2024 IEEE 9th International Conference for Convergence in Technology (I2CT)
Pune, India. Apr 5-7, 2024

Evaluating Teachers' Performance through Aspect-Based Sentiment Analysis

Abhijit Bhowmik
Faculty of Computing
Universiti Malaysia Pahang Al-Sultan Abdullah
Pekan, Malaysia
ovi775@gmail.com

Noorhuzaimi Mohd Noor
Faculty of Computing
Universiti Malaysia Pahang Al-Sultan Abdullah
Pekan, Malaysia
nhuzaimi@umpsa.edu.my

Md. Mazid-Ul-Haque
Department of Computer Science
American International Univeristy-Bangladesh
Dhaka, Bangladesh
mazid.ul.hqe@gmail.com

Md Saef Ullah Miah
Department of Computer Science
American International University-Bangladesh
Dhaka, Bangladesh
https://orcid.org/0000-0003-4587-4636

Debajyoti Karmaker
Department of Computer Science
American International University-Bangladesh
Dhaka, Bangladesh
d.karmaker@aiub.edu

2024 IEEE 9th International Conference for Convergence in Technology (I2CT) | 979-8-3503-9447-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/I2CT61223.2024.10543706

***Abstract*—This research demonstrates a novel approach for evaluating teacher performance by conducting aspect-based sentiment analysis (ABSA) on student feedback. A large dataset of over 2 million student comments about teachers is analyzed using cutting-edge natural language processing and customized deep learning techniques. The methodology involves identifying positive, negative and neutral aspects of teaching using a BiLSTM model. Rigorous preprocessing, domain adaptation, and performance metrics ensure a robust and objective evaluation. The granular, nuanced insights obtained through this aspect-level sentiment analysis enable educational institutions to provide targeted and unbiased feedback to teachers on their strengths and areas needing improvement. Moreover, this work lays the foundation for detecting potentially fraudulent reviews in academic settings – a crucial capability for safeguarding assessment integrity. The detailed aspect-based analysis methodology presented here significantly advances subjective and holistic evaluation practices. This research has far-reaching implications for enriching teacher development while upholding the credibility of performance assessments through sentiment analysis innovations.**

***Index Terms*—Feature extraction, Fraud reviews, Fraud review in academic settings detection, Teacher Performance evaluation, Deep learning, Aspect based sentiment analysis**

I. Introduction

In modern education, assessing teacher performance through student feedback carries immense significance. This research focuses primarily on Aspect-Based Sentiment Analysis (ABSA) as a powerful tool for dissecting teacher performance while establishing the groundwork for detecting fake reviews and ratings. By harnessing advanced natural language processing techniques and deep learning models, this study seeks to elevate the precision and objectivity of teacher performance evaluations, setting the stage for future endeavors to ensure the authenticity and reliability of educational assessments through sentiment analysis.

Modern methodologies for evaluating teacher performance often rely on subjective assessments and holistic sentiment analysis, but they may fall short in addressing the challenge posed by fake reviews and ratings. This research advocates adopting ABSA as a foundational framework to address this concern. ABSA's ability to dissect sentiment on a granular level enhances teacher performance assessments and lays the groundwork for safeguarding the integrity of reviews and ratings in educational contexts. To implement ABSA for teacher performance evaluations, this study introduces an innovative model that underpins the research's core objectives. This model seamlessly integrates deep learning, specifically Long Short-Term Memory (LSTM) networks, and augments sentiment analysis through a random forest classifier. Furthermore, it incorporates aspect selection and domain specificity, providing the adaptability and precision required to lay the foundation for detecting fake reviews and ratings.

While the primary thrust of this research is to provide educational institutions with a robust instrument for comprehensive teacher performance assessment, it concurrently paves the way for ensuring the credibility of reviews and ratings. According to the Spiegel Research Center [1], over 90% of consumers consult online reviews before making a purchase. After researching evaluations on several e-commerce sites, authors in [2] found that more than ten percent of reviews on e-commerce sites are fraudulent. ABSA not only delivers detailed evaluations of teacher performance but also anticipates its crucial role in future endeavors to preserve the integrity of educational assessments and the trustworthiness of reviews.

