

Supervised Sentiment Analysis of Indirect Qualitative Student Feedback for Unbiased Opinion Mining [†]

Smitha Bidadi Anjan Prasad ^{*}  and Raja Praveen Kumar Nakka 

Department of Computer Science and Engineering, Faculty of Engineering and Technology,
JAIN (Deemed-to-be University), Bengaluru 562112, India; rajapraveen.k.n@gmail.com

^{*} Correspondence: smitha.prasaad@gmail.com

[†] Presented at the International Conference on Recent Advances on Science and Engineering,
Dubai, United Arab Emirates, 4–5 October 2023.

Abstract: In the education domain, the significance of student feedback and other stakeholders for raising educational standards has received more attention in recent years. As a result, numerous instruments and strategies for obtaining student input and assessing faculty performance, as well as other facets of education, have been developed. There are two main methods to collect feedback from students, as follows: the direct and indirect methods. In the direct method, feedback is collected by distributing a questionnaire and taking their responses. The limitation of this method is that the true experience of students is not revealed, and there is room for bias in the collection and assessment of such a questionnaire. To overcome this limitation, the indirect method can be followed where social media posts can be used to collect feedback from students as they are active on social media and use it to express their opinions as posts. To address the problem of the manual annotation of large volumes of data, this paper proposes a machine learning method that uses the sentiment 140 dataset as the training set to automate the process of annotations of tweets. The same method can be used to label any qualitative data. In total, 5000 tweets were scraped and considered for this study. Various pre-processing methods, including byte-order-mark removal, hashtag removal, stop word removal, and tokenization, were applied to the data. The term frequency-inverse document frequency (TF-IDF) trigrams technique was then used to process the cleaned data. The TF-IDF technique using trigrams captures negation for sentiment analysis. The vectorized data are then processed using various machine learning algorithms to classify the polarity of tweets. Performance parameters such as the F1-score, recall, accuracy, and precision are compared. With a 94.16% F1-score, 94% precision, 94% recall, and 95.16% accuracy, the Ridge Classifier performed better than the others.

Keywords: feedback; education; direct method; indirect method; sentiment analysis



Citation: Prasad, S.B.A.; Nakka, R.P.K. Supervised Sentiment Analysis of Indirect Qualitative Student Feedback for Unbiased Opinion Mining. *Eng. Proc.* **2023**, *59*, 15. <https://doi.org/10.3390/engproc2023059015>

Academic Editors: Nithesh Naik,
Rajiv Selvam, Pavan Hiremath,
Suhas Kowshik CS and Ritesh
Ramakrishna Bhat

Published: 11 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Depending on their results, mining technologies concentrate primarily on specific subjects. The evolution of education systems has witnessed a transformative shift from traditional teacher-centric models to progressive student-centric approaches. This transition acknowledges that learners possess diverse needs, learning styles, and interests, underscoring the importance of tailoring education to individual students. Stakeholder-centered education has been made mandatory for all higher education institutions (HEI) via accreditation organizations like the National Board of Accreditation (NBA) and the National Assessment and Accreditation Council (NAAC) in the context of standardized education. In such a system, an objective feedback collection and analysis system becomes the lynchpin for the success of education. For education to be of the highest quality and standards, student feedback is essential. There are several techniques for gathering feedback, which can be broadly categorized into direct and indirect methods. Direct techniques include both offline and online strategies, including printed materials, online surveys, customized

software, and Google Forms. Direct approaches come with inherent drawbacks, notably including the tendency to prioritize quantitative over qualitative data for streamlined assessments, potentially leading to incomplete datasets. Human engagement in crafting and evaluating questionnaires further compounds the issue, consuming both time and resources. This involvement also introduces the peril of inherent bias, which can skew results. Addressing these limitations necessitates a re-evaluation of the balance between quantitative and qualitative data, streamlining human interactions, and implementing safeguards to mitigate bias. According to Chen Xin et al. [1], students frequently discuss their experiences and look for social support on social media platforms. The most commonly used social media platforms are Facebook and Twitter; indirect approaches can make use of these channels. Given the widespread use of social media among students, these platforms serve as a natural outlet for students to express their opinions and emotions, offering a rich source of unfiltered data. Leveraging sentiment analysis on these platforms allows for the efficient extraction of insights, providing a comprehensive understanding of their experiences and perspectives. This approach capitalizes on the convenience of accessing a large volume of data while minimizing direct intervention, making it a suitable choice for capturing authentic and diverse feedback.

