International Conference on "Interdisciplinary Research in Technology & Management" (IRTM 2021)

Conference Proceedings

Edited by

Satyajit Chakrabarti
Director, IEM Kolkata, India

Rintu Nath
Scientist – F, Vigyan Prasar, Department of Science and Technology, Govt. of India

Pradipta Kumar Banerji
Dean (Management Studies), Institute of Engineering and Management, Kolkata

Sujit Datta
Institute of Engineering & Management, Kolkata, India

Sanghamitra Poddar
Institute of Engineering & Management, Kolkata, India

Malay Gangopadhyaya
Institute of Engineering & Management, Kolkata, India

First edition published 2022
by CRC Press
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

and by CRC Press
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

CRC Press is an imprint of Informa UK Limited

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalogue entry has been requested.

ISBN: 978-1-003-20224-0 (ebk)

DOI: 10.1201/9781003202240

Chapter 100

Sentiment Analysis of Student Textual Feedback to Improve Teaching

Judith Goodness Khanyisa Mabunda
School of Computer Science and Applied Mathematics
The University of the Witwatersrand, Johannesburg, South Africa

Ashwini Jadhav
Science Teaching and Learning Unit
The University of the Witwatersrand, Faculty of Science, Johannesburg, South Africa

Ritesh Ajoodha
School of Computer Science and Applied Mathematics
The University of the Witwatersrand, Johannesburg, South Africa

Abstract—Instructor evaluation based on student feedback is essential in education, it allows instructors to see if their teaching has been effective. But, it is very challenging for an instructor who teaches numerous students to analyse feedback provided by all the students. To solve this problem, this paper develops a sentiment analysis model to analyse students' feedback to assess the effectiveness of teaching and learning. In this paper, machine learning models like, Support Vector Machines, Multinomial Naive Bayes, Random Forests, K-Nearest Neighbours and Neural Networks are trained on feature engineering and re-sampling techniques to classify student feedback into three sentiment classes: negative, positive or neutral, using student dataset collected from Kaggle. From the analysis before the resampling of the data, K-Nearest Neighbours model is found to be more efficient in predicting student sentiment towards teaching practices than the other models with good accuracy of 81%. After the re-sampling of the data, Neural Networks performed better than the other models with good accuracy of 84%. The model will help institutions make effective decisions towards teaching and learning strategies.

Keywords— Sentiment analysis, Student feedback, Feature engineering, Machine learning models, Re-sampling

I. INTRODUCTION

It has become common pr actice for educational institutions to collect student textual feedback on teaching and learning practices and use it as an improvement tool and performance measure. Student feedback can help instructors learn of their strengths and weaknesses and this will motivate them to perform better, thus improve their teaching. However, due to the unavailability of automated text analytic tools, the student feedback cannot be really used to their full advantage. Instructors cannot really derive useful information from the feedback to make effective decisions and improve their teaching [2].

Taking into consideration this issue, there has been determined attempts to address it. Prior research has shown that sentiment analysis is the best automated process that can be used by educational institutions to collect and analyse student feedback to evaluate the performance of instructors and highlight their strength and weaknesses in teaching.

Previous studies have used real student feedback from educational institutions as datasets from different sources like Learning Management System, Kaggle website, etc. They used uni-grams, bi-grams, tri-grams, word length, etc. as features.

Among all the features, uni-grams is widely used because of its better performance with classifiers. They considered models like Support Vector Machines, Naive Bayes and Complement Naive Bayes, Multinomial Naive Bayes, random Forest, Maximum Entropy, Decision Trees and K Nearest Neighbours. Among the different models, Support Vector Machine is found as the best classifier followed by Multinomial Naive Bayes classifier [2], [6], [12], [14], [16].

Sentiment analysis is an automated process that collects, analyse and classify text data into different sentiments such as positive, negative, or neutral especially to understand the attitudes, opinions and emotions expressed towards a particular topic, product and or service [4]. In a similar way, this paper aims to analyse student textual feedback and track positive, negative or neutral sentiment from it by developing a sentiment analysis system that can help in assessing instructors' performance and highlighting major areas of the instructors' strength and weaknesses.

