Using sentiment analysis to evaluate qualitative students' responses

Delali Kwasi Dake[1] · **Esther Gyimah[1]**

Abstract

Text analytics in education has evolved to form a critical component of the future SMART campus architecture. Sentiment analysis and qualitative feedback from students is now a crucial application domain of text analytics relevant to institutions. The implementation of sentiment analysis helps understand learners' appreciation of lessons, which they prefer to express in long texts with little or no restriction. Such expressions depict the learner's emotions and mood during class engagements. This research deployed four classifiers, including Naïve Bayes (NB), Support Vector Machine (SVM), J48 Decision Tree (DT), and Random Forest (RF), on a qualitative feedback text after a semester-based course session at the University of Education, Winneba. After enough training and testing using the k-fold cross-validation technique, the SVM classification algorithm performed with a superior accuracy of 63.79%.

Keywords Machine learning · Smart Education · Text analytics · Sentiment analysis · Unstructured text · Educational Data Mining · NRC emotion lexicon · Opinion mining

1 Introduction

One aspect of education that remains relevant to instructors is the valuable feedback mechanism related to lesson delivery and learners' understanding. Many tertiary institutions have insisted on semester-based evaluation metrics for instructors

✉ Delali Kwasi Dake
dkdake@uew.edu.gh

Esther Gyimah
egyimah@uew.edu.gh

[1] University of Education Winneba, Winneba, Ghana

to improve their teaching styles and learners' academic performance. The instructor, with the feedback, reflects on strategies to enhance learner engagement and the teaching methodology. Reflective practice allows instructors to adjust and respond to learner issues in developing a constructive learner model (Benade, 2015). Developing an appropriate skill set in students for post-graduate studies and the job market has become necessary for the modern-day instructor using reflective practice. Freidhoff (2008) ascertains the relevance of reflective practice to include the instructor's beliefs, assumptions, and values. Such values go beyond the instructor and involve the policies of the educational institutions and respective communities. Hence the learner responses that lead to instructor reflection must have ethical consideration to adapt teaching and learning strategies through actions (Priya et al., 2017) .

The quantitative and qualitative feedback mechanisms have been widely used in education to assess instructors. The quantitative feedback is numeric with percentages, rating scales, and grades, while the qualitative feedback comes in the form of commentary, assessment tasks, audio and video files (Nouri et al., 2019). Each feedback approach has merits that guide the instructor in varying the teaching dynamics to improve learner participation and performance. Tekian et al. (2017) explain the relevance of qualitative and quantitative feedback mechanisms in the learning process and caution against using the wrong assessment feedback in promoting learning. Each input from the student must be structured and analysed to aid the instructor in understanding the learner for adequate implementation and policies. A higher percentage of tertiary institutions have adopted quantitative feedback mechanisms due to the ease of analytics and easy execution of feedback results (Leung et al., 2021). The quantitative feedback mechanism linked closely to the closed-ended questionnaire approach gives a quick, definitive, and fast response (Schnall et al., 2018). Closed-ended questions have finite answers with instances of responses and analysis, but confines learners (Schnall et al., 2018). The open-ended questionnaire approach to qualitative feedback allows the learner to provide valuable free-form answers with no list of responses. The difficulty of standardising and analysing qualitative feedback from learners has restricted its usage among academic institutions (Schnall et al., 2018). Educational authorities must embrace qualitative feedback mechanisms since learners will have opinions, utterances, and feelings beyond limited options to full expression. The relevance of qualitative feedback mechanisms is heightened even further by the increasing use of online learning platforms with social media integration for teaching and learning during the pandemic. COVID-19 exposure necessitates a thorough investigation and analysis of all data sources that assist teachers in improving class engagement by altering their teaching techniques.

The advent of machine learning (ML) in text has tremendous application in industry, the medical field, and education. Machine learning has aspects that find frequent correlations and patterns in datasets with rules to predict item occurrence (Song & Lee, 2017). Machine learning has various methods, including pruning to improve item identification performance using rule ranking. Classification and clustering as data mining procedures can be used with Association Rule Mining (ARM) to better understand and analyse students' responses. While classification and clustering predict class labels and group itemset, the ARM determines the hidden relationship between items in the dataset (Mohapatra et al., 2021). The frequency and pattern iden-

tification of relevant keywords for more straightforward interpretation in a qualitative feedback system via open-ended questionnaires is vital in an intelligent educational environment. Deploying varying ML algorithms on the dataset will measure the correlated items' support and confidence for real-time analytics. The ML applications may be integrated as an add-on to any third-party online management system, which is particularly advantageous now that significantly more data is available online as a result of the COVID-19 outbreak.

The study aims to analyse qualitative students' feedback via an open-ended questionnaire using ML algorithms. The primary objective is to analyse feedback text from students and predict the class labels for more intelligent analytics and future predictions. In line with the goals of the study, we pose the following research questions:

(1) What categorisations of feedback comments were received after learners had filled out the questionnaire?
(2) What machine learning algorithm provides the best accuracy after sentiment analysis?
(3) What is the prediction performance of the best machine learning algorithm on new data?

2 Sentiment analysis and opinion mining

Sentiment analysis, or opinion mining, is a type of natural language processing (NLP) that helps determine the summarised view or emotional tone behind a text (Altrabsheh et al., 2013). Statistics from Taylor (2021) and Altexsoft (2020) estimated the proportion of structured and unstructured text at 20–80%, respectively. Unstructured text with no predefined systemisation forms the bulk of the data in the world today. Since unstructured text does not fit into any defined framework, finding insight for prediction and pattern identification has become problematic. The outburst of social media platforms with the need for expressing opinions on varying subjects has increased the demand for analysing text within a shorter time. From email spam classification, movie review recommendation, online assessment of learners, and customer review portfolios, the application domain of sentiment analysis is limitless. In education, the application of sentiment analysis has caught instructors' attention, especially with the surge in online platforms for teaching and learning. The need to improve the teaching and learning process has created an academic environment that needs constant review beyond quantitative formats. Learners typically express subjective feelings, views, emotions, and frustration when given a chance through open-ended questionnaire mechanisms. Instead of confining learners with closed-ended questionnaires, open-ended expressions are the only way to get a true sense of their feelings.

