JOURNAL OF INTERDISCIPLINARY PERSPECTIVES

ISSN Print: 2984-8288, ISSN Online: 2984-8385
Vol. 3 No. 9, pp. 557-567, September 2025

Designing a Teacher Recommender System: A Thematic Literature Review of Teacher Evaluation Systems

Marie Grace V. Ortiz*, Menchita F. Dumlao
Angeles University Foundation, Angeles City, Philippines

*Corresponding Author Email: ortiz.mariegrace@auf.edu.ph

Date received: June 17, 2025
Date revised: August 11, 2025
Date accepted: August 26, 2025

Originality: 99%
Grammarly Score: 99%
Similarity: 1%

Recommended citation:
Ortiz, M. G., & Dumlao, M. (2025). Designing a teacher recommender system: A thematic literature review of teacher evaluation systems. *Journal of Interdisciplinary Perspectives*, *3*(9), 557-567.
https://doi.org/10.69569/jip.2025.513

Abstract. Teacher evaluation systems are limited in their ability to provide numerical ratings, often failing to analyze qualitative feedback to provide teachers with valuable insights to enhance performance. This paper conducts a thematic literature review of teacher evaluation systems and tools in articles in Google Scholar, IEEE, and Proquest databases between 2014 and 2024 to determine the most appropriate sentiment analysis (SA) and topic modeling (TM) algorithms for analyzing student feedback. The review of 48 articles found that a lexicon-based SA approach, specifically VADER with a customized Filipino lexicon, offers a robust and practical solution for sentiment detection in a multilingual context. For TM, Latent Dirichlet Allocation (LDA) with human intervention is the recommended approach, providing a balance between thematic granularity and computational feasibility. The efficacy and efficiency of both algorithms are found to improve by increasing the size of a domain-specific corpus of words. Based on these findings, the paper proposes the design of TeachAIRs. This teacher recommender system includes word cloud visualizations, sentiment scores per topic, and, most critically, actionable insights derived from the integrated analysis. The development of this system is highly recommended to provide teachers with valuable and constructive real-time feedback, ultimately enhancing teaching practices and improving student learning outcomes.

Keywords: Recommender systems; Sentiment analysis; Student feedback; Teacher evaluation; Topic modeling

1.0 Introduction

The conduct of student-teacher evaluations is a cornerstone of efforts to improve educational quality. However, most current evaluation tools primarily rely on Likert-based questions. While these provide valuable quantitative scores, they often fall short in offering actionable insights into the specific teaching aspects that are successful or in need of improvement. Consequently, even when students provide qualitative comments, these traditional tools cannot typically synthesize this rich feedback into a clear and actionable format that truly informs and enhances a teacher's practice. Fostering continuous improvement requires a more systematic evaluation tool that incorporates comprehensive feedback mechanisms and a novel evaluative approach. Such a tool should not only measure performance but also uncover new information critical for professional growth. (C. M. Kim & Kwak, 2022; Looney, 2011; Papay, 2012; Rafiq et al., 2022). To ensure fair and reliable assessment, it is critical that these systems can effectively analyze qualitative comments, providing teachers with valuable, constructive feedback and insights. (Fernández & Martínez, 2022; Nasim et al., 2017; Rajput et al., 2016).

The application of machine learning (ML) offers a powerful solution to this analytical gap, enabling a deeper processing of the vast volumes of student feedback. This domain remains relatively under-researched within teacher evaluations. (Das et al., 2022; Pramod et al., 2022). One crucial ML application is Sentiment Analysis (SA), a subfield of Natural Language Processing (NLP). SA techniques process and categorize student feedback into positive, negative, or neutral sentiments, offering a deeper understanding of student needs, preferences, and emotions. Recent studies from 2019 to 2024 have increasingly demonstrated the effectiveness of SA in the educational arena. (Bhowmik, Mohd Noor, et al., 2023; Mamidted & Maulana, 2023; Peña-Torres, 2024; Zeng et al., 2023). This thematic review will examine three key approaches to SA: conventional machine learning, lexicon-based, and hybrid methods, each with its strengths in classifying sentiment. (Malebary & Abulfaraj, 2024)

However, a comprehensive understanding of teaching effectiveness requires more than a simple positive/negative sentiment. To gain a detailed insight into how to enhance the teaching and learning experience, researchers and educators recommend evaluation tools that can capture the various dimensions of teaching effectiveness. This need for a method to systematically uncover the specific themes within qualitative feedback makes Topic Modeling (TM) a highly appropriate and robust solution. As a subfield of machine learning, TM is broadly categorized into four approaches: probabilistic (e.g., Latent Dirichlet Allocation: LDA), fuzzy, algebraic (e.g., Non-negative Matrix Factorization: NMF, Latent Semantic Analysis: LSA), and neural (e.g., BERT-based methods). While the literature on teacher evaluation predominantly focuses on probabilistic topic modeling, a recent and crucial study by Hayat et al. (2024) provides a valuable comparative analysis of several of these techniques, including BERT, LDA, LSA, and NMF. Timely research from 2019 to 2024 reinforces TM's capacity to cluster comments effectively. (Bhowmik, Mohd Noor, et al., 2023; Mamidted & Maulana, 2023; Tian et al., 2022), providing teachers with direct insights into what worked well and what did not.

The combined insights derived from sentiment analysis and topic modeling underscore the critical need for actionable recommendations to foster continuous learning enhancements. This understanding naturally leads to the growing potential of Artificial Intelligence (AI) recommender systems. For feedback to be most useful, it must be immediate and allow for reflection, a change of practice, and adaptation to new strategies while there is still time to meet students' learning needs. Thus, a recommender system is an appropriate tool for providing immediate, formative feedback that can impact teachers and students alike. While most recommender systems have traditionally been student-focused, recent advancements in AI applications for education have shown their capability to suggest areas for improvement to teachers. (Seo et al., 2021), uncover student learning behavior patterns to aid curriculum redesign (Hashim et al., 2022) Moreover, reveal trends in classroom dynamics and instructional effectiveness. (Wang, 2025). With these premises highlighting the current limitations of traditional evaluations and the proven capabilities of advanced NLP and AI techniques, this paper performs a thematic literature review of teacher evaluation systems. It aims to determine the most appropriate SA and TM algorithms for analyzing student feedback in teacher evaluations, ultimately informing the design and development of an innovative teacher AI recommender system aimed at improving teacher performance.