To ensure that the developed system provides best results, five models such as Support Vector Machines (SVM), Multinomial Na¨ıve Bayes (MNB), Random Forest, K Nearest Neighbours (K-NN) and Neural Networks (NN) are trained with real students' feedback on pre-processing and feature engineering techniques. A comparative analysis is conducted among the first three models that are informed by the literature and the last two models that this paper claims will outperform the informed models. In light of the aim of this paper, the following questions are set out for the paper to answer: What are students' attitudes about teaching performance of the instructor? How does analysing student feedback improve instructor performance?

In addition, this paper suggests interpreting students' feedback by generating a word cloud for visualisation. A Word Cloud is a visual representation of the words used in a particular piece of text, with the size of each word indicating its relative frequency [3]. The more frequent the word is in a text, the larger it is in the visual. This paper contributes to the current body of literature by providing a predictive model to classify student textual responses using sentiment analysis and by providing possible strategies to use the comments provided by the predictive model to improve one's teaching and learning.

The rest of the paper is structured as follows: Section II discusses the related work that highlights work done by other authors. Section III discusses the proposed methodology which comprises of data, features, models and evaluation. Section IV presents the results and discussion of the findings. Finally, section V concludes the paper and provides future work recommendations.

II. RELATED WORK

This section explains some of the various research studies related to sentiment analysis of student feedback to improve teaching and learning approaches and limitations met by these research studies. There have been many attempts to deduce the sentiment of the responses in student forums.

A. Data

Different authors have used real students' feedback from different sources for instructor assessment. In this paper [6], the authors collected students' feedback using google forms. This paper [12] used student feedback gathered at the end of 63 courses. In this paper [14] students' feedback are collected from Kaggle website. This paper [9] used data from Learning Management System from computer science students.

On the other side, this paper [2] collected students' feedback from an institution and labelled data source by 3 experts, these includes; End of Unit Other Institutes, End of Unit University of Portsmouth and Real-time feedback University of Portsmouth, for these experts they considered 768, 117 and 190 instances respectively. The author in this paper [16] collected feedback from an institution, the data has 1036 records of students' feedback with 641 positive sentiments, 292 negative sentiments and 103 neutral sentiments. Maintaining the Integrity of the Specifications

B. Features

Several studies have used different features like uni-grams, bi-grams, tri-grams, word length, etc. Uni-grams is widely used. Paper [6] found that machine learning algorithms with uni-grams are better in performance than bi-grams. This paper

[2] got the highest performance with the use of uni-grams. On the other hand, the authors in this paper [12] used uni-grams and found that the feedbacks are incorrectly classified.

C. Models

Many different classifiers are considered like Support Vector Machines, Naive Bayes and Complement Naive Bayes, Multinomial Naive Bayes, random Forest, Maximum Entropy, Decision Trees and K-Nearest Neighbours. Among these classifiers, Support Vector Machines classifier is found to be the best by [2], [9], [16]. On the other hand, this paper [6] found that Multinomial Na¨ıve Bayes is better in performance.

D. Accuracy

Some of the authors concluded that support vector machine classifier performs better in predicting whether the opinion expressed by students towards teaching performance is positive, negative or neutral, with accuracy of 94%, 97% and 85% respectively [2], [9], [16]. On the other hand, this paper [6] compared Multinomial Naive Bayes, Support Vector Machines, etc. and concluded that Multinomial Naive Bayes classifier is more accurate with accuracy of 80%.

The various research studies related to sentiment analysis of student feedback to improve teaching and learning approaches are also presented in TABLE I.

III. RESEARCH METHODODLGY

This section provides the proposed methodology. In more details, this section outlines the methods of data collection, the features used, models, and the type of data evaluation functions.

A. Data

In this subsection, data collection methods are discussed, followed by two sub processes such as data pre-processing and word cloud visualization which are essential to extract meaningful insights from the students' feedback.

