A Project Report on

Developing ML-based Student Sentiment Analysis portal for Educational Institute

Submitted in partial fulfilment of the requirements for the award
of the degree of

Bachelor of Engineering

in

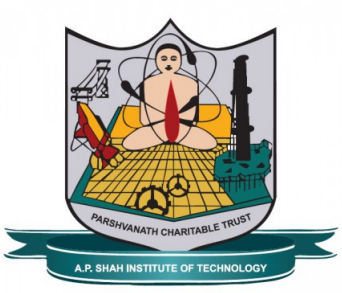
Information Technology

by

AbhayPratap Singh(19104037)
Aditya Joshi(19104044)
Bharat Singh(19104043)

Under the Guidance of

Prof. Rajashri Chaudhari



Department of Information Technology
NBA Accredited
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI

Academic Year 2022-2023

Approval Sheet

This Project Report entitled ***"Developing ML-based Student Sentiment Analysis portal for Educational Institute"*** Submitted by ***"AbhayPratap Singh"(19104037), "Aditya Joshi"(19104044), "Bharat Singh"(19104043),*** "is approved for the partial fulfilment of the requirement for the award of the degree of ***Bachelor of Engineering*** in ***Information Technology*** from ***University of Mumbai***.

(Prof. Rajashri Chaudhari)
Guide

Dr. Kiran Deshpande
Head Department of Information Technology

Place: A.P. Shah Institute of Technology, Thane
Date:

CERTIFICATE

This is to certify that the project entitled ***"Developing ML-based Student Sentiment Analysis portal for Educational Institute"*** submitted by ***"AbhayPratap Singh" (19104037), "Aditya Joshi" (19104044), "Bharat Singh" (19104043)*** for the partial fulfilment of the requirement for the award of a degree ***Bachelor of Engineering*** in ***Information Technology***, to the University of Mumbai, is a bonafide work carried out during the academic year 2022-2023.

(Prof. Rajashri Chaudhari)
Guide

Dr. Kiran Deshpande
Head Department of Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

1.

2.

Place: A.P. Shah Institute of Technology, Thane
Date:

Acknowledgement

We have great pleasure in presenting the report on **Developing ML-based Student Sentiment Analysis portal for Educational Institute** We take this opportunity to express our sincere thanks towards our guide **Prof. Rajashri Chaudhari** Department of IT, APSIT thane for providing the technical guidelines and suggestions regarding the line of work. We would like to express our gratitude towards her constant encouragement, support and guidance through the development of the project.

We thank **Dr. Kiran B. Deshpande** Head of Department, IT, APSIT for his encouragement during the progress meeting and for providing guidelines to write this report.

We thank **Prof. Sonal Jain** BE project co-ordinator, Department of IT, APSIT for being encouraging throughout the course and for guidance.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues at APSIT for their encouragement.

AbhayPratap Singh
19104037

Aditya Joshi
19104044

Bharat Singh
19104043

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

(AbhayPratap Singh - 19104037)

_____

(Aditya Joshi - 19104044)

_____

(Bharat Singh - 19104043)

Date:

Abstract

Emotional analysis of student feedback is a new field that uses NLP and ML techniques to analyze and classify students' emotions. The main aim is to provide schools with valuable information about students' understanding and behaviour in various areas of education, such as teaching quality, course content and locations. Sentiment analysis is a branch of NLP that involves identifying and classifying thoughts or emotions conveyed in positive, negative, or neutral text. This can be done by various methods such as binary classification, multi-class classification or ordinal classification. This analysis of opinions leads to a substantial estimate of all opinions expressed in the text. With the wealth of information produced by social media, it has become an important field of study with many applications such as product management, marketing, market research, and customer support. This study also examines some of the problems and limitations of emotional appraisal, as well as future research goals in this area.

Contents

List of Figures

List of Abbreviations

NLP	Natural language processing
SA	Sentiment Analysis
RFC	Random Forest Classifier
VAD	Valence Aware Dictionary
ML	Machine learning
KPI	Key Performance Indicator

Chapter 1

Introduction

In today's digital world, people express their thoughts and opinions on various topics through texts such as news, criticism, reviews and article blogs. [13]Sentiment analysis is a process that helps determine the emotion represented in this context. The purpose of opinion analysis is to evaluate whether the article is good, bad or unbiased. [14]Sentiment analysis, also known as sentiment mining, involves regularly analyzing the strength of sentiment in text, which can be classified as binary (positive or negative), multiclass, or nonstandard. It is an NLP course that aims to recognize and understand the thoughts or feelings expressed in a document. [15]Sentiment analysis is a process that aims to determine the thoughts or feelings conveyed in written texts. Specifically, it tries to determine whether the text has a positive, negative or neutral character. [12]Sentiment analysis provides insight into the thoughts and feelings of the author and/or audience by analyzing the language and content of a text. This is important in many applications such as market research, customer analysis and social media monitoring.

Covers a wide range of applications, including sentiment analysis, customer reviews, social media monitoring, and reputation management. Sentiment analysis is a multi-step process that includes pre-writing, feature extraction, and sentiment classification. In the text pre-processing stage, the text data is cleaned and normalized to remove irrelevant data and noise. Deleting actions involves accessing text in a way that machine learning algorithms can use to generate predictions. Knows the point of view of the text.

Machine learning-based techniques, on the other hand, rely on test models to show the machine learning model of previous data used to predict the hypothesis of the text. Deep learning uses neural networks to learn and recognize patterns and features in text that can be used to classify emotions. The emotional assessment uses a variety of methods, each with its own strengths and weaknesses. Opinion analysis of student feedback is the process of analyzing and classifying opinions expressed in student feedback using qualitative language techniques (NLP) and machine learning techniques. In general, schools are concerned about students' quality of teachers, classroom content, etc. It uses opinion polls to understand their views and attitudes towards various aspects of education, such as places and places. Sensitivity analysis of student feedback aims to explore specific concerns or areas of interest as well as the overall emotion in the feedback. Use a variety of methods, such as law, machine-based learning, and deep learning, to make an emotional assessment of student feedback. The process involves collecting feedback, first processing the data, and then using machine learning algorithms to determine if the feedback is positive, negative, or neutral. The rule-based approach uses a set of rules and heuristics to classify the appearances of text. [12]The

results of evaluating students' feedback can provide schools with important information on improving student learning. For example, research uncovers weaknesses in teaching quality, classroom content or infrastructure, allowing schools to take corrective action and improve the student health experience. In conclusion, emotional analysis of student feedback is an important tool for schools to understand students' thinking and attitudes. Using NLP techniques and advanced machine learning, schools can analyze large amounts of student feedback data and make decisions based on this data to improve student learning.

Chapter 2

Literature Review

The purpose of the literature review is to gain an understanding of the existing research on SA in relevant to the area of study.

- In [1], Emotions, according to Novelty Octaviani Faomasi Daeli, differ widely from person to person and might be unreasonable at times. When analysing a vast amount of relevant data, it is critical to consider emotions. Because an individual's judgement may be impacted by external influences, no single data point is more relevant than another. A person having a poor day, for example, may tweet a harsh comment on something they previously had no strong sentiments about. When analysing a big enough sample size, exceptions to feelings can be disregarded. Furthermore, because emotions can shift over time owing to personal situations or global events, it is critical to examine Twitter users' attitudes.

- In [2], Sanjeev Dhawan, Kulvinder Singh, and Priyanka Chauhan presented a dataset retrieval approach using the Twitter API to analyse user sentiment. Before beginning the sentiment analysis, the algorithm checks to see if the user is authorised. The programme analyses the first characteristics of each tweet and evaluates the sentiment polarity to analyse the sentiment. The tweet is neutral if the polarity is zero. If the polarity is higher than zero, the tweet is regarded as positive; otherwise, the tweet is considered negative. As a consequence, the suggested system categorises tweets depending on each user's sentiment polarity.

- In [3], R Raja Subramanian, Nukala Akshith, Gougula Narasimha Murthy, Manchala Vikas, Srikar Amara, and Karnam Balaji proposes a five-step emotional analysis process: data separation, prioritization, extraction, classification, and evaluation of results. Another logical thinking process is dictionary analysis, which involves the use of algorithms or hybrid methods. Machine learning algorithms are used to train different ideas and sort them into different groups. Probabilistic Multinomial Naive Bayes is used to assign each input to the appropriate group and examine the data for patterns. The IMDB dataset is an example of data that can be used for sentiment analysis.