2.0 Methodology

2.1 Research Design

This thematic literature review synthesized insights from 48 carefully selected academic papers to gain a more in-depth understanding of the existing knowledge on SA and TM applied in analyzing student textual feedback in teacher evaluation. This study employed a thematic literature review. This approach was chosen to systematically explore and summarize key findings of the selected papers rather than providing a comprehensive summary of the literature. (Creswell & David Creswell, 2018) And to find patterns and themes that are of great relevance to the development of future research context (Bhana, 2014). Unlike highly structured systematic reviews or meta-analyses, which are primarily designed to answer precise, often quantitative research questions and evaluate the efficacy of interventions, this study's objective was to synthesize diverse insights and approaches from the literature to inform the design of a novel teacher recommender system. A thematic review provided the necessary flexibility to identify recurring concepts, methodologies, findings, and recommendations across various studies related to teacher evaluation, sentiment analysis, and topic modeling. This qualitative and interpretive synthesis was crucial for establishing a robust conceptual framework and identifying the specific algorithms and components suitable for a teacher recommender system's development, which would not have been adequately served by the more restrictive scope of other review methodologies.

2.2 Data Gathering Procedure

A comprehensive and systematic search strategy was developed to identify pertinent academic papers, aligning with principles of transparent literature review. The primary databases consulted included Google Scholar, IEEE, and ProQuest, chosen for their extensive coverage of computer science, education, and interdisciplinary research. The literature search specifically targeted articles published between 2014 and 2024 to ensure the inclusion of contemporary research and technological advancements in the rapidly evolving fields of AI and education. Initial search strings used were any of the following or a combination of two or more terms: teacher evaluation, student feedback, sentiment analysis, and topic modeling. For instance, common search combinations included "teacher evaluation and sentiment analysis and student feedback" or "topic modeling and teacher evaluation and student feedback". Following the initial searches, a total of 405 articles were identified across these databases. Duplicates identified across databases were systematically discarded, resulting in 299 unique articles.

2.3 Screening and Selection Criteria

To ensure the **r**elevance and quality of the reviewed literature, a two-stage screening process was applied based on specific inclusion and exclusion criteria—the process aimed to enhance the transparency and reproducibility of the review.

Title and Abstract Screening

Initially, the 299 unique articles were screened based on their titles and abstracts. Articles were included for full text review if they appeared to discuss teacher evaluation systems employing sentiment analysis and/or topic modeling on student feedback, and articles that provide insights into the implementation of recommender systems for enhancements in educational settings. Empirical studies that utilized qualitative or textual student feedback employing sentiment analysis and/or topic modeling were included. Titles and/or abstracts that employed literature reviews and analyzed non-English student feedback were discarded. This stage resulted in 155 articles selected for review.

Full-Text Review and Final Inclusion

The 155 articles underwent a thorough full-text review if the article was published in English and considered English comments, published in 2014 – 2024 in journals and conference papers with a Digital Object Identifier (DOI). Books, book chapters, dissertations, theses, and informal reports were excluded from this study. Purely theoretical papers, literature reviews, and comparisons of algorithms without empirical data were excluded. Papers that do not mention or utilize specific sentiment analysis or topic modeling algorithms were also discarded. Ultimately, 48 articles met all inclusion criteria and were included in this thematic literature review. This structured approach, while not a full PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) literature review, adopted its core principles of explicit search strategy, defined selection criteria, and quantitative reporting to ensure the rigor and transparency of the literature selection process.

2.4 Ethical Considerations

As this study is a literature review, it involves the analysis of publicly available scholarly works. Therefore, direct interaction with human participants and the collection of primary data were not conducted. Ethical considerations primarily revolve around academic integrity and proper attribution. All sources consulted are duly cited and referenced.

3.0 Results and Discussion

3.1 Sentiment Analysis

As established in the introduction, this thematic literature review is organized around three primary methodological approaches to sentiment analysis: machine learning, lexicon-based, and hybrid methods. The following sections provide a detailed summary and discussion of key studies within each category, particularly those from 2019–2023, to determine their applicability in a teacher recommender system.

Machine Learning Approaches

The review of machine learning algorithms for sentiment analysis reveals a diverse landscape of methods, from conventional models to deep learning architectures—these data-driven approaches train models to classify sentiment based on patterns learned from labeled datasets. Table 1 summarizes various studies in teacher evaluation feedback, employing both conventional and deep learning approaches. For clarity, the following standard acronyms are used: SVM, NB, RF: Random Forest; DT: Decision Trees; J48 DT: J48 Decision Tree; MNB: Multinomial Naïve Bayes; ME: Maximum Entropy; KNN: K-Nearest Neighbors; LR: Logistic Regression; ABSA: Aspect-Based Sentiment Analysis; CNB: Complement Naïve Bayes; PART: Partial Decision Trees; LSTM: Long Short-Term Memory; CNN: Convolutional Neural Network; GRU: Gated Recurrent Unit; RoBERT: Robustly optimized BERT; DL-RNN: Deep Learning Recurrent Neural Network; ANN: Artificial Neural Network; SGD: Stochastic Gradient Descent; and MLP: Multilayer Perceptron.

Table 1. *Summary of Studies Employing Machine Learning Approaches for Sentiment Analysis in Teacher Evaluation Feedback*

