International Journal of Research in Engineering, Science and Management
Volume 8, Issue 2, February 2025
https://www.ijresm.com | ISSN (Online): 2581-5792

Sentiment Analysis in EdTech: A Study on Course Review Feedback Using NLP

Vaibhav Mahale[1*], Ronak Matolia[2], Deep Mehta[3], Anand Godbole[4]
[1,2,3]*Student, Department of Computer Engineering, Sardar Patel Institute of Technology, Mumbai, India*
[4]*Professor, Department of Computer Engineering, Sardar Patel Institute of Technology, Mumbai, India*

***Abstract*: This paper presents an EdTech platform prototype that enables users to create, edit, and consume educational content while providing sentiment analysis of course reviews. By employing Natural Language Processing (NLP) techniques, the system categorizes reviews into positive, negative, and neutral sentiments to offer educators actionable insights on course quality. We implement various NLP models, including traditional machine learning and transformer-based models, to analyze student feedback effectively. Results demonstrate that this system can accurately identify sentiment trends, supporting continuous course improvements and enhancing the educational experience.**

***Keywords*: Edtech, NLP, sentiments, models.**

1. Introduction

Background - Student feedback is an essential component of the educational process, offering insights into how effectively courses are meeting learners' needs and expectations. As education increasingly shifts to online platforms, the volume of student feedback has grown substantially, especially on EdTech platforms that facilitate interactive learning experiences. These platforms provide students with the opportunity to share their opinions on various aspects of a course, such as content quality, teaching style, and overall course structure. This feedback can be invaluable for educators aiming to adapt and improve their teaching methods and for institutions striving to maintain high educational standards.

However, the sheer volume of feedback generated on EdTech platforms presents challenges in timely and effective analysis. Manually reviewing and interpreting thousands of reviews is labor-intensive and time-consuming, making it difficult to identify patterns and respond to common concerns promptly. Automated sentiment analysis using Natural Language Processing (NLP) can streamline this process by analyzing large volumes of text-based feedback, categorizing sentiments, and revealing insights that might otherwise be overlooked. This capability allows educators and administrators to make data-informed adjustments to courses, improving the learning experience for students.

Problem Statement: Despite the availability of substantial feedback on EdTech platforms, there remains a gap in systematically analyzing this information for actionable insights. Traditional methods of interpreting feedback are not scalable in a digital learning environment with diverse and extensive data inputs. Without effective sentiment analysis, valuable patterns in student feedback may go unnoticed, and potential areas for course improvement may not be addressed. This gap limits the ability of educational institutions and EdTech platforms to leverage feedback for enhancing course quality and aligning content with student needs.

Objective: This paper presents a sentiment analysis solution that applies NLP techniques to educational reviews, with the goal of helping educators and administrators better understand student feedback. By implementing an automated system for analyzing sentiments within course reviews, this research aims to identify trends and areas for improvement, enabling educators to make informed decisions that enhance course content and delivery.

2. Related Work

Sentiment analysis in educational contexts is a growing field that leverages NLP to process student feedback, allowing educators and administrators to gain insights into student experiences and improve course quality. With the shift to online learning, EdTech platforms have become crucial in gathering large volumes of student feedback, which can be effectively analyzed through sentiment analysis.

Applications of Sentiment Analysis in Education: Several studies highlight the potential of sentiment analysis for understanding student sentiment and guiding educational improvements. Altrabsheh et al. (2014) applied sentiment analysis to real-time classroom interactions and online discussions, demonstrating that timely analysis of student feedback helps educators make proactive adjustments during class to enhance engagement [1]. Nassif et al. (2016) focused on MOOCs (Massive Open Online Courses), analyzing feedback to identify emotional patterns that influence student retention and satisfaction, which in turn guides course design for large-scale online learning environments [2]. Ochoa and Romero (2013) expanded sentiment analysis to social media, capturing student interactions to analyze informal learning patterns and gauge overall engagement [3]. These studies underscore sentiment analysis as a valuable tool for capturing student perceptions and improving teaching strategies in online

*Corresponding author: vaibhav.mahale@spit.ac.in

and traditional educational settings.

NLP Models for Sentiment Analysis in EdTech - Various machine learning models have been used for sentiment analysis in educational contexts, each with distinct advantages. Na¨ıve Bayes and Support Vector Machines (SVM) are commonly used for their simplicity and efficiency. Hutto and Gilbert (2014) demonstrated the efficacy of VADER, a lexicon-based model that assigns sentiment scores to social media and educational text with minimal computational resources, making it suitable for EdTech platforms with limited infrastructure [4]. However, while traditional models like Na¨ıve Bayes and SVM offer good performance in binary sentiment classification, they often struggle with nuanced feedback where context plays a significant role.