- In [4], C.J. Hutto and Eric Gilbert discussed his research on the creation and testing of VADER, which stands for VAD for Emotional Reasoning. VADER aims to identify and measure sentiment in short content such as tweets. In their study, the researchers used a combination of quantitative and qualitative methods to improve the speech

process and their effect on emotions. This allowed them to develop a gold-standard collection of lexical features specifically tuned for sentiment analysis in Weibo-like content. In addition, they develop five general rules for the presentation and emotional development of the text, summarizing the grammatical and syntactic structure. By doing this, they can improve the accuracy and efficiency of the emotional appraisal process.

- In [5], Y. Chandra and A. Jana adopt a voting-based classification system where the polarity-based sentiment analysis is done on live tweets which are fetched using Twitter API. Various Machine Learning Classifier Algorithms were employed to assess performance, and their results were compared. A confidence measure has been used by combining several machine learning algorithms. The Machine Learning Models, Deep Learning Models or Polarity-based approaches are taught using features since feature engineering is at the heart of machine learning. They have used several combinations of deep learning models and compared the relative performance of different algorithms.

- In [6], Y. Woldemariam uses the tweeter dataset which is used in this study after going through many stages of pre-processing, cleaning, stemming, part-of-speech tagging, and tokenization. Support Vector Machine, Multinomial Naive Bayes, and K-Nearest Neighbor are the algorithms used for lexical-based sentiment analysis. The process is then further broken down into Data Pre-processing, Lexicon-based Algorithms like RNTN, and Stanford Sentiment Treebank, which offers a labelled parse tree for the entire analysis. When comparing the results of RNTN and Lexicon-based Analysis, RNTN performs worse than Lexicon-based Analysis.

- In [7], Bahrawi and Bahraw summarise the classification and analysis of Flipkart product reviews. The RFC is used in the study to categorise customer star ratings. This method's accuracy is compared to that of the Support Vector Machine algorithm, with Random Forest showing to be more successful. The measurement of the Receiver Operating Characteristic (ROC) curve for multiclass analysis is also included in the study.

- In [9], In this paper, they conclude that Random Forest is a good classifier for classifying the multi-class Kannada sentiments having an accuracy of 72 which is an improvement over the previous score which was 65 which was achieved using Naïve Bayes Classifier. Some limitations of the model developed using Naïve Bayes Classifier have also been addressed in this paper like handling multi-class labels, and identification of sentiment polarity of comparative and conditional statements. However, working on a large corpus of data is not possible due to a limited number of product review articles which results in lower accuracy as well.

2.1 Objectives

Objectives refer to specific goals which are aimed to achieve within a given period. These goals are usually specific, measurable, achievable, relevant, and time-bound and serve as a guide to focus efforts and resources towards a desired outcome. They provide direction and a clear sense of purpose for individuals and teams and are used to evaluate progress and success towards achieving a desired outcome.
We intend to do this project implementation to meet the following objectives:

- To accurately extract students' opinions from a large number of structured review texts and classify them into sentiment classes:
  This involves utilizing natural language processing techniques to identify and extract the relevant feedback and sentiment expressed within the text. The ultimate goal is to provide an accurate and meaningful analysis of the feedback to facilitate decision-making and improvements in the learning process.

- To read, analyze, understand and tag every single line of feedback:
  The task is to carefully read and examine each line of feedback provided, in order to fully comprehend the content and meaning of the feedback. This includes analyzing the language, identifying key points or issues mentioned, and assessing the overall sentiment of the feedback and should be tagged with relevant categories or themes to help organize and categorize the information for further analysis and action.

- To create a dashboard for the faculty to monitor the records using Flask:
  These will be done using the Flask framework and web designing technologies like HTML and CSS. Routes in the flask can be used to redirect Students and admins to their desired pages and provide the utmost usability of the interface to them. With the help of the interface, it would become easier for them to explore various functionalities in an easy manner.

Chapter 3

Project Design

The project's key features, structure, criteria for success and major deliverables are all planned out in this step. The aim is to develop the design in a way that can differ from the existing systems that can be used to achieve the desired project goals.

3.1 Existing System

- The existing system architecture for sentiment analysis of student feedback systems typically involves several components, including data collection from student feedback forms, pre-processing of the data, feature extraction, sentiment analysis using machine learning algorithms, and visualization of the results.
- The data collection process involves gathering feedback from students through various channels such as online forms, surveys, or interviews. The collected data is then pre-processed to remove noise, and irrelevant information, and to convert the data into a suitable format for analysis.
- Feature extraction is the process of identifying important features from the pre-processed data that can help in sentiment analysis. These features may include keywords, phrases, or other linguistic features that can help in identifying the sentiment of the feedback.
- Sentiment analysis is performed using machine learning algorithms such as Naive Bayes, Support Vector Machines (SVM), or Convolutional Neural Networks (CNN). These algorithms are trained on labelled data to classify feedback into positive, negative, or neutral categories.
- Finally, the results of sentiment analysis are visualized using various techniques such as bar charts, pie charts, or heat maps. These visualizations can help in identifying trends in student feedback and areas for improvement in educational institutions.

3.2 Proposed System

- Data Gathering - In this, the feedback dataset will be created by gathering data from different sources. This data will be used as the Input data.

- To build and test a sentiment analysis machine learning model, data is often divided into two groups: training and testing. [13]The training process is used to teach the machine learning algorithm to recognize patterns in data and make predictions based on these patterns. On the other hand, the test method is used to evaluate the performance of the training model by comparing its predictions with the known text in the test text. This strategy helps ensure that the model not only remembers the training data but also makes predictions on new data.

- Algorithm - Once the dataset is split, the next step is to determine which machine learning (ML) algorithm to use.

- Supervised - It is a machine-learning algorithm that trains a model using tags. A data list contains input data (also called attributes) and corresponding values (also called tags or targets). The model uses this data to learn the art of data entry and exit and ensures it is correct on new previously unknown data. Pre-processing and making it ready before it can use be used to perform the SA on it. The pre-processing generally includes understanding the words and sentences and is done with the help of NLP. NLP has the following phases:

  - Lexical Analysis - The initial step in NLP is to break down the text into its separate components, such as words, punctuation, and special characters. The goal of lexical analysis is to extract meaningful information from the text and to identify the relationships between words and phrases.
  - Syntactic Analysis - This is the second phase in NLP, and it assists in analysing language structure and determining sentence meaning. The phrase is split down into its basic pieces, such as nouns, verbs, adjectives, and so on, and the connections between these parts are determined in syntactic analysis. These data may then be utilised to deduce the sentence's meaning and extract information from the text.
  - Semantic Analysis - Semantic analysis is the third phase of NLP (NLP) that plays a crucial role in understanding the meaning of a sentence. It builds upon the results of the previous NLP phases, namely lexical and syntactic analysis, and establishes the connections between words and phrases to comprehend the sentence's overall meaning. Semantic analysis leverages context and the relationships between words and phrases to enable NLP models to accurately comprehend and interpret human language.
  - Pragmatic Analysis - The fourth and final stage of the NLP process involves pragmatic analysis, which helps determine the context in which a sentence is used to understand its meaning. This step is to determine the meaning of the sentence, taking into account many factors such as the speaker's purpose, the situation in which the sentence is being spoken, cultural background, doing, and leadership. Through pragmatic analysis, NLP models can interpret and understand real-world human speech.

- Algorithm Used - Random Forest is a powerful and popular machine learning algorithm that belongs to the family of ensemble learning methods. It is widely used for classification tasks, particularly when working with high-dimensional data and complex relationships between features.

- Working of Random Forest Classifier - The algorithm works by constructing multiple decision trees using subsets of the features and data points and then combining them to make a final prediction. The trees are built using a bootstrap sample, meaning that some data points may be selected multiple times, and feature selection is randomized, meaning that different features may be considered for each tree. The final prediction is made by taking the majority vote on the predictions made by each tree. This method of combining multiple decision trees helps to reduce the risk of overfitting and can produce more accurate predictions than using a single decision tree. Additionally, Random Forest can be used to assess feature importance and select the most relevant features for the classification task.
- Polarity of the feedback - Now that all the pre-processing and classification algorithms have been applied it is time to get the results. Because RFC is a classification algorithm we will get one of the three values from one student feedback -1, 0 and 1 which means Negative, Neutral and Positive respectively.