Author	ML Approaches	Findings	Recommendations
Conventional/Traditional Machine Learning Approaches			
Dake & Gyimah (2023)	SVM, J48, DT, NB, RF	SVM classification algorithms have the highest accurate detection of 92%	Improve accuracy by considering a large amount of data
Nawaz et al. (2022)	SVM, MNB, RF,	SVM yielded the best predictive accuracy of 50% macro-average across classes	Increase training data, and apply rigorous annotation of training data.
Abiodun Ayeni et al. (2020).	SVM	The system development and implementation were successful.	Expand the dataset; keep numbers and punctuations.
Omran et al. (2020)	MNB, ME, SVM	MNB performs best with an accuracy, recall, precision, and F1-score of 85%, 85%, 86%, and 87%, respectively.	Test on a larger corpus
Al Bashaireh et al. (2019).	SVM, NB, KNN, DT	SVM achieved the best accuracy of 92%.	Increase the dataset's data and add a grammar and spell checker.
Lalata et al. (2019).	NB, LR, DT, SVM, RF, ensemble (combination of the 5 ML algorithms)	The ensemble approach achieved the highest accuracy, 90.32%, and predictive accuracy, 93.80%.	Implement multi-label sentiment analysis.
Sivakumar & Reddy (2018).	ABSA tested in NB, CNB, Weka PART (Partial Decision Trees)	PART achieved 100% recall and F-measure for positive comment detection, and 99.4% recall and 99.7% F-score for negative comments.	Improve preprocessing.
Kumar & Jain (2016).	NB	NB achieved a maximum accuracy of 89.7% for unigrams in root forms.	Use advanced ML techniques on a dataset with domain-specific knowledge.
Borromeo & Toyama (2015)	Manual SA and Automatic SA API (based on NB);	Results of the automatic method were 64.59% similar to the results of the manual procedure (the manual process was assumed to be correct).	Increase the number of contributors to increase the accuracy of the sentiment analysis.
Altrabsheh et al. (2014)	NB, CNB, ME, SVM-Linear, SVM-Radial basis, SVM-polynomial.	When the neutral class is considered, SVM linear and CNB achieved 95% and 85.89% accuracy, respectively.	Use CNB for uneven training classes.
Deep Learning Machine Learning Approaches			
Lin et al. (2024)	BERT, NN	The tool's accuracy was 95.53% for positive comments, 88.93% for negative comments, and 43.33% for neutral comments, respectively.	Fine-tune the approach with a larger dataset.
Koufakou (2024)	NB. SVM, CNN, LSTM, BERT, RoBERT, XLNet	RoBERT achieved the highest polarity accuracy of 95.5%, then BERT and XLNet follow at 83%	Explore the use of additional pre-trained models like EduBERT.
Ahmed et al. (2023)	LSTM, MNB, RF	LSTM achieved a 95.75% F1 score over MNB and RF.	Improve accuracy on a large training dataset to avoid overfitting.
Bhowmik et al. (2023)	LSTM,\ CNN, GRU	LSTM achieves the highest F1 score of 86%	Delve deeper for contextual nuances.
Edalati et al. (2022)	RF, SVM, DT, 1D-CNN, BERT	RF achieved the highest F1 score of 99.43% in aspect sentiment classification.	Use text generation techniques to balance the dataset.
Onan (2020)	NB, SVM, LR, KNN, RF, Ensemble, DL-RNN	DL-RNN achieved the highest accuracy of 98.29%	N/A
Katragadda, S. et al. (2020)	NB, SVM, ANN	ANN hits 88.2% far from NB, SVM classification accuracy.	Test for various student profiles, age, and socioeconomic status.
Rakhmanov, O. (2020).	RF, ANN	ANN accuracy for the 3-class and 5-class datasets was 97% and 92%, respectively.	Test for bigrams and trigrams.
Sutoyo et al. (2021).	CNN	"CNN achieved accuracy, precision, recall, and F1-score of 87.95%, 87%, 78%, and 81%, respectively."	Test the model with a larger dataset to increase sentiment accuracy.
Ali Kandhro et al. (2019).	MNB, SGD, SVM, RF, Multilayer Perception	MNB and MLP achieved the 87% classification accuracy.	Work shall be extended to implement multilingual feedback.
Kandhro et al. (2019).	LSTM	LSTM achieved 98% and 98.5% precision and recall, respectively.	Extend to multilingual sentiment analysis.
Moreno-Marcos, P. M. et al. (2018)	LR, SVM, DT, RF, NB, Dictionaries of words, SentiWordNet	RF performed best between AUC values of 0.71 and 0.85 and with kappa values between 0.38 and 0.61.	Train the model on a broader variety of messages.

Conventional ML models, such as SVM, NB, and RF, are widely adopted due to their interpretability and relatively low computational cost. Studies by Dake & Gyimah (2023) and Al Bashaireh et al. (2019) Highlight SVM's high accuracy, often reaching or exceeding 92%. Similarly, Omran et al. (2020) Found that MNB performed best for their dataset. A key finding from Lalata et al. (2019) Is that an ensemble approach, combining multiple traditional ML algorithms, can achieve even higher accuracy (90.32%), suggesting that a single model may not always be sufficient. The primary limitation of these methods, as noted by several authors (e.g., Nawaz et al. (2022); Abiodun Ayeni et al. (2020)There is their dependence on large, rigorously annotated training datasets to improve accuracy and generalizability.

DL methods, a subfield or an advanced approach within the broader domain of machine learning, including LSTM, CNN, and BERT, have demonstrated a clear advantage in handling the complexity and nuances of natural language. Ahmed et al. (2023) and Bhowmik, Nur, et al. (2023) Showed that LSTM models consistently achieved high F1 scores (95.75% and 86%, respectively), outperforming traditional ML models like MNB and RF on the same datasets. The ability of DL models to capture contextual information is a significant strength, addressing the "contextual nuances" that traditional methods may miss. (Bhowmik, Nur, et al., 2023). Onan (2020) achieved an impressive 98.29% accuracy with a DL-RNN model, illustrating the potential of these advanced techniques with BERT models (Koufakou, 2024; Lin et al., 2025).The trade-off, however, is the increased computational resources required and the need for even larger datasets to prevent overfitting, a point emphasized by Ahmed et al. (2023) and Sutoyo et al. (2021). While deep learning methods offer a powerful way to analyze the complexities of language, their computational demands and reliance on large datasets can be a significant limitation. An alternative approach is the use of lexicon-based methods, which rely on pre-existing word dictionaries rather than trained models.

Lexicon-Based Approach

Unlike machine learning methods that require labeled training data, lexicon-based approaches classify text by relying on pre-existing dictionaries of words annotated with sentiment scores. These methods are valued for their simplicity, interpretability, and the fact that they do not necessitate extensive, domain-specific training data. Table 2 details various research efforts, highlighting the specific lexicons used, their key findings, and the recommendations made by the authors to enhance their performance within the context of teacher evaluation. The review focuses on popular lexicons such as VADER (Valence Aware Dictionary and sEntiment Reasoner), AFINN, Bing, NRC Emotion Lexicon, and MPQA (Multi-Perspective Question Answering), examining their effectiveness in capturing the nuanced sentiments expressed in student comments.

Lexical methods, such as VADER and SentiWordNet, rely on pre-defined dictionaries of words and their associated sentiment scores. This approach is straightforward, transparent, and does not require a training dataset. Faizi (2023) and Neumann & Linzmayer (2021) Showed that VADER can produce results very close to human annotation, with Faizi (2023) Demonstrating an improvement from 77.65% to 86.45% accuracy by integrating a domain-specific lexicon. The major drawback of this method is its limited ability to handle complex grammatical structures, irony, or sarcasm. As highlighted by H. Kim & Qin (2023)VADER's performance for a diverse dataset was only a moderate F1 micro-average of 57%. Its effectiveness is highly dependent on the quality and domain-specificity of the lexicon, leading to a common recommendation to develop "language-specific education-related dictionaries." (Fargues et al., 2023).

While individual machine learning and lexicon-based methods have their respective merits, a growing body of research has explored hybrid approaches to overcome their limitations. These models combine the strengths of two or more techniques, such as integrating machine learning algorithms with sentiment lexicons or deep learning models with aspect-based analysis.

