Improving Student Engagement Using NLP Techniques
Project Proposal Report

Rimika Dhara & Akansha Kamineni
CSCI 5541 NLP: Rimika & Akansha
dhara015@umn.edu, kamin143@umn.edu

1 Introduction

Student feedback plays a critical role in evaluating the effectiveness of teaching practices, course design, and academic resources, offering actionable insights to improve educational experiences. However, high-enrollment courses generate large volumes of feedback, making it challenging to analyze responses efficiently. This issue is compounded when feedback contains nuanced sentiments expressed through open-ended text, which are difficult to quantify using traditional methods. In response to these challenges, this project aims to develop a structured framework to collect, analyze, and derive meaningful insights from student feedback, enabling data-driven improvements in teaching and learning.

Our primary focus will be on collecting data in the form of student feedback from Computer Science students at the University of Minnesota, with the potential scope limited to introductory courses, through a survey. This focus ensures a manageable participant group and allows us to assess engagement at the foundational level of the curriculum, where teaching quality and course content play a crucial role in shaping students' learning experiences. In addition to survey-based data collection, we plan to augment the dataset by analyzing public reviews from platforms such as RateMyProfessor and SRS data, which contain valuable sentiment-based evaluations of instructors and courses.

Then, we hope to perform Sentiment Analysis that will allow us to extract meaningful insights from student feedback beyond surface-level statistics. Pre-trained NLP models like BERT will be used to capture the contextual nuances within student responses in order to understand larger sentiments and themes. We also plan to experiment with lighter or optimized variations, such as DistilBERT and RoBERTa, to balance performance with efficiency. These models will be evaluated for their ability to classify positive, negative, and neutral sentiments accurately across the collected feedback data. Additionally, clustering techniques will be used to identify recurring themes and patterns within the feedback, highlighting key areas of satisfaction and potential improvement.

A critical aspect of this project will involve addressing specific research questions relevant to improving student engagement: What additional engagement factors—beyond teaching and course content—can influence student satisfaction? Which aspects of student feedback offer the most actionable insights for professors? By answering these questions, we hope to provide practical recommendations for improving teaching practices and course design.

Our findings will be visualized to highlight sentiment distributions, common feedback themes, and areas requiring improvement. These visualizations aim to provide educators and administrators with accessible, data-driven insights that can guide decision-making. Ultimately, this project seeks to offer a replicable framework that educational institutions can adopt to better understand and respond to student feedback, driving continuous improvement in teaching and learning environments.

2 Literature Survey

The initial topic idea came from a 2017 paper by Sujata Rani and Parteek Kumar which discussed the use of natural language processing techniques, specifically sentiment analysis, to improve both teaching and learning in universities (5).

The sentiment analysis approach that's taken in the paper is relatively straightforward. After preprocessing and cleaning data, the authors suggest essentially scoring student sentiments from various data sources, such as SRS statements, in the categories of anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. This can then be used to identify if a statement is positive or negative. Some emotions are given higher weight than others due to potential biases or misclassifications that could occur within the model. For example, Rani and Kumar write that joy could be positive, but it could also be a negative emotion depending on what the statement is expressing joy about. If it's about teaching, it can be positive, but if it's joy at skipping a class because it's boring, then it is no longer positive, so anticipation and trust are given a higher weight than joy when calculating satisfaction. This is an interesting metric which is based off of Robert Plutchik's wheel of emotions. The results are eventually visualized into word clouds that highlight common language that was seen, such as "boring" or "cool".

In order to evaluate the model, Rani and Kumar compared the satisfaction rates with the average grade that was received in the class. The difference between these two metrics was at most 20%. A few courses were outliers where the satisfaction was higher than the average grade, or vice versa, which could potentially be attributed to course content difficulty, but it was not further explored.

The topic of NLP in education, especially in the context of improving teaching and learning, has become more relevant. A 2023 paper by Delali Kwasi Dake and Esther Gyimah continues to tackle the use of sentiment analysis on student qualitative data (3). This paper, however, was less focused on presenting results to professors, and instead aimed to evaluate the performance of different algorithms.

Dake and Gyimah used four different classifiers on the student evaluation data, including the J48 Decision Tree, Support Vector Machine (SVM), Random Forest, and Naïve Bayes. 10-fold cross-validation was used to develop the model and 5-fold cross-validation was sued to experiment on and train it. The results were then analyzed using a confusion matrix that was generated from the model.

The authors found that SVM performed the best out of the model with an accuracy of 63.79% at the 10-fold cross-validation level. The interesting thing to note here, however, is that the Random Forest algorithm performed equally well, however when further analyzed using the confusion matrix, it was learned that it predicted the class of most feedback text wrongly from students. SVM also showed promising results for building a predictive model on new data. On a new dataset of 31 different statements, the model performed with 92% accuracy, showing a positive trajectory and potential use case.

A recent survey by Shaik et al. (2022) explores the role of sentiment analysis in education, emphasizing how opinion mining can extract valuable insights from student feedback to enhance pedagogy and decision-making. The study highlights sentiment analysis techniques at document, sentence, entity, and aspect levels, using AI-driven methods such as machine learning, deep learning, and transformers. It also discusses challenges like multipolarity and opinion spam detection that must be addressed for effective adoption of sentiment analysis in education (6).

We also learned from Kasumba and Neumman (2024) who explore the use of student crowdsourcing to address the limitations of traditional annotation methods in sentiment analysis. They highlight challenges such as bias in large language models and the inefficiencies of manual labeling in high-enrollment courses. By comparing student-crowdsourced, self-reported, and expert annotations, their study demonstrates that Random Forest models trained on crowdsourced labels significantly improve the identification of negative sentiments. This dual-purpose approach not only enhances predictive accuracy but also offers a hands-on learning experience through integration into an educational course (4).