As depicted in Fig. 1, the instructor considers a new strategy for lesson delivery based on the determined feelings or continues with the current delivery mode based on great feedbacks.

1. Students in a class with different emotions. They have the opportunity to send feedback after the course

2. The feedback text is analysed for sentiments using Classification

Excellent

Poor

Good
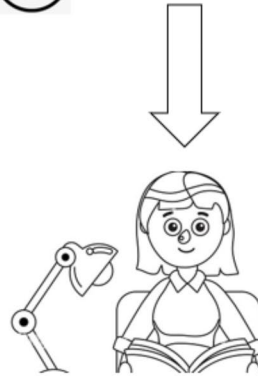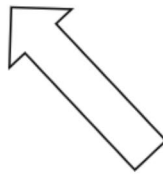
3. Reflective Practice. Teacher receives the feedback.

Fig. 1 Sentiment Analysis

3 Literature review

The analysis of sentiments has different approaches based on the data source, application, and the convenience of use. The NRC Emotion Lexicon (Mohammad & Turney, 2013) supports 40 languages and annotations for 8,116 unigram words for Hindi and 14,182 for English. The Lexicon uses negative or positive sentiments to associate with words in its database. A word match returns an emotion vector from the Lexicon. The sentiment parameters in the Lexicon include anger, anticipation, disgust, fear, joy, sadness, surprise, trust, negative and positive. The second approach uses machine learning algorithms (Ray, 2019) to train the dataset and build a classifier for future prediction of text. This approach requires manual labelling of the students' text before training begins. The review only looks at machine learning methods, such as classification algorithms for opinion mining or sentiment analysis.

Duwairi & Qarqaz (2014) built a classification model for sentiment analysis on Arabic text. After preprocessing, 2591 text out of 10,500 were used to train the model. The Naïve Bayes, SVM, and KNN classifiers under the 10-fold cross-validation technique were used to detect a given review's polarity by building a model. SVM performed with the highest accuracy of 75.25%.

Altrabsheh et al., (2014) labelled 1036 instances of data from the University of Portsmouth by verifying the reliability of the labels using Krippendorff's alpha and Fleiss kappa to generate a good confidence percentage. The SVM, Naïve Bayes, Complement Naïve Bayes, and Maximum Entropy classifiers were trained using the data. The results of training the classifiers show that SVM has the highest accuracy and is best suited for predicting students' sentiments.

Gottipati et al., (2018) compared statistical classifiers to rule-based methods to extract explicit qualitative suggestions from learners in the School of Information Systems, Singapore Management University. The decision tree (DT) from the classification results gave the highest accuracy and performance, with a 78.1% F-score. The study concluded that rule-based mining methods are time-consuming and that detecting responses that do not conform to rules is difficult.

In a related study, Gottipati et al., (2017) proposed a conceptual framework for students' feedback on selected courses. The study applied text analytics algorithms to analyse qualitative feedback from the students by implementing a prototype system and Student Feedback Mining Systems (SFMS). The sentiment analysis gave a precision of 80.1%, a recall of 86.4%, and an F-Score of 83.5%. The results showed that applying text analytics algorithms is useful for discovering knowledge patterns from qualitative feedback.

Guleria et al., (2015) developed a web-based feedback system and applied Association Rule Mining (ARM) using the Apriori Algorithm. They also proposed a web-enabled educational ARM tool for collecting feedback from students. The proposed ARM classification tool for association rules would assist instructors in identifying associations that may be used to predict which courses students want to study.

Another paper (Hashim et al., 2018) presented a model for association rule mining based on the Apriori algorithm and a feature selection algorithm called ReliefF to analyse students' responses to a questionnaire. The results show that the ReliefF algorithm is good at picking out features and finding relationships between them. It can be used with students' responses to help teachers find out how well they did in school.

Using the K-means clustering algorithm to analyse students' feedback on a university course, Abaidullah et al., (2014) developed a model for assessing and analysing students' feedback data. The results showed three (3) clusters, with some of the clusters having a percentage of one less or more than a hundred, which could be attributed to the sum of squared errors within clusters, 61.8243.

Dhanalakshmi et al., (2016) focused on using opinion mining tool RapidMiner to classify students' feedback from a module evaluation survey at Middle East College in Oman. The data was tested using various supervised learning algorithms such as Support Vector Machines, Naïve Bayes, K Nearest Neighbour, and Neural Networks. The results showed that Naïve Bayes was the best algorithm for accuracy and recall, whereas K-Nearest Neighbour was the best algorithm for precision.

Thi & Giang (2021) proposed a system for categorising students' feedback using opinion mining from students' feedback data collected from a University in Vietnam over a two (2) year period in 2017 and 2018. The data was organised into three (3) classes: positive, negative and neutral, and a sentiment dataset of 5000 classified sentences was built from the dataset. The researchers then applied Naïve Bayes, Maximum Entropy, and Support Vector Machines algorithms on the dataset. The results
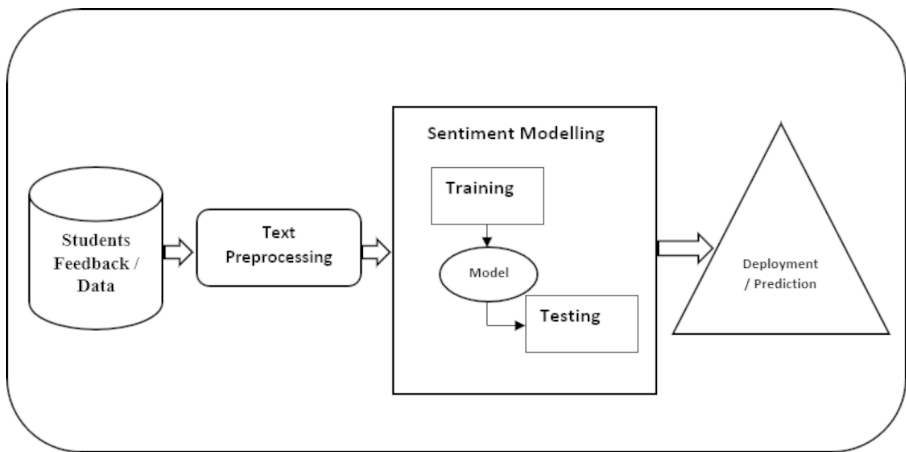
Fig. 2 Sentiment Mining Architecture

indicated that Maximum Entropy was the best classifier with an accuracy of 91.36%. Based on the results, a student's feedback system could be developed to help detect students' opinions.

