Business, Education, Social Sciences, and Technology
Vol. 1, no. 2, July 2025, pp. 149-154
https://doi.org/10.69478/BEST2025v1n2a022

Towards Effective Distance Education Implementation: Utilizing Descriptive Analytics, Opinion Mining, and Sentiment Analysis for Online Education Mentors and Learning Materials

Mary A. Soriano[1] and Amy Lyn M. Maddalora[2]

[1] College of Computer Studies, AMA University, Quezon City, Philippines
[2] College of Computing Studies, Information and Communication Technology, Isabela State University, Isabela, Philippines
masoriano@amauonline.com (Corresponding Email)

Abstract. This study utilizes sentiment analysis, an application of Natural Language Processing (NLP), to automate and improve the evaluation of online student feed-back. Student comments from academic years 2018–2025, were cleaned, preprocessed, and analyzed using the VADER sentiment tool to label feedback as positive, negative, or neutral. These labeled data were further used to train a neural network model that uses Long Short-Term Memory (LSTM) to improve sentiment classification. Tokenization, stopword elimination, lemmatization, and con-traction handling were all parts of the preparation step. VADER proved effective in detecting sentiment polarity and intensity in short student comments, while LSTM achieved an overall accuracy of 88.6%, particularly strong in classifying positive and neutral sentiments. A confusion matrix was used to assess model performance, measuring precision, recall, and F1-score. The descriptive method of textual data analysis provided insightful information about the issues that concerned students in various Online departments. The findings underscore the value of automated sentiment analysis as a feedback tool to continuously improve men-tor performance and course delivery in online learning environments. The study also highlights the need for balanced training datasets to enhance the classification of all sentiment types.

Keywords: NLP, Sentiment Analysis, VADER, LSTM, Confusion Matrix, Descriptive Analytics, Opinion Mining.

1. Introduction

Over the past few years, online learning has significantly changed traditional schooling. As a result, online learning platforms are now indispensable resources for educating diverse audiences. The primary component of instructional quality is the teachers' performance, which directly impacts the competitive and conducive environment of academic institutions [1]. Assessing the performance of faculty members or mentors [2] is now an essential part of the educational system, as it aims to evaluate the performance of mentors. It is often used in the promotion and annual appraisal process. Moreover, open feedback typically is not included in the overall evaluation of performance due to a lack of automated text analytics methods [4]. In order to sustain mentor effectiveness and course quality and, ultimately, create a

Received [May 20, 2025]; Revised [June 5, 2025]; Accepted [June 30, 2025]

Published by InnoCon Publishing
ISSN xxxx-xxxx

more successful learning environment, online student feedback is essential. However, it became more difficult to filter and analyze the survey responses as the number of students increased. These problems include prone to biases [5], labor-intensive due to the large datasets [6], high risk of human error increases [7], and more difficult, leading to inconsistent results.

With the problems stated above, this study will utilize sentiment analysis, which can effectively address many of the issues associated with manual feedback analysis in various ways. Sentiment Analysis will be employed to analyze students' feedback from the CME survey results. This will ensure that evaluations are not just focused on the numerical rating but also the comments of students to the mentors and course learning materials such as video lectures, modules, presentations, virtual class recordings, the content of the course written by Subject Matter Experts (SMEs), and mentors' involvement and support on the online learning of the students. The result will be used to evaluate the mentors' performance, determine the mentor's teaching quality, and inform appraisal for mentor re-alignment requests. Additionally, it will provide recommendations for course learning materials redevelopment and suggestions on how to improve the processes as a whole.

2. Methodology

The primary dataset was first extracted, a Course and Mentor Evaluation (CME) survey result in Online Education's Learning Management (LMS) covering the AY of 2018-2029 up to 2024-2025 3rd Trimester. All the extracted files are integrated into a single file, and then data cleaning and formatting are applied. A row with blank cells is filled with the word "No Comment" to complete the entire row. If the comment is "N/A" or "NA," it is replaced with "Not Applicable." All duplicate rows and comments with inconsistent values or wrong spelling are also removed.

Natural Language Processing (NLP) is then applied to the cleaned and formatted dataset, utilizing preprocessing techniques such as tokenization, stop word filtering, lemmatization, and contraction handling. The preprocessed data, or the final dataset, is then divided into three categories: 70% for training, 10% for validation, and 20% for testing.

