Developing ML-based Student Sentiment Analysis portal for Educational Institute

AbhayPratap Singh
Departrment of Information Technology
A.P. Shah Institute of Technology
Thane, India
abhay9769.as@gmail.com

Bharat Singh
Departrment of Information Technology
A.P. Shah Institute of Technology
Thane, India
bharat2808singh@gmail.com

Aditya Joshi
Departrment of Information Technology
A.P. Shah Institute of Technology
Thane, India
Adityajoshi334@gmail.com

Assistant Prof.Rajashri Chaudhari
Departrment of Information Technology
A.P. Shah Institute of Technology
Thane, India
rrchaudhari@apsit.edu.in

***Abstract*—Sentiment Analysis on Student Feedback Forms is a growing field that utilizes Natural Language Processing (NLP) and machine learning techniques to analyze student feedback and classify the sentiment expressed in the feedback. The goal of Sentiment Analysis on Student Feedback Forms is to provide educational institutions with valuable insights into student perceptions and attitudes towards various aspects of the educational experience, such as teaching quality, course content, and facilities. Sentiment analysis, often known as opinion mining, is a branch of natural language processing (NLP) that seeks to discern the sentiment represented in text automatically. The sentiment can be positive, negative, or neutral, and it can range from a binary classification to a multi-class or a continuous scale classification. The rise of social media and the vast amount of data generated through it has made sentiment analysis an important and active area of research. The paper also highlights some of the most common applications of sentiment analysis, including product and brand management, market research, and customer service. Moreover, the study examines some of the obstacles and limits of sentiment analysis, as well as future objectives for research on this subject.**

Index Terms - NLP, SA

I. Introduction

In today's digital world, people express their opinions and emotions about various topics through text, such as social media posts, reviews, comments, and blog posts. Sentiment analysis is a technique that assists in determining the sentiment represented in such content. The purpose of sentiment analysis is to evaluate whether a given text is good, negative, or neutral. In addition, the intensity of the sentiment can also be analyzed, which can range from a binary classification (positive or negative) to a multi-class or a continuous scale classification.

Opinion Mining, also known as sentiment analysis, is a branch of research concerned with evaluating and recognising the sentiment or emotion portrayed in a piece of text. This is a Natural Language Processing (NLP) area that seeks to assess whether a given text exhibits a positive, negative, or neutral attitude. Sentiment Analysis can be applied to a wide range of applications, including customer feedback analysis, social media monitoring, and brand reputation management.

The process of Sentiment Analysis involves several steps, including text pre-processing, feature extraction, and sentiment classification. The text pre-processing step involves cleaning and normalizing the text data to remove any irrelevant information or noise. The feature extraction stage includes encoding the text input in a manner that can be utilised by a machine learning algorithm to generate predictions. Lastly, in the sentiment classification stage, a machine learning model is trained on the processed data to predict the sentiment of fresh text input.

There are several techniques used in Sentiment Analysis, including rule-based methods, machine learning-based methods, and deep learning-based methods. To categorise the sentiment of a text, rule-based approaches use a collection of pre-defined rules and heuristics.

Sentiment Analysis is a growing field with a wide range of applications in various industries. The increased availability of vast volumes of text data has enabled the development of increasingly complex sentiment analysis models, resulting in greater accuracy and dependability in sentiment analysis predictions.

Sentiment Analysis on Student Feedback Forms is the process of analysing and classifying the sentiment expressed in student feedback forms using natural language processing (NLP) and machine learning techniques. Typically, sentiment analysis is utilised at educational institutions to acquire insight into students' opinions and attitudes concerning various aspects of the educational experience, such as teacher quality, course content, and facilities.

Sentiment Analysis on Student Feedback Forms seeks to discover the general sentiment indicated in the feedback as well as particular areas of concern or appreciation. Collecting student feedback forms, pre-processing text data, and then applying machine learning algorithms to identify the sentiment as positive, negative, or neutral are all part of the process.

Sentiment Analysis on Student Feedback Forms use a variety of methodologies, including rule-based methods, machine learning-based methods, and deep learning-based ways. To categorise the sentiment of a text, rule-based approaches use a collection of pre-defined rules and heuristics.

The findings of Sentiment Analysis on Student Feedback Forms can give educational institutions with useful information on how to improve the educational experience for their students. For example, the research may uncover areas of weakness in teaching quality, course content, or infrastructure, allowing institutions to take corrective action and increase student happiness.

