Analysis of Textual Feedback of Students for Course Evaluation in Universities Through Machine Learning Algorithms.

Naseer Ahmed[1], Mah Gul Bizanjo[1], Afrasiyab Khan[2], Saba Gull[1], Saleem khaliq[2], Nooruddin[2]
[1]Department of Computer Science, Balochistan University of Engineering & Technology, Khuzdar, 89120, Pakistan
[2]Department of Software Engineering, Balochistan University of Engineering & Technology, Khuzdar, 89120, Pakistan

E-mails: naseerbajoi@gmail.com, mahgulbizanjo@gmail.com, afrasiiyabkhan@gmail.com, sabanaseerbajoi@gmail.com, saleemkhaliq1075@gmail.com, nooruddin4201sn@gmail.com

Corresponding Author: Afrasiyab Khan, afrasiiyabkhan@gmail.com

Received: 27-04-2024; **Accepted:** 05-04-2024; **Published**: 10-05-2024

Abstract: Many educational institutions worldwide make significant efforts to collect student feedback to understand their perspectives on the courses and faculty. This feedback is used to enhance the institution's environment. In this modern world, institutions use data collection techniques to gather feedback. However, they lack the proper techniques to analyze and utilize this data to improve the educational quality of the institute using textual feedback. This study presents techniques for analyzing the written feedback from students, which was collected for course evaluation over a year. This paper focuses on techniques including Multinomial Naive Bayes Classifier, Long Short-Term Memory(LSTM), and Random Forest to enhance the outcomes of sentiment analysis. Ultimately, our efforts resulted in the LSTM achieving 97.45% accuracy during model testing for three types of sentiments: positive, neutral, and negative. This paper also aims to identify a clear research gap in this field and discusses the work of other researchers, including their less accurate models from the past. We also discuss the processes of collecting a sufficient amount of data to train this model, and then utilize a set of 25,689 data points for training. Furthermore, this paper primarily focuses on enhancing the quality of education. The initial model has been implemented at Balochistan UET Khuzdar, and it has produced satisfactory results. In the future, efforts will be made to find the perfect way to enhance the quality of education.

Index Terms: Sentiment analysis, Course Evaluation, Machine learning, Student Textual feedback, Educational quality enhancement.

1. Introduction

A prominent topic in education circles worldwide is the efficacy of teachers as a deciding factor in educational quality. As a result, precise and effective teacher assessment has grown in importance as a subject of academic study [1]. Nowadays, Sentiment analysis has become more popular as more companies pay close attention to reviews[2]. In reality, the assessments and written feedback from students and courses are not just connected to administrative choices regarding teachers' promotions and salary increases, but they also offer teachers a comprehensive view of their teaching effectiveness and aid in enhancing the quality of instruction. These assessments may improve the bonds between teachers and students while also promoting the growth and development of the pupils.

Opposition mining is another name for sentiment analysis. Sentiment analysis examines, evaluates, and extracts opinions from textual content [3]. It uses natural language processing (NLP) technology to identify a document's emotional tone [4]. Since the 1960s, a lot of theories on sentiment analysis—the identification and categorization of emotions—have been developed [5]. Opinion mining is a vast field of research that combines natural language processing, machine learning, psychology, and sociology to uncover users' and customers' fundamental ideas. Views from users may be available on a number of websites, such as Instagram, Facebook, and Twitter. Scholars started looking through these websites around ten years ago to find sentiments or viewpoints [6]. Delivering polarity results and assessing people's emotions both depend on sentiment analysis. Reviews on different themes, such as goods and people, carry tremendous weight in the eyes of enterprises, which is why suitable text arrangement becomes beneficial for interpreting the sentiment state [7]. Feedback provided by students can be categorized into two forms: textual feedback and grading feedback based on Likert scale scores[8]. In the case of Likert scale scoring, students are presented with questions and asked to rate their responses using a predefined scale. This approach primarily concentrates on gathering feedback related to specific topics, but it may not accurately capture the perfect/present sentiments of the students[9]. The intelligent analysis of student behavior and the LMS in connection to course results is being done with the help of several ML algorithms [10]-[11]. Textual feedback is used to determine the actual sentiment of the students. Students are given a series of questions to reply in sentences in this textual style. It benefits both the academic administration and the instructor in overcoming organizational challenges. As a result, automatic models and methods are required to effectively handle textual feedback[12].

Surveys of students' opinions are used by educational institutions after each semester to get their thoughts on the courses they took[13]. Both qualitative and quantitative data, such as comments, course evaluations, and student demographics, are included in the feedback. While qualitative data analysis uses methods like natural language processing (NLP) to interpret textual comments, quantitative data analysis provides statistical insights into course feedback [14]. Listening to students' opinions about classes, material, and instruction is made possible by this analysis.

Sentiment analysis is categorized into five types which are listed in Table 1 . But our paper focuses on Sentence-level sentiment analysis.

Type of Analysis	Definition
Document-Level Sentiment Analysis	This study primarily aims to classify written works, such as articles or reviews, as Positive, negative, and neutral depending on the overall tone they reflect.
Sentence-level sentiment analysis	When looking at this situation the examination centers, on the emotions conveyed in each sentence of a document giving an understanding of the feelings expressed throughout parts of the text.
Aspect-Based Sentiment Analysis	This approach aims to pinpoint and capture the feelings linked to elements or subjects referenced in the content. For instance, in a review of a product the attitudes, towards aspects of the product (, like its performance, appearance and ease of use) can be assessed separately.
Entity-Level Sentiment Analysis	The purpose of this investigation is to determine the attitudes stated about the persons, businesses, or goods that are mentioned in the text. It facilitates understanding the feelings associated with topics covered in a literary work.
Comparative Sentiment Analysis	Evaluating the attitudes expressed toward topics or ideas included in the literature is one method. Finding out what kinds of feelings or preferences people have for certain things or traits is the goal.

Table 1 Definitional table of sentiment analysis based on five types.

Google Forms, QEC semester feedback collection, and online sentiment datasets were used in this study to collect student responses that reflect a wide range of opinions. The objective is to extract opinion statements and use machine learning techniques to assign a positive or negative classification to them. This will play a significant role in course evaluation. Machine learning approaches may employ supervised or unsupervised learning[15]. Classification issues can be resolved using various techniques, including Long Short-Term Memory (LSTM), Naive Bayes, and Random Forests. The lexicon-based method detects sentiment polarity in textual information by utilizing a sentiment lexicon, which is essentially a compilation of terms with corresponding sentiment polarities[16].