The 70% dataset allotted for training is then applied in sentiment analysis. Starting with the use of VADER. In VADER, the student feedback is labeled as positive, negative, or neutral [9] based on the compound value from CME. Each word in the CME and is matched in the VADER lexicon dictionary, associated with a sentiment polarity score ranging from -4 to +4. After the data has been labeled using VADER, the next step is to apply the Long Short-Term Memory (LSTM) algorithm to classify the labeled data. In LSTM, sequence formatting is applied, where the student comments or feedback are converted into sequences. Afterward, sequence formation techniques are employed, such as tokenization and embedding. The final step is applying the confusion matrix into the trained dataset where the model's functionality is evaluated. The evaluation also includes the accuracy, precision, recall, and F1-score. These metrics provide important information about how well the machine classifies sentiment from student feedback. Through this step the detailed examination of misclassifications by categorizing the data into true positives, true negatives, false positives, and false negative.

3. Results and Discussion

From the CME dataset, the course initially consisted of 41,201 entries across 16 columns. Upon inspection, the researcher found 29,433 entries with missing values in the course content column. Additionally, 563 feedback entries were flagged as invalid due to containing fewer than 3 characters, special characters, or numeric-only responses. After data cleaning, a total of 29,996 entries were removed, resulting in 11,205 valid course feedback entries for analysis. Similarly, the mentor evaluation dataset also included 41,201 entries across 9 columns. The researcher identified 29,576 missing values in the mentor feedback column and 614 entries that were in-valid for the same reasons being too short, containing only special characters, or numeric values. After cleaning, 11,011 valid mentor feedback entries remained. Overall, a total of 22,206 valid feedback entries from the Course and Mentor (CME) datasets are retained for modeling.

To determine the polarity of relevant tokens from the structured feedback data, the researcher uses a Natural Language Processing (NLP) technique called sentiment analysis, applying the VADER (Valence Aware Dictionary and Sentiment Reasoner) tool. VADER is especially effective in analyzing short text, such as student feedback, because it not only detects whether a sentiment is positive, negative, or neutral it also considers the strength or intensity of that sentiment.

	Feedback	Category	Feedback_Cleaned	Sentiment Compound	Sentiment
8	Please dont confuse us some questions have the...	Course Feedback	please dont confuse u question choice chose an...	-0.0346	Neutral
9	The modules are really readerfriendly and info...	Course Feedback	module really readerfriendly informative quiz ...	0.0000	Neutral
10	good	Course Feedback	good	0.4404	Positive
18	I loved it Better than plain modules	Course Feedback	loved better plain module	0.7783	Positive
22	modules were full of imcomplete sentences with...	Course Feedback	module full imcomplete sentence lot repetetive...	-0.2960	Negative

Figure 1. Sentiment Analysis

Figure 1 Shows the Sentiment Analysis results using VADER. Each feedback entry is automatically analyzed using VADER, which assigns a compound score ranging from -1 to +1. This score reflects the overall emotional tone of the text: (1) A compound score of 0.05 or higher is labeled as positive; (2) A score of -0.05 or lower is considered negative; and (3) Scores between -0.05 and 0.05 fall into the neutral category. By using this method, the system efficiently tags each piece of feedback with a sentiment label based on how strongly the words express emotion. This allows the researcher to group feedback by sentiment and identify which words or phrases (tokens) are most commonly associated with each type of emotion.

After labeling the dataset as positive, negative or neutral using VADER, the LSTM algorithm was employed as a deep learning approach to analyze and classify sequential data, particularly in the context of textual feedback while confusion matrix is used to evaluate the LSTM as classification model.

Figure 2 shows the performance of the classification model showing the classification report, and learning curves for both training and validation phases. Positive samples were predicted most accurately, with 2089 correctly classified and only 134 misclassified as Neutral. Neutral samples also had good performance, with 1100 correctly classified and only a small number (38 and 134) being confused with

Positive or Negative classes. Negative samples were not predicted correctly at all 239 negative samples were misclassified, primarily as Neutral (191) and to a lesser extent as Positive (48). This reveals a significant class imbalance issue or model bias, where the model fails to recognize the Negative class entirely.