In conclusion, Sentiment Analysis on Student Feedback Forms is an important tool for educational institutions to gain insights into the perceptions and attitudes of their students. By using advanced NLP and machine learning techniques, institutions can analyze large volumes of student feedback data and make data-driven decisions to improve the educational experience for their students.

II. Literature review

The goal of a literature review is to get an overview of existing Sentiment Analysis research and controversies pertinent to the field of study. The examination of the literature aided in the selection of an appropriate algorithm and feature extraction procedure for obtaining efficient results.

In [1], Emotion is inherently passionate from individual to individual, and can even be completely senseless, according to Novelty Octaviani Faomasi Daeli. When mining a massive and relevant case of data, it is critical. attempting to evaluate emotion. No single data point is fundamentally important. An individual's feeling towards a brand or product may be influenced by at least one indirect cause; for example, someone may have a terrible day and tweet a hostile remark about something they, for the most part, had a completely unprejudiced opinion about. Exceptions are weakened in the aggregate when a large enough example is used. Furthermore, because sentiment fluctuates over time due to an individual's state of mind, world events, and so on, it is important to analyse the feelings of Twitter customers.

In [2], Sanjeev Dhawan, Kulvinder Singh and Priyanka Chauhan propose an algorithm where the Twitter dataset is fetched from Twitter API for analysis of the sentiments of users. First, the user is checked for authorization then the algorithm starts sentiment analysis. To analyze the sentiments, the first features of tweets are extracted to check the sentiment polarity of each tweet. If the polarity of the feeling is equal to zero, the tweet is neutral; if it is more than zero, the tweet is positive; otherwise, the tweet is negative. As a result, the suggested algorithm distinguishes tweets based on the sentiment polarity of each user's tweet.

In [3], According to R Raja Subramanian, Nukala Akshith, Gougula Narasimha Murthy, Manchala Vikas, Srikar Amara, and Karnam Balaji, may be divided into five separate phases. Data separation, pre-processing, feature extraction, classification, and outcome assessment are the stages involved. Lexicon analysis, using an algorithm and a hybrid technique, is another process used for sentiment analysis. The different inputs will be trained and divided into different classes using this machine-learning algorithm. In order to assign each input to the appropriate class and to analyse the data's pattern, probabilistic multinomial naive bayes are utilised. The IMDB dataset is a dataset that they used to categorise sentiment analysis.

In [4], C.J. Hutto and Eric Gilbert present their findings on the methodical development and evaluation of VADER (Valence Aware Dictionary for sEntiment Rea-soning). VADER generates and empirically validates a gold-standard collection of lexical features (together with their associated sentiment intensity measures) that are specifically attuned to sentiment in microblog-like contexts using a combination of qualitative and quantitative methodologies. It then considers five general rules that encapsulate grammatical and syntactical patterns for expressing and intensifying sentiment intensity. The correlation coefficient demonstrates that VADER (r = 0.881) matches ground truth as well as individual human raters (r = 0.888). (aggregated group mean from 20 human raters for sentiment intensity of each tweet). Further examination of the classification accuracy reveals that VADER (F1 = 0.96) surpasses individual human raters (F1 = 0.84) in correctly identifying the sentiment of tweets into positive, neutral, or negative categories. VADER outperformed (and in most cases outperformed) eleven other well-rated sentiment analysis technologies. It highlights the benefits of incorporating humans as a fundamental part of the development process in computer science.

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.

III. Proposed System and Architecture

Proposed sentiment analysis (SA) system architecture. After preprocessing input data-student feedback was obtained from both formal sources such as course surveys and informal sources such as blogs and forums-To classify thoughts and emotions, the system employs natural language processing in conjunction with the NRC Emotion Lexicon. Positive, negative, and neutral emotions are the three types of emotions.