Deep learning models, particularly recurrent neural networks (RNNs) and transformer-based models, have been shown to handle complex, context-rich data more effectively. Huang et al. (2020) used Long Short-Term Memory (LSTM) networks for sentiment analysis in discussion forums, achieving better results than traditional models by capturing the sequential nature of text and understanding context within sentences [5]. Rios et al. (2018) applied Bidirectional Encoder Representations from Transformers (BERT) to analyze student feedback, showing that BERT's contextualized embeddings allow for more accurate sentiment classification in cases where the sentiment depends on sentence structure and context [6]. These studies reveal the effectiveness of deep learning and transformer-based models in educational sentiment analysis, particularly when the feedback is detailed and context-sensitive.

Challenges in Sentiment Analysis of Educational Feedback: Despite advancements, educational sentiment analysis faces unique challenges. Sarcasm detection is a significant issue, as students often use sarcasm or humor in feedback, which can lead to misinterpretation by sentiment classifiers. Bouazizi and Ohtsuki (2016) addressed this by integrating sarcasm detection features into traditional sentiment analysis models, enhancing accuracy for social media and educational contexts [7]. Handling sarcasm is particularly relevant in education, where students may use humor to express frustration, which if misinterpreted, could skew sentiment results.

Another challenge is the variation in student language, which includes colloquial expressions, abbreviations, and domain-specific terminology. This variability makes it difficult for traditional sentiment analysis models to accurately interpret feedback without fine-tuning. Jain et al. (2017) emphasized the need for domain adaptation, noting that fine-tuning models like BERT on educational data can capture unique linguistic patterns more effectively, thus improving sentiment classification accuracy [8]. This approach aligns with Medhat et al. (2014), who highlight the importance of customizing sentiment lexicons and adapting models to specific domains to ensure accurate classification in text-rich fields like education [11].

Multilingual Sentiment Analysis in EdTech - With global EdTech platforms catering to diverse linguistic groups, multilingual sentiment analysis is also a relevant area. Balahur and Turchi (2014) explored sentiment analysis using machine translation to analyze feedback in multiple languages, finding that multilingual models help EdTech platforms gain insights across language barriers [9]. Such models, however, require complex architectures and may face limitations in accuracy due to translation nuances. Advances in multilingual models, such as Multilingual BERT, offer promise for handling non- English educational feedback, broadening the applicability of sentiment analysis in global EdTech settings.

Summary of Techniques and Findings: Overall, the existing body of research underscores the effectiveness of sentiment analysis in EdTech, especially when using advanced NLP models like LSTM and BERT. These models are capable of handling the complex linguistic patterns and context-dependent nature of educational feedback. However, challenges remain in detecting sarcasm, accommodating language variability, and scaling models for multilingual data. By addressing these challenges, sentiment analysis in education can provide actionable insights, enabling educators to tailor their courses to student needs and improving the overall learning experience on EdTech platforms.

3. Framework

The proposed framework of the research work is conducted in three different modules.

A. Dataset and its Features

The first module includes data collection and pre- processing of data. A sample of online reviews is collected from our edtech platform. It includes six features as explained in table 1.

B. Proposed Approach

The proposed sentiment analysis framework for course reviews on our educational platform is a comprehensive approach designed to provide instructors with valuable feedback by analyzing student reviews. The process begins with data collection, where reviews, including fields like user ID, course ID, review text, and ratings, are gathered from the platform's feedback system. The collected data undergoes preprocessing, which involves converting text to lowercase, removing punctuation and special characters, eliminating stopwords, and tokenizing the text into individual words.

For feature extraction, we use the Term Frequency-Inverse Document Frequency (TF-IDF) technique to identify key phrases, assigning weights to terms based on their relevance in each review. Sentiment analysis is performed using two methods: the VADER algorithm, a rule-based tool designed for straightforward sentiment classification, and BERT, a transformer-based model capable of capturing complex language nuances. Each review is classified as positive, negative, or neutral based on sentiment scores from both models, enabling a comparative analysis of performance and accuracy. Reviews are then annotated with sentiment tags and key phrases to highlight strengths and areas for improvement. The results are presented in an interactive dashboard, allowing instructors to view sentiment distribution, analyze key phrases, and gain actionable insights to enhance their courses.