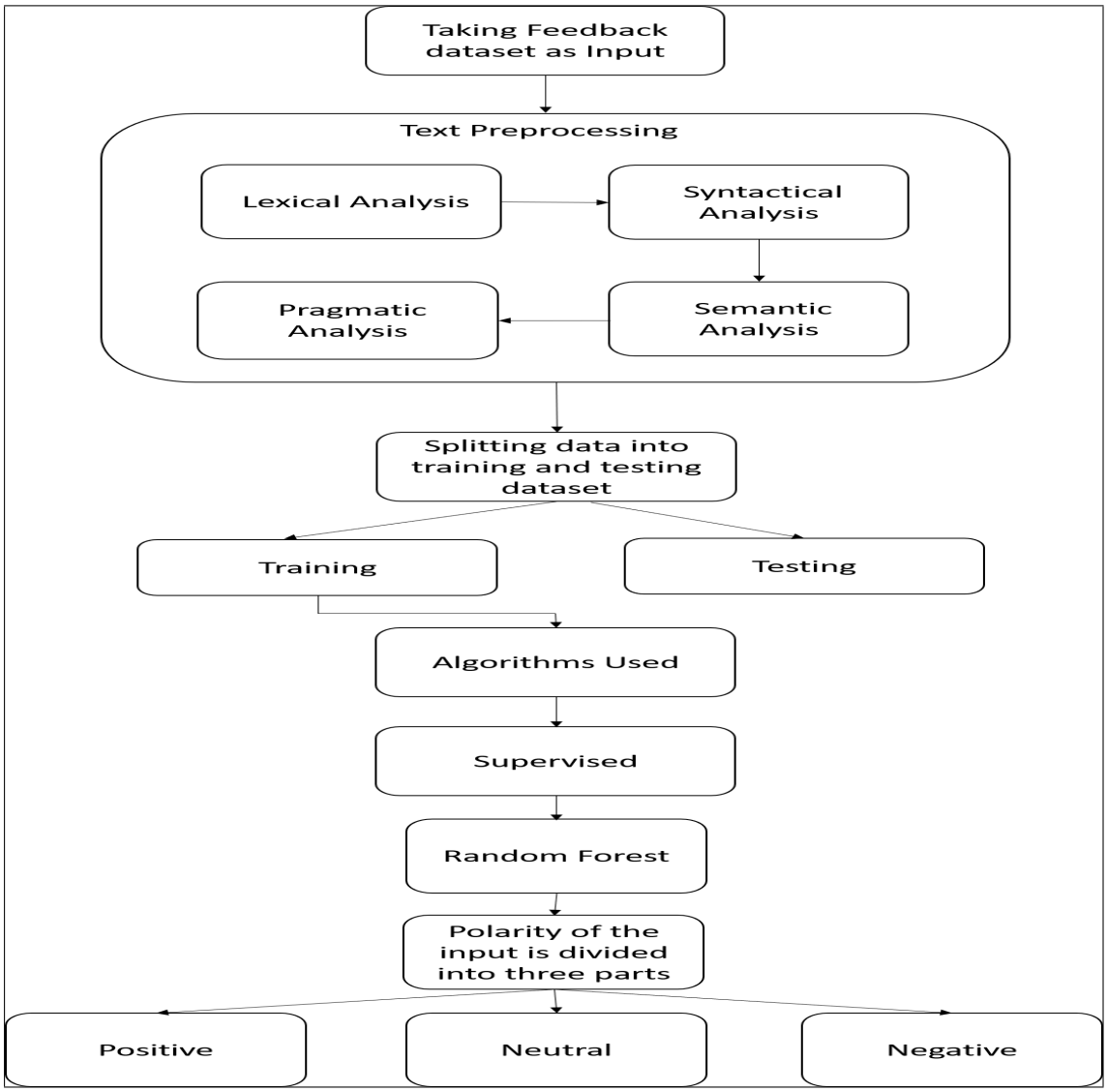


Figure 3.1: Activity Diagram

3.3 System Diagram

A system diagram is a graphical depiction of the various components that make up a system and the relationships between them. It helps to visualize the overall structure of the system, its processes, inputs, outputs, and the flow of data or materials. System diagrams can be used to design, analyze, or communicate complex systems, including engineering systems, business processes, software applications, and more. They provide a high-level view of the system, helping readers to understand its architecture and identify potential problems or opportunities for improvement.

- Activity Diagram

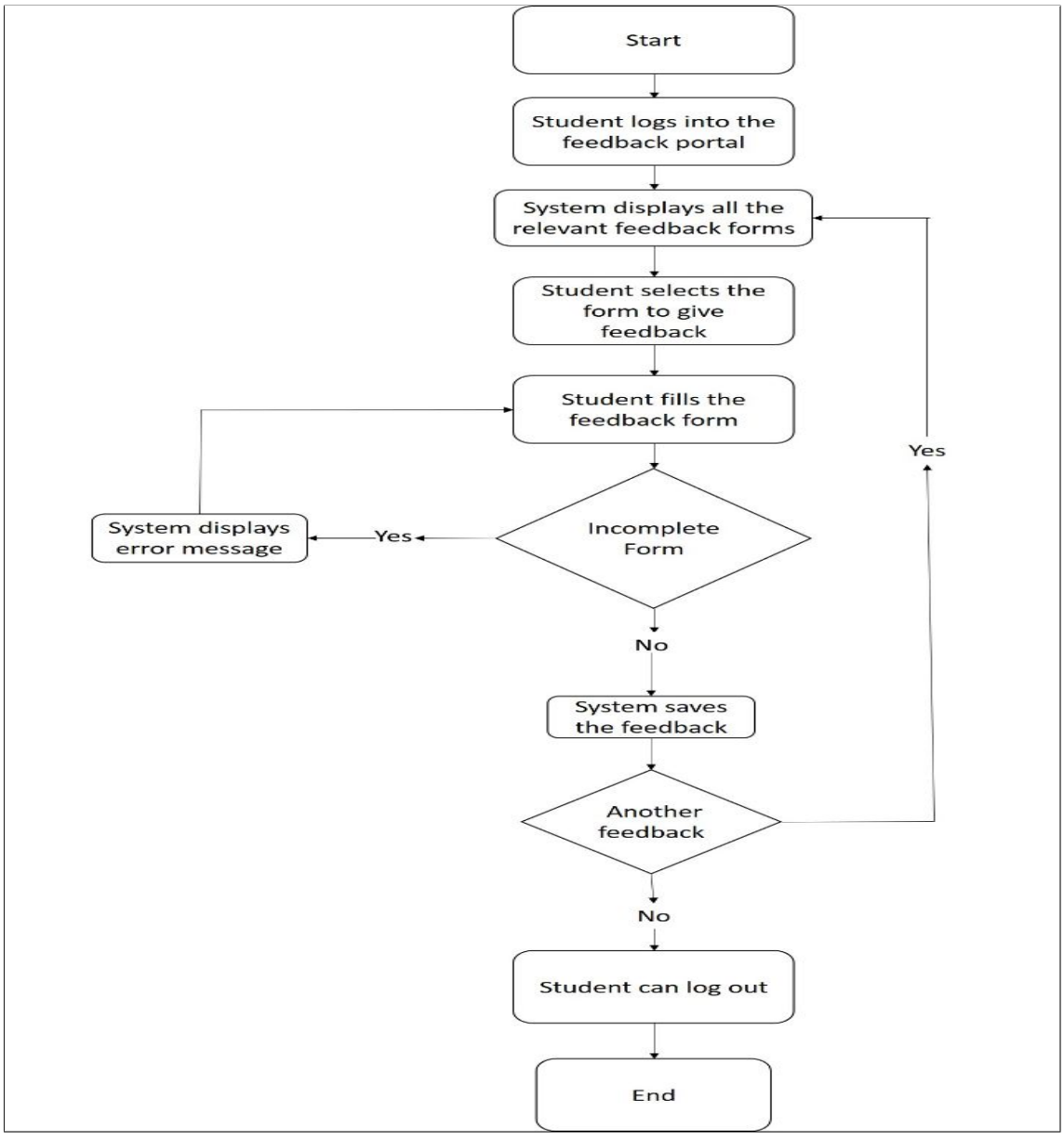


Figure 3.2: Activity Diagram

The figure shown is the complete system diagram of the proposed system. First, the Student will log in by entering their credentials. The Student can register for the events of their subjects and give feedback for the same. The feedback will then be pre-processed and applied to the RFC which will predict and give it a sentiment value of 1, 0, -1 which means Positive, Neutral and negative respectively.

- Use Case Diagram
  A use case diagram is a visual representation of actors, use cases, and their interactions in a system, used to describe system functionalities and requirements.