Table 2. *Summary of Studies Employing Lexical-Based Approaches for Sentiment Analysis in Teacher Evaluation Feedback*

Author	Lexicon-Based Approach	Findings	Recommendation
Faizi, R. (2023)	VADER, VADER + education sentiment lexicon	Integration of the education lexicon with VADER improved the sentiment accuracy from 77.65% to 86.45%.	N/A
Fargues et al. (2023).	VADER lexicon, customized emotion dictionary	Implementation facilitates an automated student feedback analyzer (no accuracy mentioned)	Generate language-specific education-related dictionaries.
H. Kim & Qin (2023)	VADER	VADER is pretty good at analyzing students' sentiments with an F1 micro-average of 57% and an F1 macro 55%.	Improve VADER's performance for a larger dataset.
Neumann, M. & Linzmayer, R. (2021)	VADER	VADER produces sentiment scores very close to human annotators with a mean absolute difference of 0.95.	Fine-tune the VADER lexicon with larger populations.
Tzacheva & Easwaran (2021).	NRC Emotion lexicon	The trust emotion has increased from 2015 to 2020	Apply NRC to diverse groups.
Wook et al. (2020)	VADER	With analysis on capitalization and emojis, the teacher's performance was better analyzed alongside quantitative results.	Add more words to the lexicon to enhance the classification accuracy with a spell and grammar checker.
Praveenkumar et al. (2020).	Syuzhet, Bing, AFINN, NRC	All approaches calculated positive sentiments, and trust is the most highly mentioned word.	Include content and delivery to be added for evaluation of teaching, increasing the dataset.
Kaur et al. (2020).	Dictionaries of words, SentiWordNet, LR, SVM, DT, RF, NB	Using AUC (Area Under Curve), and kappa, dictionaries of words achieved a 0.78 and 0.54, higher degree of agreement, than SentiWordNet on the posts with both types of datasets.	Increase the number of raters to a broader range of messages.
Aung & Myo (2017)	AFINN, lexicon-based sentiment word database (manual annotation & dictionary-based)	There is a variation in the average polarity scores for AFINN and the proposed lexicon database.	N/A
Rajput et al. (2016).	Modified MPQA sentiment dictionary	Modified subjectivity sentiment scores are comparable with the Likert-based score.	Use a modified subjectivity corpus in the academic domain to achieve better results.

Hybrid Approaches

The fusion of ML and lexicon-based approaches allows for a more nuanced and accurate understanding of sentiment, particularly in complex or domain-specific texts like teacher evaluation feedback. Table 3 presents a summary of studies that have successfully employed hybrid approaches, detailing the specific combinations of methods, their superior performance metrics, and the recommendations for further enhancing these integrated systems.

Table 3. *Summary of Studies Employing Hybrid Approaches for Sentiment Analysis in Teacher Evaluation Feedback*

Author	Hybrid Approach	Findings	Recommendation
Melba Rosalind & Suguna (2022)	Proposed hybrid approach: Customized Sentiment lexicon with seeded LDA and ME	The proposed approach achieved an enhanced accuracy of 80.67% compared to existing models.	Aspects to be investigated using linguistic methods.
Qu & Zhang (2020)	ABSA & AFINN dictionary	Other data analysis can be conducted with the aspect sentiment scores.	Gather more data from students to develop a domain-specific lexicon.
Kaur et al. (2020)	NB, SVM, NRC	NB achieved a higher F score of 75% compared to SVM's 70%. NRC revealed positive feedback.	Expand to different subjects for a larger dataset.
Sindhu et al. (2019).	Online sentiment analyzer, SVM, NB, SVM Linear, Lexicon + NB, ABSA- LSTM	ABSA LSTM achieved 93%, the highest accuracy in sentiment detection.	Expand the model to Roman Urdu and other natural language comments.
Nasim et al. (2017)	RF and Lexicon (modified MPQA); SVM and Lexicon (modified MPQA) trained with various learning algorithms	RF with MPQA trained with TF-IDF achieved 93.4% accuracy and 92.6% F-measure compared to the SVM model training.	Analyze feedback at the aspect level of teachers' performance.
Borromeo & Toyama (2015)	Manual SA; Automatic SA API (based on NB); Crowdsourcing (paid vs. Volunteer)	Compared to manual SA, pay-based crowdsourcing SA achieved a 76.32% similarity index compared to the automatic SA's similarity of 64.59%.	Increase the number of contributors to increase the accuracy of the sentiment analysis.

The review of hybrid models presents a compelling argument for combining the strengths of different methodologies. These models, which integrate ML algorithms with lexical dictionaries or other linguistic features, consistently demonstrate superior performance. Nasim et al. (2017) Found that combining Random Forest with a modified MPQA lexicon yielded a high accuracy of 93.4%, outperforming a purely ML-based model. Similarly, Sindhu et al. (2019) Achieved the highest accuracy of 93% with an ABSA-LSTM model, which combines aspect-based sentiment analysis with a deep learning architecture. This fusion leverages the structured, rule-based nature of lexicons for a foundational sentiment score while allowing the ML or DL component to learn from the data and capture more complex patterns.

While hybrid and deep learning models demonstrate superior performance in controlled environments, the computational limitations and extensive data requirements they entail present a significant barrier to practical implementation. Therefore, for designing a robust recommender system with a focus on ease of implementation and resource efficiency, a lexicon-based approach, especially when enhanced with a domain-specific lexicon, is the most judicious choice. This method offers a strong balance between performance and practicality, providing nuanced and reliable sentiment analysis without the high computational cost. This literature analysis concludes that a lexicon-based approach is the most practical choice for sentiment analysis. However, a truly robust recommender system requires more than just sentiment scores; it needs to identify the specific areas of praise or criticism. Therefore, to complement sentiment analysis and provide actionable data for teachers, the subsequent section will delve into the utility of topic modeling for uncovering the key themes within student feedback.

3.2 Topic Modeling

TM is a key solution in educational contexts primarily due to its ability to efficiently analyze large volumes of unstructured textual data and extract meaningful insights (Hujala et al., 2020; Karunya et al., 2020; Sun & Yan, 2023). This thematic review of topic modeling literature examines various algorithms applied to student evaluations, providing the foundation for a data-driven recommender system. Table 4 summarizes the key findings from studies that have successfully used topic modeling to extract themes from student feedback.

The reviewed literature highlights the increasing application of topic modeling techniques to student feedback, with a primary focus on probabilistic and neural methods. Studies predominantly utilize Latent Dirichlet Allocation (LDA) to classify comments into thematic topics. For instance, Nawaz et al. (2022) found that LDA successfully identified salient features robust to noise, while Karunya et al. (2020) demonstrated that it can uncover significant information not captured by traditional Likert-scale feedback. However, a recurring recommendation across these studies is to apply LDA to larger datasets to improve its efficiency and performance (Nawaz et al., 2022; Sun & Yan, 2023). Clarizia et al. (2018) also suggest enhancing LDA by introducing annotated lexicons or filtering suggestions, pointing to the need for a more refined approach. While LDA is a popular choice, a crucial comparative study by Hayat et al. (2024) Reveals the emergence of more powerful techniques. Moving beyond the individual studies, a clear synthesis of the findings indicates that the BERT-based approach and NMF were more effective at identifying fine-grained, detailed topics than LDA and LSA. For a recommender system, this is a significant finding, as granular topics (e.g., "internet connection issues" vs. a general "technology" topic) are essential for providing actionable, specific recommendations to teachers.

