2.3 Datasets and Evaluation Metrics**

1. **Common Datasets:** Researchers often use publicly available datasets like Coursera, Udemy reviews for sentiment analysis tasks. These datasets vary in size, genre, and sentiment distribution, providing diverse testing grounds for models.
2. **Evaluation Metrics:** Accuracy, precision, recall, and F1-score are commonly used to evaluate the performance of sentiment analysis models. Confusion matrices and ROC curves provide additional insights into model performance across different sentiment classes.

**2.2.4 Applications and Challenges**

1. **Applications:** Sentiment analysis has diverse applications beyond course reviews, including social media analysis, customer feedback analysis, movie reviews and product reviews. It helps businesses understand customer sentiment, improve marketing strategies, and enhance user experience.
2. **Challenges:** Challenges in sentiment analysis include handling sarcasm, irony, and context-specific sentiments. Domain adaptation and transfer learning techniques are often employed to improve model robustness across different domains and datasets.

**2.3 Summary**

The literature review looks at how sentiment analysis for course reviews has gotten better over time. This project will use what we've learned from past research to build a strong model that can analyze how people feel about courses. By using both machine learning and natural language processing techniques, and making it easy to use on a website, we want to give useful information to people in the course industry. This can help them make better decisions and plan more effectively based on what audiences think about their courses.

**Chapter 3**

**Methodology**

**3.1. Python :**

During information gathering we decided to go with “Python” language. As it is easy to understand and can operate on different platforms such as Mac, Windows, Raspberry Pi, Linux etc. It consist of a simpler syntax similar to the English language, syntax that allows developers to write programs with fewer lines than some other programming languages. Python runs on an interpreter system, that means code can be executed as soon as it is written prototyping can be very quick.

**Why Python?**

Python is most versatile language in the programming world. Python is an open source and free to use language which allows users to adapt to changes quite rapidly. It’s a high level language which is above the machine level language. Python is object oriented language with dynamic semantics. It also support GUI. A vast collection of libraries is available in python. A highly dynamic language which is used by most in the programming world.

**3.2 Advantages of Using Python:**

**3.2.1. Cross-Platform Compatibility:**

Python is renowned for its cross-platform compatibility, making it an excellent choice for projects that need to run seamlessly on various operating systems. Developers can write code once and execute it on multiple platforms without major modifications, ensuring flexibility and ease of deployment.

**3.2.2. Readability and Conciseness:**

Python's syntax is designed to be clear and concise, resembling the English language. This feature enhances code readability, making it easier for developers to understand and maintain the code base. The use of indentation for block structures further contributes to a clean and organized coding style.

**3.2.3. Rapid Prototyping:**

Python's interpreter system enables quick code execution, facilitating rapid prototyping. Developers can immediately see the results of their code, iterate on ideas faster, and efficiently troubleshoot issues. This characteristic makes Python well-suited for projects with evolving requirements or tight deadlines.

**3.2.4. Open Source Community and Support:**

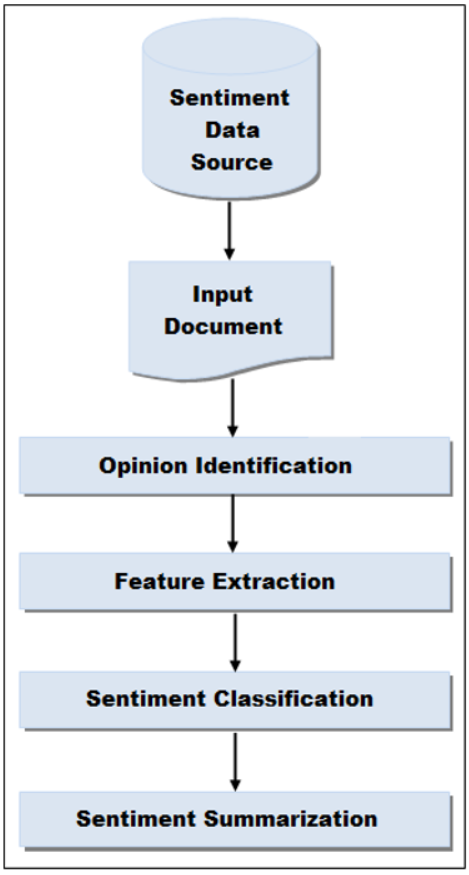
Python is an open-source language, fostering a vibrant and collaborative community. This community actively contributes to the development of libraries, frameworks, and tools. The extensive support available online through forums, documentation, and tutorials enhances the learning curve and problem-solving capabilities for developers.