3 Novelty of Approach

While the use of NLP in educational institutions is being furthered explored, there's still a clear lack of research and experimentation in the topic area, especially when it comes to understanding and analyzing student feedback to better teaching practices. There are a few key ways that our proposed approach differentiates itself from previous research.

First, we are proposing more advanced sentiment analysis techniques through the use of models like BERT. Prior studies, such as the one from Dake and Gyimah, use traditional machine learning algorithms for analysis (3). BERT uses bidirectional context representation which allows it to better understand language. It not only processes the meaning of a word, but also it's meaning within the greater context of a sentence, unlike traditional algorithms. The use of pre-trained transformer models will likely lead to a more accurate and nuanced model.

The model will also feature an emotional weight system, based on the wheel of emotions similar to that proposed by Rani and Kumar (5). We want to implement a much more dynamic way to handle biases by experimenting with contextual weight adjustments that depend on additional course factors, such as difficulty or grading expectations. To further reduce bias, the data that is collected will also go through student-crowdsourcing, as suggested by Kasumba and Neumman (4). The combination of these techniques introduces less bias in the initial data processing stages, which can help reduce misclassification, and also reduces misclassification in the training stage, which is different from previous research where there was only one area of focus.

Finally, we want to present actionable insights for instructors. Many papers suggest different models that can be used or different ways to process the data, but very few focus on the practical application in day-to-day activities. We not only suggest word clouds and graphs, but want to provide personalized feedback, generated by the model, to each instructor to help them understand how they can continue or improve on their current teaching methods to better student learning. The real-world application is completely different from the research that we've seen and add another aspect of novelty to our project.

4 Future Work

Looking ahead, there are some clear steps that we will need to take for the final project. First, we must acquire data through student surveys, web-scraping on RateMyProfessor.com, and SRS data (if accessible). Following this stage, we will label the data using student crowdsourcing techniques before using a pre-trained transformer to train the model. The final stage will require visualizing the data in a usable format.

There are a few things that we will be considering as we progress, based on feedback from the initial proposal and presentation. In the training stage, it's important to consider any biases that could impact the model. There may be some strong data outliers which completely skew the model, like a very positive or very negative piece of feedback which doesn't fit the average response. We plan to address this by weighting the emotional scoring and reducing bias as much as possible in the initial training stage. Some emotions could and should have higher weights than other based on their potential for bias. We plan to use emotional weights, similar to Rani and Kumar, as a baseline and adapt as needed when training the model to decrease misclassification (5). We are also looking into crowdsourcing the labelling of data to reach a common consensus and reduce bias, as suggested by Kasumba and Neumman (4). Where available, the course difficulty score will also be used to interpret the results. If a student thought that course content was extremely difficult, they may have a completely different opinion on teaching skills than someone who thought the course content was easy, which is a metric that must be considered and could be used to reduce data bias.

For the final visualizations, it was suggested that we get a better understanding of what the end-user may want. We will be talking to instructors at the University of Minnesota to understand what would be most beneficial when viewing the feedback. Our current plan is to visualize the data using charts and word clouds, but also provide personalized feedback created by the model. Our team wanted to focus on practical application of the model, so we realized that it would be important to understand the needs of the instructors that it was meant to be used by. We'll be looking to understand what kind of visualizations would be most helpful, such as graphs or something more abstract, as well as how suggestions from the model could be presented. This will hopefully lead to a formal implementation of the model in current processes.

5 Contributions

For the proposal, Rimika and Akansha divided the work to utilize their strengths effectively. Rimika wrote the introduction, two sections of the literature review, the contributions section, and the bibliography. Akansha contributed to two sections of the literature review and authored the novelty of work and future work sections.

In the future, both members will collaborate closely, particularly on data collection, data preprocessing, and sentiment analysis. As of now, we are planning for Rimika to focus on survey-based data collection, while Akansha will handle web scraping tasks. Since web scraping may require more technical work, responsibilities may switch to ensure a balanced workload and skill development.

References

[1] Baker, R. S., Ocumpaugh, J., Andres, J. M. A. L., Marino, M. T. (2017). Neurodiversity in education: Consequences for learning and teaching. ACII 2017. Retrieved from https://learninganalytics.upenn.edu/ryanbaker/ACII2017_183.pdf

[2] Brar, R. (n.d.). *Student Feedback Dataset* [Data set]. Kaggle. Retrieved from https://www.kaggle.com/datasets/brarajit18/student-feedback-dataset

[3] Dake, D.K., Gyimah, E. (2023). Using sentiment analysis to evaluate qualitative students' responses. *Education and Information Technologies, Vol. 28*, 4629–4647. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9581765/

[4] Kasumba, R., Neumman, M. (2024). Practical Sentiment Analysis for Education: The Power of Student Crowdsourcing. *Proceedings of the AAAI Conference on Artificial Intelligence, 38*(21), 23110-23118. https://doi.org/10.1609/aaai.v38i21.30356

[5] Rani, S., Kumar, V. (2017). A sentiment analysis system to improve teaching and learning. *International Journal of Computer Applications, 174*(24), 28-33. Retrieved from https://ieeexplore.ieee.org/document/7924253

[6] Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., Galligan, L. (2022). Sentiment analysis and opinion mining on educational data: A survey. *Data-Centric AI*. Retrieved from https://www.sciencedirect.com/science/article/pii/S2949719122000036