The structure of this document is as follows: In the "Literature Review" section, we will provide an overview of prior studies on sentiment analysis and machine learning techniques. The "Methodology" section describes how materials are categorized based on student comments. The "Performance Analysis" section compares machine learning techniques based on F-score, accuracy, and final results. The "Conclusion" section summarizes the findings and offers final observations.

2. LITERATURE REVIEW

The field of sentiment analysis has been extensively researched. There haven't been a lot of studies, in text classification that categorize texts as negative, positive, or neutral. [17]-[18]. In the realm of education, instructors are often assessed through both data, like percentages and rating scales (feedback) and commentary, assessment tasks as well, as audio and video files (qualitative feedback) [19]. A comprehensive investigation was conducted to assess emotions in three domains: sentiment analysis, feature extraction, and framework [20]. In methods supervised learning techniques, like Long Short Term Memory (LSTM) and Naive Bayes are explored. The findings show that Long Short Term Memory outperforms better than classifiers in terms of accuracy. As per the research Naive Bayes does better than LSTM with datasets whereas LSTM excels, with datasets. As per the research Naive Bayes does better than LSTM with datasets whereas LSTM excels, with datasets. [21]. In a study conducted at Middle East College in Oman, researchers classified student responses from a module assessment survey using the RapidMiner opinion mining program. They used neural networks, K closest neighbor, support vector machines, naïve Bayes, and other learning methods to analyze the data. Among the examined algorithms, K Nearest Neighbor demonstrated precision while Naïve Bayes had the highest accuracy and recall, according to the results [15]. The effectiveness and exactness of the sentiment model rely on the methods employed. A strategy, in data mining was created to categorize faculty ratings at an institution, from 1 to 5 using characteristics[22]. The research utilized the Naive Bayes classifier and text-mining techniques to evaluate student feedback. Nonetheless, a limitation of this investigation was its inability to precisely grasp the sentiments of the students [23]. It can take time and effort to process a significant amount of comments gathered at the conclusion of the semester. As described in Long Short-Term Memory (LSTM), Naive Bayes, and Maximum Entropy (ME), the machine learning techniques are [9]-[10]. Decision trees and Multinomial Naive Bayes are utilized for sentiment analysis of Twitter data [21]. The study utilized the n-gram approach [24] to extract features from 1150 documents. The classification model's performance was assessed using recall, F-score, precision, and accuracy measures [25]. The study addresses the issue of sentiment polarity categorization using input data from online product reviews on Amazon. [26]-[27]. A sentiment analysis categorization model for Arabic text was created in 2014. Out of 10,500 texts, 2591 were used for model training after preprocessing. To ascertain the sentiment of the reviews, the 10-fold cross-validation method was applied with the SVM, KNN, and Naive Bayes classifiers. At 75.25% SVM classifier had the best accuracy [28].

The Literature Review section of this paper focuses on finding a significant gap in existing research. After conducting a thorough search and analysis, we have discovered various sentiment analysis models and their corresponding accuracy rates, as outlined in Table 2. Our research indicates that our sentiment analysis model outperforms the accuracy levels of current models. This conclusion supports the results of the literature review, confirming the existence of a substantial research gap.

Different models with their accuracy Rate.			
Reference. No	Research Paper Title	Author	Accuracy Rate
[1]	Automatic scoring of student feedback for teaching evaluation based on aspect-level sentiment analysis.	Ping Ren[1]	79%
[19]	Using sentiment analysis to evaluate qualitative students' responses.	Delali Kwasi Dake[1]	63.70%
[28]	Arabic Sentiment Analysis using Supervised Classification.	Rehab M. Duwairi[1]	75.99
[29]	Opinion mining from student feedback data using supervised learning algorithms.	Dhanalakshmi V.,Dhivya Bino[1]	94.67%

Table 2 Finding a Research gap by getting the Accuracy of different researched models

3. METHODOLOGY

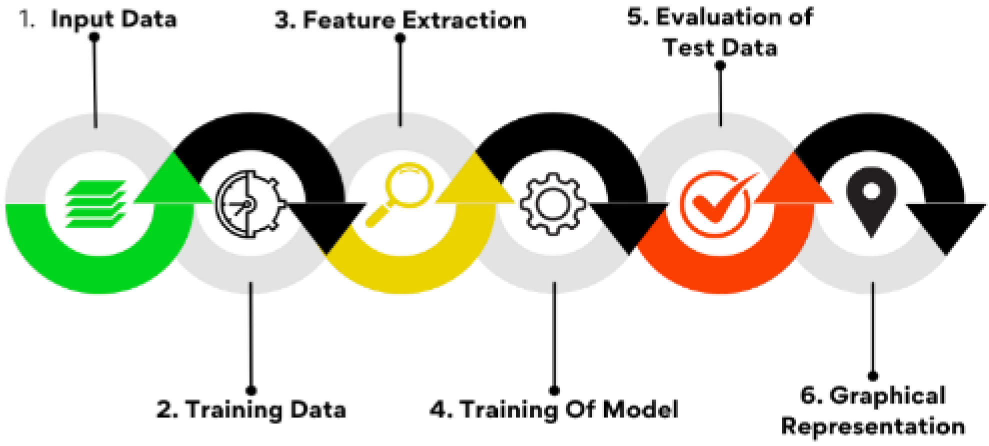


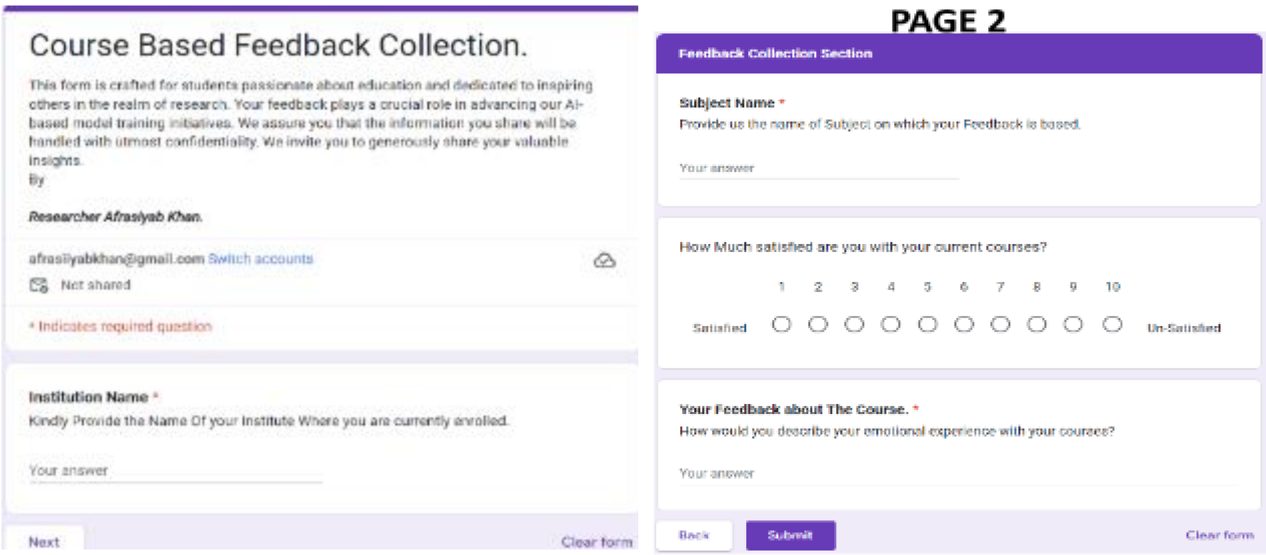
Figure 1 Sentiment Analysis Working Methodology.