Table 4. *Summary of Studies Employing Topic Modeling Approaches in Teacher Evaluation Feedback*

Author	Technique	Findings	Recommendations
Neural Topic Modeling			
Hayat et al. (2024)	BERT, LDA, LSA, NMF	BERT-based approach identified more fine-grained topics than LDA, LSA, and NMF.	Apply BERT to a larger dataset to confirm its effectiveness.
Kandhro et al. (2023).	LDA, LSA	LDA outperforms LSA for aspects extraction.	Explore the use of more powerful models like BERT to capture aspect information.
Probabilistic Topic Modeling Approaches			
Gencoglu et al. (2023).	Unsupervised LDA	LDA identified topics not detected by humans.	Apply LDA to a larger, different cultural dataset to improve and accommodate multilingual feedback.
Ishmael et al. (2023).	Unsupervised LDA	LDA correctly categorized 92% of the sentiments on topics of assessment delivery and language used.	Scale the model to larger datasets.
Kandhro et al. (2023).	LDA, LSA	LDA outperforms LSA for aspects	Explore the use of more

		extraction.	powerful models like BERT to capture aspect information.
Sun & Yan (2023).	Unsupervised LDA	The number of topics, 8, achieved a coherence score of 0.593.	Extend the use of LDA in larger datasets with longer phrases and more words.
Melba Rosalind & Suguna (2022)	Unsupervised LDA; Supervised LDA; customized sentiment lexicon; ME	Customized sentiment lexicon with supervised LDA method and ME achieved an enhanced accuracy of 80.67%.	Use supervised LDA over unsupervised LDA.
Nawaz et al. (2022).	Unsupervised LDA	Manual interpretation is needed to produce human-understandable issues and detect salient topics.	Increase LDA efficiency by testing a large dataset.
Hujala et al. (2020).	Unsupervised LDA	Combining LDA, thematic analysis, and multilevel regression analysis increases speed and depth of analysis.	Apply a method to accommodate multilingual massive feedback.
Karunya et al. (2020)	Unsupervised LDA	LDA identified unique and significant information not captured by Likert-scale feedback	Integrate LDA into a broader automated system.
Unankard & Nadee (2020).	Unsupervised LDA	Comments were classified into different topics.	Evaluate the performance of the analysis of more student course feedback.
Pyasi et al. (2019)	Semi-supervised LDA (manual labeling of topics generated by the LDA algorithm)	LDA provided multiple and more relevant topics for comments.	Explore methods to solve the grammatical structure of comments.
Clarizia et al. (2018).	Unsupervised LDA	LDA revealed topics that were more complex to students.	Introduce an annotated lexicon with the LDA approach.

3.3 Synthesis and Recommendations for a Recommender System

The thematic review reveals a clear trajectory in the application of topic modeling to student feedback. While traditional probabilistic methods like LDA are effective and robust, they often require manual interpretation and may lack the granularity needed for a sophisticated recommender system. In contrast, newer neural and algebraic methods, particularly the BERT-based approach and NMF, show great promise. Their ability to uncover more fine-grained and detailed topics, as demonstrated by Hayat et al. (2024) It is a critical advantage.

However, a crucial consideration for a practical recommender system is the trade-off between model performance and computational cost, especially in a multilingual context. Deep learning approaches in both sentiment analysis and topic modeling, while achieving the highest accuracy, often demand significant computational resources and longer processing times. (Andrewson et al., 2023; Nawaz et al., 2022; P. M. Moreno-Marcos et al., 2018; Ren et al., 2023). For a system designed to provide immediate, actionable insights to teachers, the development and deployment of these complex models may not be a practical or cost-effective solution due to the persistent need for human interpretation validation and the nuances of the qualitative feedback. (Gencoglu et al., 2023; Sindhu et al., 2019; Sun & Yan, 2023; Unankard & Nadee, 2020).

Given these practical constraints, a more balanced and feasible approach is recommended. This review recommends a two-pronged strategy for the design of the teacher recommender system: (a) For **Sentiment Analysis**: A lexical-based approach is highly suitable for its easy deployment mechanisms, cost-effectiveness, and less data training requirement (Gunasekaran, 2023; Nandakumar et al., 2022; Ren et al., 2023) . The use of an approach like VADER combined with a customized, domain-specific lexicon (such as a Filipino lexicon) offers a robust balance of high accuracy, interpretability, and computational efficiency. (Hixson, 2019). This hybrid method leverages the strengths of rule-based systems while adapting to the unique linguistic nuances. (Nasim et al., 2017; Qu & Zhang, 2020; Sunar & Khalid, 2024), (b) For **Topic Modeling**: A refined probabilistic model like LDA or a semi-supervised variant is the most pragmatic choice. This approach provides a balance between thematic granularity (Ishmael et al., 2023; Nawaz et al., 2022; Pyasi et al., 2018; Sun & Yan, 2023) and computational feasibility (Gottipati et al., 2018; Hujala et al., 2020; Karunya et al., 2020), making it a viable option for an immediate and deployable system.

This recommendation acknowledges that while cutting-edge models represent the future, the most appropriate choice for an immediate, deployable system is one that balances high performance with practical considerations of computational cost and development complexity. This choice would ensure that the system can generate timely, actionable, and data-driven recommendations, directly addressing the need for continuous improvement in teacher performance.

4.0 Conclusion

Based on the review, the SA lexicon-based approach with the VADER lexicon is more appropriate for sentiment analysis to determine the polarity of comments. A domain-specific corpus of words, manually labelled, can be added to the VADER lexicon, thus increasing the accuracy and polarity detection of the system. Since students use different languages or even their dialects to express their feelings and provide feedback to their teacher, the VADER lexicon is the appropriate technique for sentiment analysis. For the topic modeling approach, the LDA approach with human intervention is widely used in topic modeling to identify topics students comment on in teacher evaluations (Nawaz et al., 2022; Pyasi et al., 2018). With the language nuances that are present in the Filipino language, human intervention is an everyday activity for both VADER and LDA. With these results, the development of TeachAIRs, a teacher's recommender system, is highly recommended.

The practical implications of such a system extend beyond individual teachers to impact broader institutional policies. For school administrators, TeachAIRs can serve as a powerful tool for strategic decision-making. By aggregating the topic modeling and sentiment analysis data across departments in higher education, administrators can identify institution-wide trends in teaching effectiveness. This data could inform professional development programs, allocate resources more effectively, and highlight areas where curriculum or pedagogical support is most needed (Nandakumar et al., 2022; Ren et al., 2023; Sunar & Khalid, 2024; Zayed Alatwah, 2022). For policy frameworks, the system provides a data-driven basis for a more equitable and effective teacher evaluation process, moving beyond subjective ratings to evidence-based insights that can be used to set clear, objective standards for teaching excellence and inform ongoing quality assurance initiatives (Rakhmanov, 2020; Zeng et al., 2023).