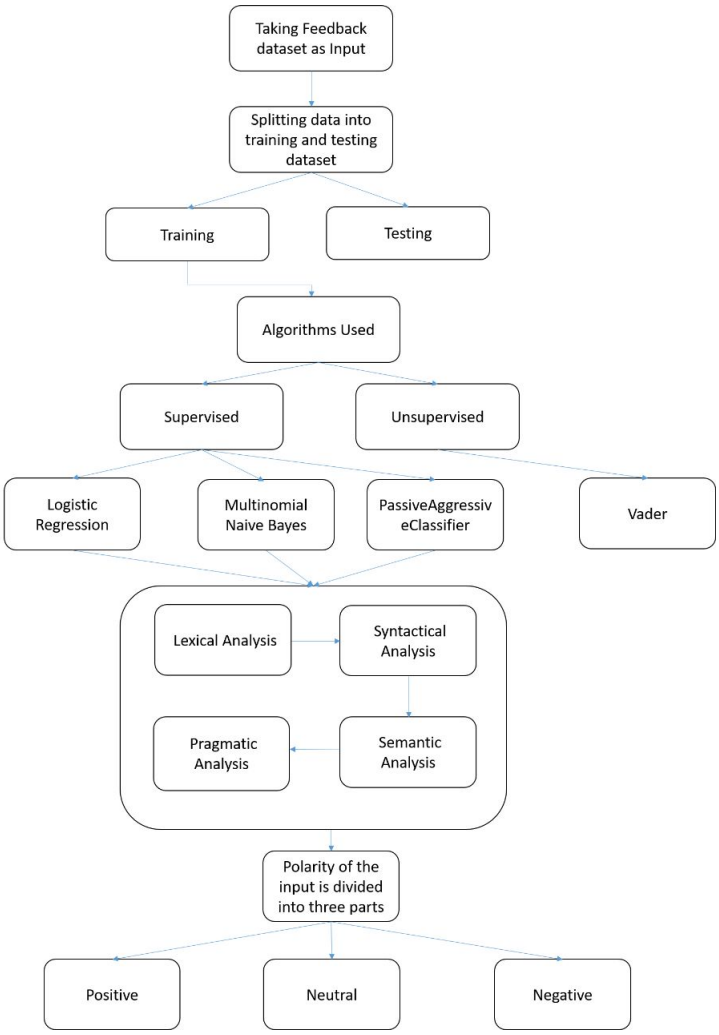

Fig. 1. Proposed System

- 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 on which we will be performing the Sentiment Analysis and giving it a score after processing it and performing some operations on it.
- Splitting the Dataset into training and testing - Now we split the data into two parts that are the training and the testing data.
  - Training - A portion of the data is set aside for training the model, which may vary depending on the type of Machine Learning Model used and the amount of data available.
  - Testing - The other part is reserved for testing purposes which be used to evaluate how well our model is created.
- Algorithm - After the splitting of the dataset is completed we have to choose which type of ML algorithm we can apply. There are two types of Supervised and Unsupervised Learning.
- Supervised - It is a machine learning algorithm that trains the model using a labelled dataset. The labelled dataset is made up of input data (also known as features) and the values that correspond to them (also known as labels or targets). This data is used by the model to learn the mapping between the input data and the output values, allowing it to make accurate predictions on new, previously unknown data. We will have to pre-process and make it ready before we can use it 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 - This is the third phase in NLP and it assists in determining the meaning of a sentence. The semantic analysis uses the results of previous NLP phases, such as lexical and syntactic analysis, to establish the links between words and phrases and the meaning of the sentence.
  - Pragmatic Analysis - The fourth and final step of NLP helps analyse the context in which a sentence is used to understand its meaning. The goal of the pragmatic analysis is to determine the intended meaning of a sentence by considering factors such as the speaker's intention, the situation in which the sentence was spoken, and the social and cultural context.
- Unsupervised learning - It is a machine learning algorithm that trains the model using an unlabeled dataset. Unlike supervised learning, there are no predefined output

values or labels associated with the input data. The goal of unsupervised learning is to identify patterns, structures, and relationships in the data without being explicitly told what to look for.

  - Vader - It is a natural language processing lexicon and sentiment analysis tool with rules. It is designed to analyze the sentiment of a text and provide a score based on the emotional content of the text. The score lies between -1 to +1, where -1 represents negative sentiment, +1 represents positive sentiment, and 0 represents neutral sentiment. VADER employs a mix of lexicon-based and rule-based approaches to determine the sentiment of a text. It analyzes the text for the presence of words and phrases that have a positive or negative connotation, as well as rules for handling negation and punctuation. VADER also considers the context of the text by taking into account the valence of the surrounding words.

- Polarity of the feedback - Now that all the pre-processing and classification algorithms have been applied it is time for us to get the results. Because a classification algorithm has been used we will get one of the three values from one student feedback -1, 0 and 1 which means Negative, Neutral and Positive respectively.