Applying sentiment analysis via a method of machine learning is the focus of this work. It involves collecting textual data (Input Data), labeling a portion with positive, negative, or neutral sentiments (Training Data), and then using this labeled data to train an algorithm to recognize these emotions in new text. The algorithm extracts and analyzes features such as word choice, punctuation, and grammar to enhance its ability to accurately predict sentiment(Feature Extraction). The findings are then presented visually for clarity (Graphical Representation). This approach involves six key steps, as depicted in Figure 1.

So, we follow Figure 1 and perform our methodology phases step by step. The first step is:

3.1. Dataset

To develop an advanced sentiment analysis model, a substantial amount of data is required [30], and collected


Course Based Feedback Collection.

This form is crafted for students passionate about education and dedicated to inspiring others in the realm of research. Your feedback plays a crucial role in advancing our AI-based model training initiatives. We assure you that the information you share will be handled with utmost confidentiality. We invite you to generously share your valuable insights.
By

Researcher Afrasiyab Khan.

afrasiiyabkhan@gmail.com Switch accounts
Not shared

* Indicates required question

Institution Name *
Kindly Provide the Name Of your Institute Where you are currently enrolled.

Your answer

Next Clear form

PAGE 2

Feedback Collection Section

Subject Name *
Provide us the name of Subject on which your Feedback is based.

Your answer

How Much satisfied are you with your current courses?

	1	2	3	4	5	6	7	8	9	10	
Satisfied	○	○	○	○	○	○	○	○	○	○	Un-Satisfied

Your Feedback about The Course. *
How would you describe your emotional experience with your courses?

Your answer

Back Submit Clear form

Figure 2 Snapshot of Google Form used to Collect students Feedbacks.

through diverse methods and sources, as outlined in this study, is necessary:

3.1.1. Google Form survey

To gather valid and effective input from students, we designed a Google Form survey. That includes a variety of questions regarding their experiences in courses, with instructors, and their overall satisfaction. The survey included open-ended questions to gather detailed written feedback. We distributed the Google Form to students from various academic disciplines to ensure a comprehensive dataset. Using Google Forms, we have collected 1,248 responses from students. The Form is shown in Figure 2.

3.1.2. QEC Department of Balochistan UET Khuzdar

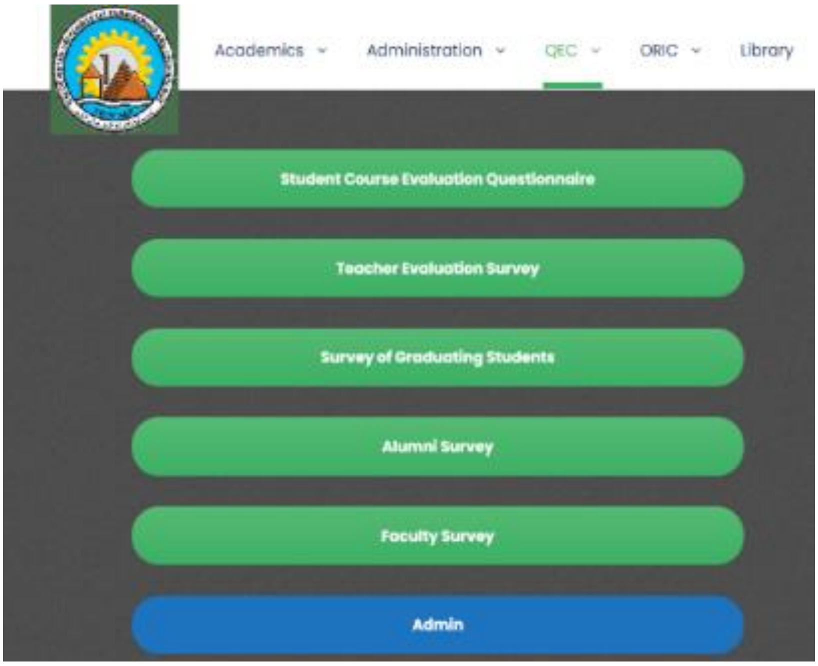


Figure 3 Data Collected From QEC Department BUET

This is our second source where we gather a large amount of data. The "Balochistan University of Engineering and Technology Khuzdar" has assigned its Quality Enhancement Cell (QEC) the duty of gathering survey responses from students via the university website at the conclusion of each academic year, as shown in Figure 3.

We received permission from the Vice Chancellor of the institution and gathered student feedback from 2019 to 2023. The feedback from QEC was collected in a Microsoft Excel spreadsheet. We have obtained 16,374 feedback from this source.

3.1.3. Open Source Datasets

To enrich our analysis and incorporate external perspectives, we utilize a third-party provider to collect online data from open-source platforms. This data includes public comments and opinions on education, courses, and informative topics. This additional source provides a broader perspective and captures sentiments expressed beyond the academic realm. Additionally, we used an open-source dataset from https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment to enhance the model's accuracy. The Kaggle dataset contained social feedback that enabled us to achieve superior performance in sentiment analysis.

Table 3 below shows the total number of datasets collected from three main sources for a research project, amounting to 25,699 datasets. The sources include Google Forms, where we collected 1,248 datasets; from the QEC Department of UET Khuzdar (2019-2023), we obtained 16,374 datasets; and from an open-source online platform, we gathered the dataset of 8,067 datasets.