1) *Data Collection*: Data consisting of students' feedback is collected from Kaggle. This is an existing dataset that is collected from a prominent university in India and it is used to create an overall institutional report. It consists of 185 records with 12 columns of which 6 columns are features like teaching, course content, examination, etc. with students' feedback and 6 columns are ratings of the students' feedback [1]. In this paper, only the teaching feature with student feedback about how they feel towards the teaching practices and its corresponding ratings column are considered.
2) *Data Pre-processing:* The following pre-processing phases are performed in Python Jupyter Notebook:
   a) *Basic Data Cleaning:* The first step in cleaning the data is to remove stop words, words that do not add much meaning to the students' feedback, words like, have, the, etc. in order to improve classification accuracy. Punctuation are removed and the data is converted into lowercases. The feedbacks are then split into individual tokens such as words, this is because much processing of raw text is done at token level.
   b) *Vocabulary:* The tokens are then used to prepare a vocabulary. Vocabulary refers to the set of unique words in the corpus (collection of text data) [5]. Vocabulary is used to create the tokenized input sentences. Each word in the vocabulary is treated as a unique feature.
   c) *Conversion of text into numeric:* Texts are converted into numeric, this is because some machine learning libraries do not take categorical variables as input. The conversion of text into the corresponding numerical form can be achieved in various approaches. In this paper, we used the bag of words model to convert text to numbers.
   d) *Feature Engineering:* In this phase, raw text feedback is transformed into feature vectors and new features are generated using the existing dataset. The following approaches are implemented in order to obtain relevant features from our dataset: Count Vectorizer and TF-IDF. They use vocabulary as features.

- Count Vectors as features – used to get the texts' bag-of-words counts as a vector.
- TF-IDF Vectors as features - Bag of words approach has a drawback because it assigns a score to a word based on its occurrence in a particular document. It doesn't take into account the fact that the word might also be having a high frequency of occurrence in other documents as well. TFIDF feature resolves this issue by multiplying the term frequency of a word by the inverse document frequency. The TF stands for "Term Frequency" while IDF stands for "Inverse Document Frequency". TF-IDF is a numerical statistic that reflects how important a word is to a document in a collection [13].

e) *Train Test Split:* To overcome overfitting, the dataset is divided into training and testing splits, this allows us to see how the algorithms performed during the testing phase. The split ratio is 80:20.

3) *Word Cloud Visualization:* A word cloud is generated from the words in the students' feedback in Python. The more frequent the word is in the students' feedback, the larger it is in the word cloud visual.

B. FEAUTURES

Features like, CountVectorizer, TFIDF and SMOTE are implemented in this paper. The CountVectorizer is used to create a bag of words for the conversion of student feedback into numerical form while the TFIDF is used to show the importance of each word in the vocabulary. Resampling technique, such as, SMOTE technique is applied in this paper because the distribution of the sentiment classes is not balanced, so this technique is used to balance the distribution of the classes to enhance the performance of the classification models.

C. MODELS

Three machine learning classifiers, Support Vector Machines (SVM), Multinomial Na¨ıve Bayes (MNB) and Random Forests (RF) informed by the literature are used in this paper due to their better performance with good accuracy in earlier research. A comparative analysis is conducted between these three classifiers and other two machine learning classifiers, such as Neural Networks and K-Nearest Neighbours that we think will outperform the literature and give good results based on the data.