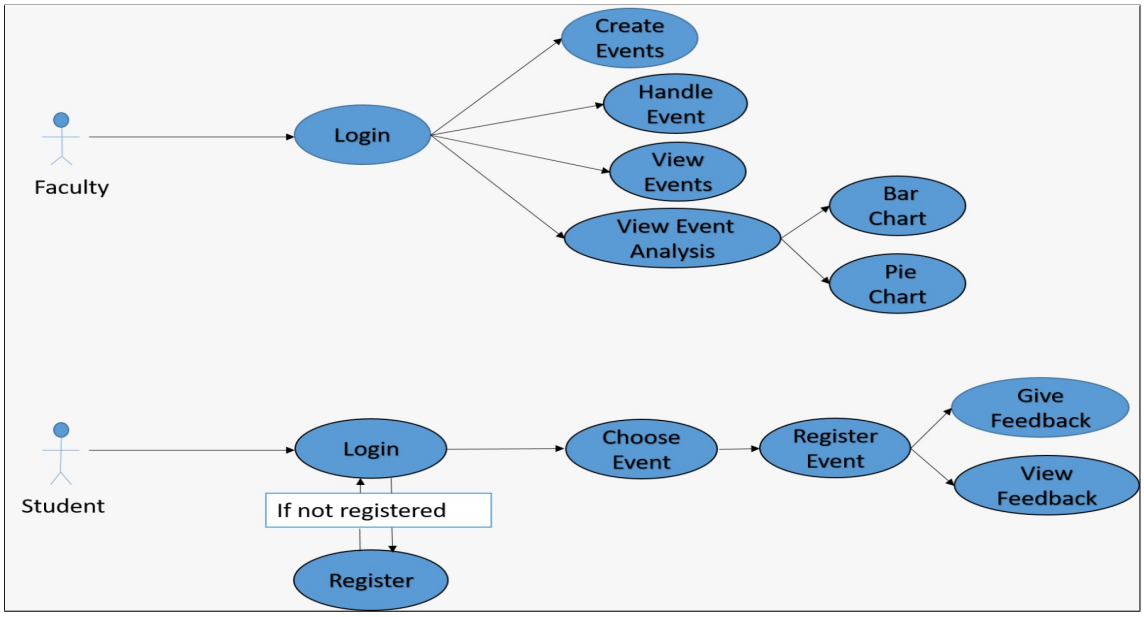


Figure 3.3: Usecase Diagram

In this use case diagram, there are three main actors: student users, faculty users, and the feedback analysis system.
The student can log in to the portal and register for an ongoing event whose registration will be opened by the faculty and submit their feedback through the "Submit Feedback" use case. The system will then perform sentiment analysis on the feedback to determine the overall sentiment. The faculty user can view the feedback through the "View Feedback" use case, and they can also access the feedback analysis data through the View Event use case. They can also create new events and view all the created events and also open and close the registration for the existing events through the faculty page. The feedback analysis system performs sentiment analysis on the submitted feedback through the "Sentiment Analysis" use case. The analysis performed by the system will be displayed in the form of various types of plots like bar charts, and pie plots which will make it easier for the faculty to understand the overall sentiment of the students regarding the event.

- Sequence Diagram
  The sequence diagram outlines the data flow in the system. First, students provide feedback for the courses they are enrolled in. The input data is pre-processed using NLP techniques to understand words and sentences. Once pre-processing is complete, the feedback is classified into positive, neutral, or negative categories.

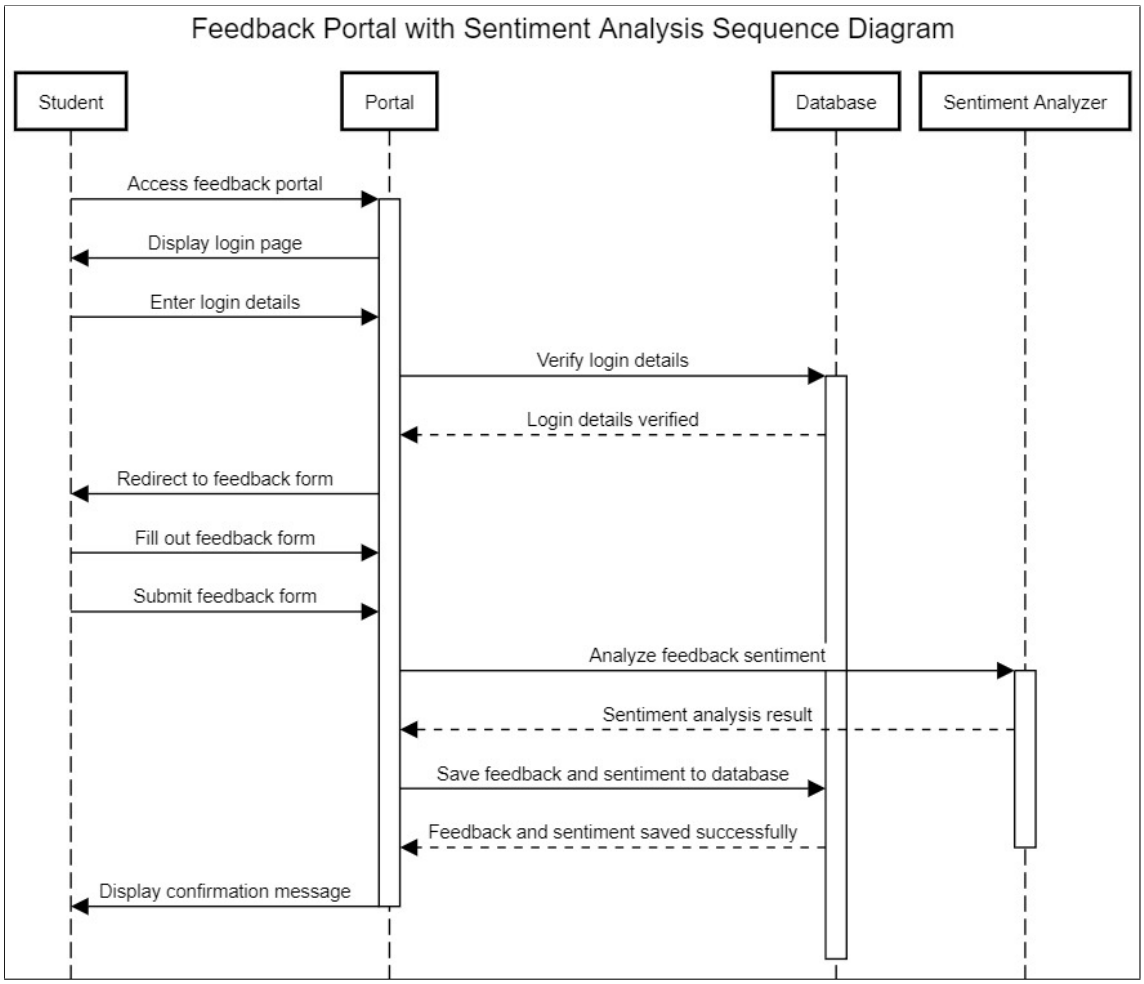


Figure 3.4: Sequence Diagram

The classification algorithm outputs one of three values (-1, 0, or 1) based on the feedback. Finally, the results are displayed in graphical form using bar graphs and pie charts, providing a clear representation of the feedback. To improve the efficiency of the system, we could consider using automated pre-processing tools, real-time feedback analysis, machine learning algorithms for classification, advanced data visualization techniques, and optimizing system resources.

Chapter 4

Project Implementation

Project implementation consists of visions and plans with which we are supposed to build the end product. This includes the logical conclusion, after evaluating, deciding, visioning, planning and finding the other resources for the project. Technical implementation is one of the major aspects of executing a project.