S.no	Data Collection Platforms	Number of Feedback Collected
1	Google Forms	1,248
2	QEC Department of UET Khuzdar (2019-2023)	16,374

3	Open source Online Platform	8,067
Total		25,689

Table 3 Total Number of Dataset collected from 3 main sources.

3.2. Preparing Training Data

This section also presents a detailed explanation of the steps to create a training dataset for the examination of written feedback received from the evaluation of students and courses. It gives two machine learning processes known as supervised learning, which relies on labeled data points, and unsupervised learning which works, devoid of labeled data. It applies learning because it combines the Sentiment Intensity Analyzer language model of the VADER. This tool assigns sentiment scores ranging from -1 to 1 to each phrase to show the degree of negativity or positivity. Additionally, it gives a sentiment score between -4 and 4 to every word, within the phrase. [9].

The training dataset is constructed using 25,000 feedback samples from three sources; Google Forms, the QEC Department, and Open Source. These feedback submissions are categorized as positive, neutral, or negative through a labeling process. To ensure consistency and minimize biases multiple researchers independently label a segment of the data resulting in a rater reliability score of 97%.

Subsequent data cleaning and pre-processing steps involve normalization, tokenization, stemming/lemmatization, stop word removal, and feature engineering. Normalization consists of converting text to lowercase and removing diacritics to ensure a consistent data format. Tokenization breaks down sentences into words while stemming and lemmatization help analyze words accurately. Removing stop words that don't add meaning improves training efficiency. Feature engineering adds elements like sentiment scores word frequencies and document length to enhance feedback representation, for machine learning models.

This carefully prepared and labeled dataset plays an important role based on subsequent machine learning analyses. Hand labeling, extensive pre-processing, and integration of several sources aim to build a robust and accurate training dataset for accurate textual analysis of student and course assessment comments.

Creating training data is a time-consuming job that involves a lot of work. In Table 4 you can see an outline of how we created the training data, for our sentiment analysis model. We carefully sorted through written feedback from places like surveys and online platforms classifying them as positive, negative, and neutral. This thorough labeling forms the basis for teaching the model how to identify and understand patterns, in new text.

S.no	Feedback	Source	Sentiment
1	Professor of Networking subject creates an excellent learning environment. His encouragement and support have boosted my confidence in the course. The way he provides constructive feedback on assignments has been incredibly helpful in my understanding of the material	QEC Department of Balochistan UET khuzdar	Positive
2	I find the education system to be rigid and inflexible. The one-size-fits-all approach doesn't cater to diverse learning styles, and the pressure of exams can be overwhelming. There's a need for a more personalized and adaptive approach to accommodate the unique needs of students.	Google Form	Negative
3	Unfortunately, I've had a frustrating experience in the class of organizational behavior. The grading seems inconsistent, and it's not always clear what is expected in assignments. I've tried seeking clarification, but the communication is lacking.	QEC Department of Balochistan UET khuzdar	Negative
4	The flight with PIA was okay. Nothing extraordinary but also no major issues. The seating was comfortable, and the in-flight entertainment was decent. A neutral experience overall.	Open-source online Dataset	Neutral

Table 4 Manually Labeling of Feedback used for training the model.

Table 5 presents the breakdown of sentiment labels manually assigned to the 25,699-point training dataset. The data obtained from Google Forms, QEC Khuzdar (2019-2023), and an open-source platform demonstrates that 45% of the sentiments are positive, 38% are negative, and 17% are neutral. This detailed categorization serves as the basis for developing our sentiment analysis model.

S.no	Type Of Sentiment	Number of Feedback Based on Sentiments
1	Positive	11,502
2	Negative	9,805
3	Neutral	4,382

Table 5 Total number of Three Classified Sentiments Collected After Labeling.

3.3. Feature Extraction

This approach involves the extraction of features from datasets to create a format that is well-suited for machine learning algorithms. Both for train and the test sets of data, feature extraction is applied. The Scikit-learn package contains tools for feature extraction and tokenization of textual data. Tokenization divides textual data into discrete words or tokens. Tokenizing text documents and creating a lexicon of recognized phrases are done with the scikit-learn Count Vectorizer tool.

3.4. Model Training

Our study centered on the usage of three unique models: long short-term memory (LSTM), random forest (RF), and

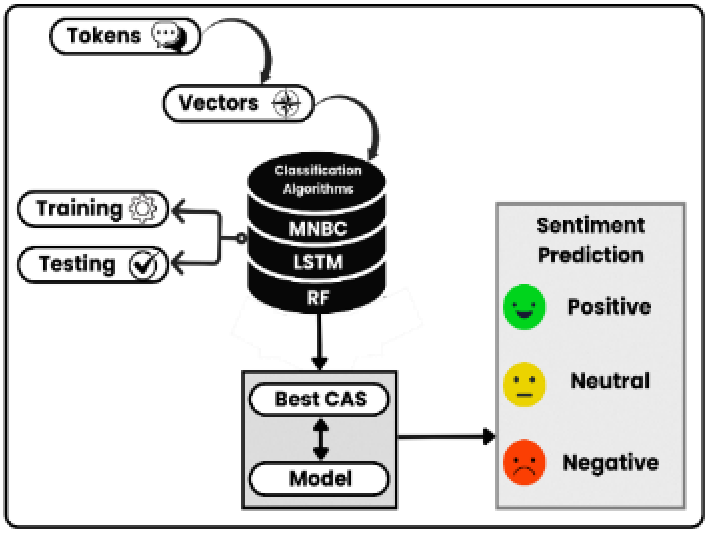


Figure 4 Sentiment Analysis Modeling Process.

Multinomial naive Bayes (MNBC) to better sentiment analysis. This section presents a full discussion of the difficult approach utilized to train these models to accurately identify text using machine learning techniques.

The graphics in Figure 4 demonstrate the techniques that are applied in constructing a sentiment analysis model. Initially, raw text data is employed as the input followed by the selection of an annotated subset. This subset, serving as the training dataset offers insights, into the model for recognizing cues through sophisticated feature extraction techniques.

The machine learning techniques employed, like LSTM, multinomial naive Bayes, and random forest intricately grasp the connections, between text features and sentiment categorizations throughout the learning phase. This assimilation of information empowers the models to make guesses, about the sentiment of text. These methods include:

3.4.1.Long Short-Term Memory (LSTM)

LSTM stands for Long Short-Term Memory, one of the famous recurrent neural networks (RNN) that is often used in machine translation jobs. LSTMs are specifically crafted to manage data, like text, and possess the ability to retain information from inputs for a longer duration compared to conventional RNNs [31].