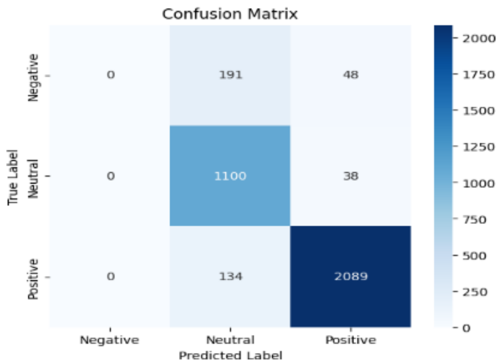

Figure 2. Confusion Matrix Analysis

```
Classification Report:

              precision    recall  f1-score   support

    Negative       0.00      0.00      0.00       239
     Neutral       0.77      0.97      0.86      1138
    Positive       0.96      0.94      0.95      2223

    accuracy                           0.89      3600
   macro avg       0.58      0.64      0.60      3600
weighted avg       0.84      0.89      0.86      3600

Test Accuracy: 0.8858
```

Figure 3. Classification Report

Figure 3 shows the overall accuracy of the model which is 88.6%, which is relatively high. However, the macro average F1-score (0.60) and precision (0.58) are much lower, reflecting poor performance across all classes, particularly due to the failure in identifying the negative class. The weighted average, which considers support class distribution, is high because the positive class dominates the dataset.

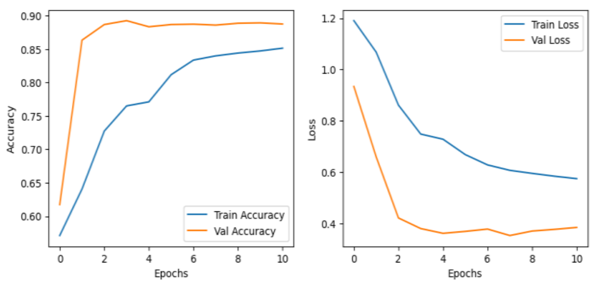

Figure 4. Training and Validation Curve

Figure 4 shows the accuracy curve where the validation accuracy rapidly increases and stabilizes around 88-89% early in training, indicating fast convergence. The loss curves for both training and validation decrease over epochs, with no signs of overfitting. Validation loss continues to decrease and does not diverge from training loss. These patterns suggest that the model is learning effectively and generalizes well on the validation data. However, the imbalance in class recognition (particularly the failure to learn the Negative class) is not reflected in the loss and accuracy trends alone.

4. Conclusion and Recommendations

Meaningful and trustworthy analysis depends heavily on how well the features are extracted and data is prepared. The use of the VADER in analyzing student feedback through sentiment analysis successfully determined the polarity of relevant tokens within the structured student feedback data. This method enabled the researcher to automatically classify each feedback entry as positive, negative, or neutral based on its compound sentiment score. VADER's ability to analyze short texts and account for the intensity of sentiment made it well-suited for processing brief student comments. The LSTM model, has proven effective in classifying student feedback sentiments, particularly in identifying positive and neutral sentiments. With an overall classification accuracy of 88.6%, the model demonstrated strong general performance and stable learning behavior, as evidenced by the consistent decrease in training and validation loss, as well as the convergence of accuracy during training. How-ever, despite the high accuracy, the model revealed a significant limitation in classifying negative sentiment. The confusion matrix showed that the LSTM model failed to correctly identify any negative samples, with most being misclassified as neutral or positive. This indicates a probable class imbalance within the dataset or model bias, where the negative class was underrepresented or under-learned during training.

The LSTM model performed well, however its inability to accurately detect negative feedback highlights a critical gap in this study. To address this, other researchers may use future models and should be able to train with more balanced datasets or enhanced using techniques like class weighting or oversampling. Institutions should also prioritize collecting and preserving critical feedback, as it provides the most actionable insights for improvement.

References

[1] Q. Rajput, S. Haider, S. Ghani, "*Lexicon-Based Sentiment Analysis of Teachers' Evaluation*," Applied Computational Intelligence and Soft Computing, vol. 2016, no. 1, October 2016, pp. 1-12, https://doi.org/10.1155/2016/2385429.