All imported modules like pandas, numpy, flask, matplotlib, etc., are required by the model in Fig. 4.1.

```
import pandas as pd
import nltk
from nltk import word_tokenize
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import nltk
import gensim
from gensim.utils import simple_preprocess
from gensim.parsing.preprocessing import STOPWORDS
```

Figure 4.1: Imported Modules

Here the Stop words are being removed from the Dataset and also the Student Input that we will be getting in Fig 4.2

```
# Obtain additional stopwords from nltk
from nltk.corpus import stopwords
stop_words = stopwords.words('english')
stop_words.extend(['from', 'subject', 're', 'edu', 'use']) # add additional stop words to remove



# Remove stopwords and remove words with 2 or less characters using gensim
def preprocess(text):
    result = []
    for token in gensim.utils.simple_preprocess(text):
        if token not in gensim.parsing.preprocessing.STOPWORDS and len(token) > 3 and token not in stop_words:
            result.append(token)
    return result


df['clean'] = df['feedback'].apply(preprocess)
```

Figure 4.2: Stop Words Removal

Here the splitting of the Dataset in Training and Testing parts is happening for the Model building in Fig 4.3

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(df.feedback, df.sentiment, test_size = 0.2)
```

Figure 4.3: Splitting the Dataset into 2 parts i.e. Training and Testing

TFIDF Vectorizer is used after all the pre-processing is done on the texts to give it a score between 0 to 1 such that it tells us about the importance of it in that particular Student Feedback and saving it in a Pickle File in Fig 4.4

```
from sklearn.feature_extraction.text import TfidfVectorizer
tfvect = TfidfVectorizer(stop_words='english',max_df=0.7)
tfid_x_train = tfvect.fit_transform(x_train)
tfid_x_test = tfvect.transform(x_test)
import pickle
with open('tfidf.pickle', 'wb') as f:
    pickle.dump(tfvect, f)
```

Figure 4.4: TFIDF Vectorizer in use

Applying RFC and getting the Accuracy of the model and saving the model afterwards in a Pickle File in Fig 4.5

```
from sklearn.ensemble import RandomForestClassifier
rfc_classifier=RandomForestClassifier()
rfc_classifier.fit(tfid_x_train,y_train)
y_pred = rfc_classifier.predict(tfid_x_test)
score = accuracy_score(y_test,y_pred)
rfc_score = round(score*100,2)
print(f'Accuracy: {rfc_score}%')