Table I This table showcases work done by other authors on sentiment analysis of student feedback

Authors	Data	Features	Models	Evaluation
[Nabeela Altrabsheh, Cocea, and Fallahkhair 2014]	Labelled by 3 experts: End of Unit Other Institutes, 768 instances. End of Unit University of Portsmouth, 117 instances. Real-time feedback University of Portsmouth, 190.	Unigrams	Naive Bayes, Complement Naive Bayes (CNB), Maximum Entropy and Support Vector Machine (SVM) were trained using real students' feedback. 10-fold cross-validation was used to test learning performance.	Accuracy: SVM is of 94% and CNB is 84%.
[Ullah 2016]	Institution data, 1036 data with 641 positive sentiments, 292 negative sentiments and 103 neutral sentiments.	Uni-gram, Bi-gram, Tri-gram and Opinion words	Support Vector machines, Naïve Bayes and complement Naïve Bayes and Maximum Entropy and apply of neutral class.	Accuracy: 97%, 89%, 84% and 87%.
[Rajput 2016]	1748 students' feedback provided at the end of 63 courses conducted during 2010 and 2014.	Type, length of word and part of speech of the word	Sentiment analysis was performed using Knime workflow	To measure its performance, accuracy, recall, precision, and F-measure were computed and the results were found to be very positive.
[Shinde 2019]	kaggle website, 186 records.	Student Response System	For analysis, Microsoft Excel tool was used. With R programming.	Neutral reviews(105) surpass those of positive(8) and negative(8).
[Adesh 2019]	Google forms	ngram method	Support Vector Machine (SVM), Naïve Bayes (NB), Random Forests	Accuracy: NB is 80%.
[Kavitha 2019]	Learning Management System (LMS), 200 samples from computer science students.	Performance in statistical learning Chi-square, IG, Mutual Information and Symmetric Uncertainty.	Naïve Bayes, Support Vector machine, Decision Tree and K-Nearest Neighbour. NLTK and Valence Aware Dictionary and sentiment reasoner for analysis.	Accuracy: 85% for SVM and Information Gain performed better.

Therefore this paper uses five classifiers. The classifiers are trained using the training dataset on TF-IDF vectors and tested with the testing dataset. These models are briefly described in the following subsections.

Support Vector Machines

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. The model extracts a best possible hyperplane line that separates two classes [10].

Multinomial Naïve Bayes

Multinomial Naïve Bayes (MNB) has been widely used in text classification. Given a set of labelled data, MNB often uses a parameter learning method called Frequency Estimate (FE), which estimates word probabilities by computing appropriate frequencies from data [15].

Random Forests

Random forest algorithm is a supervised classification model. This model creates the forest with number of trees [7]. Random forest classifier works well with missing values. It does not overfit the model when there are more trees in the forest.

K-Nearest Neighbours

K-Nearest Neighbour is a non-parametric classification algorithm. This algorithm works by finding the distances between a new data point and all the labelled datasets in the data, it reads through the whole dataset to find out the k nearest neighbours closest to the new data point, then the votes for the most frequent label in classification will be the class for the new data point [11].

Neural Networks

Neural Network is a complex model which tries to imitate the way the human brain develops classification rules. A neural net consists of different layers of neurons. Each layer receives inputs from previous layers and pass outputs to further layers. Its advantage is the ability to create complex prediction functions and emulate human thinking in a way that no other algorithm can [17].

D. Evaluation

Confusion matrix is used for evaluation, accuracy is calculated from the confusion matrices for both classifiers. The accuracy of all the classifiers is evaluated using 5-fold cross validation method on the training dataset. The goal for cross validation is to test the model in the training phase and then provide insight on how the specific model adapts. Cross validation is very efficient in mitigating overfitting. Overfitting refers to a model that performs too well on the training dataset but underperforms on new data (test dataset).

The interpretation in the Confusion Matrix is as follows:

- True Positive (TP), when both the actual and predicted values are positive.
- True Neutral (TNeu), when both the actual and predicted values are neutral.
- True Negative (TNeg), when both the actual and predicted values are negative.
- False Positive (FP), when the actual value is either negative or neutral and the predicted value is positive.
- False Neutral (FNeu), when the actual value is either positive or negative and the predicted value is neutral.
- False Negative (FNeg), when the actual value is either positive or neutral and the predicted value is negative.

Table II A confusion matrix to describe the performance of a classification model on a set of test data for which the true values are known.