This research is driven by the dual imperative of enhancing teacher performance evaluation and establishing the foundation for authentic educational review and rating systems. By deploying advanced natural language processing techniques and deep learning, this study aims to offer a nuanced evaluation of teacher performance and lay the essential groundwork for detecting fake reviews and ratings, benefiting educators, institutions, and students.

II. Related Work

Thousands of reviews are published online every dayc [3]. When used effectively, this content can comfort users of the service's quality or warn them of the reverse if the reviews are poor. Review fraud is when individuals or organizations use user-generated information, such as by fabricating fake reviews, to misrepresent their company or competitors. Analyzing user sentiment and identifying fake reviews and ratings is crucial for businesses to progress. Contextual and behavioral features can help improve the accuracy and depth of sentiment analysis when used appropriately and are the most common features used to identify fraud reviews. Review-centric characteristics, service-centric attributes, and reviewer-centric attributes are all used to extract both types of features [2].

A significant amount of research has been done on user reviews of various services they have received. Authors in [4] review the literature on fake review detection on online platforms. Their study covers both basic research and commercial solutions, as well as the reasons for the present approaches and legislation's low performance in mitigating damage caused by false reviews [4]. The findings of [5] have significant implications for increasing consumer decision-making effectiveness and the reliability of online user reviews. Using deception and attitude-behavior consistency theories, [5] characterizes review inconsistency from many perspectives, including rating sentiment, content, and language and presents hypotheses regarding their implications on detecting fraudulent online user reviews. The authors used the bag of words model and the glove embedding matrix focusing on fake reviews in [6]. Their approach used two different feature extraction techniques and three new deep-learning algorithms on text classifications. The experimental analysis with an existing public dataset produced similar or better results than traditional machine-learning models.

The authors in [7] concentrate on developing a machine learning model for fake review detection and comparing the performance of three different algorithms. The random forest algorithm outperforms the other two algorithms due to this research. An overview of the work that investigates the market for fake product reviews on Amazon.com, where reviews are purchased in large private internet groups on Facebook and other sites are provided in [8]. It has been discovered that many products, including those with many reviews and high average ratings, purchase fake reviews. Purchasing fake reviews on Facebook results in a significant but temporary increase in average ratings and reviews. Factors distinguishing between false and legitimate reviews have been identified in [9]. The author develops a novel detection method by distinguishing authentic and fraudulent word patterns relevant to a domain (e.g., hotel services). The proposed methodology in [10] used four typical multidomain fake review datasets: hotels, restaurants, Yelp, and Amazon. The suggested CNN-LSTM model was examined in two types of experiments, in-domain, and cross-domain tests, in their work. A comparison of the outcomes of in-domain tests with current methodologies was undertaken, and it was discovered that the suggested model outperformed the compared methods. The authors investigate whether a user's vulnerability to fake reviews influences user engagement, intention to adopt information and proclivity to purchase on online review sites to protect the online customer experience from malicious marketing actions in [11].

The authors of their study [12] address creating and detecting fake reviews using two language models, ULMFiT and GPT-2, to generate fake product reviews based on an Amazon e-commerce dataset. The findings of the study have implications for consumer protection, firm defense against unfair competition, and review platform responsibility. The authors propose a machine learning method for detecting fake reviews [13]. In addition to the review features extraction process, their study employs several features engineering techniques to extract various reviewer behaviors. The authors compare the performance of several classifiers in both cases: KNN, Naive Bayes (NB), SVM, Logistic Regression, and Random Forest. Different n-gram language models, specifically bi-gram and trigram, are also considered during the evaluations. The results [13] show that their model performs well when the extracted behavioral features of the reviewers are taken into account. Because of the high cost of manual labeling, it is not feasible to label reviews on a large scale. To address this issue, in [14], researchers propose a fake reviews detection model based on vertical ensemble tri-training and active learning (VETT-AL) to improve detection performance by utilizing unlabeled reviews. The experimental results show that the proposed model performs well in classification. There is still a need for a survey that can analyze and summarize existing approaches.