import pickle
with open('RFC.pickle', 'wb') as f:
    pickle.dump(rfc_classifier, f)
```

Figure 4.5: Model Creation and Model Saving

Here, is the home page of the Feedback system which has 2 parts Admin Side and the Student Side as shown in Fig 4.6

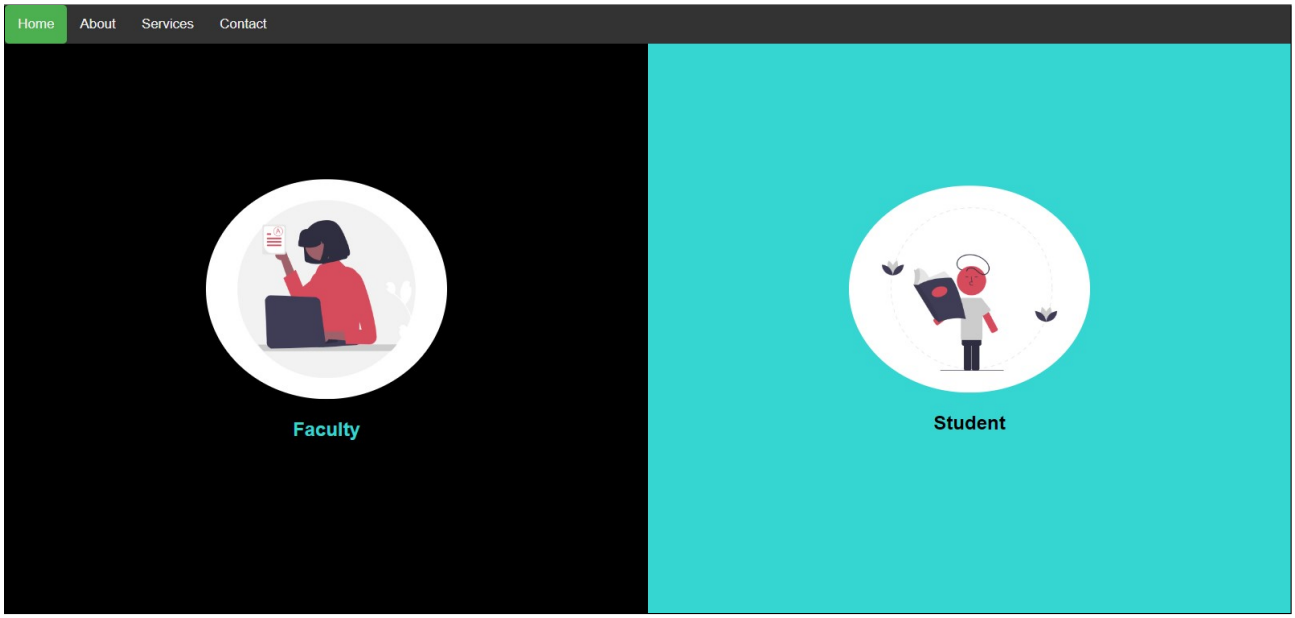


Figure 4.6: Home Page

This is the Admin login page where an Admin can use valid credentials and access the system as shown in Fig 4.7

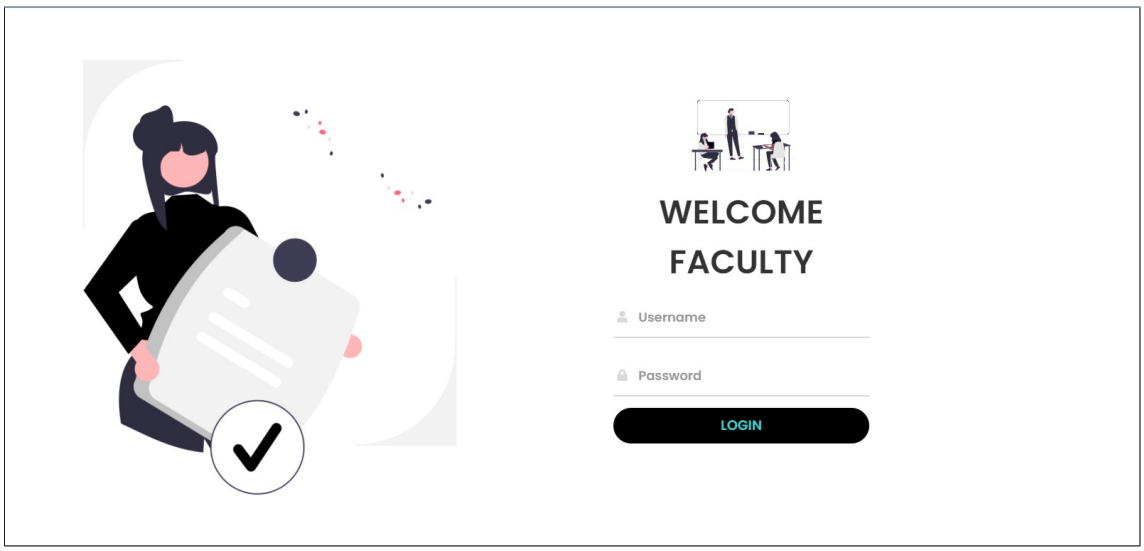


Figure 4.7: Admin Login Page

This is the Home Page of the Admin where in he can create a New Feedback Form for a Seminar or Event in Fig 4.8

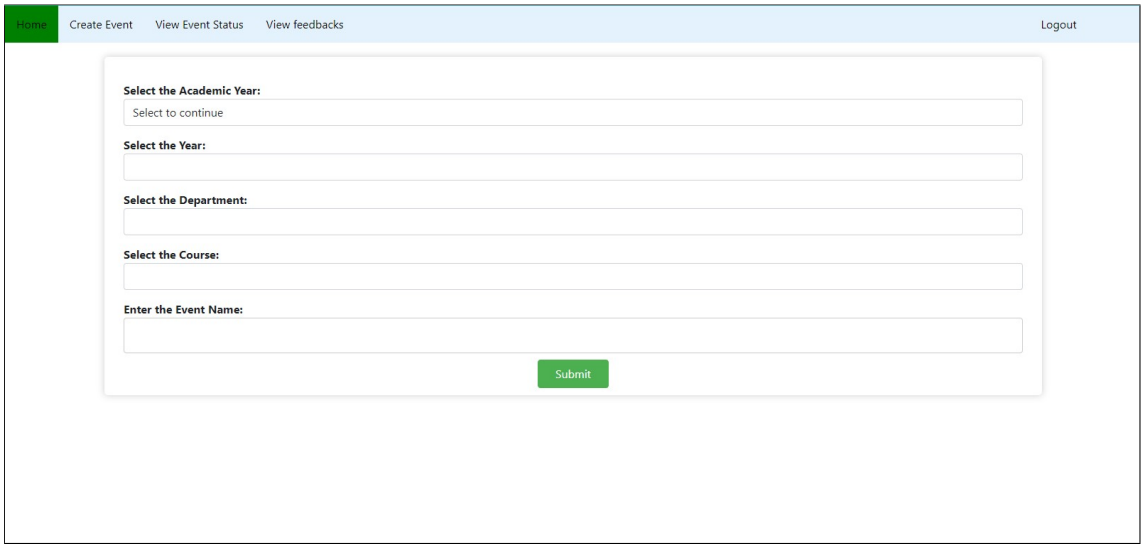


Figure 4.8: Feedback Form Creation by Admin

Admin can handle the Feedback Form in different ways such as Closing the Registration for the Form or even Deleting the Feedback Form in Fig 4.9

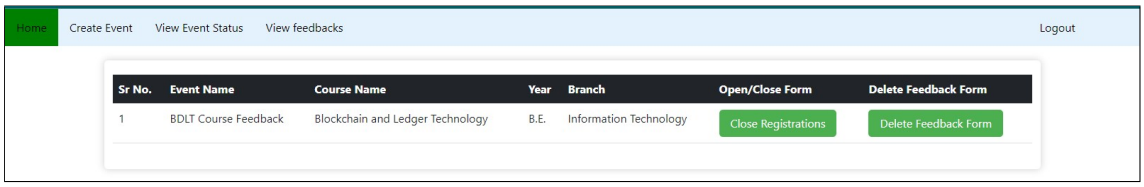


Figure 4.9: Admin Action on Feedback Form

Admin can see the Review of every Form and have different types of graphs to visualise the data for that Form or Event in Fig 4.10

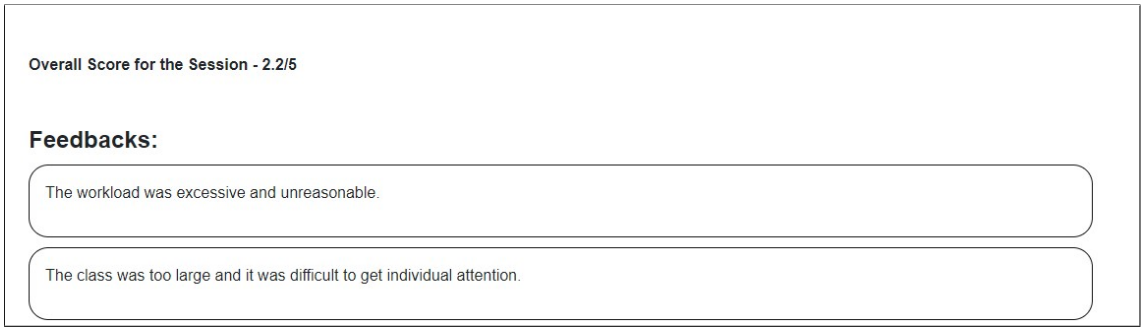


Figure 4.10: Reviewing the Feedback Form

This is the Student login page where a Student can use valid credentials and access the system as shown in Fig 4.11

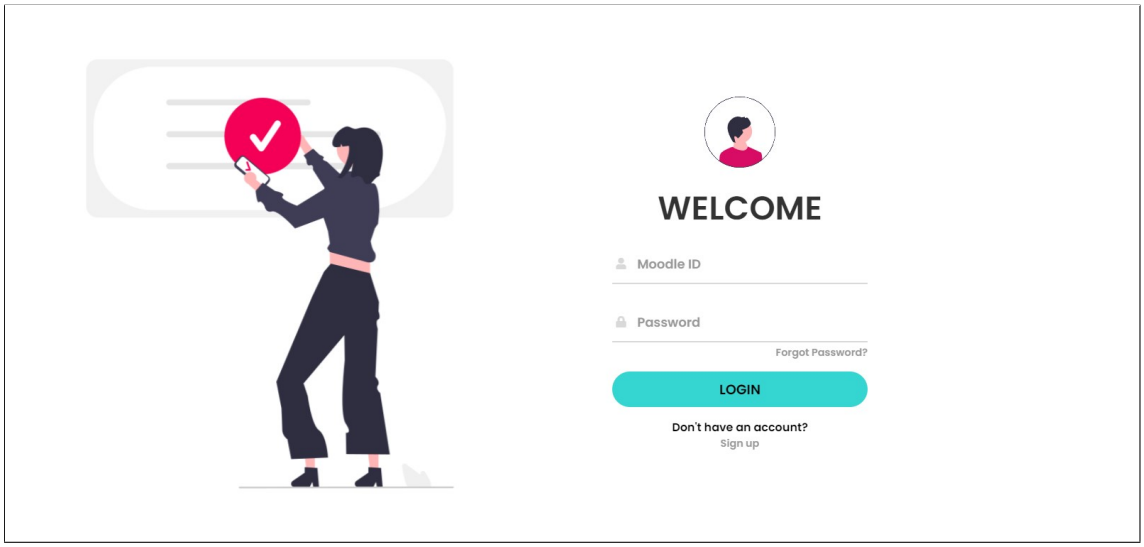


Figure 4.11: Student Side Login

Students can view all the Feedback Form that have been opened for them and Register themselves for the same in Fig 4.12

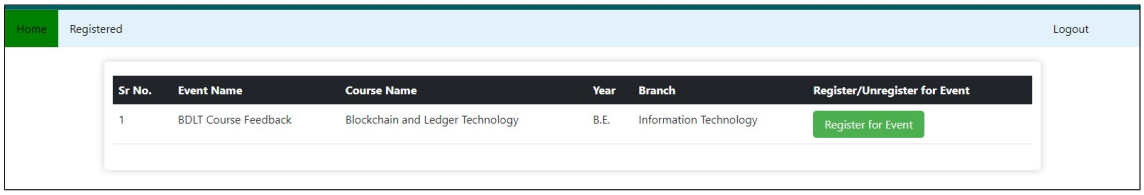


Figure 4.12: Student Home Page and Feedback Form Registration

Feedback Form with the Questionnaire for the Event or Seminar in Fig 4.13

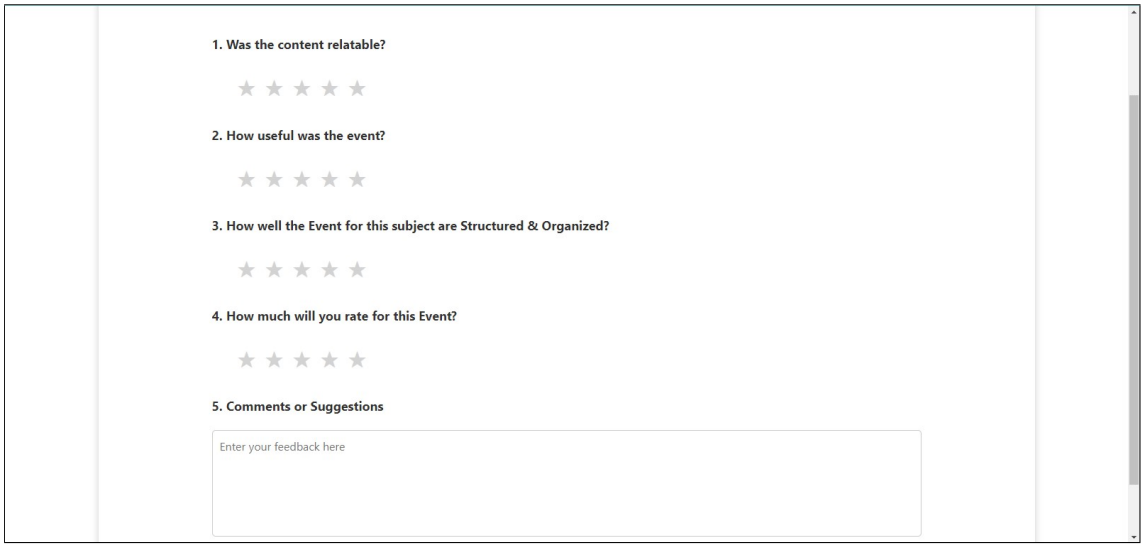


Figure 4.13: Feedback Form with the Questionnaire

Chapter 5

Testing

Testing is an organized summary of testing objectives, activities, and results. It is created and used to help stakeholders (product managers, analysts, testing team, and developers) understand product quality and decide whether a product, feature or defect resolution is on track for release. Test documentation includes all files that contain information on the testing team's strategy, progress, metrics, and achieved results. The combination of all available data serves to measure the testing effort, control test coverage, and track future project requirements.

5.1 Functional Testing

5.2 Unit Testing

Unit testing is the first level of testing, which is typically performed by the developers themselves. It helped us understand the desired output of each module, which we had broken down into separate units and in classifying the Student feedbacks on the basis of the algorithm that we have used.

5.3 Various Testcases

Here are some possible test cases for a sentiment analysis system:

Test Case	Sentence	Expected Result	Actual Result
Positive	I absolutely loved this Seminar	Positive	Positive
Negative	The Resources were terrible, I hated it	Negative	Negative
Neutral	Seminar was good and resources were terrible.	Neutral	Neutral

Chapter 6

Result

A machine learning-based student feedback portal can provide insight and improve learning outcomes. Such systems can help teachers identify areas of improvement, improve experiences, and increase student engagement by harnessing the power of data analysis and predictive modelling. However, it is important to remember that the success of the system depends on the quality and quantity of data collected, the accuracy and stability of the algorithms used, and the ethics of student information. Therefore, it is necessary to ensure that transparency, integrity and confidentiality are taken into account when designing and using the system. Overall, a well-designed machine-based student feedback portal can be a valuable tool for teachers to enhance learning, successfully encourage student action, and support the development of continuing education.

6.1 Pie Plot

A pie chart is a way of summarizing a set of nominal data or displaying the different values of a given variable. The Negative, Positive and Neutral Sentiments of the Feedback are taken and given plotted on the Pie Chart and the Percentage of each is shown in Fig 6.1. Pie Plots are a great way to represent a set of data and how they are divided into a set of values.