3.4.2. Multinomial Naïve Bayes

Mainly MNBC focuses on the classification approach to classify texts by considering both class and word probabilities. The classifier relies on the frequency of terms, in the document as a factor, for making predictions [32][33] [34].

3.4.3. Random Forest

Random forest is a versatile classifier used in machine learning, which has high performance in both classification and regression problems. Its key benefits include its parametric nature, excellent accuracy, in classification, and ability to understand the importance of variables [35].

3.5. Evaluation of the test data

It is the act of gathering and analyzing data to determine how well an organization is carrying out its intended operations. Evaluation, within a model's context, is the term used for the last action undertaken directly following the training. At this point, it is crucial to assess the model's performance and generalizability. Evaluation utilizes a separate test set to measure accuracy and the ability to execute the trained model. Via the test set, the model's anticipated accuracy is validated using new untested data. The assessment presents findings on the model's precision in delivering predictions, and general performance and usability outside a training set. Thus, evaluation is crucial for determining whether a proposed model is viable and reliable for practical usage, which, in turn, shows model functionality, areas that may need further adjustment, and whether a model is suitable for a real-life application.

3.6. Performance Analysis

I consider three machine learning algorithms; Random Forest, Multinomial Naive Bayes, and Long Short Term Memory as the three considered in this work. I will quantify the performance of the machine learning algorithms using the bigram and unigram features, which are based on accuracy, and the F score. Bigrams entail two parts or tokens from the string whereas unigrams involve the singular part or token in the string. Bigram is one of the features that compare how other Machine learning systems work. We run the models on a test dataset and assess their performance. Other factors are considered before determining how accurate it is, and hence the F-score. Accuracy involves accurately doing our model state predictions. There is another statistic that is a balance between precision and recall, called F-score, or F1 score. It takes into account the balance between false positive predictions and false negative predictions and it also takes into account the precision of predictions when the class numbers in the dataset are unbalanced.

Accuracy: Equation 1 specifies that accuracy is measured by dividing the total number of rows in the dataset by the number of accurate projections of the model made [36].

$$Accuracy = \frac{\text{characters/words correctly recognized}}{\text{All characters/words}} \quad (1)$$

Calculating the F-score - an integral measure when evaluating models- requires accounting for both recall and precision together. This vital metric utilizes equation 2 as its formula and considers a weighted average approach that enables an accurate determination of how well models perform[37]

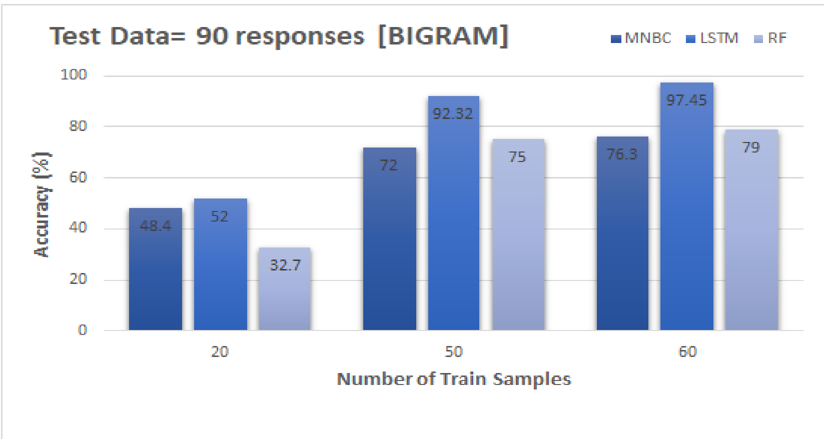


Figure 5 The Bigram LSTM, MNBC, and RF algorithms' performance for different train samples

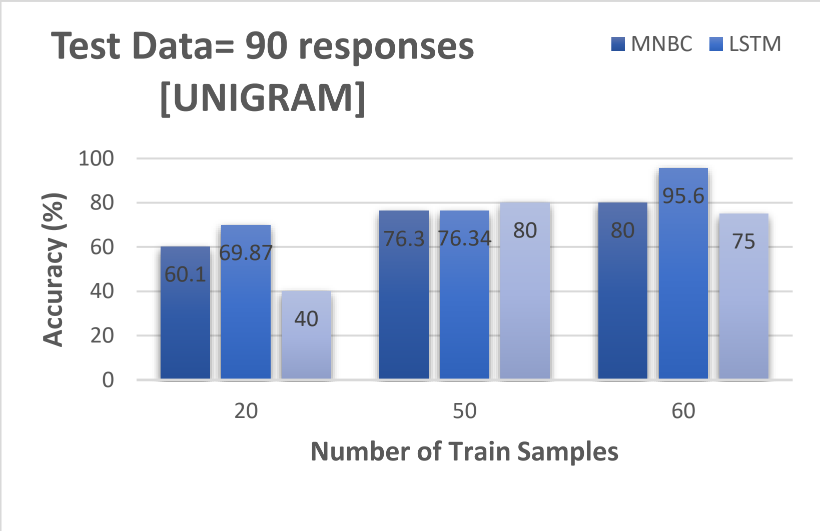


Figure 6 Performance of the Unigram LSTM, MNBC, and RF algorithms for various train samples

$$F\ Score = f\ 1 = f = 2 * \frac{precision \times Recall}{Precision + Recall} \qquad (2)$$

Figures 5,6 and Table 6 display the accuracy results for both unigrams and bigrams using the Random Forest, Long Short-Term Memory, and Multinomial Nave Bayes Classifier. The training set size was varied while the test data was kept constant for the research. Notably, the models' accuracy increases as the training data increases.

Model	Accuracy	Precision	Recall	F-Score
LSTM	97.45%	92.5%	90%	80%
MNBC	80%	82.5%	70.5%	70.5%
RF	79%	80.5%	38.6%	60%

Table 6 Accuracy of different methods

The LSTM method consistently yields improved accuracy as more data is utilized to train the model. measured in terms of accuracy against Random Forest (RF), Multinomial Naive Bayes Classifier (MNBC), and other methods. However, there is no improvement in RF and MNBC's accuracy. Compared to MNBC and RF, the LSTM offers more precision.