With this holistic and rigorous approach to understanding students' comments in teacher evaluation, the Students' voices are heard through their comments that will drive the teachers to take data-driven actions. Thus, not only will the proposed recommender system enhance teachers' performance, but it will also significantly improve student outcomes and experiences. Future research could address the current gap in the literature by exploring the use of less-common topic modeling approaches, such as fuzzy or neural methods, on teacher evaluation data. Investigating whether these techniques can uncover different thematic structures or offer greater precision in a multilingual context would be a valuable contribution to the field.

5.0 Contribution of Authors

Ortiz, Marie Grace – conceptualization, writing, and submission.
Dumlao, Mencita – writing, editing.

6.0 Funding

No funding agency funded this study.

7.0 Conflict of Interest

The authors declare no conflict of interest.

8.0 Acknowledgment

We, the authors, are grateful to all contributors to this paper and their respective families.

9.0 References

Abiodun Ayeni, O., Mercy, A., et al (2020). Web-based student opinion mining system using sentiment analysis. International Journal of Information Engineering and Electronic Business, 12(5), 33–46. https://doi.org/10.5815/ijieeb.2020.05.04

Adinolfi, P., D'Avanzo, E., Lytras, M. D., Novo-Corti, I., & Picatoste, J. (2016). Sentiment analysis to evaluate teaching performance. International Journal of Knowledge Society Research, 7(4), 86–107. https://doi.org/10.4018/ijksr.2016100108

Ahmed, N., Khouro, M. A., Khan, A., Dawood, M., Dootio, M. A., & Jan, N. U. (2023). Student textual feedback sentiment analysis using machine learning techniques to improve the quality of education. Pakistan Journal of Engineering, Technology & Science, 11(2), 32–40. https://doi.org/10.22555/pjets.v11i2.1039

Al Bashaireh, R., Sabeeh, V., & Zohdy, M. (2019). Towards a new indicator for evaluating universities based on twitter sentiment analysis. Proceedings - 6th Annual Conference on Computational Science and Computational Intelligence, CSCI 2019, 1398–1404. https://doi.org/10.1109/CSCI49370.2019.00261

Ali Kandhro, I., Ameen Chhajro, M., Kumar, K., Lashari, H. N., & Khan, U. (2019). Student feedback sentiment analysis model using various machine learning schemes: A review. Indian Journal of Science and Technology, 14(12), 1–9. https://doi.org/10.17485/ijst/2019/v12i14/143243

Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014). Sentiment analysis: Towards a tool for analysing real-time students feedback. Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, 2014-December, 419–423. https://doi.org/10.1109/ICTAI.2014.70

Andrewson, S., Mason, J., & Joel, R. (2023). Leveraging NLP for personalized learning: Adaptive feedback systems in higher education. https://www.researchgate.net/publication/391319413

Aung, K. Z., & Myo, N. N. (2017). Sentiment analysis of students' comment using lexicon based approach. Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017. https://doi.org/10.1109/ICIS.2017.7959985

Balahadia, F. F., Fernando, M. C. G., & Juanatas, I. C. (2016). Teacher's performance evaluation tool using opinion mining with sentiment analysis. Proceedings - 2016 IEEE Region 10 Symposium, TENSYMP 2016, 95–98. https://doi.org/10.1109/TENCONSpring.2016.7519384

Bhana, V. M. (2014). Interpersonal skills development in generation y student nurses: A literature review. In Nurse Education Today (Vol. 34, Issue 12, pp. 1430–1434). Churchill Livingstone. https://doi.org/10.1016/j.nedt.2014.05.002

Bhowmik, A., Mohd Noor, N., Saef Ullah Miah, M., Mazid-Ul-Haque, M., & Karmaker, D. (2023). A comprehensive data set for aspect-based sentiment analysis in evaluating teacher performance. AIUB Journal of Science and Engineering, 22(2), 200–213. https://doi.org/10.53799/AJSE.V22I2.862

Bhowmik, A., Nur, N. M., Saef, M., Miah, U., & Karmekar, D. (2023). Aspect-based sentiment analysis model for evaluating teachers' performance from students' feedback. AIUB Journal of Science and Engineering, 22(3), 287–294. https://doi.org/10.53799/ajse.v22i3.921

Borromeo, R. M., & Toyama, M. (2015). Automatic vs. crowdsourced sentiment analysis. ACM International Conference Proceeding Series, 0(CONFCODENUMBER), 90–95. https://doi.org/10.1145/2790755.2790761

Chakravarthy, V. J., Kameswari, M., Mydeen, H. D., & Seenivasan, M. (2021). Opinion mining from student text review for choosing better online courses. IOP Conference Series: Materials Science and Engineering, 1070(1), 012067. https://doi.org/10.1088/1757-899x/1070/1/012067

Clarizia, F., Colace, F., De Santo, M., Lombardi, M., Pascale, F., & Pietrosanto, A. (2018). E-learning and sentiment analysis: A case study. ACM International Conference Proceeding Series, 111–118. https://doi.org/10.1145/3178158.3178181

Creswell, J. W., & David Creswell, J. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Sage Publications, Inc. https://tinyurl.com/mryt9hvm

Dake, D. K., & Gyimah, E. (2023). Using sentiment analysis to evaluate qualitative students' responses. Education and Information Technologies, 28(4), 4629–4647. https://doi.org/10.1007/s10639-022-11349-1

Das, S., Roy, S., Bose, R., Acharjya, P. P., & Mondal, H. (2022). Analysis of student sentiment dynamics to evaluate teachers performance in online course using machine learning. 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 668–673. https://doi.org/10.1109/ICAAIC53929.2022.9792958

Edalati, M., Imran, A. S., Kastrati, Z., & Daudpota, S. M. (2022). The potential of machine learning algorithms for sentiment classification of students' feedback on MOOCs. Lecture Notes in Networks and Systems, 296, 11–22. https://doi.org/10.1007/978-3-030-82199-9_2

Faizi, R. (2023). Using sentiment analysis to explore student feedback: A lexical approach. International Journal of Emerging Technologies in Learning, 18(9), 259–267. https://doi.org/10.3991/ijet.v18i09.38101

Fargues, M., Kadry, S., Lawal, I. A., Yassine, S., & Rauf, H. T. (2023). Automated analysis of open-ended students' feedback using sentiment, emotion, and cognition classifications. Applied Sciences (Switzerland), 13(4). https://doi.org/10.3390/app13042061

Fernández, M. P., & Martínez, J. F. (2022). Evaluating teacher performance and teaching effectiveness: Conceptual and methodological considerations. https://doi.org/10.1007/978-3-031-13639-9_3

Gencoglu, B., Helms-Lorenz, M., Maulana, R., Jansen, E. P. W. A., & Gencoglu, O. (2023). Machine and expert judgments of student perceptions of teaching behavior in secondary education: Added value of topic modeling with big data. Computers and Education, 193. https://doi.org/10.1016/j.compedu.2022.104682