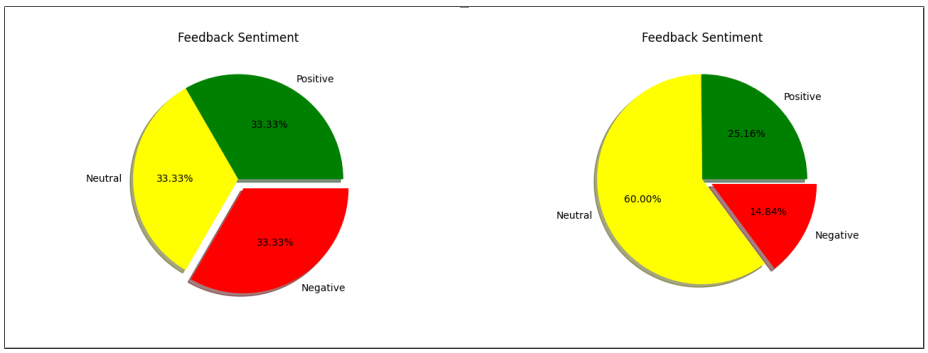


Figure 6.1: Multiple Pie Plots that shows Data Distribution

6.2 Bar Chart

A Bar chart is a diagram that uses narrow bands of different heights to show different amounts so that they can be compared. It gives the count of Negative, Positive and Neutral in the whole Feedback Form in Fig 6.2. It gives a better understanding of the Different types of values in a set and can be compared easily.

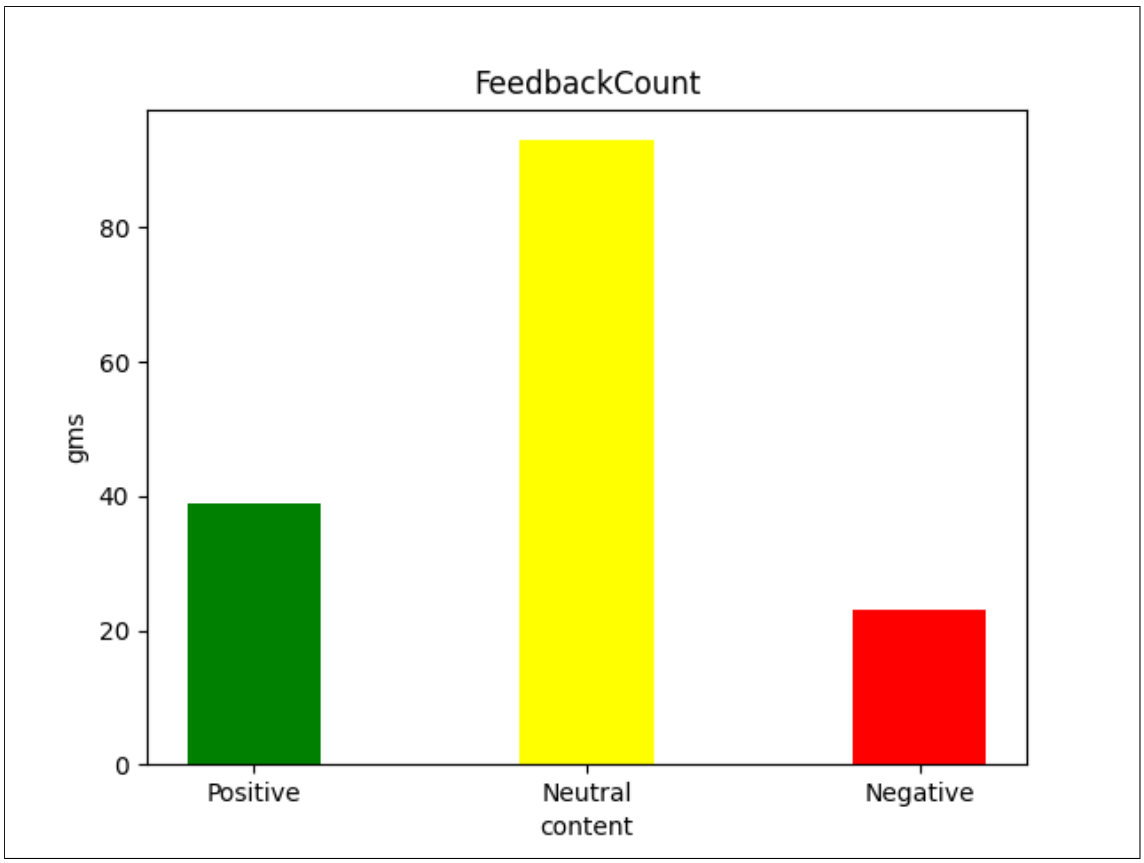


Figure 6.2: Bar Graphs Showing the Count of Positive, Negative and Neutral Feedbacks

6.3 Gauge Meter

A Gauge Meter is a data visualization type used to display a single value of data in a quantitative way. By using a coloured data range, gauge charts enable us to understand the progress of a KPI against a set target. It helps to get an overall Sentiment from the Feedback form that the Admin has chosen.

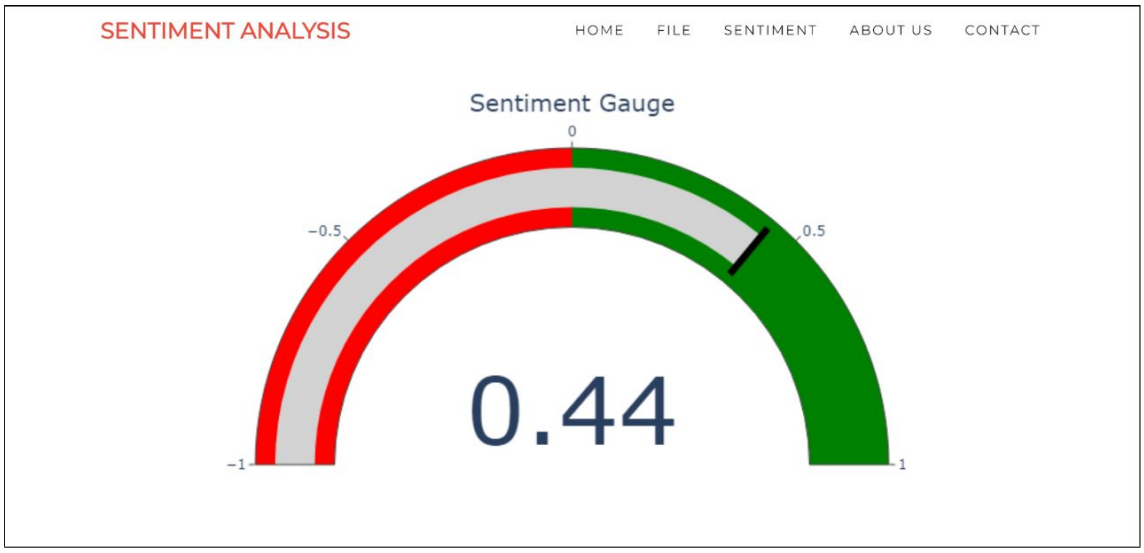


Figure 6.3: Gauge meter to give the overall Sentiment of the Event

6.4 Feedbacks and Overall Score

The last that the Admin can look at is the Actual Feedback that the Students have given but without the identity of that Student which maintains anonymity. It represents the Feedback that all the Students have given and an Overall Score out of 5 from the whole Student Feedback.

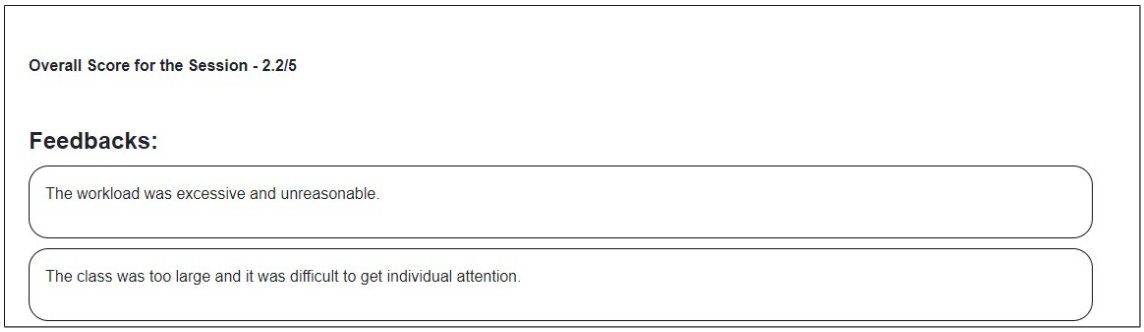


Figure 6.4: All the Comments and Overall Rating of the Event is shown

Chapter 7

Conclusions and Future Scope

In conclusion, a student feedback portal is designed that utilizes sentiment analysis can provide valuable insights into the overall satisfaction of students with their educational experience. By analyzing the sentiment of student feedback, administrators can quickly identify areas that need improvement and take action to address any issues. This can lead to improved student satisfaction, increased retention rates, and a better reputation for the educational institution. However, it's important to note that sentiment analysis is not a foolproof method of analyzing feedback. It can be influenced by factors such as sarcasm, irony, and cultural differences. Additionally, it's important to consider that students may not always provide honest feedback, or may not feel comfortable expressing negative sentiments in a public forum. Overall, a student feedback portal that incorporates sentiment analysis can be a useful tool for educational institutions to gain insights into the experiences of their students. However, it should be used in conjunction with other methods of data collection and analysis to provide a more comprehensive understanding of student satisfaction.

In future work, it would be interesting to explore ways to overcome the challenges and limitations that may be to improve Text Pre-processing which involves the development of advanced techniques to handle complex and unstructured data, including slang and misspelt words. This can lead to more accurate analysis and interpretation of the text data. Multi-Modal Analysis will help us Integrate multimedia data, such as images or audio, into sentiment analysis models to offer a comprehensive analysis of student feedback. Taking Personalized Feedback which will provide personalized feedback to students based on their feedback data is an exciting area of research in the field of sentiment analysis. By analyzing each student's feedback data, NLP models can identify specific areas where the student may need additional support or resources, and provide personalized feedback and recommendations to help them improve their learning experience. This can ultimately lead to better academic performance and greater student satisfaction. Automated feedback generation with the help of an algorithm that generates feedback using student feedback data and assists them in saving time and effort.

Bibliography

[1] Daeli, N. O. F., & Adiwijaya, A. (2020, May 11). Sentiment Analysis on Movie Reviews using Information Gain and K-Nearest Neighbor. Journal of Data Science and Its Applications, 3(1), 1-7. https://doi.org/https://doi.org/10.34818/jdsa.2020.3.22.

[2] S. Dhawan, K. Singh and P. Chauhan, "Sentiment Analysis of Twitter Data in Online Social Network," 2019 5th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India, 2019, pp. 255-259, doi: 10.1109/ISPCC48220.2019.8988450.