2. Related Works

This section looks at several studies that apply several machine learning methods, such as Naive Bayes, Support Vector Machines (SVM), Random Forest, and lexicon-based approaches, to student comments through sentiment analysis and mining. The emotional content in social media posts is evaluated using sentiment analysis. This process employs sentiment extraction techniques to derive insights from the text. Initially, the text is pre-processed, removing noise and tokenizing it. Next, methods like lexicon-based analysis or machine learning models classify the sentiment as positive, negative, or neutral. These techniques discern the context and account for nuances, yielding valuable emotional insights from vast textual data. Aung & Myo [2] used a lexicon-based approach for sentiment analysis, and they found that the Afinn dictionary has limitations. According to [3] 2021, the most commonly used techniques for sentiment analysis in the last decade, i.e., from 2014, are Naïve Bayes, SVM, Decision Tree, and Logistic Regression when using supervised learning, and Vader for lexicon-based sentiment analysis. Ref. [4] suggests utilizing OMFeedback, a specially designed software system, to gather and analyze student input using a lexicon-based method and the Vader Sentiment Intensity Analyzer. Ref. [5] also uses the Vader Sentiment Analyzer for annotations. Ref. [6] used SentiWordNet for the annotation of opinions. Ref. [7] proposed a Bing lexicon CSL to perform sentiment analysis. Refs. [8,9] used a lexicon dictionary for sentiment analysis, using terms such as joy, anger, fear, sadness, and disgust. The lexicon-based approaches can label the dataset with the polarity of opinions, but labeling a large amount of data in the dataset can be challenging. Ref. [10] used LMS to collect student feedback and used six classifiers of machine learning methods to annotate the sentiment polarity. Multinomial Logistic Regression, the Gaussian Naive Bayes classifier, Multilayer Perceptron, K nearest neighbors, Decision Tree, and Support Vector Machines are the machine learning techniques that were employed. The results revealed that the Logistic Regression performed better than the others. Ref. [11] performed clustering using K-means to cluster and then classified using supervised machine learning algorithms like Logistic Regression, Random Forest, and Support Vector Machine. Ref. [12] used Decision Tree, Naïve Bayes, and SVM (Support Vector Machine) to annotate the opinions on student feedback data and examine student social media posts. Ref. [13] used a Support Vector Machine and Naïve Bayes on 5000 review data. They performed document-level sentiment analysis with an accuracy of 72.80% using a Support Vector Machine and 81% using Naïve Bayes. Ref. [14] used a multi-class classification model to analyze and annotate student speech. Ref. [15] proposed an aspect-based model for analyzing student feedback with the highest accuracy of 80.67%. Ref. [16] used Random Forest, Support Vector Machine, and Decision Tree, and deep learning

methods to annotate student feedback data. In [17], Random Forest was used, which showed a better performance for labeling and finding fine the grain sentiment classification for sad, anger, happiness, surprise, and disgust. Ref. [18] used the CNN learning model for annotating MOOC-related data with 82.10% of the F-measure. Ref. [19] performed a manual annotation for reviews provided by 181 students. They annotated using the CNN model. Ref. [20] created a multi-head fusion model for sentiment analysis, utilizing LSTM for learning with Glove and Cove embedding. For sentiment analysis, BERTs, or bidirectional encoder representation transformers, were employed. BERT's accuracy using the CNN model was 92.8%. Researchers have examined the application of sentiment analysis in understanding the attitudes and emotions that students share on social media. Educational institutions may improve the quality of education they offer by analyzing these attitudes to learn more about students' viewpoints, spot areas that need work, and make data-driven choices. The paper emphasizes how social media analysis and algorithm choice might enhance the educational process. In order to improve the precision and granularity of sentiment analysis models in the educational context, this study also identifies future research goals.