4 Research methodology

The study modified the opinion mining architecture by Gottipati et al., (2018) as the process model for the evaluative feedback from the students. As shown in Fig. 2, the five-phase data science process involves students' qualitative data, text preprocessing, sentiment modelling, and deployment.

4.1 Students feedback data

We conducted this study to evaluate qualitative students' feedback in an introductory course taught by one of the researchers at the Department of ICT Education, University of Education, Winneba, Ghana. Approval was sort from the University of Education, Winneba Research Ethics Committee to collect data from the students. The authors can confirm that the study complied with all ethical regulations and students' consent were duly obtained. Google Forms were used to collect 280 instances of Level 100 students who offered Computer Systems and Applications during the 2020–2021 academic year. The respondents range from ages 18 to 26, with 85% males. At the time, the pandemic necessitated some course sessions being held online.

Research question 1 What categorisations of feedback comments were received after learners had filled out the questionnaire?

In response to Research Question 1, students were required to anonymously provide objective feedback and evaluate the lecturers' handling of the course to col-

	Qualitative Comment	Sentiment	Excellent	Good	Poor
Table 1 Sample Qualitative Feedback Comments from Students	*Everything was excellent to me and I wish it continue that way.Thank you*	Positive		]	
	An excellent mode of lesson delivery with adequate demonstrations and easy to understand notes	Positive	]		
	The course is nice and also understandable but the delivery mode for the online part was not effective for me thanks	Negative			]
	Very clear, understandable ,friendly ,not harsh on students and general am happy	Positive		]	
	Well I enjoy the course and enjoy about it. It's my favorite course so far	Positive	]		
	I like his way of delivery and I like the course very much	Positive		]	
	The delivery mode was poor because of poor gaget (microphone & speaker) used at the lab. Course content explanation was okay by the TA but was unfair to meet the main lecturer(Mr Dake) once by our group (6) while other groups experienced him enough	Negative			]

lect opinion data. As shown in Table 1, the opinions or sentiments of students are either positive or negative. Additionally, the sentiments have also been categorised as excellent, good, and poor. Excellent comments are ideal feedback because they demonstrate the learner's positive sentiments throughout the constructed sentence. The good comments are rated next, with less enthusiasm from the student. The poor comments imply that the learner is wholly dissatisfied with the lesson, and the learner is emphatic about the lesson's outcome.

Fig. 3 : Text Preprocessing
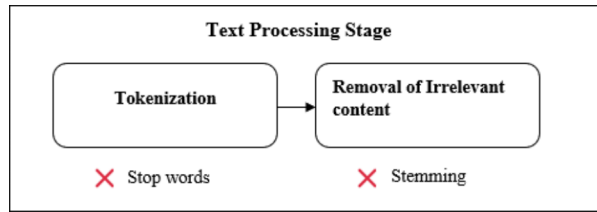
4.2 Text preprocessing

During the text preprocessing stage, missing values and unidentified terms were removed from the dataset. Of the 280 instances of text data, 48 occurrences contained many unidentified characters that had little or no relevance to the classification labels. For the model's construction, 232 text data points were employed, with training and test data separation using the k-fold cross-validation. The string to word vector tokenisation in Weka was used to break up the sentences into units called tokens. Tokenisation is relevant in unstructured data analytics (Wongkar & Angdresey, 2019) and forms the basis for natural language processing, especially with opinion mining and sentiment analysis. Stop words and stemming are not implemented in the pre-processing stage, as illustrated in Fig. 3. This allows for a complete and non-edited representation of the students' feedback.

4.3 Sentiment modelling

The proposed approach is based on sentiment analysis or classification. Four supervised learning algorithms, including Random Forest, J48 Decision Tree, Naive Bayes, and Support Vector Machine (SVM), were utilised to construct the predictive model in this study. After analysing the classification metrics, the dataset was divided using the k-fold cross-validation technique to get the training and test data automatically separated. As shown in Fig. 4, the best performing classification algorithm is selected after enough training and testing by varying the number of k in the k-fold cross-validation. The highest algorithm based on accuracy, precession, and recall is then used to forecast students' future sentiments or opinions.

Random Forest: The Random Forest algorithm (Shaik & Srinivasan, 2019) is a combination of decision tree algorithms that is learned by the bagging technique. In addition to adding randomness to the model with each new growth tree, the random forest also selects the best feature from a random subset of characteristics to be used.

J48 Decision Tree: By inferring simple decision rules from the dataset, the Decision Tree (DT) (Breşfelean, 2007) technique generates a training model. The decision tree divides a node based on information gain until it reaches the terminal node.

Naïve Bayes: A Naive Bayes (NB) (Berrar, 2018) classifier uses the Bayes theorem to calculate a posterior probability. Naive Bayes assumes that the occurrence of one feature is independent of the occurrence of other features and performs better with categorical inputs.