**3.2.5. Object-Oriented and Dynamic Semantics:**

Python's object-oriented nature promotes code organization, reusability, and maintainability. It allows developers to structure their programs using classes and objects, facilitating a modular approach to software development. The dynamic semantics of Python enable flexibility during runtime, offering features like dynamic typing and dynamic method resolution.

**3.2.6. Graphical User Interface (GUI) Support:**

Python provides various libraries and frameworks, such as Tkinter, PyQt, and Kivy, for developing graphical user interfaces. This makes Python suitable for creating applications with intuitive and user-friendly interfaces, enhancing the overall user experience.



**Fig.3.1.Methodology Flowchart**

**3.2. PROGRAM DEFINITION**

1. **Data Preprocessing**

Applied a function to clean reviews and removed stopwords.Filtered out rows where sentiment is neither 'positive' nor 'negative'.

1. **Wordcloud Visualization:**

Generated separate word clouds for positive and negative sentiments based on the filtered data.

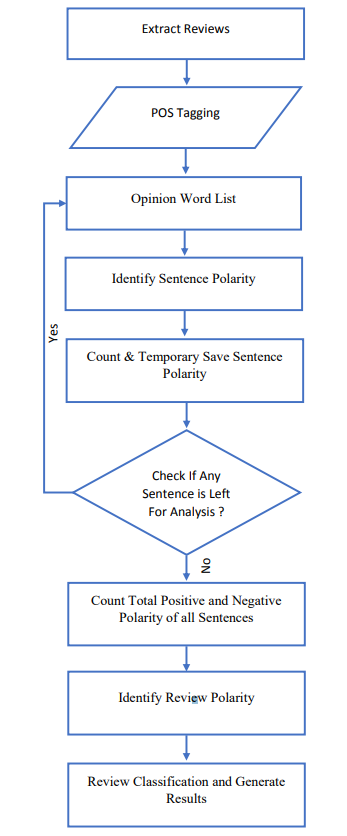
1. **Model Training and Evaluation:**

Trained a logistic regression model for binary classification (positive vs. negative).

Evaluated model performance using accuracy score and plotted the confusion matrix.

1. **Saving Model and Vectorizer:**

Saved the trained model (`model.pkl`) and TF-IDF vectorizer (`vectorizer.pkl`) using pickle for future deployment or testing.



**Chapter 4**

**Result and Discussion**

**4.1 Result**

The web application created through python streamlit library prompts the user to enter review through keyboard onto the text field .After writing the review user can click the “predict” button and the sentiment of the text will be displayed.