3. System Model

The methodology used for the system model is depicted in Figure 1. The methodology of the system encompasses the following steps:

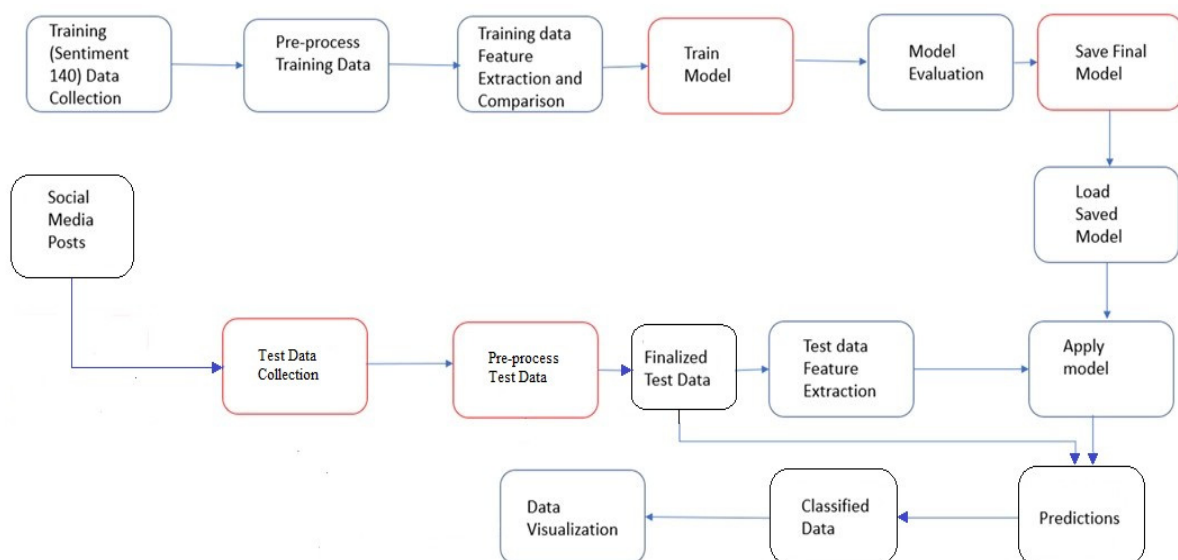


Figure 1. Methodology.

3.1. Training Data Collection (Sentiment 140 Data Collection)

The utilization of the Sentiment 140 dataset proves pertinent for training and annotating models in the realm of student feedback sentiment analysis. This dataset, containing a diverse array of tweets labeled with sentiments, can be harnessed to enhance the accuracy and efficacy of sentiment analysis models tailored to student feedback. The first phase entails gathering the training data; in this instance, 1,600,000 rows of the sentiment140 dataset were used. Tweets were selected based on different topics, hashtags, and user demographics. Additionally, efforts were made to encompass a balanced mix of positive, negative, and neutral sentiments. The sentiment analysis model was trained using this dataset as its basis.

3.2. Pre-Process Training Data

In order to clean and prepare the training data for analysis, preprocessing techniques were applied to it. To standardize the text data, the following operations were

performed: erasing URLs, the removal of null rows, tokenization, and lowercase conversion, along with the removal of hashtags, removal of the @ symbol, removal of the URL, and stop-word removal.

3.3. Extraction and Comparison of Training Data Features

The preprocessed training data were subjected to feature extraction algorithms. Different techniques were used to extract pertinent characteristics from the text, including count vectorization and TF-IDF vectorization. To choose the best strategy, the accuracy of each feature extraction method was compared.

3.4. Train Model

The sentiment analysis model was trained using the training data after feature extraction. To create predictions regarding sentiment, the model discovered patterns and correlations within the data.

3.5. Model Evaluation

The trained model progresses through an iterative process of evaluation to enhance its performance. This can involve fine-tuning the model parameters or adjusting the training process to achieve better accuracy.

3.6. Saving the Final Model

The model is stored for further use if it has performed satisfactorily. This makes deployment and reuse simple.

3.7. Test Data Collection

Several preprocessing procedures were applied to make the dataset standardized for analysis, such as souping to remove HTML markup or tags from the text, the byte-order-mark (BOM), URL address, number, special character, and Twitter ID removal. Converting to lower-case, dropping duplicates, tokenizing, and joining were also conducted.

3.8. Load-Saved Model

Whenever sentiment analysis is necessary, the system can be loaded using the saved pre-trained model.

3.9. Applying the Model

The preprocessed test data are subjected to the same feature extraction technique as the training data. Sentiment predictions are generated using the training model and the test data's extracted attributes. Each data point in the test dataset is given a sentiment label by the model (such as positive or negative). Based on the assigned sentiment labels, the test dataset and the sentiment predictions are categorized.