Support Vector Machine (SVM): The Support Vector Machine (SVM) algorithm (Noble, 2006) distinctly classifies data points by finding a hyperplane in an N-dimen-
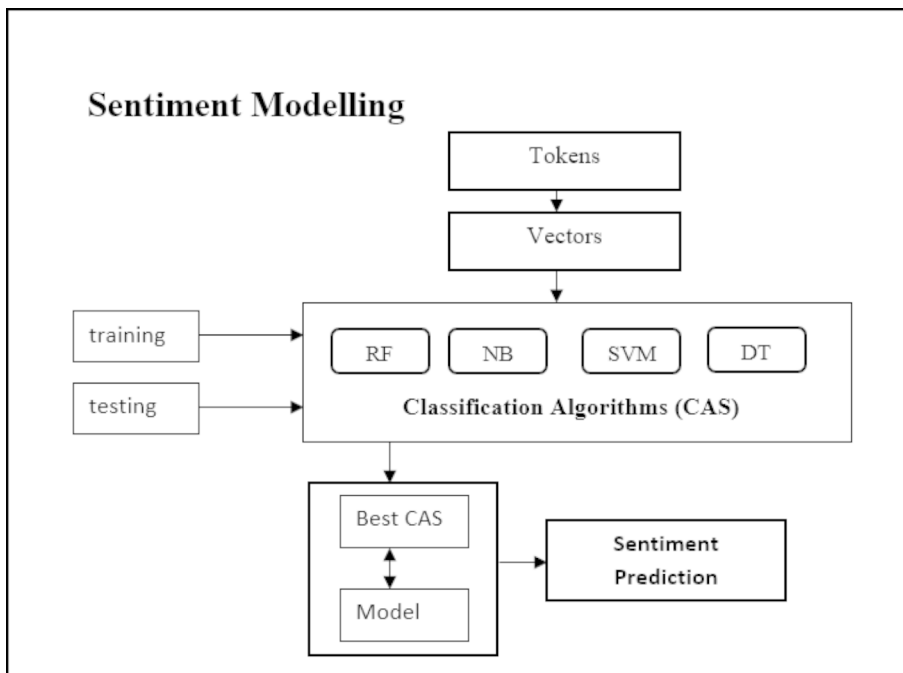
Fig. 4 : Sentiment Modelling

sional space, where N represents the number of features. As hyperplanes are the decision boundaries in the SVM algorithm, different classes are predicted based on the data points and the side they fall to the hyperplane.

5 Results and analysis

The student's feedback data is analysed with the Weka modelling tool (Hall et al., 2009) developed by the University of Waikato and implemented using the Java programming language. The four supervised learning algorithms, including the J48 Decision Tree, Support Vector Machine (SVM), Random Forest, and Naïve Bayes, were used to build the model and compared them. We used 10-fold cross-validation to develop the model and then ran another experiment on the training data using 5-fold cross-validation. The four classification algorithms are also analysed using the confusion matrix generated from the classification model.

5.1 String to word vector

The StringtoWordVector unsupervised learning filter is applied on the data to create a vector of words from the feedback text from students. As shown in Figs. 5 and 1073 words were generated from the 232 instances of string texts.
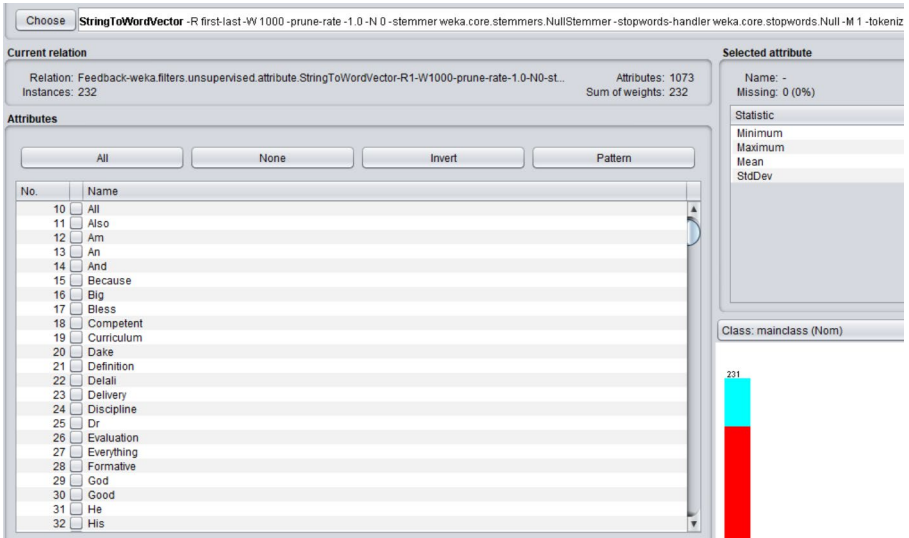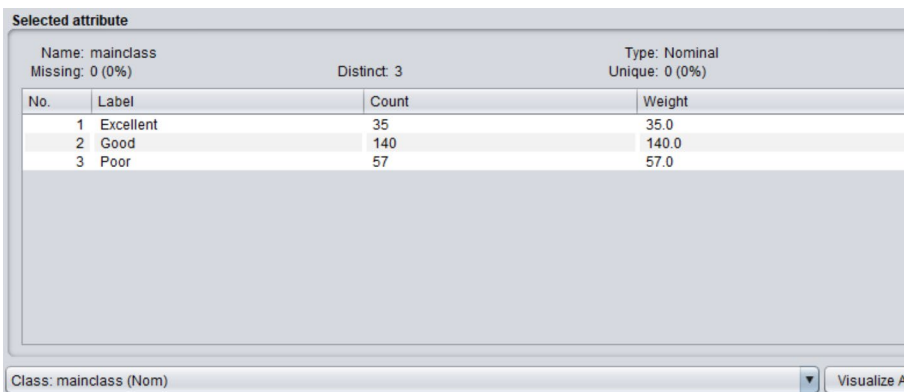
Fig. 5 StringToWord Vector Filter