Gottipati, S., Shankararaman, V., & Lin, J. R. (2018). Latent dirichlet allocation for textual student feedback analysis. ICCE 2018 - 26th International Conference on Computers in Education, Main Conference Proceedings, 220–227. https://ink.library.smu.edu.sg/sis_research/4215

Gunasekaran, K. P. (2023). Exploring sentiment analysis techniques in natural language processing: A comprehensive review. https://doi.org/10.48550/arXiv.2305.14842

Hariyani, C. A., Hidayanto, A. N., Fitriah, N., Abidin, Z., & Wati, T. (2019). Mining student feedback to improve the quality of higher education through multi-label classification, sentiment analysis, and trend topic. 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2019. https://doi.org/10.1109/ICITISEE48480.2019.9003818

Hashim, S., Omar, M. K., Jalil, H. A., & Sharef, N. M. (2022). Trends on technologies and artificial intelligence in education for personalized learning: Systematic literature review. International Journal of Academic Research in Progressive Education and Development, 11(1). https://doi.org/10.6007/IJARPED/v11-i1/12230

Hayat, F., Shatnawi, S., & Haig, E. (2024). Comparative analysis of topic modelling approaches on student feedback. International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K - Proceedings, 1, 226–233. https://doi.org/10.5220/0012890400003838

Hixson, T. (2019). Reactions vs. reality: Using sentiment analysis to measure university students' responses to learning ArcGIS. Journal of Map and Geography Libraries, 15(2–3), 263–276. https://doi.org/10.1080/15420353.2020.1719266

Hujala, M., Knutas, A., Hynninen, T., & Arminen, H. (2020). Improving the quality of teaching by utilizing written student feedback: A streamlined process. Computers & Education, 157, 103965. https://doi.org/10.1016/J.COMPEDU.2020.103965

Ishmael, O., Kiely, E., Quigley, C., & McGinty, D. (2023). Topic modelling using latent dirichlet allocation (LDA) and analysis of students sentiments. 2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE), 1–6. https://doi.org/10.1109/JCSSE58229.2023.10201965

Kandhro, I. A., Wagan, A. A., Kumar, K., & Shaikh, Z. U. (2023). An efficient LSTM based cross domain aspect based sentiment analysis (CD-ABSA). Mehran University Research Journal of Engineering and Technology, 42(3), 89. https://doi.org/10.22581/muet1982.2303.10

Kandhro, I. A., Wasi, S., Kumar, K., Rind, M., & Ameen, M. (2019). Sentiment analysis of students comment by using long-short term model. Indian Journal of Science and Technology, 12(8), 1–16. https://doi.org/10.17485/ijst/2019/v12i8/141741

Karunya, K., Aarthy, S., Karthika, R., & Jegatha Deborah, L. (2020). Analysis of student feedback and recommendation to tutors. Proceedings of the 2020 IEEE International Conference on Communication and Signal Processing, ICCSP 2020. https://doi.org/10.1109/ICCSP48568.2020.9182270

Kaur, W., Balakrishnan, V., & Singh, B. (2020). Improving teaching and learning experience in engineering education using sentiment analysis techniques. IOP Conference Series: Materials Science and Engineering, 834(1). https://doi.org/10.1088/1757-899X/834/1/012026

Kim, C. M., & Kwak, E. C. (2022). An exploration of a reflective evaluation tool for the teaching competency of pre-service physical education teachers in Korea. Sustainability (Switzerland), 14(13). https://doi.org/10.3390/su14138195

Kim, H., & Qin, G. (2023). Summarizing students' free responses for an introductory algebra-based physics course survey using cluster and sentiment analysis. IEEE Access, 11, 89052–89066. https://doi.org/10.1109/ACCESS.2023.3305260

Koufakou, A. (2024). Deep learning for opinion mining and topic classification of course reviews. Education and Information Technologies, 29(3), 2973–2997. https://doi.org/10.1007/s10639-023-117362

Kumar, A., & Jain, R. (2016). Sentiment analysis and feedback evaluation. Proceedings of the 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education, MITE 2015. https://doi.org/10.1109/MITE.2015.7375359

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

Latika Tamrakar, M., Shrivastava, P., & Ghosh, S. M. (2021). An analytical study of feature extraction techniques for student sentiment analysis. In Turkish Journal of Computer and Mathematics Education (Vol. 12, Issue 11). https://doi.org/10.1007/978-3-030-80216-5_20

Lin, F., Li, C., Lim, R. W. Y., & Lee, Y. H. (2025). Empower instructors with actionable insights: Mine and visualize student written feedback for instructors' reflection. Computers and Education: Artificial Intelligence, 8. https://doi.org/10.1016/j.caeai.2025.100389

Lin, F., Li, C., Wei, R., Lim, Y., Yew, & Lee, H. (2024). Developing a feedback analytic tool to support instructor reflection. https://doi.org/10.58459/icce.2024.4859

Looney, J. (2011). Developing high-quality teachers: Teacher evaluation for improvemente jed_1492 440..455. http://dx.doi.org/10.2307/41343393

Malebary, S. J., & Abulfaraj, A. W. (2024). A stacking ensemble based on lexicon and machine learning methods for the sentiment analysis of tweets. Mathematics, 12(21). https://doi.org/10.3390/math12213405

Mamidted, A. D., & Maulana, S. S. (2023). The teaching performance of the teachers in online classes: A sentiment analysis of the students in a state university in the Philippines. Randwick International of Education and Linguistics Science Journal, 4(1), 86–95. https://doi.org/10.47175/rielsj.v4i1.639

Melba Rosalind, J., & Suguna, S. (2022). Predicting students' satisfaction towards online courses using aspect-based sentiment analysis. 20–35. https://doi.org/10.1007/978-3-031-11633-9_3ï

Nandakumar, R., Pallavi, M. S., Pramath, P. H., & Hegde, V. (2022). Sentimental analysis on student feedback using NLP & POS tagging. International Conference on Edge Computing and Applications, ICECAA 2022 - Proceedings, 309–313. https://doi.org/10.1109/ICECAA55415.2022.9936569

Nasim, Z., Rajput, Q., & Haider, S. (2017). Sentiment analysis of student feedback using machine learning and lexicon based approaches. International Conference on Research and Innovation in Information Systems, ICRIIS. https://doi.org/10.1109/ICRIIS.2017.8002475

Nawaz, R., Sun, Q., Shardlow, M., Kontonatsios, G., Aljohani, N. R., Visvizi, A., & Hassan, S. U. (2022). Leveraging AI and machine learning for national student survey: Actionable insights from textual feedback to enhance quality of teaching and learning in UK's higher education. Applied Sciences (Switzerland), 12(1). https://doi.org/10.3390/app12010514

Neumann, M., & Linzmayer, R. (2021). Capturing student feedback and emotions in large computing courses: A sentiment analysis approach. SIGCSE 2021 - Proceedings of the 52nd ACM Technical Symposium on Computer Science Education, 541–547. https://doi.org/10.1145/3408877.3432403

Omran, T., Sharef, B. T., Hadjar, K., & Subramanian, S. (2020). Machine learning for improving teaching methods through sentiment analysis. Applied Mathematics and Information Sciences, 14(2), 309–317. https://doi.org/10.18576/amis/140215

Onan, A. (2020). Mining opinions from instructor evaluation reviews: A deep learning approach. Computer Applications in Engineering Education, 28(1), 117–138. https://doi.org/10.1002/cae.22179

Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino,P. J., Estévez-Ayres, I., & Kloos, C. D. (2018). Sentiment analysis in MOOCs: A case study. 2018 IEEE Global Engineering Education Conference (EDUCON), 1489–1496. https://doi.org/doi:10.1109/EDUCON.2018.8363409.

Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. In Harvard Educational Review (Vol. 82, Issue 1). https://doi.org/10.17763/haer.82.1.v40p0833345w6384

Peña-Torres, J. A. (2024). Towards an improved of teaching practice using sentiment analysis in student evaluation. Ingeniería y Competitividad, 26(2). https://doi.org/10.25100/iyc.v26i2.13759

Pramod, D., Vijayakumar Bharathi, S., & Raman, R. (2022). Faculty effectiveness prediction using machine learning and text analytics. 2022 IEEE Technology and Engineering Management

Conference (TEMSCON EUROPE), 40–47. https://doi.org/10.1109/TEMSCONEUROPE54743.2022.9801997
Praveenkumar, T., Manorselvi, A., & Soundarapandiyan, K. (2020). Exploring the students feelings and emotion towards online teaching: Sentimental analysis approach. 137–146. https://doi.org/10.1007/978-3
Pyasi, S., Gottipati, S., & Shankararaman, V. (2018). SUFAT - An analytics tool for gaining insights from student feedback comments. 1–9. https://doi.org/10.1109/FIE.2018.8658457
Pyasi, S., Gottipati, S., & Shankararaman, V. (2019). SUFAT - An analytics tool for gaining insights from student feedback comments. Proceedings - Frontiers in Education Conference, FIE, 2018-October. https://doi.org/10.1109/FIE.2018.8658457
Qu, W., & Zhang, Z. (2020). An application of aspect-based sentiment analysis on teaching evaluation. 89–104. https://doi.org/10.35566/isdsa2019c6
Rafiq, S., Afzal, A., & Kamran, F. (2022). Exploring the problems in teacher evaluation process and its perceived impact on teacher performance. Gomal University Journal of Research, 38(04), 482–500. https://doi.org/10.51380/gujr-38-04-08
Rajput, Q., Haider, S., & Ghani, S. (2016). Lexicon-based sentiment analysis of teachers' evaluation. Applied Computational Intelligence and Soft Computing, 2016, 1–12. https://doi.org/10.1155/2016/2385429
Rakhmanov, O. (2020). A comparative study on vectorization and classification techniques in sentiment analysis to classify student-lecturer comments. Procedia Computer Science, 178, 194–204. https://doi.org/10.1016/j.procs.2020.11.021
Ren, P., Yang, L., & Luo, F. (2023). Automatic scoring of student feedback for teaching evaluation based on aspect-level sentiment analysis. Education and Information Technologies, 28(1). https://doi.org/10.1007/s10639-022-11151-z
Katragadda, S., Ravi, V., Kumar, P., & Lakshmi, G. J. (2020). Performance analysis on student feedback using machine learning algorithms. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 1161–1163. https://doi.org/doi:10.1109/ICACCS48705.2020.9074334
Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. International Journal of Educational Technology in Higher Education, 18(1). https://doi.org/doi.org/10.1186/s41239-021-00292-9
Sindhu, I., Muhammad Daudpota, S., Badar, K., Bakhtyar, M., Baber, J., & Nurunnabi, M. (2019). Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation. IEEE Access, 7, 108729–108741. https://doi.org/10.1109/ACCESS.2019.2928872
Sivakumar, M., & Reddy, U. S. (2018). Aspect based sentiment analysis of students opinion using machine learning techniques. Proceedings of the International Conference on Inventive Computing and Informatics, ICICI 2017, 726–731. https://doi.org/10.1109/ICICI.2017.8365231
Sun, J., & Yan, L. (2023). Using topic modeling to understand comments in student evaluations of teaching. Discover Education, 2(1). https://doi.org/10.1007/s44217-023-00051-0
Sunar, A. S., & Khalid, M. S. (2024). Natural language processing of student's feedback to instructors: A systematic review. IEEE Transactions on Learning Technologies, 17, 741–753. https://doi.org/10.1109/TLT.2023.3330531
Sutoyo, E., Almaarif, A., & Yanto, I. T. R. (2021). Sentiment analysis of student evaluations of teaching using deep learning approach. Lecture Notes in Networks and Systems, 254, 272–281. https://doi.org/10.1007/978-3-030-80216-5_20
Tian, X., Tang, S., Zhu, H., & Xia, D. (2022). Real-time sentiment analysis of students based on mini-xception architecture for wisdom classroom. Concurrency and Computation: Practice and Experience, 34(21). https://doi.org/https://doi.org/10.1002/cpe.7059
Tzacheva, A., & Easwaran, A. (2021). Emotion detection and opinion mining from student comments for teaching innovation assessment. International Journal of Education (IJE), 09(02), 21–32. https://doi.org/10.5121/IJE2021.9203
Unankard, S., & Nadee, W. (2020). Topic detection for online course feedback using LDA. Emerging Technologies for Education. SETE 2019. Lecture Notes in Computer Science, 11984, 133–144. https://doi.org/10.1007/978-3-030-38778-5_16
Wang, J. (2025). The impact of AI teaching on teaching quality. International Journal of Web-Based Learning and Teaching Technologies, 20(1), 1–22. https://doi.org/10.4018/IJWLTT.376489
Wook, M., Razali, N. A. M., Ramli, S., Wahab, N. A., Hasbullah, N. A., Zainudin, N. M., & Talib, M. L. (2020). Opinion mining technique for developing student feedback analysis system using lexicon-based approach (OMFeedback). Education and Information Technologies, 25(4), 2549–2560. https://doi.org/10.1007/s10639-019-10073-7
Zayed Alatwah, S. (2022). Using sentiment analysis to evaluate first-year engineering students teamwork textual feedback. 2022 ASEE Annual Conference & Exposition. https://doi.org/10.18260/1-2--41460
Zeng, J., Luo, K., Lu, Y., & Wang, M. (2023). An evaluation framework for online courses based on sentiment analysis using machine learning. International Journal of Emerging Technologies in Learning, 18(18), 4–22. https://doi.org/10.3991/ijet.v18i18.42521