Actual	Predicted		
	Negative (–1)	**Neutral (0)**	**Positive (1)**
Negative (-1)	TNeg	FNeu	FP
Neutral (0)	FNeg	TNeu	FP
Positive (1)	FNeg	FNeu	TP

The accuracy is computed as:

$$Accuracy = \frac{TP + TNeu + TNeg}{TP + TNeu + TNeg + FP + FNeu + FNeg} * 100$$

IV. RESULTS AND DISCUSSION

This section presents in details the experimental results of predicting the sentiment of students using existing dataset that is designed based on students' feedback acquired from a prominent university in India. The dataset comprises of 185 records of students' feedback about how good they find their instructors' teaching practices. Sentiment analysis is performed using Python programming language. 5-fold cross validation method is used to evaluate the accuracy of all the five classifiers in the training dataset. Experiments are conducted in two ways: Firstly, the training dataset is split into 5 parts, this is known as 5-fold cross validation.

The models are trained on the five parts on TF-IDF vectors before and after SMOTE. Secondly, after cross validation on training dataset, the analysis is validated on the testing dataset on TFIDF vectors before and after SMOTE technique to see if the models are able to create accurate predictions on data they haven't been trained on.

A. Word cloud visualization

A word cloud from the complete 185 students' feedback is shown in Fig. 1. The larger the word is in the visual the more frequent it appears in the students' feedback. The most frequent words are good, lecture, delivery, and so forth. Word clouds can help in identifying patterns that are difficult to identify from reading the students' feedback [8].

Fig. 1. Word Cloud of student feedback about teaching.

A. Class Distribution before and after Smote

This section shows how the data is distributed before and after applying SMOTE technique to balance the class distribution from the dataset. Fig. 2 shows the class distribution before applying SMOTE technique, it can be seen that data is mostly distributed to the positive class 1. On the other hand, Fig. 3 shows the class distribution after applying SMOTE technique, it is clear now that the classes are equally distributed.

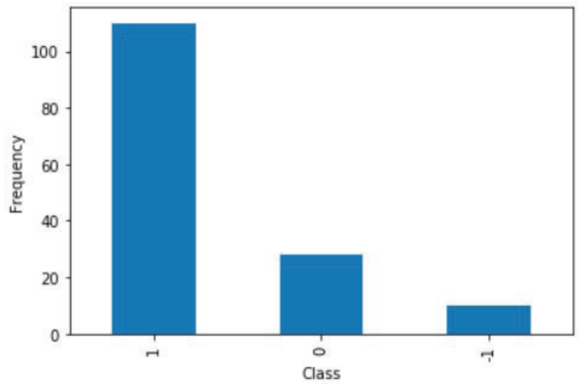

Fig. 2. This figure shows the distribution of each sentiment (positive, negative or neutral) on the students' feedback before the data was balanced using SMOTE technique

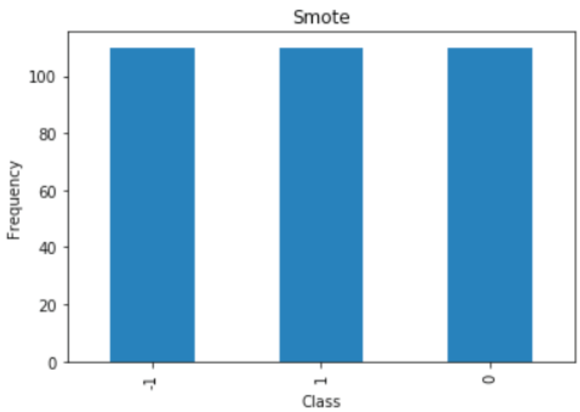

Fig. 3. This figure shows the distribution of each sentiment (positive, negative or neutral) on the students' feedback after the data was balanced using SMOTE technique.

B. 5-Fold Cross Validation before and after SMOTE

Training dataset is split into 5-folds cross-validation and then an average of the results are taken for each model. TABLE III shows the accuracy of the analysis done on each fold and the average (mean) accuracy before and after SMOTE technique is applied. It can be seen that SVM and RF performed better than the other classifiers with mean accuracies of 80% in the training dataset before the class distribution is balanced. After balancing the class distribution, the NN performed better with mean accuracy of 80% and SVM mean accuracy remained constant, this is because of the smaller dataset used in this work.