Fig. 6 Class Attributes

As shown in Fig. 6, the class attributes consist of 35 counts of excellent, 140
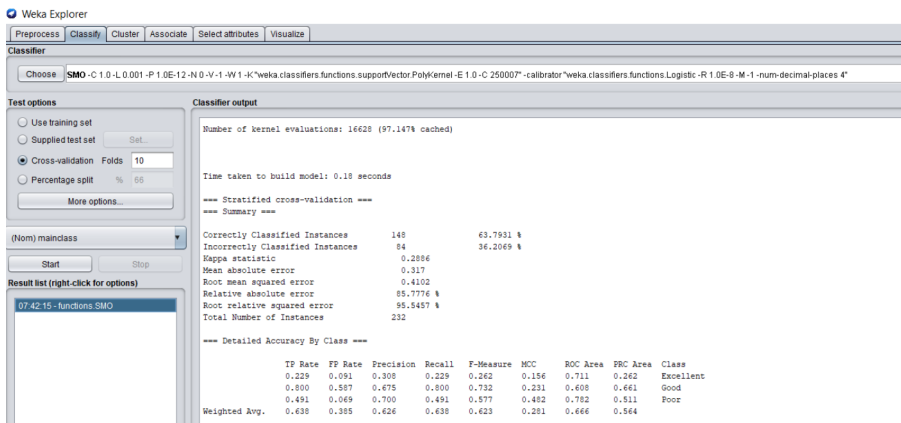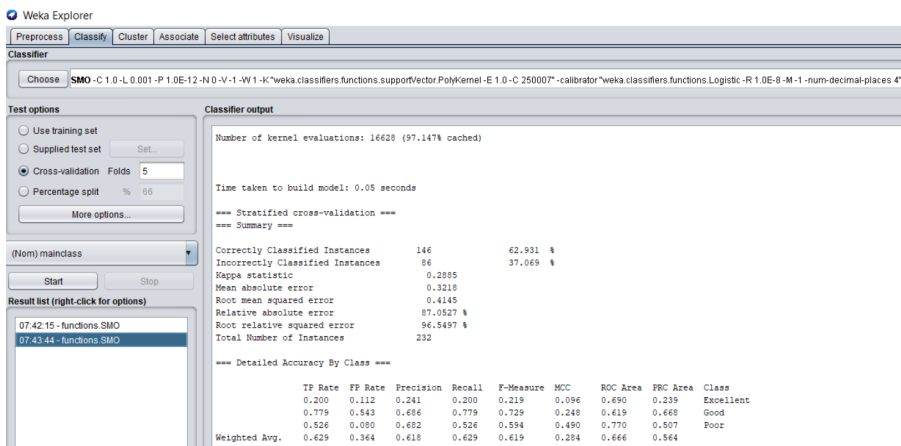
Fig. 7 10-Fold Cross Validation for SVM

Fig. 8 5-Fold Cross Validation for SVM

counts of good and 57 counts of poor. The count is from the instances of correctly labelled class attributes in the training data.

5.2 K-fold cross validation

The k-fold cross-validation approach is a common technique for estimating a machine learning algorithm's performance on a given dataset (kumar & Sahoo, 2012). The technique divides a dataset into folds and then tests one piece at a time while training is done on the remaining nine. The procedure generates results that are averaged. In Weka, the default standard is the 10-fold cross-validation that has achieved good results in literature. We compared the 10-fold standard cross validation to a 5-fold cross validation procedure to ascertain the performance difference in building the model based on the dataset. As shown in Figs. 7 and 8, the 10 and 5 cross-validations

Table 2 Accuracy of the Classifiers

K-Fold Cross Validation	SVM	J48 DT	NB	RF
10-Fold	63.79	57.76	57.76	63.79
5-Fold	62.93	61.21	58.19	62.01

Fig. 9 Classifier Accuracy
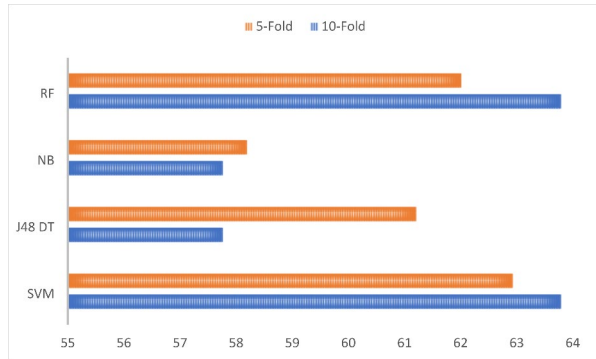
Table 3 10-Fold Cross Validation

Classifier	Precision	Recall	F-Measure	ROC
SVM	0.626	0.638	0.623	0.666
J48 DT	0.551	0.578	0.558	0.580
NB	0.581	0.578	0.578	0.696
RF	-	0.638	-	0.759

as an example were applied to the SVM dataset in building the model. From the results, the 10-fold outperforms the 5-fold for correctly classified instances.

5.3 Model accuracy

Research question 2 What machine learning algorithm provides the best accuracy after sentiment analysis?

All four supervised learning algorithms were subjected to both 10-fold and the 5-fold cross-validations. As shown in Table 2; Fig. 9, the accuracy of the classifiers based on the cross-validation is listed. As the results show, the performance of the classification model is impacted by the k-fold cross-validation process based on the number of k. In terms of accuracy, SVM performed better when compared with J48 DT, NB, and RF classification algorithms, with an accuracy of 63.79%. From Table 2, even though SVM and RF had the same accuracy, the confusion matrix results based on the precision and recall in Table 3 show that the RF algorithms predicted the class of most feedback text wrongly from students.

5.3.1 10-Fold cross validation

From the 10-Fold Cross-Validation in Table 3, SVM, with its classification metrics including accuracy, precision, and recall, gave the model a higher chance of accurately predicting each text's class label.

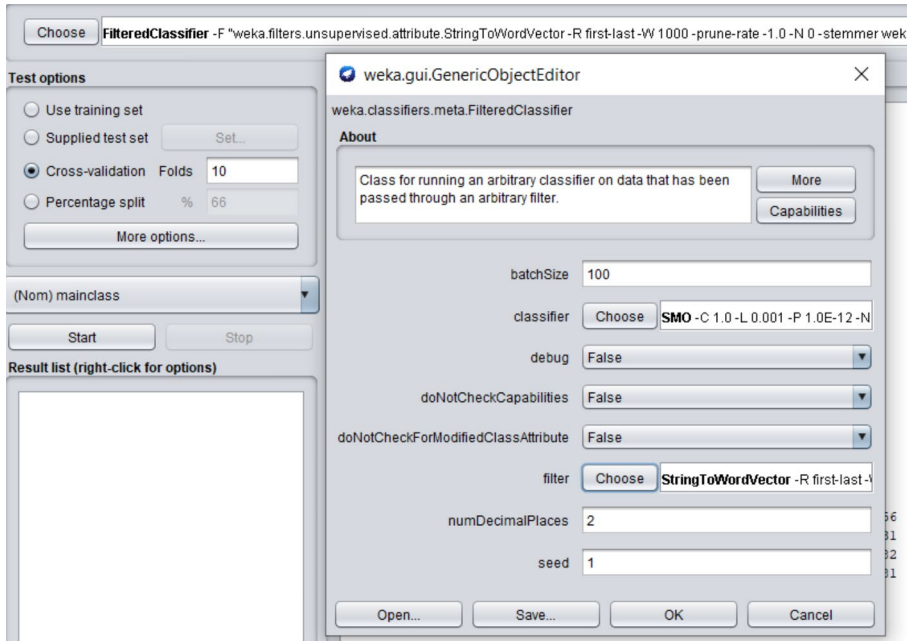| Table 4 5-Fold Cross Validation	Classifier	Precision	Recall	F-Measure	ROC
	SVM	0.618	0.629	0.619	0.666
	J48 DT	0.580	0.612	0.592	0.585
	NB	0.588	0.582	0.583	0.695
	RF	0.582	0.621	0.504	0.725