4. Results and Discussion

4.1. Feature Extraction

Various approaches to extracting characteristics from the training data were compared. The selection of TF-IDF trigrams as the feature extraction method for the analysis of sentiments is grounded in its ability to comprehensively capture contextual information. Unlike individual words or bigrams, TF-IDF trigrams can effectively capture negation cues, such as "not so good" or "did not like", by considering the presence of negating terms alongside sentiment-bearing words. Figure 2 shows the accuracy comparison graph for the count vectorizer and TF-IDF vectorizer. Unigrams, bigrams, and trigrams of the TF-IDF vectorizer and the count vectorizer were both used. According to the accuracy of the findings, TF-IDF trigrams outperformed all other methods on the training set. The TF-IDF unigram (between 60,000 and 100,000 features with 79.84% validation accuracy),

TF-IDF bigram (100,000 features with 82.04% validation accuracy), and TF-IDF trigram (90,000 features with 82.22% validation accuracy) were all taken into consideration.

Out[46]: <matplotlib.legend.Legend at 0x1ddf5a98c10>

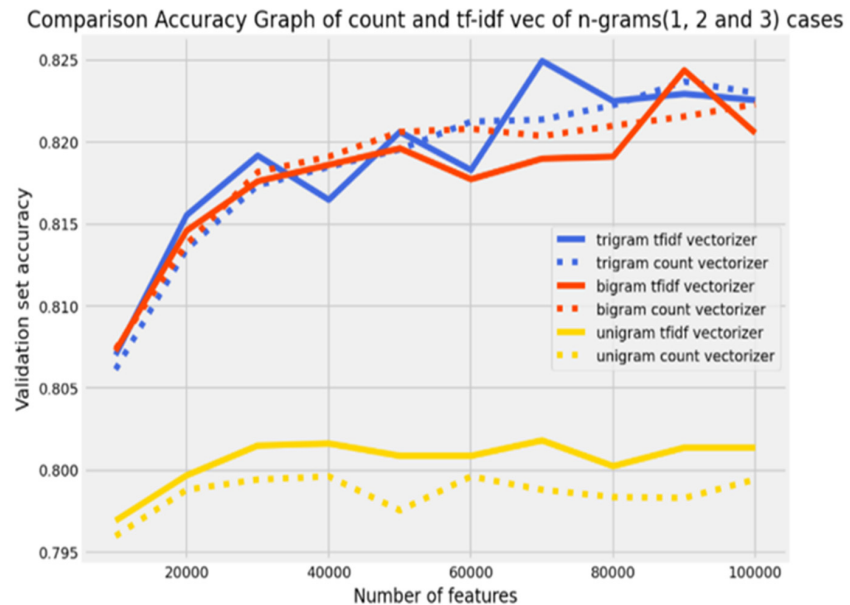


Figure 2. Accuracy comparison graph of TFIDF and count vectorizer Uni-, Bi- and Trigram cases.

4.2. Modeling and Comparing Various Classification Model Results

Various classification models were employed to predict attitudes in the test dataset. The supervised learning approach was used to build these models. Among the classification models used in this study were Linear SVC and Linear SVC with L1-based feature selection, Logistic Regression, the Vader emotion analyzer, AdaBoost, Perceptron, Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Ridge Classifier, Passive-Aggressive, Nearest Centroid, and AdaBoost. Each classifier was trained using the labeled training data before predicting data from the test dataset. To ascertain how well each classifier performed in accurately predicting these attitudes, its accuracy was assessed.

4.3. Accuracy Comparison on Tweet Dataset

The accuracy of the comparison of different approaches is displayed in Table 1. In the context of sentiment analysis, classification model accuracy is an important evaluation metric that gauges how well the model predicts the sentiment of text data.

Table 1. Accuracy comparison.