To bridge the gap, the survey paper presented in [15] details the task of fake review detection, summarizing existing datasets and collection methods. It examines the current feature extraction techniques. It also critically summarizes and analyzes existing techniques to identify gaps based on two groups: traditional statistical machine learning methods and deep-learning methods. In addition, the authors conduct a benchmark study to examine the performance of various neural network models and transformers that have not yet been used for fake review detection. For the deception dataset, experimental results [15] on two benchmark datasets show that RoBERTa outperforms state-of-the-art methods in a mixed domain. Like in business sectors, education sectors are also focusing on reviews available over the internet about various components like acceptance of degrees, quality of teachers, quality of teaching, educational resources, and many more relevant parameters. It is crucial to know whether the students are providing actual reviews and ratings, as various important factors depend on these reviews and ratings from an organizational perspective. However, in the research domain, there is a significant gap in this important matter. Also, datasets containing this sort of review and ratings are scarce. In the existing literature, there is close to no work available on analyzing student reviews and ratings of the teachers. Most educational institutes interpret the surveys in their own ways, and no significant methodical approach is available to do so efficiently. Most of the research works can be found in the educational domain are done on students' feedback and their sentiment analysis on various matters such as on learning environments and platforms [16]–[19], tools [20], teaching quality [21]–[23], and the relativity of their topics [24].

The authors in [25] address the need for a more comprehensive examination of student feedback to enhance teaching and learning effectiveness. They propose aspect-based sentiment analysis to delve into specific components contributing to student opinions, including course content, teaching approach, assessment methods, and student support. Their methodology involves collecting feedback data, preprocessing it, and utilizing linguistic patterns, machine learning models, and lexicon-based approaches to extract aspects and determine sentiment polarity. This targeted strategy enables educators to identify areas for improvement or alteration in teaching tactics, course material, and student support systems. While highlighting the benefits, the study acknowledges challenges such as extensive feedback collection, precise aspect categorization, and potential subjectivity in sentiment analysis algorithms, underscoring the need for further refinement.

In [26], the authors explore the application of aspect-based sentiment analysis to analyze student feedback gathered from the Twitter API. They emphasize categorizing this feedback based on various criteria, highlighting key aspects like teaching, placement, facilities, sports, events, fees, and transportation. The study employs classification methods such as decision tree induction, support vector machines, and the Naive Bayes classifier to classify feedback into positive and negative sentiments. Additionally, they use K-means clustering to group feedback based on multiple aspects. This aspect-based approach offers educational institutions valuable insights into student emotions and concerns, aiding in informed decision-making and enhancing the teaching-learning process. However, the study acknowledges data preparation challenges and suggests potential preprocessing improvements to enhance sentiment analysis accuracy for student feedback at the sentence level.

The preceding discussions have illuminated a notable void in efficient fake review and rating detection, specifically focusing on its underrepresentation in educational evaluations. The extant body of research primarily centers its efforts on detecting fraudulent reviews and ratings within the context of e-commerce customer feedback. These methodological approaches have been developed to suit platforms such as Amazon, Yelp, or hotel review datasets. While these established methodologies have demonstrated commendable performance within their corresponding domains, a discernible research gap exists – notably, the scarcity of comprehensive investigations dedicated to applying sentiment analysis techniques for uncovering fake reviews within the domain of teacher performance evaluations. In response to this research gap, our study endeavors to introduce a pioneering approach.

This research diverges from the existing literature by zeroing in on the previously underexplored realm of education. Within this terrain, teacher performance evaluations are derived from a newly formed dataset encompassing student comments. By expanding the perspective of sentiment analysis into this untapped educational landscape, our research aims to provide a novel perspective on the detection of fake reviews and ratings, thereby opening promising avenues for enhancing the authenticity and integrity of educational assessments.