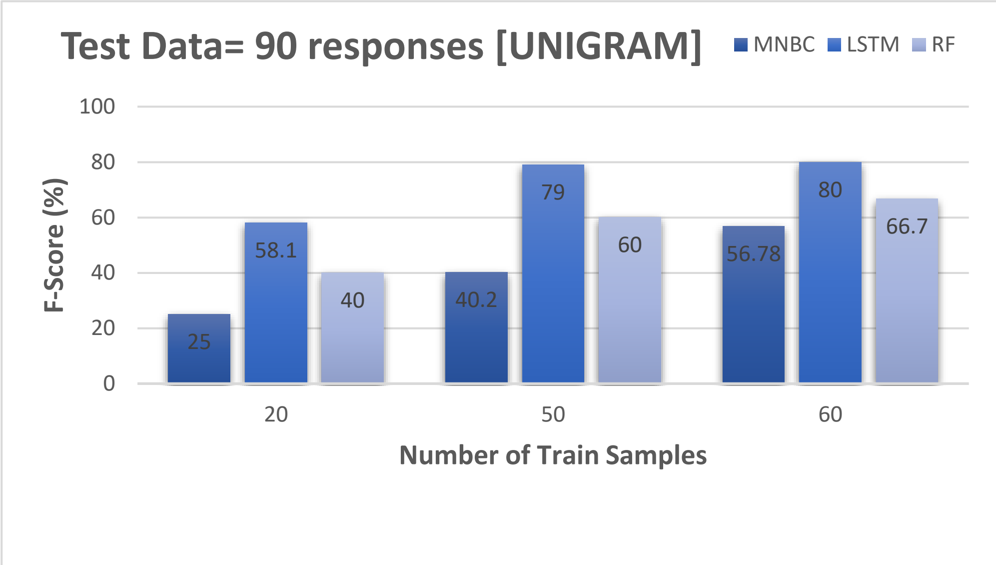


Figure 8 shows a graph of the F-Score for a Unigram v/s the number of train samples.

The MNBC and LSTM algorithms' F Scores rise linearly with the amount of training data shows in Figures 7 and 8 respectively.

4. RESULTS

This part will cover our sentiment analysis study's results, with an emphasis on performance indicators like accuracy and F-score as well as the broader effects that follow from our research. Our main objective was to assess the Model's suitability for sentiment classification. The results that are offered here consist of both quantitative evaluations and insightful observations that were made after extensive testing.

Figure 9 shows how well the three algorithms, in our sentiment analysis system perform; Multinomial Naive Bayes Classifier (MNBC), Random Forest (RF), and Long Short Term Memory (LSTM). The data shows that the accuracy of

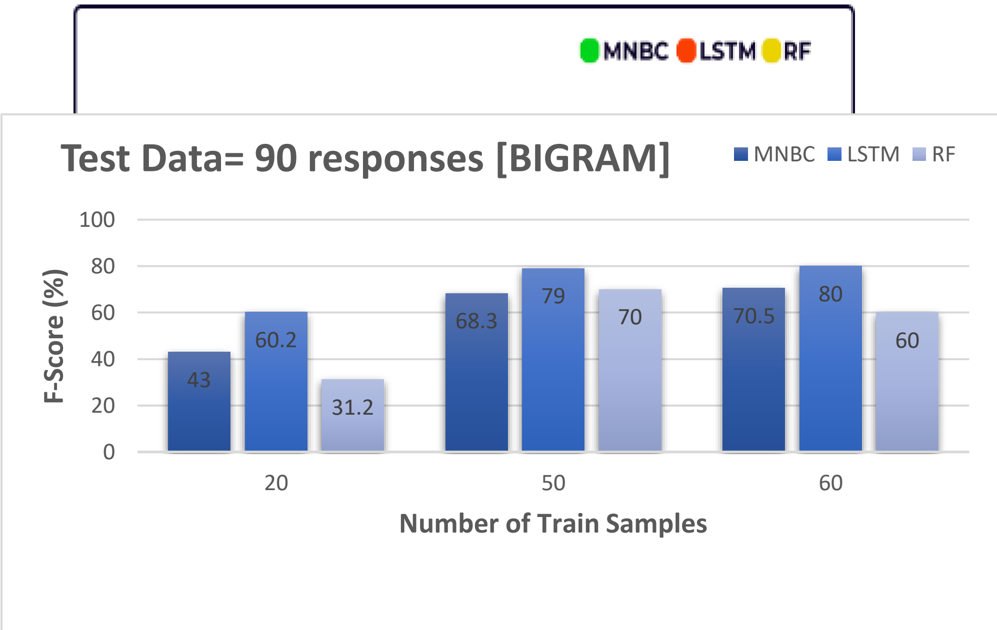


Figure 7 Shows a graph of Bigram's F Score as a function of the quantity of train samples.

all three algorithms increases with the amount of training samples. Interestingly the LSTM algorithm consistently outperforms the others achieving the accuracy regardless of training sample sizes. While MNBC and RF show performance RF slightly falls behind. These results indicate that for our dataset and task, LSTM provides dependable and

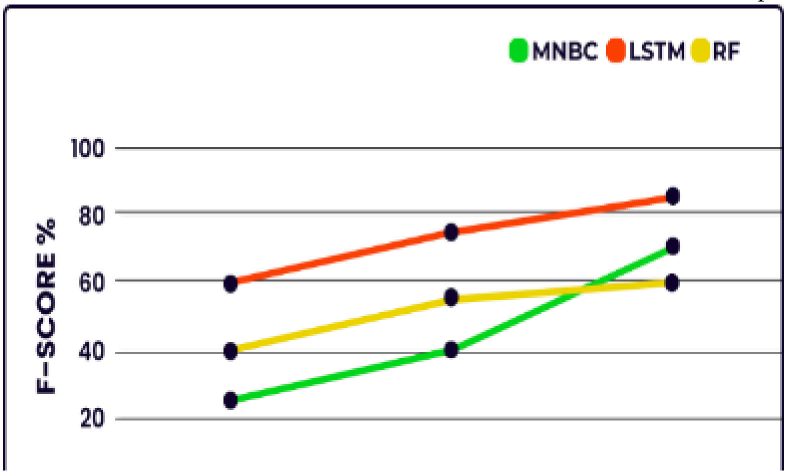


Figure 10 shows the f-scores results of MNBC, LSTM, and RF Algorithms.

accurate sentiment analysis capabilities.