C. Analysis on test dataset before and after SMOTE

In the experiments, the five models are used to analyse the students' feedback to predict the sentiment of students towards teaching. Therefore, TABLE IV depicts the confusion matrices of all the models before balancing the class distribution using SMOTE technique. It can be seen from these matrices that there are 3 students' feedback in the negative class for SVM, MNB, RF and NN, then the models were successful in predicting 0 of those correctly in the negative class, but 3 were classified as positive. In contrast, there are 3 students' feedback in the negative class in K-NN, 0 of them was classified correctly, but 1 of them was classified in the neutral class and 2 were classified in the positive class.

In contrast, there are 7 students' feedback in the neutral class for all the classifiers, but 1 of them was classified correctly by SVM and NN. These two classifiers predicted 6 in the positive class. On the other hand, MNB and RF classified 0 of these 7 feedback correctly, but 7 were classified on the positive class. However, K-NN classified 5 of these students' feedback correctly, but 2 were classified as positive.

There are 27 students' feedback in the positive class for all the classifiers and all of them are classified correctly by SVM and MBN. However, RF and NN classified 26 of the 27 students' feedback correctly, but RF classified 1 of them in the negative class and NN classified 1 of them in the neutral class. On the other hand, K-NN classified 25 of them correctly, but 2 of them were classified in the neutral class.

Therefore the confusion matrices shows that the best performance was for positive class. The lower number of records in the dataset affected the performance of the negative and neutral classes.

TABLE V depicts the confusion matrices of all the models after balancing the class distribution using SMOTE technique. It can be observed from the confusion matrices after SMOTE that the performance on negative and neutral classes improved. SVM, MNB, K-NN and NN improved significantly in classifying all the students' feedback correctly for the negative class. But, for the neutral class, only the RF improved in classifying the feedback correctly. However, the best performance is still for the positive class like before SMOTE technique was applied, this is because of the smaller number of records from the dataset.

TABLE VI shows the prediction accuracy of the five models before and after balancing the class distribution, from the analysis before SMOTE technique is applied, K-Nearest Neighbors model is found to be more efficient in predicting student sentiment towards teaching practices than the other models with good accuracy of 81%. After SMOTE technique was applied, Neural Networks performed better than the other models with good accuracy of 84%.

V. DISCUSSION

The five models performed well when classifying the sentiment of students towards teaching practices, because their properties match with the properties of the dataset. This means they are able to accurately classify small amount of data. Amongst these models, two of them, the K-NN and NN were hypothesised to work better than the other three models (MNB, SVM and RF) informed by the literature. This hypothesis was made because the models informed by the literature are mostly considered in text classification as they are proven to be best by many studies, therefore, this research wants to see how other models perform in predicting sentiment from text data.

Both the K-NN and NN were chosen because they were not mostly used in prior research and they can be trained on small dataset. As hypothesised by this research, the K-NN performed better than the other models before the class distribution could be balanced. On the other hand, the NN model performed better than the other models after the class distribution was balanced.

Table III The accuracy of each model for each part in 5-fold cross validation on training dataset before and after the class distribution was balanced and the average accuracy.

K-Fold	SVM		MNB		RF		K-NN		NN	
	Before	After	Before	After	Before	After	Before	After	Before	After
1-fold	83%	80%	73%	76%	83%	80%	90%	88%	73%	78%
2-fold	83%	83%	73%	76%	83%	80%	70%	80%	73%	76%
3-fold	80%	79%	73%	76%	77%	78%	70%	79%	77%	79%
4-fold	76%	76%	76%	80%	76%	76%	72%	70%	76%	87%
5-fold	79%	80%	76%	79%	83%	79%	79%	70%	76%	80%
Mean	80%	80%	74%	77%	80%	79%	76%	77%	75%	80%