III. Methodology

The core of this research is the Aspect-Based Sentiment Analysis (ABSA) methodology, which is incorporated here to conduct a detailed analysis of teacher performance using student comments. ABSA is essential for producing a comprehensive evaluation by untangling the complex web of sentiments woven throughout student input, thus boosting the quality of insights for teacher performance. Figure 1 represents the proposed method of this research.

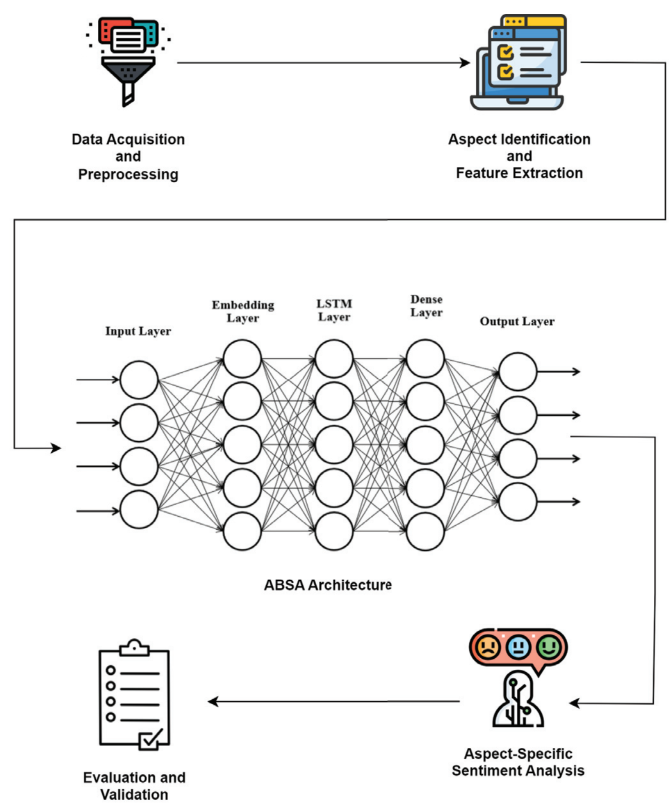

Fig. 1. Overview of the proposed methodology.

A. Input Data Acquisition and Preprocessing

The initial phase involves gathering data from the university's administrative system, encapsulating student assessments and evaluations. A dataset comprising 2,204,523 rows was meticulously processed, each representing a unique record within the university's teacher performance evaluation corpus [27], [28]. This involved rigorous data cleansing, including handling null values, text cleansing, noise reduction, and word count calculation, to ensure data quality and relevance for subsequent analysis.

B. Aspect Identification and Feature Extraction

ABSA hinges on identifying relevant aspects within student comments, such as teaching style, course content, and communication skills. Aspect extraction is achieved through natural language processing (NLP) techniques, combining rule-based and machine learning-driven approaches to discern the facets underpinning each comment. Feature extraction encompasses textual, semantic, and contextual attributes conducive to sentiment analysis, forming the foundation for subsequent modeling phases.

C. Aspect-Based Sentiment Analysis Model Architecture

The heart of ABSA lies in its deep neural network architecture, thoughtfully designed to capture the intricate nuances of sentiment embedded within student comments. This architecture encompasses the following layers.

1) Input Layer: The entry point for textual data, accepting student comments in their original form or preprocessed sequences.
2) Embedding Layer: Employing word embeddings or pretrained embeddings, this layer converts textual data into fixed-size dense vectors, enabling seamless information propagation through subsequent layers.
3) LSTM Layer: The Long Short-Term Memory (LSTM) layer captures the temporal dependencies between words within student comments. This recurrent layer can discern subtle shifts in sentiment across sequences, contributing to a comprehensive sentiment analysis.
4) Dense Layer: Following LSTM processing, the dense layer further refines the extracted features, facilitating the abstraction of salient sentiment-related attributes.
5) Output Layer: The ultimate layer, the output layer, performs sentiment classification based on the ascertained features, categorizing each review into predefined sentiment classes.