Fig. 10 Filtered Classifier in SVM

5.3.2 . 5-Fold cross validation

From the 5-Fold Cross-Validation in Table 4, the SVM still outperforms J48, NB and RF classification algorithms even though its performance accuracy is reduced to 62.93%. J48, NB, and RF performed significantly better with the 5-Fold cross-validation than with the 10-Fold cross-validation.

5.4 Filtered classifier with SVM

Research question 3 What is the prediction performance of the best machine learning algorithm on new data?

The classification metrics from the support vector machine significantly show a good trajectory for building the prediction model. In deployment, the Filtered Classifier (V, 2014), which allows for the execution of an arbitrary classifier on filtered data, is applied. The Filtered Classifier is relevant in matching the number of attributes in the training set data to the test data. This is done after the StringToWordVector filter is applied on the dataset, as shown in Fig. 10.
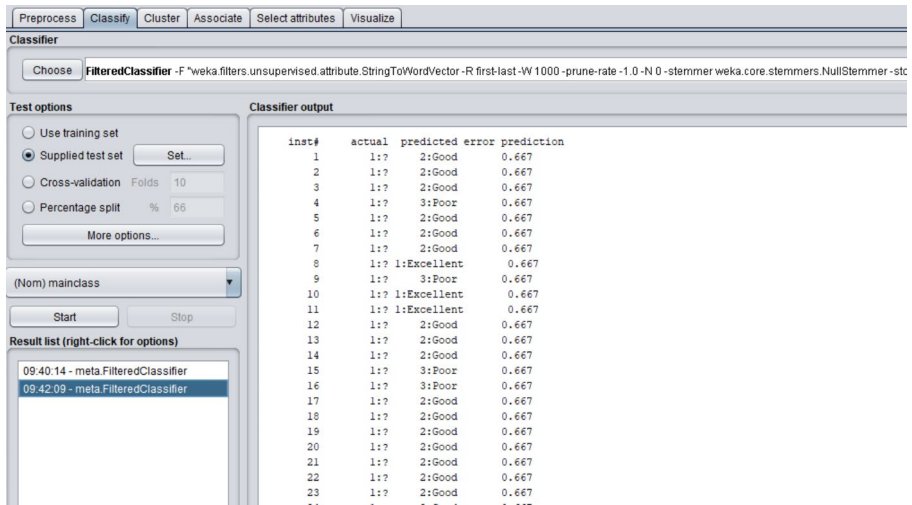
Fig. 11 Predicted Results using SVM

Figure 11 shows the predicted results of the new 31 data instances supplied to the model in Weka. As shown in Table 5, the class labelling of test data was primarily accurate. The distinction between Excellent and Good, to a large extent, was problematic to the classifier.

6 Discussion of results

As shown in Table 5, the test predictions using the SVM classification algorithm show 92% accurate detections from the 31 test data supplied. Significantly, most of the student's comments were accurately predicted by the SVM model to the suitable class. The SVM model clearly distinguishes poor from good or excellent comments when tested with sample data. The difficulty arises when differentiating between good and excellent data. Labelling students' feedback in training the model affects the prediction accuracies in test data. In supervised learning, correctly labelling training data before implementing an algorithm is relevant in accurately predicting test samples. In labelling the training data for the study, excellent comments are ideal and highly positive feedback from the learner who wholly understood the lesson. Good comments though positive, are not outstanding. The SVM model in differentiating a comment that is excellent or good relied on the training data. A class label of poor or good comments only will have resulted in a higher predictive percentage for training data and test data model accuracies.

Comparative analysis was conducted between the findings of this study and those of similar studies evaluating the qualitative responses of learners using machine learning. Thi & Giang (2021) used opinion mining to categorise learner feedback into positive, negative and neutral sentiments. With this model, the maximum entropy classifier compared to the SVM was the best, with an accuracy of 91.36%. Duwairi & Qarqaz (2014) proposed a sentiment analysis model based on students' feedback

Table 5 Prediction Results using Support Vector Machine (SVM)