Classifiers Comparison					
Sl. No.	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	Ridge Classifier	95.16	94	94	94
2	Linear SVC	94.73	94	95	95
3	Linear SVC with L1-based feature selection	94.62	93	94	93
4	Logistic Regression	93.75	90	92	95
5	Passive-Aggressive	91.03	81	88	84
6	Multinomial Naïve Bayes	90.62	81	89	84
7	Bernoulli Naïve Bayes	89.43	79	89	90
8	AdaBoost	84.56	70	73	72
9	Nearest Centroid	83.02	71	80	73
10	Perceptron	76.15	67	79	68
11	Lexicon Analyzer				
	Vader sentiment analyzer	84.83	-	-	-

Linear SVC with L1-based feature selection aids in dimensionality reduction and enhancing interpretability. Bernoulli Naïve Bayes and Multinomial Naïve Bayes handle text data effectively. The Ridge Classifier stands out as the focal point of performance assessment due to its ability to mitigate multicollinearity issues and maintain model stability, thus yielding reliable results in sentiment analysis tasks. The classification models' accuracy score for this project is 98.16 from the Ridge Classifier, which is the highest compared to other classifiers. This model performed remarkably well in predicting the sentiment on the test dataset, as seen by the high accuracy score.

The classification report computes metrics like precision, recall, F1-score, and support for each sentiment class (positive and negative) to offer additional insights into the effectiveness of the classification models. Let us explain these metrics in detail:

- **Precision:** The precision of a test is determined by dividing its true positives by the total of its true positives and false positives. Few incorrect positive predictions are indicated by a high precision score.
- **Recall:** The ratio of true positives to the total of true positives and false negatives is called recall, which is sometimes referred to as sensitivity or the true positive rate. Recall scores that are high suggest fewer incorrect negative predictions.
- **F1-score:** The harmonic mean of recall and precision is known as the F1-score. It takes into account both precision and recall, providing an equitable assessment of the model's performance. When there is an uneven distribution of classes, the F1-score is helpful.
- **Support:** The number of occurrences in every sentiment class is represented by a support. It shows how many occurrences of each sentiment the model has predicted.

These aid in evaluating the model's efficacy for every sentiment class independently and offer a more comprehensive grasp of the model's benefits and drawbacks.

4.4. Data Visualization

The performance of the various sentiment analysis models on the dataset of tweets is gauged by Figure 3, which shows the comparison of various classifiers with their accuracy, precision, recall, and F1-score. Figure 4 shows the accuracy comparison graph of the test data for tweets. It offers insightful information on how well the algorithms anticipate sentiment from real-time social media data.

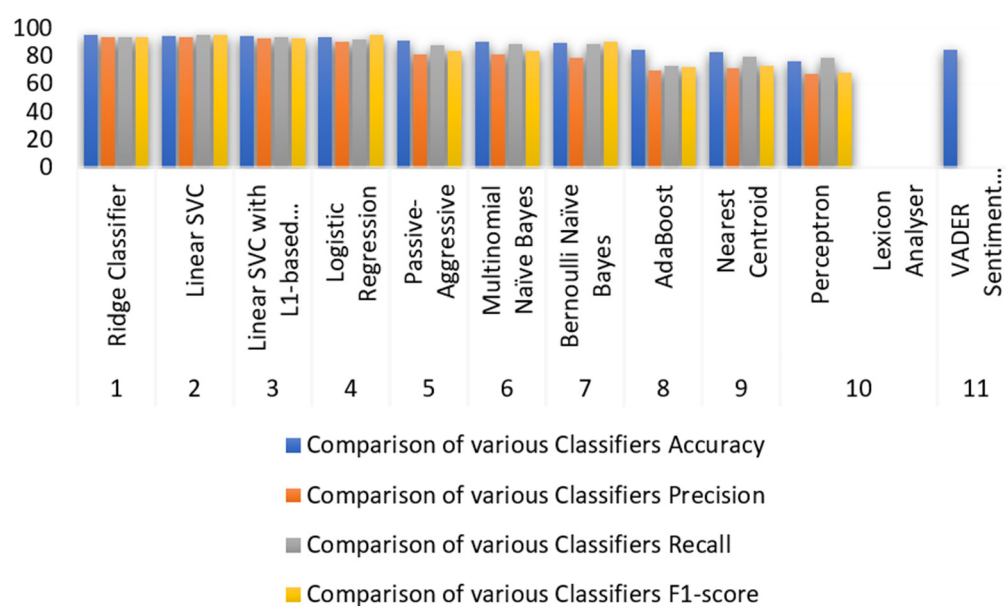


Figure 3. Comparison of various classifiers.