D. Aspect-Specific Sentiment Analysis

ABSA is further customized to enable aspect-specific sentiment analysis. This entails the application of the ABSA architecture to each identified aspect independently, allowing for a nuanced evaluation of teacher performance across various dimensions. This aspect-level granularity augments the depth of insights, revealing specific strengths and areas for improvement.

E. Evaluation and Validation

The ABSA models undergo rigorous evaluation and validation, employing performance metrics like precision, recall, F1-score, and accuracy. Robust validation methods, including cross-validation and holdout validation, ensure their reliability and generalization to unseen data.

IV. Results and Discussion

This section presents and discusses the outcomes of this research on teachers' performance assessment through Aspect-Based Sentiment Analysis (ABSA). Advanced natural language processing techniques and deep learning models were employed to explore the intricate dimensions of ABSA, primarily focusing on teacher evaluations derived from student feedback. The aim was to enhance teacher performance evaluations' precision and impartiality while providing educators with constructive feedback. Additionally, the research set the stage for future endeavors related to fake reviews and rating detection through sentiment analysis.

The analysis of loss and accuracy curves for the proposed model is illustrated in Figure 2. These curves provide insights into the model's performance during the training and validation phases. Notably, the absence of significant gaps or irregularities between the training and validation loss curves indicates that overfitting or underfitting issues were not observed. Similarly, the consistency between the training and validation accuracy curves suggests that stable and reliable performance was achieved across the dataset.

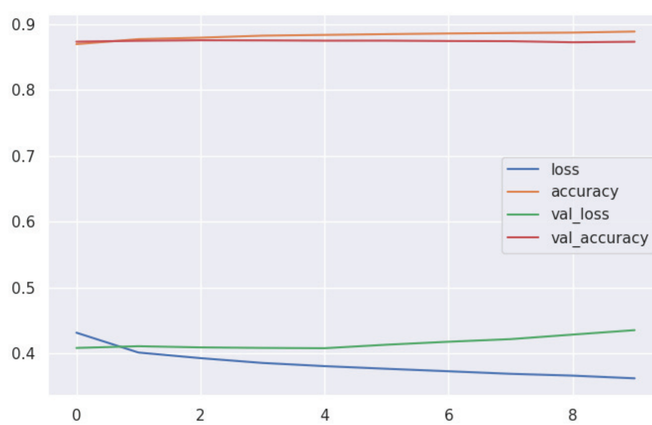

Fig. 2. Accuracy and Loss curves for the proposed model.

A confusion matrix was constructed to evaluate the model's performance in classifying sentiment categories. This matrix offers a comprehensive view of the model's classification accuracy and highlights its ability to differentiate between sentiment categories. The confusion matrix metrics are presented in Figure 3.

The confusion matrix offers a comprehensive view of the model's classification accuracy and ability to distinguish between sentiment categories. The key metrics derived from the confusion matrix are presented in Table I 1140 instances were correctly classified as "Negative", indicating successful

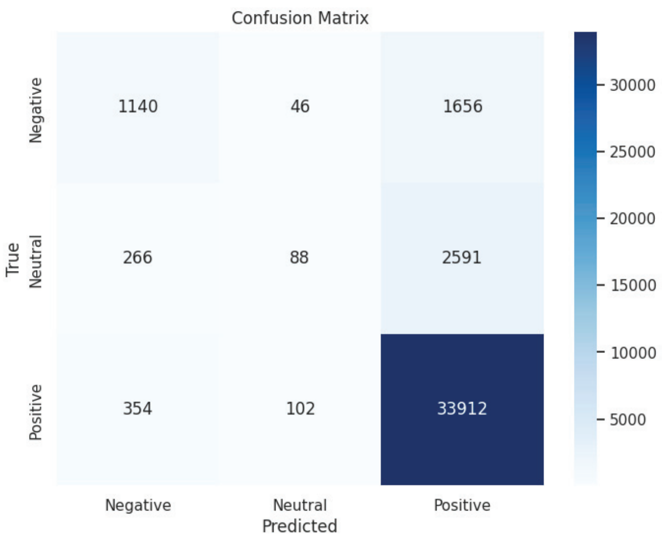

Fig. 3. Confusion Matrix of the proposed model.