SVM			
Actual Values	Predicted Values		
	Negative (-1)	Neutral (0)	Positive (1)
Negative (-1)	0	0	3
Neutral (0)	0	1	6
Positive (1)	0	0	27
MNB			
Actual Values	Predicted Values		
	Negative (-1)	Neutral (0)	Positive (1)
Negative (-1)	0	0	3
Neutral (0)	0	0	7
Positive (1)	0	0	27
RF			
Actual Values	Predicted Values		
	Negative (-1)	Neutral (0)	Positive (1)
Negative (-1)	0	0	3
Neutral (0)	0	0	7
Positive (1)	1	0	26
K-NN			
Actual Values	Predicted Values		
	Negative (-1)	Neutral (0)	Positive (1)
Negative (-1)	0	1	2
Neutral (0)	0	5	2
Positive (1)	0	2	25
NN			
Actual Values	Predicted Values		
	Negative (-1)	Neutral (0)	Positive (1)
Negative (-1)	0	0	3
Neutral (0)	0	1	6
Positive (1)	0	1	26

Table IV This table shows the confusion matrices that describe the performance of the classification models before SMOTE technique was applied to balance the class distribution. The diagonal numbers are correctly classified students' feedback.

SVM			
Actual Values	Predicted Values		
	Negative (-1)	Neutral (0)	Positive (1)
Negative (-1)	3	0	0
Neutral (0)	3	3	1
Positive(1)	2	1	24
MNB			
Actual Values	Predicted Values		
	Negative (-1)	Neutral (0)	Positive (1)
Negative(-1)	3	0	0
Neutral(0)	4	3	0
Positive(1)	1	2	24
RF			
Actual Values	Predicted Values		
	Negative (-1)	Neutral (0)	Positive (1)
Negative (-1)	1	1	1
Neutral (0)	1	2	4
Positive (1)	0	0	27
K-NN			
Actual Values	Predicted Values		
	Negative (-1)	Neutral (0)	Positive (1)
Negative (-1)	3	0	0
Neutral (0)	1	2	4
Positive (1)	1	2	24
NN			
Actual Values	Predicted Values		
	Negative (-1)	Neutral (0)	Positive (1)
Negative (-1)	3	0	0
Neutral (0)	2	3	2
Positive (1)	1	1	25

Table V This table shows the confusion matrices that describe the performance of the classification models after SMOTE technique was applied to balance the class distribution.

VI. CONCLUSION ANF FUTURE WORK

This section presents the conclusion of this paper and suggestions for future work in sentiment analysis.

A. Conclusion

In this paper, a combination of pre-processing phases with different feature engineering techniques, re-sampling technique and machine learning classification models is applied to analyse students' textual feedback. It is found that, using pre-processing methods, feature engineering techniques and machine learning models properly increased accuracy when predicting sentiment of students' feedback towards teaching practices. It is also observed that, analysis on text dataset with an unbalanced class distribution affects the accuracy of some classes. The use of re-sampling techniques to balance class distribution shows an improved performance of models in sentiment analysis.

From the results after resampling technique, SMOTE, it can be seen that as a result of the smaller dataset used in this paper, the accuracy of the negative and neutral classes are still affected. Therefore, from this paper, it is suggested that large datasets be used for significant learning like sentiment analysis. The results proves true the hypothesis proposed by this paper that K-NN and Neural Networks models will outperform the models informed by the literature, Neural Networks performed better than the other models. This paper indicates that Neural Network model gives very good results; therefore, it could be used for real-time students' feedback analysis.

Table VI Prediction accuracy of the models before and after the class distribution was balanced

Models	Accuracy before SMOTE	Accuracy after SMOTE
Support Vector Machine	76%	81%
Multinomial Naıve Bayes	73%	81%
Random Forest	70%	81%
K-Nearest Neighbors	81%	78%
Neural Networks	73%	84%

From the analysis, it can be observed that the accuracy measured against the 5-fold cross validation of the training dataset is very good and accuracy measured against a test dataset is also good, therefore, this paper concludes that the models generalize well from the training dataset to unseen data(test dataset). The sentiment analysis model built by this paper is able to analyse and classify students' feedback' into positive, negative or neutral. Educational institutions can utilize this model to analyse and understand how students feel about teaching practices. From the analysis, effective decisions can be made to improve teaching.