Figure 10 shows the F score results of three machine learning methods utilized in our sentiment analysis system; Random Forest (RF), Multinomial Naive Bayes Classifier (MNBC), and Long Short Term Memory (LSTM). As the number of training examples grows from 20 to 60 all three algorithms exhibit enhanced F scores. Notably, LSTM consistently outperforms the rest by achieving the F scores. This implies that for our dataset and objective LSTM emerges

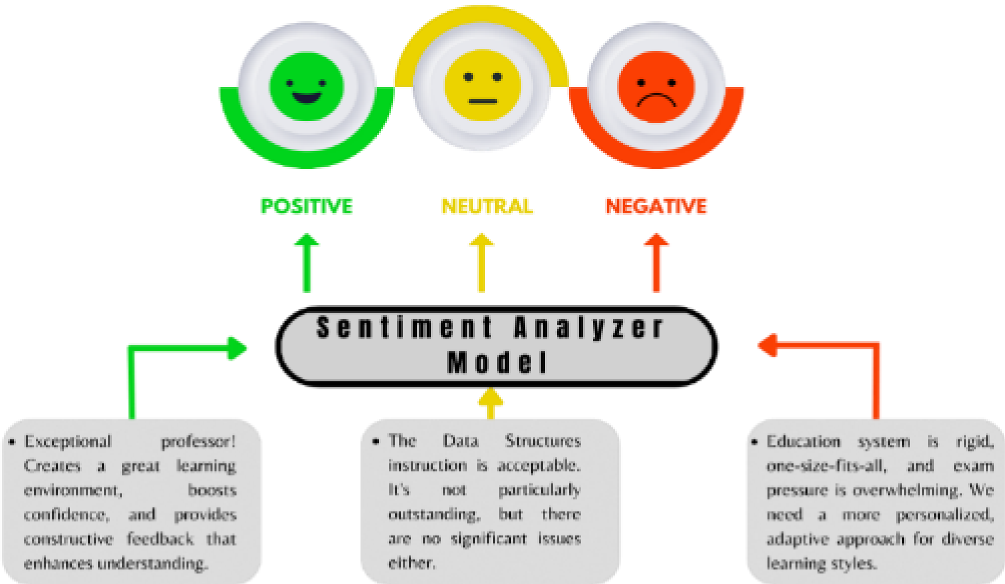


Figure 11 Shows the Final working method of the sentiment analysis model with the example of sentiments they generated.

as the efficient algorithm, for sentiment analysis displaying exceptional accuracy, in detecting positive, negative, and neutral sentiments.

Figure 11 demonstrates how a sentiment analysis model is created using an LSTM algorithm to categorize the sentiment of text data. By training the model with labeled data it successfully captures the connections, within sentences resulting in sentiment predictions for each case. The visualization showcases how sentences are classified into negative categories showcasing the LSTM's capacity to grasp subtle emotional nuances in the data. This emphasizes the efficiency of LSTM-based models, in assessing sentiment in written material.

5. CONCLUSION AND FUTURE WORK

The downside of institutional grading feedback is that it does not accurately reflect students' feelings and fails to utilize them to achieve positive outcomes. The analysis of this textual feedback and the development of an accurate model fill this gap by enabling universities and other educational institutions to utilize this research and improve the quality of education. The feedback is collected and then provided to the trained model in this study, where the sentiments are categorized as positive, negative, and neutral. According to this study, the long short-term memory (LSTM) algorithm yields higher accuracy compared to the Multinomial Naive Bayes and Random Forest Classifier algorithms. Moreover, the long short-term memory method outperforms the other two methods with an accuracy of 95.75%. Future research in this area will likely focus on enhancing the model's accuracy by incorporating a large volume of training data.

References:

[1] P. Ren, L. Yang, and F. Luo, "Automatic scoring of student feedback for teaching evaluation based on aspect-level sentiment analysis," Educ Inf Technol (Dordr), vol. 28, no. 1, pp. 797–814, Jan. 2023, doi: 10.1007/s10639-022-11151-z.

[2] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," IEEE Intell Syst, vol. 28, no. 2, pp. 15–21, 2013, doi: 10.1109/MIS.2013.30.

[3] Mahāwitthayālai Songkhlānakharin. College of Computing, C. Electrical Engineering/Electronics, IEEE Thailand Section, and Institute of Electrical and Electronics Engineers, The 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology : ECTI-CON 2020 : 24-27 June 2020, virtual conference hosted by College of Computing, Prince of Songkla University.

[4] T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, and L. Galligan, "Sentiment analysis and opinion mining on educational data: A survey," Natural Language Processing Journal, vol. 2, p. 100003, Mar. 2023, doi: 10.1016/j.nlp.2022.100003.

[5] Z. Kastrati, F. Dalipi, A. S. Imran, K. P. Nuci, and M. A. Wani, "Sentiment analysis of students' feedback with nlp and deep learning: A systematic mapping study," Applied Sciences (Switzerland), vol. 11, no. 9. MDPI AG, 2021. doi: 10.3390/app11093986.

[6] W. Shen, International Working Group on Computer Supported Cooperative Work in Design, Institute of Electrical and Electronics Engineers, and M. IEEE Systems, Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD) : June 27-29, 2013, Whistler, BC, Canada.

[7] M. Ali and A. Imdad, "Sentiment Summerization and Analysis of Sindhi Text," International Journal of Advanced Computer Science and Applications, vol. 8, no. 10, pp. 296–300, 2017, doi: 10.14569/ijacsa.2017.081038.

[8] N. Ahmed, M. A. Khouro, A. Khan, M. Dawood, M. A. Dootio, and N. U. Jan, "Student textual feedback sentiment analysis using machine learning techniques to improve the quality of education," Pakistan Journal of Engineering, Technology & Science, vol. 11, no. 2, pp. 32–40, Dec. 2023, doi: 10.22555/pjets.v11i2.1039.

[9] D. D. Dsouza, Deepika, D. P. Nayak, E. J. Machado, and N. D. Adesh, "Sentimental analysis of student feedback using machine learning techniques," International Journal of Recent Technology and Engineering, vol. 8, no. 1 Special Issue 4, pp. 986–991, 2019.

[10] IEEE Computer Society. Malaysia Chapter, IEEE Computer Society, Institute of Electrical and Electronics Engineers, S. IEEE Conference on Open Systems (2013 : Kuching, and S. IEEE Conference on Wireless Sensors (2013 : Kuching, IC3e 2013 : 2013 IEEE Conference on e-Learning, e-Management and e-Services : 2-4 December 2013, Kuching, Sarawak, Malaysia.

[11] IEEE Communications Society and Institute of Electrical and Electronics Engineers, 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) : 13-16 Sept. 2017.