TABLE I
CONFUSION MATRIX SUMMARY OF THE PROPOSED MODEL.

	Negative	Neutral	Positive
True Negatives (TN)	1140	88	33912
False Positives (FP)	46	266	102
False Negatives (FN)	1656	2591	354

identification and avoidance of false negatives. 46 instances originally labeled as "Neutral" were inaccurately classified as "Negative". 1656 genuinely "Positive" instances were misclassified as "Negative". 88 instances were correctly identified as "Neutral," signifying accurate predictions. 266 instances initially labeled as "Negative" were wrongly classified as "Neutral". 2591 instances that were truly "Positive" were mistakenly categorized as "Neutral". 33912 instances were accurately predicted as "Positive", indicating strong classification performance. 102 instances originally labeled as "Negative" were falsely identified as "Positive." 354 instances that were indeed "Neutral" were erroneously classified as "Positive". These comprehensive insights from the confusion matrix allow us to evaluate the model's strengths and weaknesses in sentiment classification across various categories, guiding potential areas of improvement.

This comprehensive representation enables the assessment of the model's efficacy in assigning sentiment labels to text data, providing valuable insights into the strengths and shortcomings of its performance for each sentiment category.

The performance of the LSTM model was evaluated using various metrics, including precision, recall, and F1-score for each sentiment class. Additionally, weighted average precision, recall, and F1-score values were calculated to assess the model's performance across all classes. Test loss and test accuracy were observed as Test 0.403 and 0.877. The performance metrics of the proposed model are shown in Table II.

TABLE II
ACHIEVED EXPERIMENTAL RESULTS FOR THE PROPOSED MODEL.

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.70	0.36	0.48	2842
Neutral	0.54	0.02	0.04	2945
Positive	0.88	0.99	0.94	34368
Accuracy			0.88	40155
Macro Avg	0.71	0.46	0.48	40155
Weighted Avg	0.85	0.88	0.84	40155

The results indicate that the model excels in classifying "Positive" sentiment, achieving high precision, recall, and F1-score. However, room for improvement exists in classifying "Negative" and "Neutral" sentiments, where recall and F1-scores are relatively lower. The weighted average metrics suggest that, on average, the model performs well, with high precision, recall, and F1-score across all classes, emphasizing its overall effectiveness in sentiment analysis.

In conclusion, the results of this study suggest that Aspect-Based Sentiment Analysis, coupled with deep learning techniques, can be a valuable tool for assessing teachers' performance based on student feedback. Strong performance is observed in identifying positive sentiments, while further refinements may be needed to improve classification accuracy for negative and neutral sentiments. These findings contribute to ongoing efforts to enhance performance evaluations in educational settings and open avenues for addressing challenges related to fake review detection and sentiment analysis in the future.

V. CONCLUSION AND FUTURE WORK

This research explored teachers' performance assessment through Aspect-Based Sentiment Analysis (ABSA), specifically focusing on teacher evaluations derived from student feedback. The intricate dimensions of ABSA were delved into, employing advanced natural language processing techniques and deep learning models, aiming to elevate the precision and impartiality of performance evaluations. Additionally, constructive feedback to educators was sought while setting the stage for future endeavors related to fake reviews and rating detection through sentiment analysis.