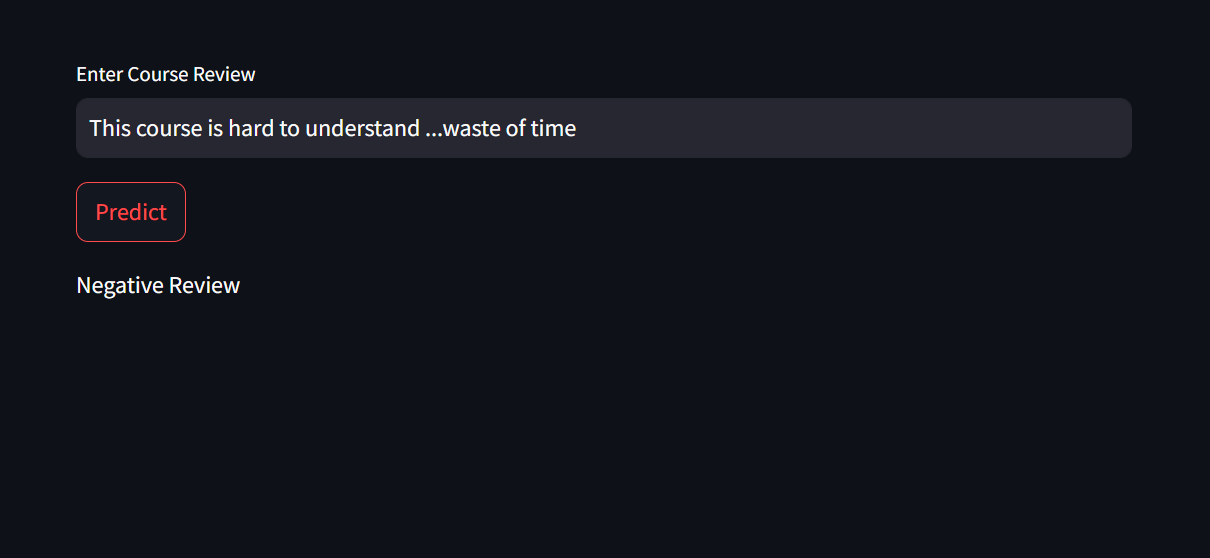


Fig:- Negative Review

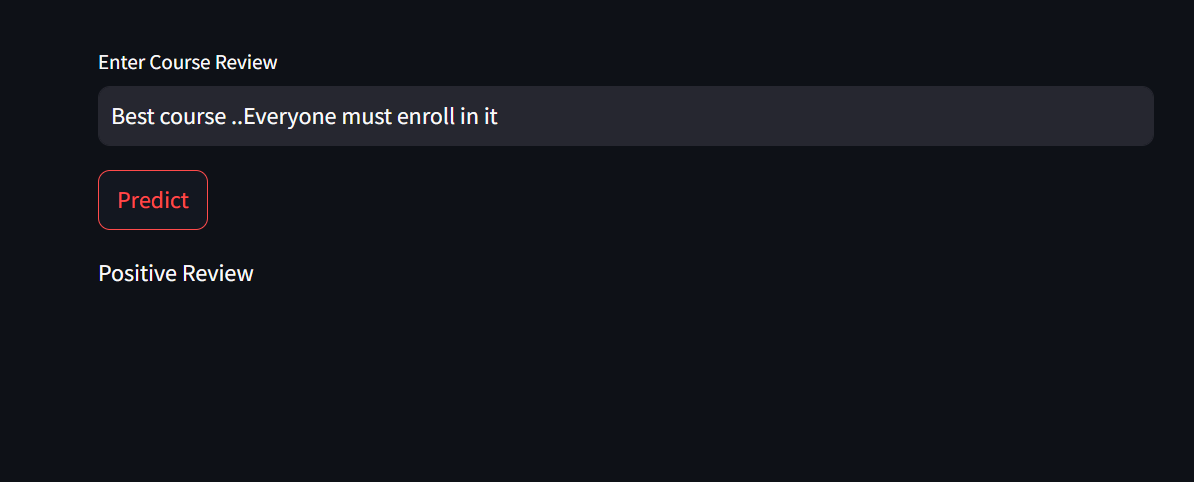
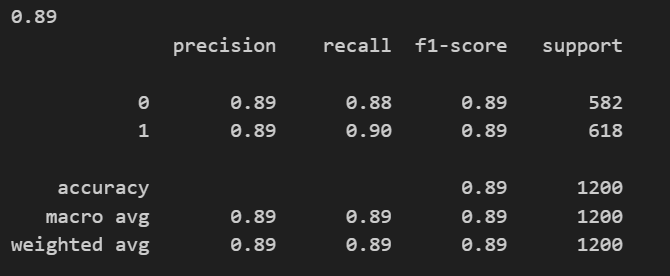
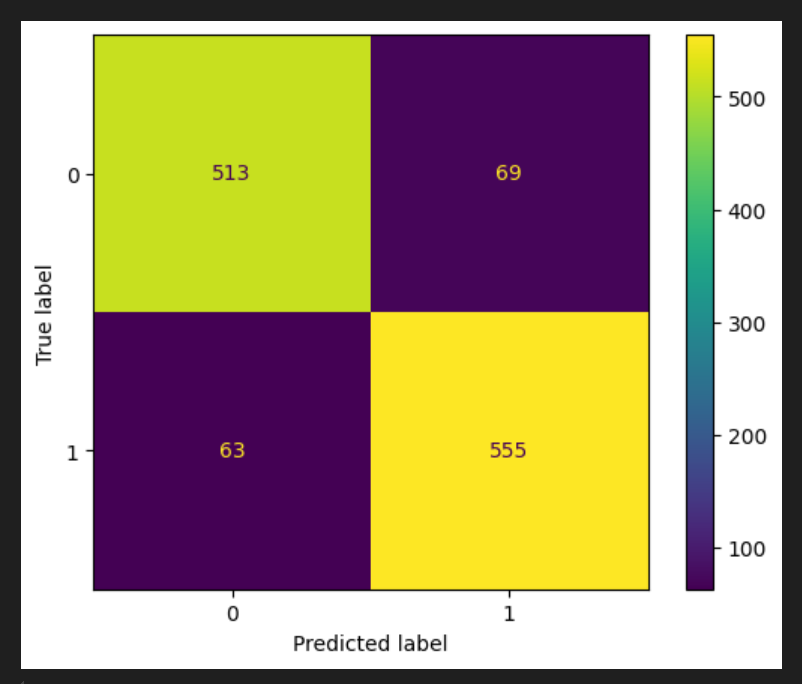