[12] Z. Kastrati, B. Arifaj, A. Lubishtani, F. Gashi, and E. Nishliu, “Aspect-Based Opinion Mining of Students’ Reviews on Online Courses,” in ACM International Conference Proceeding Series, Association for Computing Machinery, Apr. 2020, pp. 510–514. doi: 10.1145/3404555.3404633.

[13] I. Ali Kandhro, M. Ameen Chhajro, K. Kumar, H. N. Lashari, and U. Khan, “Student Feedback Sentiment Analysis Model Using Various Machine Learning Schemes A Review,” Indian J Sci Technol, vol. 14, no. 12, pp. 1–9, Apr. 2019, doi: 10.17485/ijst/2019/v12i14/143243.

[14] H. Zhao, Z. Liu, X. Yao, and Q. Yang, “A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach,” Inf Process Manag, vol. 58, no. 5, Sep. 2021, doi: 10.1016/j.ipm.2021.102656.

[15] N. Ahmed, “A review of existing Machine Translation Approaches, their Challenges and Evaluation Metrics,” Pakistan Journal of Engineering, Technology & Science, vol. 11, no. 1, pp. 29–44, Dec. 2023, doi: 10.22555/pjets.v11i1.1002.

[16] M. M. Almosawi and S. A. Mahmood, “Lexicon-Based Approach For Sentiment Analysis To Student Feedback.” [Online]. Available: http://www.webology.org

[17] Z. Nasim, Q. Rajput, and S. Haider, “Sentiment analysis of student feedback using machine learning and lexicon-based approaches,” International Conference on Research and Innovation in Information Systems, ICRIIS, pp. 1–6, 2017, doi: 10.1109/ICRIIS.2017.8002475.

[18] G. S. Chauhan, P. Agrawal, and Y. K. Meena, “Aspect-based sentiment analysis of students’ feedback to improve teaching–learning process,” in Smart Innovation, Systems and Technologies, Springer Science and Business Media Deutschland GmbH, 2019, pp. 259–266. doi: 10.1007/978-981-13-1747-7_25.

[19] D. K. Dake and E. Gyimah, “Using sentiment analysis to evaluate qualitative students’ responses,” Educ Inf Technol (Dordr), vol. 28, no. 4, pp. 4629–4647, Apr. 2023, doi: 10.1007/s10639-022-11349-1.

[20] P. Mata, J. X. Rita, A. Batista, and J. X. Rita, “Sentiment analysis – a literature review,” Academy of Entrepreneurship Journal, vol. 27, no. SpecialIssue 2, pp. 1–10, 2021.

[21] B. K. Bhavitha, A. P. Rodrigues, and N. N. Chiplunkar, “Comparative study of machine learning techniques in sentimental analysis,” Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2017, no. Icicct, pp. 216–221, 2017, doi: 10.1109/ICICCT.2017.7975191.

[22] IEEE Computational Intelligence Society, International Neural Network Society, Institute of Electrical and Electronics Engineers, and B. C.) IEEE World Congress on Computational Intelligence (2016 : Vancouver, 2016 International Joint Conference on Neural Networks (IJCNN) : 24-29 July 2016, Vancouver, Canada.

[23] K. S. Krishnaveni, R. R. Pai, and V. Iyer, “Faculty rating system based on student feedbacks using sentimental analysis,” 2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017, vol. 2017-Janua, pp. 1648–1653, 2017, doi: 10.1109/ICACCI.2017.8126079.

[24] P. P. A., “Performance Evaluation and Comparison using Deep Learning Techniques in Sentiment Analysis,” Journal of Soft Computing Paradigm, vol. 3, no. 2, pp. 123–134, Jul. 2021, doi: 10.36548/jscp.2021.2.006.

[25] M. Baygin, “Classification of Text Documents based on Naive Bayes using N-Gram Features,” 2018 International Conference on Artificial Intelligence and Data Processing, IDAP 2018, pp. 1–5, 2019, doi: 10.1109/IDAP.2018.8620853.

[26] X. Fang and J. Zhan, “Sentiment analysis using product review data,” J Big Data, vol. 2, no. 1, 2015, doi: 10.1186/s40537-015-0015-2.

[27] “SENTIMENTAL ANALYSIS OF STUDENT FEEDBACK USING MACHINE LEARNING TECHNIQUES.”

[28] R. M. Duwairi and I. Qarqaz, “Arabic sentiment analysis using supervised classification,” in Proceedings - 2014 International Conference on Future Internet of Things and Cloud, FiCloud 2014, Institute of Electrical and Electronics Engineers Inc., Dec. 2014, pp. 579–583. doi: 10.1109/FiCloud.2014.100.

[29] 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC). IEEE, 2016.

[30] Jāhāṅgīranagara Biśvabidyālaẏa. Department of Computer Science and Engineering, Jāhāṅgīranagara Biśvabidyālaẏa, South Asian University, Institute of Electrical and Electronics Engineers. Bangladesh Section, and Institute of Electrical and Electronics Engineers, 2016 International Workshop on Computational Intelligence : 12-13 December, Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh.

[31] M. Zhou, N. Duan, S. Liu, and H. Y. Shum, "Progress in Neural NLP: Modeling, Learning, and Reasoning," Engineering, vol. 6, no. 3, pp. 275–290, 2020, doi: 10.1016/j.eng.2019.12.014.

[32] G. Qiang, "Research and improvement for feature selection on naive bayes text classifier," Proceedings of the 2010 2nd International Conference on Future Computer and Communication, ICFCC 2010, vol. 2, pp. 156–159, 2010, doi: 10.1109/ICFCC.2010.5497362.

[33] S. Xu, Y. Li, and Z. Wang, "Bayesian multinomial naïve bayes classifier to text classification," Lecture Notes in Electrical Engineering, vol. 448, no. 15, pp. 347–352, 2017, doi: 10.1007/978-981-10-5041-1_57.

[34] N. Sharma and M. Singh, "Modifying Naive Bayes classifier for multinomial text classification," 2016 International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2016, 2016, doi: 10.1109/ICRAIE.2016.7939519.

[35] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 67, no. 1, pp. 93–104, 2012, doi: 10.1016/j.isprsjprs.2011.11.002.

[36] M. Junker, R. Hoch, and A. Dengel, "On the evaluation of document analysis components by recall, precision, and accuracy," Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, pp. 717–720, 1999, doi: 10.1109/ICDAR.1999.791887.

[37] B. Marie, A. Fujita, and R. Rubino, "Scientific credibility of machine translation research: A meta-evaluation of 769 papers," ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference, pp. 7297–7306, 2021, doi: 10.18653/v1/2021.acl-long.566.