The findings have been analyzed, and the model's performance has been assessed. The absence of significant gaps or irregularities between the training and validation loss curves, as demonstrated in Figure 2, indicates that the model's susceptibility to overfitting or underfitting issues was minimal. Moreover, the consistency observed between the training and validation accuracy curves suggests that stable and dependable performance was achieved across the dataset. The construction and examination of the confusion matrix enabled a comprehensive evaluation of the model's capacity to classify sentiment categories, revealing strengths and areas for improving classification accuracy. Although the model excelled in accurately identifying positive sentiments, further enhancements are necessary to address the classification of negative and neutral sentiments.

The performance metrics of the LSTM model, as illustrated in Table II, demonstrate the model's proficiency in achieving high precision, recall, and F1-scores for positive sentiment classification. Nonetheless, the model exhibited lower recall and F1-scores for negative and neutral sentiment classes, indicating areas needing further optimization. The weighted average metrics underscore the overall effectiveness of the model in sentiment analysis. In conclusion, Aspect-Based Sentiment Analysis, coupled with deep learning techniques, has been shown to be a valuable tool for assessing teachers' performance based on student feedback. While the model excels in recognizing positive sentiments, addressing challenges related to negative and neutral sentiment classification will be the focus of future work. Additionally, applying sentiment analysis techniques in detecting fake reviews and ratings is an avenue that holds promise and will contribute to the broader field of sentiment analysis and its practical applications.

Acknowledgment

The work has been supported by UMPSA RDU GRANT RDU230352, titled "A fraud detection model for instructor's evaluation based on semantic keyword extraction using machine learning".

References

[1] H. Paul and A. Nikolaev, "Fake review detection on online e-commerce platforms: a systematic literature review," *Data Mining and Knowledge Discovery*, vol. 35, no. 5, pp. 1830–1881, 2021.

[2] N. Jindal and B. Liu, "Analyzing and detecting review spam," in *Seventh IEEE international conference on data mining (ICDM 2007)*. IEEE, 2007, pp. 547–552.

[3] M. S. U. Miah, J. Sulaiman, T. B. Sarwar, N. Ibrahim, M. Masuduzzaman, and R. Jose, "An automated materials and processes identification tool for material informatics using deep learning approach," *Heliyon*.

[4] C. Sun, Q. Du, and G. Tian, "Exploiting product related review features for fake review detection," *Mathematical Problems in Engineering*, vol. 2016, 2016.

[5] G. Shan, L. Zhou, and D. Zhang, "From conflicts and confusion to doubts: Examining review inconsistency for fake review detection," *Decision Support Systems*, vol. 144, p. 113513, 2021.

[6] D. Baishya, J. J. Deka, G. Dey, and P. K. Singh, "Safer: Sentiment analysis-based fake review detection in e-commerce using deep learning," *SN Computer Science*, vol. 2, no. 6, pp. 1–12, 2021.

[7] V. Sumathi, S. Pudhiyavan, M. Saran, and V. N. Kumar, "Fake review detection of e-commerce electronic products using machine learning techniques," in *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*. IEEE, 2021, pp. 1–5.

[8] S. He, B. Hollenbeck, and D. Proserpio, "Exploiting social media for fake reviews: evidence from amazon and facebook," *ACM SIGecom Exchanges*, vol. 19, no. 2, pp. 68–74, 2021.

[9] S. Moon, M.-Y. Kim, and D. Iacobucci, "Content analysis of fake consumer reviews by survey-based text categorization," *International Journal of Research in Marketing*, vol. 38, no. 2, pp. 343–364, 2021.

[10] S. N. Alsubari, S. N. Deshmukh, M. H. Al-Adhaileh, F. W. Alsaade, and T. H. Aldhyani, "Development of integrated neural network model for identification of fake reviews in e-commerce using multidomain datasets," *Applied Bionics and Biomechanics*, vol. 2021, 2021.

[11] L. Lo Presti and G. Maggiore, "Vulnerability on collaborative networks and customer engagement: defending the online customer experience from fake reviews," *Quality & Quantity*, pp. 1–15, 2021.