B. Future Work

For experimental purposes, this paper used small dataset of 185 records, however, the five models are bias towards the negative and neutral classes in terms of accuracy. Therefore, from this paper, it is suggested that large datasets be used for significant learning like sentiment analysis to avoid biasness. This paper also suggests that Neural Networks be used for further studies because it gives efficient accuracy in prediction

ACKNOWLEDGMENT

This work is based on the research supported in part by the national research foundation of south africa (grant number: 121835).

REFERENCES

[1] AjitBrar. Student feedback dataset institutional academic performance evaluation through student feedback. Springer, 2017.

[2] Nabeela Altrabsheh, Mihaela Cocea, and Sanaz Fallahkhair. Learning sentiment from students' feedback for real-time *interventions in classrooms. In International conference on adaptive and intelligent systems,* pages 40–49. Springer, 2014.

[3] Rob Atenstaedt. Word cloud analysis of the bjgp. *Br J Gen Pract,* 62(596):148–148, 2012.

[4] Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik Van Der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. Sentiment analysis in the news. *arXiv preprint arXiv:*1309.6202, 2013.

[5] Wenhu Chen, Yu Su, Yilin Shen, Zhiyu Chen, Xifeng Yan, and William Wang. How large a vocabulary does text classification need? a variational approach to vocabulary selection. *arXiv preprint arXiv:*1902.10339, 2019.

[6] Daneena Deeksha Dsouza, Divya P Nayak Deepika, Elveera Jenisha Machado, and ND Adesh. Sentimental analysis of student feedback using machine learning techniques.

[7] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition,* volume 1, pages 278–282. IEEE, 1995.

[8] Thidaporn Jumpakate. Word clouds and english language teaching in thai classroom. *RMUTK Journal of Liberal Arts,* 2(1):15–26, 2020.

[9] R.K. Kavitha. Sentiment research on student feedback to improve experiences in blended learning environments. In *International Journal of Innovative Technology and Exploring Engineering (IJITEE),* volume 8, pages 693–700. IJITEE, 2019.

[10] Chinmayee Kinage, Abhishek Kalgutkar, Amruta Parab, Sejal Mandora, and Sunita Sahu. Performance evaluation of different machine learning based algorithms for flood prediction and model for real time flood prediction. In *2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), pages* 1–7. IEEE, 2019.

[11] Leif E Peterson. K-nearest neighbor. *Scholarpedia,* 4(2):1883, 2009.

[12] Quratulain Rajput, Sajjad Haider, and Sayeed Ghani. Lexicon-based sentiment analysis of teachers' evaluation. *Applied Computational Intelligence and Soft Computing,* 2016, 2016.

[13] W Scott. Tf-idf from scratch in python on real world dataset. *Towards Data Science,* 15, 2019.

[14] Prashali S Shinde, Asmita R Kanase, Rutuja S Pawar, and Yamini U Waingankar. Sentiment analysis of feedback data. *Management,* pages 160–163, 2019.

[15] Jiang Su, Jelber S Shirab, and Stan Matwin. Large scale text classification using semi-supervised multinomial naive bayes. In *Proceedings of the 28th international conference on machine learning (ICML-11),* pages 97–104. Citeseer, 2011.

[16] Mohammad Aman Ullah. Sentiment analysis of students feedback: A study towards optimal tools. In *2016 International Workshop on Computational Intelligence (IWCI),* pages 175–180. IEEE, 2016.

[17] Zhan Wang, Yifan He, and Minghu Jiang. A comparison among three neural networks for text classification. In *2006 8th international Conference on Signal Processing,* volume 3. IEEE, 2006.