Table 1
Features included in the dataset

Feature	Description
Course Name	Name of course
Instructor Name	Name of teacher who published the course
Price	Price of the course in Rupees
Rating	User rating between 1 to 5
Reviews	User reviews provided for every course
Review Votes	Number of people who found this review helpful

1) Data Collection

The dataset for sentiment analysis consists of student reviews for different courses on the platform. Each dataset includes the following key fields: user ID, course ID, review text, and rating. These reviews are collected through the platform's feedback system and stored in a database.

2) Data Preprocessing

To ensure the quality of the data for sentiment analysis, we perform several text preprocessing steps on the collected reviews. The preprocessing pipeline follows these major steps:

Lowercasing: The review text is converted to lowercase to maintain uniformity and avoid case sensitivity.

Removing Punctuation and Special Characters: Punctuation marks, digits, and special symbols are removed from the reviews to focus solely on meaningful text content.

Stopword Removal: Common stopwords (e.g., "and", "the", "is") that do not carry significant meaning are removed using the stopword library to reduce the noise in the data.

Tokenization: The cleaned text is tokenized, breaking the review into individual words for further analysis.

3) Feature Extraction Using TF-IDF

The Term Frequency-Inverse Document Frequency (TF-IDF) method is employed to extract significant terms from each student review. This approach allows us to isolate words that are relevant to the unique content of a review while downplaying words that are commonly found across multiple reviews. By identifying and weighting these unique keywords, we can provide instructors with targeted insights into specific areas that may require attention or improvement.

TF-IDF Works by Combining Two Main Components—Term Frequency (TF) and Inverse Document Frequency (IDF)—as outlined below:

Term Frequency (TF): For each review, we begin by calculating the frequency of every word within the document. Term frequency measures how often each word appears in an individual review, emphasizing words that may be important within that specific feedback. Higher frequency terms in a review may represent central themes or areas of emphasis, guiding instructors on what aspects students focus on in their feedback.

Inverse Document Frequency (IDF): To determine how unique or common each word is across all reviews in the dataset, we calculate its inverse document frequency. IDF assigns a lower weight to words that are widespread throughout the entire set of reviews (e.g., general words like "good" or "course"), reducing their impact. Conversely, less common words that appear in fewer reviews are weighted more heavily, as they may highlight specific aspects of a course that are noteworthy or unique to individual student experiences. By doing so, IDF helps filter out commonly used terms and emphasizes distinctive keywords that are relevant to each review's content.

Keyword Extraction: After calculating TF-IDF scores, we extract key terms for each review based on a predefined TF-IDF threshold. Terms exceeding this threshold are designated as significant keywords, capturing the primary topics or sentiments expressed in the review. For example, terms such as "interactive," "challenging," or "well-structured" may emerge as keywords, each offering insights into specific areas of the course. These extracted keywords allow instructors to quickly identify areas where students perceive strengths or weaknesses, making it easier to address specific feedback points.

By tagging each review with both its sentiment score and these extracted keywords, the system presents a concise summary that is both quantitative (sentiment) and qualitative (key phrases). This combination provides instructors with actionable insights, allowing them to understand and respond effectively to student feedback, ultimately leading to targeted course enhancements.

4) Sentiment Analysis Using VADER and BERT

To determine the sentiment of each review, we utilize two approaches: the VADER (Valence Aware Dictionary and Sentiment Reasoner) algorithm and the BERT (Bidirectional En- coder Representations from Transformers) model. This dual approach allows us to compare the effectiveness of a rule-based versus a deep learning-based sentiment analysis method, ensuring that our framework captures both straightforward and nuanced sentiments in student feedback.

VADER for Sentiment Scoring and Classification: VADER is a rule-based sentiment analysis tool optimized for social media and short text, making it suitable for informal reviews. VADER assigns each review a sentiment score from -1 to +1, where scores closer to -1 indicate negative sentiment, scores closer to +1 indicate positive sentiment, and a score of 0 represents neutral sentiment. Each review is categorized into one of three sentiment classes:

- *Positive*: Reviews with a score greater than 0.
- *Negative*: Reviews with a score less than 0.
- *Neutral*: Reviews with a score equal to 0.

BERT for Contextual Sentiment Analysis: To capture more complex and context-dependent sentiments that may not be fully addressed by rule-based methods, we also employ BERT, a deep learning model capable of understanding the nuanced context within each review. BERT performs sentiment analysis by fine-tuning on a labeled sentiment dataset, enabling it to recognize subtleties in language, such as mixed sentiments or context-specific phrases. BERT assigns a probability distribution over sentiment classes (positive, neutral, negative) based on the contextual meaning of each review, making it more adaptable to varied and detailed feedback.