IV. Implementation

- Student Input - In this, the student will give feedback for the respective course that they are enrolled in. This data will be used as the Input data on which we will be performing the Sentiment Analysis and giving it a score after processing it and performing some operations on it.
- Data Pre-Processing - Now that we have the input data from the Student we will have to pre-process and make it ready before we can use it to perform the SA on it. The pre-processing generally includes understanding the words and sentences and is done with the help of NLP steps mentioned in the proposed system.
- Applying Algorithm on Feedback - After all the pre-processing steps have been performed on the feedback it is ready to be put into a Classification Algorithm to determine if the feedback is positive, negative or neutral.
- Getting the Polarity of the feedback - Now that all the pre-processing and classification algorithms have been applied it is time for us to get the results. Because a classification algorithm has been used we will get one of the three values from one student feedback -1, 0 and 1 which means Negative, Neutral and Positive respectively.
- Graphs - The findings will be produced to be shown in graphical forms as graphs provide a better knowledge of things and it is simple to understand rather than reading texts.

V. Results

VI. Conclusion

In conclusion, a machine learning-based student feedback portal can provide valuable insights and improve the qual-

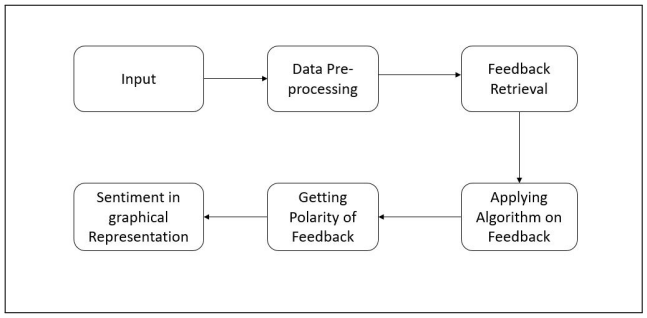

Fig. 2. Implementation Flowchart

Fig. 3. To enter a sentence for performing SA

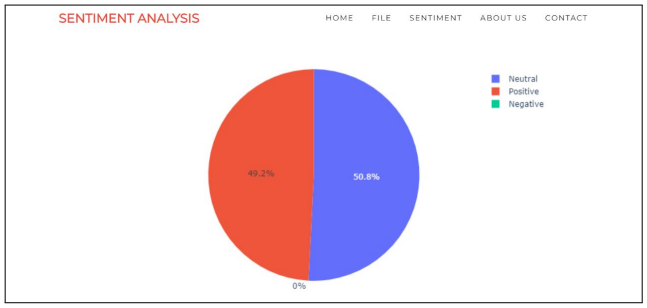

Fig. 4. Pie Chart

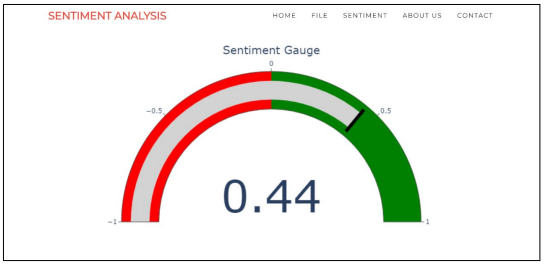

Fig. 5. Gauge Image

ity of education. A system like this may help instructors discover areas for improvement, tailor learning experiences, and increase student engagement by using the power of data analysis and predictive modelling. However, it is important to note that the success of the system depends on the quality and quantity of data collected, the accuracy and robustness of the algorithms used, and the ethical use of student data. Therefore, it is essential to ensure that the system is designed and implemented with transparency, fairness, and privacy in mind. Overall, a well-designed machine learning-based student feedback portal can be a valuable tool for educators to enhance the learning experience, foster student success, and drive continuous improvement in education.

VII. Future Scope

- Improved Text Pre-processing: Developing more sophisticated text pre-processing techniques to better handle complex and unstructured data, such as misspelt words or slang.
- Multi-Modal Analysis: Integrating other forms of data, such as images or audio, into sentiment analysis models to provide a more complete picture of student feedback.
- Personalized Feedback: Developing models that can provide personalized feedback to students based on their feedback data, which can help to improve the educational experience for each student.
- Automated Feedback Generation: Developing algorithms that automatically generate feedback based on student feedback data can help educators save time and effort in analyzing and categorizing feedback.

Acknowledgment

This project would not have come to fruition without the invaluable help of our guide Assistant Prof. Rajashri Chaudhari. Expressing gratitude towards our HoD, Dr Kiran Deshpande, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project.

References

[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.