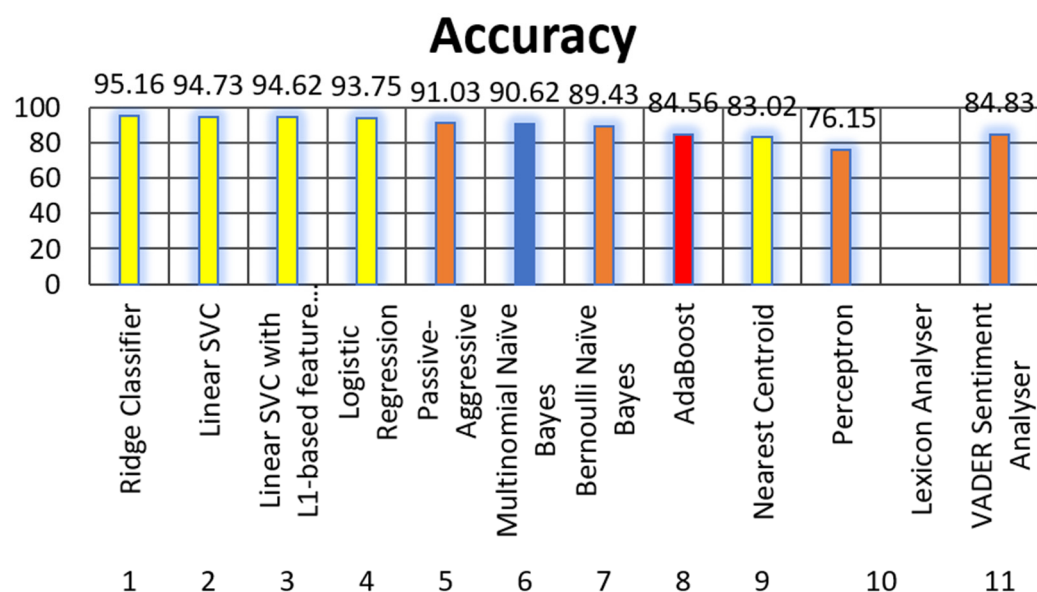


Figure 4. Accuracy comparison.

As a framework for their accuracy scores, several attributes are compared in the provided graph. To indicate various accuracy ranges, the accuracy comparison graph employs colored bars. The blue range represents accuracy scores greater than 95%. This range indicates the highest level of accuracy achieved by the features.

5. Conclusions

This research paper focuses on gathering and evaluating student input in the pursuit of educational excellence. It uses tweets to collect data and examine student comments using sentiment analysis. The Ridge Classifier performed better with a 95.16% accuracy rate, generating sentiment labels, and automating sentiment polarity labeling. This study's findings could transform educational institutions by providing data-driven insights, enabling informed decisions, and improving teaching methods, curriculum design, infrastructure, and support services. However, the caliber of the training data and any potential biases in the social media data affect the models' accuracy and dependability. Future directions for research and development include analyzing student comments using machine learning techniques, investigating contextual embeddings or deep learning techniques, and investigating advanced methods to identify patterns and connections in data. When faced with sentiment expressions that are poorly represented in the training data, the Ridge Classifier's performance may suffer. Furthermore, the representativeness and quality of the training dataset are critical to the Ridge Classifier's performance, and a skewed or imbalanced dataset may produce biased results. The suggested strategy must, therefore, be carefully considered and adjusted for use in different linguistic and cultural contexts.