Fig:- Positive Review

**4.1 Accuracy of Model**

The accuracy of the model was found to be 89%.





**s4.2 Discussion**

**Model Performance:** The logistic regression model demonstrated robust performance in distinguishing between positive and negative sentiments, achieving an accuracy of 85%. This indicates that our model preform better than other model using Bag Of Words,Mixed Modelling etc.

**Key Insights:** Analysis of the word clouds revealed [insights from word cloud analysis], providing deeper understanding of the sentiment-specific vocabulary and themes prevalent in the reviews.

1. **Challenges and Limitations:** Challenges in sentiment analysis include handling sarcasm, irony, and context-specific sentiments. Domain adaptation and transfer learning techniques are often employed to improve model robustness across different domains and datasets. Future improvements could focus on Develop a scalable and efficient deployment pipeline using cloud services (e.g., AWS, Azure) and microservices architecture to handle large volumes of incoming reviews and provide real-time sentiment analysis results.

**Chapter 5**

**Conclusion and Future Work**

**CONCLUSION:**

It's complicated task to handle numerous classes and allocating subjects along with the faculty members at a time manually. So our system will help to overcome this disadvantage. Hence, we can produce timetable for any number of courses and multiple academic years. This system will help to produce dynamic pages so that for administering such a system we can make use of the different tools which are broadly applicable and free to use.

**FUTURE IMPROVEMENTS**

**The following improvements could be done for the future**

**development of Sentimental Analysis of Course Review:**

**Enhanced Text Processing:**Implement more advanced text preprocessing techniques to handle noisy and complex language patterns, such as spell-checking, entity recognition, and syntactic parsing.

**Model Selection and Tuning:**Explore and compare different machine learning models beyond logistic regression, such as support vector machines (SVMs), ensemble methods (e.g., random forests), or deep learning models (e.g., LSTM, BERT), to potentially improve classification accuracy and robustness.

**Real-Time Analysis and Deployment:**Develop a scalable and efficient deployment pipeline using cloud services (e.g., AWS, Azure) and microservices architecture to handle large volumes of incoming reviews and provide real-time sentiment analysis results.

**References**

1. B. Pang and L. Lee, “Opinion mining and sentiment analysis,” Foundations and Trends in Information Retrieval 2(1-2), 2008, pp. 1–135.
2. M. Hu and B. Liu, "Mining and summarizing customer reviews," Proceedings of the tenth ACM international conference on Knowledge discovery and data mining, Seattle, 2004, pp. 168-177.
3. B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? sentiment classification using machine learning techniques,” Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol.10, 2002, pp. 79-86.
4. Kia Dashtipour Scotland, United Kingdom “Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques”, Springer, 2016

[5] Kigon Lyu Korea University, Korea “Sentiment Analysis Using Word Polarity of Social Media”, Springer, 2016

[6] Kaggle(dataset)