Test Feedback No.	Qualitative Comment	Test Prediction
1	The lecture is good in class & his slides are well explained. His good, you can read his slides & understand	Good
2	Moderate	Good
3	Infact among other courses, ICTE 111 is the best so far. Kudos to Mr Delali and his able TA	Good
4	Normal	Poor
5	The lecturer's mode of delivery was perfect. I give it a five star. thanks	Good
6	It was good	Good
7	Delivery mode was very good, l for one enjoyed the lectures class, explanations were vividly done. ICTE111 gives an overview of ICT and some branches of ICT.	Good
8	The lecturer used excellent delivery method in presenting content to the learners. Course content explanation was absolutely perfect. How the notes of the course was presented was marvelous and reading it was very understandable. ICTE 111 course has helped me to know a lot about Technology.	Excellent
9	Course content was intact and exact but the explanation of some areas of the content was not all that clear especially the database and the operating system softwares. But all the same as we where moving on we got to understand it. The delivery of the content was on point and explanation of the lecture also clear and in all l love the lecturer.	Poor
10	The delivery mode ,course content and everything about this course is fine just that we were suppose to go a little deeper into the microsoft office suit. But i will say this is one of the best courses. Thank you	Excellent
11	I think the delivery mode and content so far is been good as slides were self explanatory. And my understanding of the course is that,it is a course that will equip students with the knowledge or understanding of how the computer system works or what goes on in the computer.	Excellent
12	It was very interesting	Good
13	The delivery mode is clear and easy to understand, and the course had given me a broad knowledge in ICT	Good
14	The delivery mode was superb, course content explanation was understandable	Good
15	The mode of delivery of the lecturer is good but the problem i have is that,the mode of delivery needs a little improvement. The issue is that, most of us are slow learners and may need a little detailed explanation for a better understanding. I sometimes lag behind due to the speedy delivery mode.this makes the course quite difficult for me	Poor
16	The delivery was nice. I now understand that, the course is the stepping stone for anyone that wishes to be in the technological world	Poor
17	Truly it was great and everything was normal to my understanding	Good
18	I will say delivery mode was good, general understanding on my part wasn't all that bad but I was compelled to go an extra extra mile to understand the ERD because I wasn't able to get the concept very well in class. But in all, you've done a really great job. Kudos	Good
19	Everything throughout the semester was good and fantastic As the lectures made us understand every concept being taught	Good
20	I like his delivery and explanation of subject matters and the course content is simple and straight forward	Good

Table 5 (continued)

Test Feedback No.	Qualitative Comment	Test Prediction
21	*First of all, I want to say thank you very much for the time you have spent on us. I really appreciate it. If anyone should ask me to select my favorite course for the semester, I will choose ICTE111 even though the examination wasn't all that good for me.Please hope you will slow down a little bit next semester. Thank you*	*Good*
22	*Icte 111 is about the introduction of some core Ict courses like networking, DBMS, Etc. Mode of delivery was fine and the explanation from Dr Dakeh was very Good… Thanks to Him and The Ict Department*	*Good*
23	*Was quiet challenging Concerning first online quiz*	*Good*
24	*With regards to the course contents, it was very helpful. It's had removed the computer phobia from some students who think learning of ICT is difficult.Teacher and students interaction was awesome.It's good to learn such course to which has given us hope to the next level.The lesson delivery mode was perfect and understandable*	*Good*
25	*The lesson delivering mode was detailed, engaging. Concrete evidence of learning resources to draw the learner close to the contents, were helpful*	*Good*
26	*Perfect explanations. I always enjoy the class*	*Good*
27	*Well delivered and It was quite understandable. It was to know more about computers and how to operate with application software*	*Good*
28	*It's awesome*	*Good*
29	*Everything was cool. It was kk and I hope it continues like this till our final year*	*Good*
30	*Per my own assertion, I enjoy the delivery mode, the course content is very good but the little problem is my understanding of the course. Sometimes I don't get it well when some of the contents are explained. Thank you*	*Poor*
31	*The teaching is ok some how. Sometimes, u are harsh and not friendly to the learners which most of us found it difficult to cooperate with others. So far, so good. I edge u to be a little bit friendly when lecturing. Bravo*	*Poor*

in Arabic text. They found that SVM outperforms the Naïve Bayes classifier with an accuracy of 72.25%. Altrabsheh et al., (2014) compared the Complement Naïve Bayes (CNB) and the SVM classifiers in building a model for real-time interventions in the classroom based on learner feedback. In their results, the SVM outperforms CNB with an accuracy of 94%. Lalata et al., (2019) compared the SVM and the NB algorithms in building an ensemble machine learning model to evaluate feedback comments received from learners in faculty evaluation. For the 5-Fold cross-validation technique, the SVM outperforms the NB with an accuracy of 90.17%. Umair & Hakim (2021) developed a sentiment analyser model for students' feedback before and after the COVID-19 pandemic. The SVM algorithm utilised performed better than the NB with an accuracy of 85.62%.

The SVM machine learning algorithm has shown promising results compared to other classification algorithms for sentiment analysis in evaluating students' feedback. Even though some literature (Lalata et al., 2019) and (Umair & Hakim, 2021) utilised feature selection mechanisms for sentiment analysis, it resulted in low classification accuracy.

7 Conclusions and Future Works

This research work considered a qualitative feedback dataset from students at the University of Education, Winneba, for sentiment analyses. After preprocessing, the dataset was subjected to four supervised machine learning algorithms in building the model for prediction and deployment. After implementing the k-fold cross-validation for k=10 and k=5, results show that the Support Vector Machine (SVM) has the highest accuracy of 63.79% with k=10. For complete modelling of the student's feedback and to prevent misconstrued representation, the model avoided the removal of stop words and stemming during training and testing. Practical prediction of 31 text instances shows a high percentage accuracy of 92% using the trained model.

The research forms a blueprint for implementing sentiment analysis via machine learning for auto-generation analytics for learning management systems (LMS), especially during forum discussions and qualitative student feedback reviews. Such implementations will improve teaching and learning and form the bases for SMART campus architectures.

Competing interests　The authors declares they have no competing interest.

References

Abaidullah, A. M., Ahmed, N., & Ali, E. (2014). Identifying Hidden Patterns in Students' Feedback through Cluster Analysis. *International Journal of Computer Theory and Engineering*, *7*(1), 16–20. https://doi.org/10.7763/ijcte.2015.v7.923

Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014). Learning sentiment from students' feedback for real-time interventions in classrooms. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8779 LNAI*, 40–49. https://doi.org/10.1007/978-3-319-11298-5_5

Altrabsheh, N., Gaber, M. M., & Cocea, M. (2013). SA-E: Sentiment analysis for education. *Frontiers in Artificial Intelligence and Applications*, *255*, 353–362. https://doi.org/10.3233/978-1-61499-264-6-353

Benade, L. (2015). Teachers' Critical Reflective Practice in the Context of Twenty-first Century Learning. *Open Review of Educational Research*, *2*(1), 42–54. https://doi.org/10.1080/23265507.2014.998159

Berrar, D. (2018). Bayes' theorem and naive bayes classifier. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, *1–3*(2018), 403–412. https://doi.org/10.1016/B978-0-12-809633-8.20473-1