Author Contributions: S.B.A.P. and R.P.K.N.; conceptualization; software; validation; formal analysis; investigation; resources; data curation; writing—preparation of the original draught; writing—review and editing; writing—visualization; S.B.A.P. and R.P.K.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, X.; Vorvoreanu, M.; Madhavan, K. Mining Social Media Data for Understanding Students' Learning Experiences. *IEEE Trans. Learn. Technol.* **2014**, *7*, 246–259. [\[CrossRef\]](#)
2. Aung, K.Z.; Myo, N.N. Sentiment analysis of students' comment using lexicon-based approach. In Proceedings of the IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 24–26 May 2017; pp. 149–154.
3. Dalipi, F.; Zdravkova, K.; Ahlgren, F. Sentiment Analysis of Students' Feedback in MOOCs: A Systematic Literature Review. *Front. Artif. Intell.* **2021**, *4*, 728708. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Wook, M.; Razali NA, M.; Ramli, S.; Wahab, N.A.; Hasbullah, N.A.; Zainudin, N.M.; Talib, M.L. Opinion mining Technique for developing student feedback analysis system using lexicon-based approach (OMFeedback). *Educ. Inf. Technol.* **2020**, *25*, 2549–2560. [\[CrossRef\]](#)
5. Dsouza, D.D.; Deepika, D.P.N.; Machado, E.J.; Adesh, N.D. Sentimental analysis of student feedback using machine learning techniques. *Int. J. Recent Technol. Eng.* **2019**, *8*, 986–991.
6. Baccianella, S.; Esuli, A.; Sebastiani, F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta, 17–23 May 2010.
7. Rosalind, J.M.; Suguna, S. Predicting Students' Satisfaction towards Online Courses Using Aspect-Based Sentiment Analysis. In *Computer, Communication, and Signal Processing*; Neuhold, E.J., Fernando, X., Lu, J., Piramuthu, S., Chandrabose, A., Eds.; IFIP Advances in Information and Communication Technology; Springer: Cham, Switzerland, 2022; p. 651.
8. Ortony, A.; Turner, T.J. What's basic about basic emotions? *Psychol. Rev.* **1990**, *97*, 315. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Mohammad, S.M.; Kiritchenko, S.; Zhu, X. Nrc-canada: Building the state-of-the art in sentiment analysis of tweets. In Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, GA, USA, 14–15 June 2013; pp. 321–327.
10. Ömer Osmanoğlu, U.; Atak, O.N.; Çağlar, K.; Kayhan, H.; Can, T. Sentiment analysis for distance education course materials: A machine learning approach. *J. Educ. Technol. Online Learn.* **2010**, *3*, 31–48. [\[CrossRef\]](#)
11. Lwin, H.H.; Oo, S.; Ye, K.Z.; Lin, K.K.; Aung, W.P.; Ko, P.P. Feedback analysis in outcome base education using machine learning. In Proceedings of the 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 24–27 June 2020.
12. Sivakumar, M.; Reddy, U.S. Aspect based sentiment analysis of student's opinion using machine learning techniques. In Proceedings of the 2017 International Conference on Inventive Computing and Informatics, Coimbatore, India, 23–24 November 2017.
13. Ahmad, M.; Aftab, S.; Bashir, M.S.; Hameed, N. Sentiment analysis using SVM: A systematic literature review. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*. [\[CrossRef\]](#)
14. Dehbozorgi, N.; Mohandoss, D.P. Aspect-based emotion analysis on speech for predicting performance in collaborative learning. In Proceedings of the 2021 IEEE Frontiers in Education Conference, FIE, Lincoln, NE, USA, 13–16 October 2021; IEEE: Piscataway, NJ, USA, 2021.
15. Lalata, J.P.; Gerardo, B.; Medina, R. A sentiment analysis model for faculty comment evaluation using ensemble machine learning algorithms. In Proceedings of the 2019 International Conference on Big Data Engineering, Hong Kong, China, 11–13 June 2019; pp. 68–73.
16. Edalati, M.; Imran, A.S.; Kastrati, Z.; Daudpota, S.M. The Potential of Machine Learning Algorithms for Sentiment Classification of Students' Feedback on MOOC. In *Lecture Notes in Networks and Systems*; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; pp. 11–22.
17. Wehbe, D.; Alhammedi, A.; Almaskari, H.; Alsereidi, K.; Ismail, H. UAE e-learning sentiment analysis framework. In Proceedings of the ArabWIC 2021: The 7th Annual Intl. Conference on Arab Women in Computing in Conjunction with the 2nd Forum of Women in Research, Sharjah, United Arab Emirates, 25–26 August 2021; ACM: New York, NY, USA, 2021.
18. Kastrati, Z.; Imran, A.S.; Kurti, A. Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs. *IEEE Access* **2020**, *8*, 106799–106810. [\[CrossRef\]](#)
19. Chaithanya, D.S.; Narayana, K.L.; Maheh, T.R. A Comprehensive Analysis: Classification Techniques for Educational Data mining. In Proceedings of the 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON), Bengaluru, India, 19–21 November 2021; pp. 173–176. [\[CrossRef\]](#)
20. Ramakrishna, M.T.; Venkatesan, V.K.; Bhardwaj, R.; Bhatia, S.; Rahmani, M.K.I.; Lashari, S.A.; Alabdali, A.M. HCoF: Hybrid Collaborative Filtering Using Social and Semantic Suggestions for Friend Recommendation. *Electronics* **2023**, *12*, 1365. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.