[12] J. Salminen, C. Kandpal, A. M. Kamel, S.-g. Jung, and B. J. Jansen, "Creating and detecting fake reviews of online products," *Journal of Retailing and Consumer Services*, vol. 64, p. 102771, 2022.

[13] A. M. Elmogy, U. Tariq, M. Ammar, and A. Ibrahim, "Fake reviews detection using supervised machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021.

[14] C. Yin, H. Cuan, Y. Zhu, and Z. Yin, "Improved fake reviews detection model based on vertical ensemble tri-training and active learning," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 12, no. 4, pp. 1–19, 2021.

[15] R. Mohawesh, S. Xu, S. N. Tran, R. Ollington, M. Springer, Y. Jararweh, and S. Maqsood, "Fake reviews detection: A survey," *IEEE Access*, vol. 9, pp. 65 771–65 802, 2021.

[16] Z. Kastrati, F. Dalipi, A. S. Imran, K. Pireva Nuci, and M. A. Wani, "Sentiment analysis of students' feedback with nlp and deep learning: A systematic mapping study," *Applied Sciences*, vol. 11, no. 9, p. 3986, 2021.

[17] K. Lundqvist, T. Liyanagunawardena, and L. Starkey, "Evaluation of student feedback within a mooc using sentiment analysis and target groups," *International Review of Research in Open and Distributed Learning*, vol. 21, no. 3, pp. 140–156, 2020.

[18] N. M. Dickson-Karn, "Student feedback on distance learning in the quantitative chemical analysis laboratory," *Journal of Chemical Education*, vol. 97, no. 9, pp. 2955–2959, 2020.

[19] E. Santhanam, B. Lynch, and J. Jones, "Making sense of student feedback using text analysis–adapting and expanding a common lexicon," *Quality Assurance in Education*, 2018.

[20] J. J. Bullón, A. H. Encinas, M. J. S. Sánchez, and V. G. Martínez, "Analysis of student feedback when using gamification tools in math subjects," in *2018 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 2018, pp. 1818–1823.

[21] M. Hujala, A. Knutas, T. Hynninen, and H. Arminen, "Improving the quality of teaching by utilising written student feedback: A streamlined process," *Computers & education*, vol. 157, p. 103965, 2020.

[22] J. G. Pyke and J. J. Sherlock, "A closer look at instructor-student feedback online: A case study analysis of the types and frequency," *Journal of Online Learning and Teaching*, vol. 6, no. 1, pp. 110–121, 2010.

[23] Z. Nasim, Q. Rajput, and S. Haider, "Sentiment analysis of student feedback using machine learning and lexicon based approaches," in *2017 international conference on research and innovation in information systems (ICRIIS)*. IEEE, 2017, pp. 1–6.

[24] D. E. Clayson, "Student evaluations of teaching: Are they related to what students learn? a meta-analysis and review of the literature," *Journal of marketing education*, vol. 31, no. 1, pp. 16–30, 2009.

[25] G. S. Chauhan, P. Agrawal, and Y. K. Meena, "Aspect-based sentiment analysis of students' feedback to improve teaching–learning process," in *Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2018, Volume 2*. Springer, 2019, pp. 259–266.

[26] M. Sivakumar and U. S. Reddy, "Aspect based sentiment analysis of students opinion using machine learning techniques," in *2017 international conference on inventive computing and informatics (ICICI)*. IEEE, 2017, pp. 726–731.

[27] A. Bhowmik, N. M. Noor, M. S. U. Miah, M. Mazid-Ul-Haque, and D. Karmaker, "A comprehensive dataset for aspect-based sentiment analysis in evaluating teacher performance," *AIUB Journal of Science and Engineering (AJSE)*, vol. 22, no. 2, pp. 200–213, 2023.

[28] M Saef Ullah Miah, "A Novel Dataset for Aspect-based Sentiment Analysis for Teacher Performance Evaluation," Apr. 2023. [Online]. Available: https://data.mendeley.com/datasets/b2yhc95rnx/1