Breşfelean, V. P. (2007). Analysis and predictions on students' behavior using decision trees in weka environment. *Proceedings of the International Conference on Information Technology Interfaces, ITI*, 51–56. https://doi.org/10.1109/ITI.2007.4283743

Dhanalakshmi, V., Bino, D., & Saravanan, A. M. (2016). Opinion Mining from Student Feedback Data Using Supervised Learning Algorithms. *2016 3rd MEC International Conference on Big Data and Smart City*, 183–206. https://doi.org/10.1109/ICBDSC.2016.7460390

Duwairi, R. M., & Qarqaz, I. (2014). Arabic sentiment analysis using supervised classification. *Proceedings – 2014 International Conference on Future Internet of Things and Cloud, FiCloud 2014*, 579–583. https://doi.org/10.1109/FiCloud.2014.100

Freidhoff, J. R. (2008). Reflecting on Affordances and Constraints and Their Impact on Pedagogical Practices. *Journal of Computing in Teacher Education*, 24(4), 117–122

Gottipati, S., Shankararaman, V., & Gan, S. (2017). A conceptual framework for analysing students' feedback. *Proceedings - Frontiers in Education Conference, FIE*, 2017-Octob, 1–8. https://doi.org/10.1109/FIE.2017.8190703

Gottipati, S., Shankararaman, V., & Lin, J. R. (2018). Text analytics approach to extract course improvement suggestions from students' feedback. *Research and Practice in Technology Enhanced Learning*, 13(1), https://doi.org/10.1186/s41039-018-0073-0

Guleria, P., Sharma, A., & Sood, M. (2015). Analysis and Association Rule Mining. *International Journal of Data Mining & Knowledge Management Process*, 5(6), 35–44. https://doi.org/10.5121/ijdkp.2015.5603. Web-Based Data Mining Tools: Performing Feedback

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11(1), 10–18

Hashim, S. A., Hamoud, K. A., & Awadh, A. W. (2018). ANALYSING STUDENTS' ANSWERS USING ASSOCIATION RULE MINING BASED ON FEATURE SELECTION.JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY, 53(5)

kumar, Y., & Sahoo, G. (2012). Analysis of Parametric & Non Parametric Classifiers for Classification Technique using WEKA. *International Journal of Information Technology and Computer Science*, 4(7), 43–49. https://doi.org/10.5815/ijitcs.2012.07.06

Lalata, J. A. P., Gerardo, B., & Medina, R. (2019). A sentiment analysis model for faculty comment evaluation using ensemble machine learning algorithms. *ACM International Conference Proceeding Series*, 68–73. https://doi.org/10.1145/3341620.3341638

Leung, A., Fine, P., Blizard, R., Tonni, I., & Louca, C. (2021). Teacher feedback and student learning: A quantitative study. *European Journal of Dental Education*, 25(3), 600–606. https://doi.org/10.1111/eje.12637

Mathew Priya, M. Prasanth, & Peechattu Princ, J. (2017). Reflective Practices: A Means To Teacher Development. *Asia Pacific Journal of Contemporary Education and Communication Technology*, 3(1), 126–131. www.apiar.org.au

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436–465. https://doi.org/10.1111/j.1467-8640.2012.00460.x

Mohapatra, D., Tripathy, J., Mohanty, K. K., & Nayak, D. S. K. (2021). Interpretation of Optimised Hyper Parameters in Associative Rule Learning using Eclat and Apriori. *Proceedings – 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, *Iccmc*, 879–882. https://doi.org/10.1109/ICCMC51019.2021.9418049

Umair, M., & Hakim, A. (2021). *Sentiment Analysis of Students' Feedback before and after COVID-19 Pandemic Sentiment analysis of Students Feedback before and after COVID-19 Pandemic View project*. 12(July), 177–182. www.researchtrend.net

Nouri, J., Saqr, M., & Fors, U. (2019). Predicting performance of students in a flipped classroom using machine learning: Towards automated data-driven formative feedback. ICSIT 2019 - 10th International Conference on Society and Information Technologies, Proceedings, 17(2), 79–82

Ara, Tekian Christopher J., Watling Trudie E., Roberts Yvonne, Steinert John, Norcini (2017) Qualitative and quantitative feedback in the context of competency-based education. Medical Teacher 39(12) 1245-1249 7 10.1080/0142159X.2017.1372564

Schnall, A. H., Wolkin, A., & Nakata, N. (2018). Methods: Questionnaire Development and Interviewing Techniques. In Disaster Epidemiology: Methods and Applications. Elsevier Inc. https://doi.org/10.1016/B978-0-12-809318-4.00013-7

Song, K., & Lee, K. (2017). Predictability-based collective class association rule mining. Expert Systems with Applications, 79, 1–7. https://doi.org/10.1016/j.eswa.2017.02.024

Taylor, C. (2021, May 21). Structured vs. Unstructured Data. Datamation. https://www.datamation.com/big-data/structured-vs-unstructured-data/

Structured vs Unstructured Data: Compared and Explained. (2020, December 14). Altexsoft. https://www.altexsoft.com/blog/structured-unstructured-data/

Ray, S. (2019). A Quick Review of Machine Learning Algorithms. Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019, 35–39. https://doi.org/10.1109/COMITCon.2019.8862451

Thi, N., & Giang, P. (2021). Sentiment Analysis for University Students ' Feedback SENTIMENT ANALYSIS FOR UNIVERSITY STUDENTS ' FEEDBACK. February 2020. https://doi.org/10.1007/978-3-030-39442-4

Wongkar, M., & Angdresey, A. (2019). Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019, 1–5. https://doi.org/10.1109/ICIC47613.2019.8985884

Shaik, A. B., & Srinivasan, S. (2019). A brief survey on random forest ensembles in classification model. In Lecture Notes in Networks and Systems (Vol. 56). Springer Singapore. https://doi.org/10.1007/978-981-13-2354-6_27

Noble, W. S. (2006). What is a support vector machine? Nature Biotechnology, 24(12), 1565–1567. https://doi.org/10.1038/nbt1206-1565