Comparison of VADER and BERT: Using both VADER and BERT for sentiment analysis allows us to evaluate the strengths of each approach. VADER provides efficient and quick sentiment classification, suitable for reviews with clear sentiment expressions. However, for reviews with complex or nuanced language, BERT offers a more accurate analysis by considering contextual relationships between words. By comparing the results from VADER and BERT, we assess the consistency between the two models and identify cases where BERT's contextual understanding provides additional insights. This comparison highlights the practical implications of each model and helps determine the most effective approach for analyzing student feedback in an educational context.

5) Appending Sentiment Tags and Key Phrases

Following the sentiment analysis and keyword extraction, each review is annotated with additional features to provide instructors with actionable insights:

Sentiment Tags: Based on the scores from both VADER and BERT, each review is assigned a sentiment classification (positive, negative, or neutral). In cases where VADER and BERT results differ, both sentiment tags are presented to give instructors a comprehensive view of possible interpretations. *Key Phrases*: The key phrases identified in the review using TF-IDF are appended to the dataset, allowing instructors to understand the specific topics or aspects of the course that each review addresses. Together, the sentiment tags and key phrases provide a detailed summary of student feedback, combining sentiment orientation with thematic focus.

4. Classification

Classification involves categorizing student reviews based on their sentiment into three classes: Positive, Negative, and Neutral. For this task, we leverage the BERT (Bidirectional Encoder Representations from Transformers) model, which is well-suited for understanding complex and nuanced language, including reviews containing humor or subtle sentiment cues. BERT processes each review by converting it into a tokenized input sequence, which captures contextual and semantic relationships between words. This contextual understanding allows BERT to generate sentiment scores by analyzing the entire sentence structure rather than relying solely on individual words. The resulting sentiment score ranges from -1 to +1, where scores closer to -1 indicate negative sentiment, scores closer to +1 indicate positive sentiment, and scores near 0 are classified as neutral.

Each review is then assigned to one of the three sentiment classes:

- *Positive*: Reviews with a score greater than 0 are classified as positive.
- *Neutral*: Reviews with a score close to 0 are classified as neutral.
- *Negative*: Reviews with a score less than 0 are classified as negative.

After classification, we store the sentiment score, the corresponding sentiment class, and any key phrases identified during feature extraction in the database. The key phrases are generated using the Term Frequency-Inverse Document Frequency (TF-IDF) technique, which highlights the most relevant terms within each review. These stored features provide a structured overview of each review's sentiment and primary themes, allowing instructors to identify areas of strength or potential improvement based on student feedback.

This approach not only enables efficient categorization but also preserves valuable context from complex language, offering deeper insights into student sentiment on the educational platform.

5. Results and Discussion

A. Analysis Dashboard for Sentiment Insights

The primary output of this study is an interactive dashboard designed to provide instructors with actionable insights derived from student reviews. This dashboard integrates sentiment analysis and keyword extraction to highlight critical feedback trends. By leveraging advanced techniques like BERT for sentiment classification and TF-IDF for keyword identification, the dashboard ensures both accuracy and interpretability.

1) Visualization of Sentiments

The dashboard categorizes student reviews into positive, neutral, and negative sentiments. This classification helps instructors quickly gauge the overall sentiment distribution for their course. Trends over time can also be analyzed, giving instructors insights into whether course modifications result in improved student sentiment.

2) Keyword Analysis and Contextual Insights

In addition to sentiment scores, the dashboard presents key phrases extracted using TF-IDF. These keywords provide a deeper understanding of specific topics or aspects frequently mentioned by students. For example: Positive keywords may highlight strengths like "engaging lectures" or "detailed explanations." Negative keywords could identify areas needing improvement, such as "unclear assignments" or "fast-paced content."

The combination of sentiment scores and keyword insights allows instructors to connect the sentiment classification to actionable themes within the feedback.

3) Educational Impact and Feedback Utilization

The primary goal of the dashboard is to help instructors: - Understand student perceptions of their course through structured and meaningful data. - Identify recurring themes in positive and negative feedback for targeted improvements. Monitor how changes in course design or delivery impact student sentiment over time.

By translating unstructured student reviews into structured sentiment data and key insights, the dashboard empowers instructors to enhance their courses and foster better learning outcomes. This approach not only improves the quality of education but also promotes data-driven decision-making in instructional design.