[2] A. L. Alix, D. J. Datul, M. C. Fernando-Raguro, A. C. Lagman, R. T. Adao, "*Faculty Evaluation System Platform with Decision Support Mechanism*," in the Proceedings of 2022 10th International Conference on Information and Education Technology (ICIET), Matsue, Japan, April 9-11, 2022, https://doi.org/10.1109/iciet55102.2022.9779033.

[3] Y. Kim, S. Kwak, M. Lee, M. Jeong, M. Park, Y. G. Park, "*Determination of Optimal Water Intake Layer Using Deep Learning-Based Water Quality Monitoring and Prediction*," Water, vol. 16, no. 1, December 2023, https://doi.org/10.3390/w16010015.

[4] B. Jagtap, V. A. Dhotre, "*SVM and HMM Based Hybrid Approach of Sentiment Analysis for Teacher Feedback Assessment*," International Journal of Emerging Trends & Technology in Computer Science, vol. 3, no. 3, June 2014, pp. 229-232,

https://www.researchgate.net/publication/387948295_SVM_and_HMM_Based_Hybrid_Approach_of_Sentiment_Analysis_for_Teacher_Feedback_Assessment.

[5] J. Arvai, "*Thinking, Fast and Slow. Farrar, Straus and Giroux*," Journal of Risk Research, vol. 16, no. 10, March 2013, pp. 1322-1324, https://doi.org/10.1080/13669877.2013.766389.

[6] N. Marz, J. Warren, "*Big Data: Principles and Best Practices of Scalable Realtime Data Systems*," Manning Publications Co., CT, USA, March 2015, https://dl.acm.org/doi/10.5555/2717065.

[7] S. L. Schneider, M. L. Council, "*Distance Learning in The Era Of COVID-19*," Archives of Dermatological Research, vol. 313, May 2020, pp. 389-390, https://doi.org/10.1007/s00403-020-02088-9.

[8] S. Abd El-Ghany, M. A. Mahmood, A. A. Abd El-Aziz, "*An Accurate Deep Learning-Based Computer-Aided Diagnosis System for Gastrointestinal Disease Detection Using Wireless Capsule Endoscopy Image Analysis*," Applied Sciences, vol. 14, no. 22, November 2024, https://doi.org/10.3390/app142210243.

[9] M. J. H. Al-Battbootti, I. Marin, S. Al-Hameed, R. C. Popa, I. Petrescu, C. A. Boianglu, N. Goga, "*Designing and Developing an Advanced Drone-Based Pollution Surveillance System for River Waterways, Streams, and Canals Using Machine Learning Algorithms: Case Study in Shatt al-Arab, South East Iraq*," Applied Sciences, vol. 14, no. 6, March 2024, https://doi.org/10.3390/app14062382.

[10] S. S. N. Syamsi, A. I. Hadiana, Y. H. Chrisnanto, "*Covid-19 Sentiment Analysis Using Random Forest Classification*," Enrichment Journal of Multidisciplinary Research and Development, vol. no. 6, 2023, https://doi.org/10.55324/enrichment.v1i6.53.

[11] Nehal, D. Jeet, V. Sharma, S. Mishra, C. Iwendi, J. Osamor, "*Twitter Sentiment Analysis and Emotion Detection Using NLTK and TextBlob*," in the Proceedings of 2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, December 12-13, 2023, https://doi.org/10.1109/ICCAKM58659.2023.10449540.

[12] A. Fadlurohman, N. Yunanita, F. H. N. Rohim, A. K. Wardani, A. F. Ningrum, "Classification of Mypertamina App Reviews Using Support Vector Machine," VARIANCE: Journal of Statistics and Its Applications, vol. 6, no. 2, October 2024, https://doi.org/10.30598/variancevol6iss2page223-228.

[13] L. S. Christakou, "*Depression Classification from rs-fMRI: A Local-to-Global Graph Neural Network Approach*," Master's Thesis, Department of General Psychology, University of Padova, Italy, October 2024, https://core.ac.uk/download/636345198.pdf.

[14] N. A. A. Khleel, K. Nehéz, "*Improving the Accuracy of Recurrent Neural Networks Models in Predicting Software Bug Based on Undersampling Methods*," Electrical Engineering and Computer Science, vol. 32, no. 1, October 2023, pp. 478-493, https://doi.org/10.11591/ijeecs.v32.i1.pp478-493.