[3] Alexander, V., Mantzaris., Randyll, Pandohie., Michael, Hopwood., Phuong, Pho., Dustin, Ehling., Thomas, G., Walker. (2021). Introducing Tagasaurus, an Approach to Reduce Cognitive Fatigue from Long-Term Interface Usage When Storing Descriptions and Impressions from Photographs. 9(3), 45-. Available from: 10.3390/TECHNOLOGIES9030045

[4] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text", ICWSM, vol. 8, no. 1, pp. 216-225, May 2014.

[5] Y. Chandra and A. Jana, "Sentiment Analysis using Machine Learning and Deep Learning," 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2020, pp. 1-4, doi: 10.23919/INDIACom49435.2020.9083703.

[6] Y. Woldemariam, "Sentiment analysis in a cross-media analysis framework," 2016 IEEE International Conference on Big Data Analysis (ICBDA), Hangzhou, China, 2016, pp. 1-5, doi: 10.1109/ICBDA.2016.7509790.

[7] Bahrawi, Bahrawi. (2019). SENTIMENT ANALYSIS USING RANDOM FOREST ALGORITHM ONLINE SOCIAL MEDIA BASED. 2. h.29-33.

[8] Y. Hegde and S. K. Padma, "Sentiment Analysis Using Random Forest Ensemble for Mobile Product Reviews in Kannada," 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, India, 2017, pp. 777-782, doi: 10.1109/IACC.2017.0160.

[9] P. Karthika, R. Murugeswari and R. Manoranjithem, "Sentiment Analysis of Social Media Network Using Random Forest Algorithm," 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Tamilnadu, India, 2019, pp. 1-5, doi: 10.1109/INCOS45849.2019.8951367.

[10] D'Aniello, G., Gaeta, M. & La Rocca, I. KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis. Artif Intell Rev 55, 5543–5574 (2022). https://doi.org/10.1007/s10462-021-10134-9.

[11] Kiran Kumar, P., Jahna Tejaswi, N., Vasanthi, M.L., Srihitha, L.L., Phanindra Kumar, B. (2022). Sentimental Analysis on Multi-domain Sentiment Dataset Using SVM and Naive Bayes Algorithm. In: Garg, D., Jagannathan, S., Gupta, A., Garg, L., Gupta, S. (eds) Advanced Computing. IACC 2021. Communications in Computer and Information Science, vol 1528. Springer, Cham. https://doi.org/10.1007/978-3-030-95502-1_16.

[12] Z. Kastrati, F. Dalipi, A. S. Imran, K. Pireva Nuci, and M. A. Wani, "Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study," Applied Sciences, vol. 11, no. 9, p. 3986, Apr. 2021, doi: 10.3390/app11093986.

[13] "TWITTER SENTIMENT ANALYSIS USING VADER ON PYTHON", International Journal of Emerging Technologies and Innovative Research (www.jetir.org — UGC and issn Approved), ISSN:2349-5162, Vol.7, Issue 5, page no. pp1025-1031, May 2020

[14] Sivabharathi, G., and K. Chitra. "Aspect-based Sentiment Analysis on Current State of the Market by Classifying Customer Product Reviews Using Support Vector Machine with Several Kernel Learning." International Journal of Health Sciences, no. I, 3 Apr. 2022, pp. 3168-3180, doi:10.53730/ijhs.v6nS1.5371.

[15] Abdi, Asad & Sedrakyan, Gayane & Veldkamp, Bernard & Hillegersberg, Jos & van den Berg, Stéphanie. (2023). Students feedback analysis model using deep learning-based method and linguistic knowledge for intelligent educational systems. Soft Computing. 1-22. 10.1007/s00500-023-07926-2.

[16] https://blog.lyrise.ai/supervised-unsupervised-ml

Appendices

Appendix-I: Installation of libraries

1. **pip install virtualenv**
2. **virtualenv stud_feed**
3. **stud_feed/Scripts/activate**
4. **pip install flask**
5. **pip install pandas**
6. **pip install numpy**
7. **pip install mysql**
8. **pip install Flask-WTF**
9. **pip install python-time**
10. **pip install livereload**
11. **pip install nltk**
12. **pip install matplotlib**
13. **pip install hashlib**
14. **pip install python-csv**
15. **pip install pickle5**
16. **pip install scikit-learn**
17. **pip install gensim**
18. **python app.py**

Publication

Paper entitled **"Developing ML-based Student Sentiment Analysis portal for Educational Institute"** will be presented at **"2023 International Conference on Contemporary Challenges in Science and its Engineering Applications (IC3SEA 2023)"** by **"Aditya Joshi, AbhayPratap Singh, Bharat Singh and Rajashri Chudhari".** on 5th-6th May 2023 at Study World College of Engineering Palathurai, Coimbatore, Tamilnadu, India.