B. Discussion on Results

The results demonstrate that combining sentiment analysis with keyword extraction provides a holistic view of student feedback. While VADER is effective for short and straightforward feedback, BERT excels in handling complex language and nuanced sentiments. This dual approach ensures that instructors receive accurate, context-aware insights. The inclusion of keyword analysis adds a layer of interpretability, enabling instructors to take targeted actions based on specific

Table 2
Comparison of VADER and BERT for sentiment analysis of student reviews

VADER	BERT
Model Type	
Rule-based	Deep learning (transformer-based)
Handling of Contextual Nuances	
Limited; performs well on straight-forward sentiments but may miss context-dependent meanings	Strong; captures nuanced sentiments and contextual meanings within complex language
Sentiment Score Range	
-1 to +1, where 0 is neutral; scores closer to -1 are negative, and scores closer to +1 are positive	Provides a probability distribution across sentiment classes (e.g., positive, neutral, negative), allowing for flexible interpretation
Suitability for Short/Informal Texts	
Well-suited; performs efficiently on short texts with clear sentiments	Suitable but optimized for longer or more complex texts where con- text is important
Processing Time	
Fast, due to simple rule-based calculations	Slower, due to computational complexity of deep learning architecture
Interpretability	
High; easily interpretable due to its rule-based nature	Moderate; uses hidden layers, which makes interpretability more challenging
Recommended Use	
Effective for reviews with clear, direct sentiment expressions	Recommended for reviews with complex or nuanced language, or when higher accuracy in understanding context is required

feedback themes. The integration of these methodologies into a single dashboard highlights the potential of leveraging advanced natural language processing tools for improving educational practices.

6. Conclusion

The primary objective of this study was to leverage Natural Language Processing (NLP) and sentiment analysis techniques to analyze student reviews and provide actionable insights for improving educational practices. Through the development of an interactive dashboard that integrates sentiment analysis and keyword extraction, we have successfully created a tool that offers instructors valuable feedback on their courses.

By utilizing sentiment classification methods like BERT and keyword extraction through TF-IDF, the dashboard presents a clear, structured view of student feedback. The sentiment analysis allows for the categorization of feedback into positive, neutral, and negative sentiments, while the keyword extraction identifies key themes that can guide improvements in course design and delivery.

The results demonstrate that NLP techniques can effectively support data-driven decision-making in education. By offering both high-level sentiment analysis and detailed, contextual insights through keyword analysis, the dashboard enables instructors to monitor student perceptions and make informed changes to enhance the learning experience.

In conclusion, this study has successfully addressed the objective of using NLP to provide feedback that not only high-lights areas of improvement but also reinforces strengths in the educational process. The integration of sentiment analysis into educational feedback mechanisms has the potential to foster more personalized, effective, and adaptive teaching strategies, ultimately contributing to improved student outcomes.

References

[1] Altrabsheh, N., Gaber, M. M., Cocea, M. (2014). Sentiment analysis: Towards a tool for facilitating feedback in educational environments. International Conference on Adaptive and Intelligent Systems.

[2] Nassif, A., Luo, W., Habieb, M. (2016). Sentiment analysis of student reviews in MOOCs. Journal of Educational Technology Society, 19(3), 158-169.

[3] Ochoa, X., Romero, C. (2013). Analyzing student interactions in social media as informal learning environments. International Journal of Learning and Media, 4(3-4), 33-44.

[4] Hutto, C., Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Eighth International Conference on Weblogs and Social Media (ICWSM-14).

[5] Huang, C., Shih, S., Chiu, C. (2020). LSTM networks for analyzing student feedback in discussion forums. Journal of Educational Data Mining, 12(1), 28-43.

[6] Rios, A., Knapp, B. (2018). Applying BERT for sentiment analysis in educational feedback. Proceedings of the Workshop on Innovative Use of NLP for Educational Applications.

[7] Bouazizi, M., Ohtsuki, T. (2016). A pattern-based approach for sarcasm detection on Twitter. IEEE Access, 4, 5477-5488.

[8] Jain, S., Tripathi, M., Patel, S. (2017). Fine-tuning deep learning models for educational sentiment analysis. Journal of Artificial Intelligence in Education, 27(4), 679-702.

[9] Balahur, A., Turchi, M. (2014). Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. Computer Speech Language, 28(1), 56-75.

[10] Medhat, W., Hassan, A., Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093-1113.

[11] Zhang, L., Wang, S., Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.

[12] Yadav, R., Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. Artificial Intelligence Review, 53(6), 4335-4385.