Improving Student Engagement Using NLP Techniques: Building a Better Rate My Professor

Rimika Dhara &
Akansha Kamineni
(dhara015@umn.edu)
(kamin143@umn.edu)

BACKGROUND

Student feedback is essential for evaluating teaching effectiveness, course design, and academic resources, providing actionable insights for improvement. However, high-enrollment courses generate large volumes of nuanced, open-ended feedback that traditional methods struggle to quantify. This project addresses these challenges by developing a structured framework to collect and analyze feedback, enabling data-driven improvements in teaching and learning.

We propose advanced sentiment analysis techniques using models like BERT, which employs bidirectional context representation for a nuanced understanding of language, surpassing traditional approaches used by Dake and Gyimah (3). Incorporating an emotional weight system inspired by the wheel of emotions from Rani and Kumar (5), we dynamically adjust for contextual factors like course difficulty to reduce bias, further mitigated by student-crowdsourcing during data preprocessing, as suggested by Kasumba and Neumman (4). Additionally, we provide personalized feedback to instructors, extending beyond word clouds to enhance teaching practices and student learning outcomes, introducing a novel, real-world utility to our project.

RESULTS

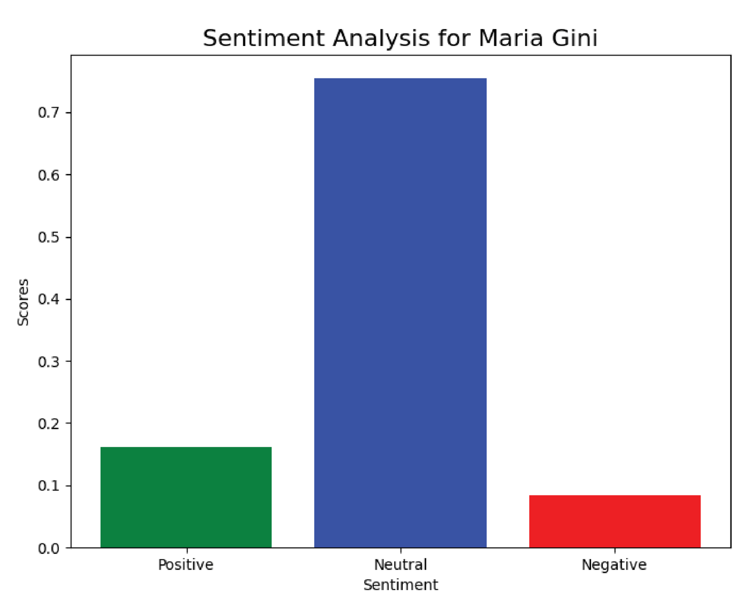

Fig 1. Basic Sentiment Analysis for One Professor

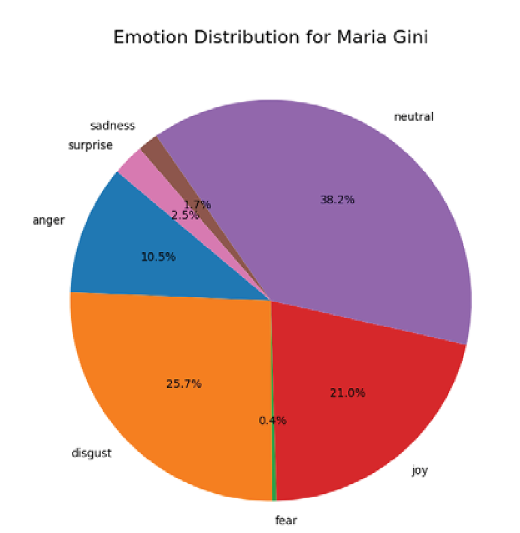

Fig 2. Robust Emotion Analysis for One Professor

Using the results that were gathered from the various NLP models, we were able to prompt an LLM, specifically Google Gemini, using an API to create a summary of the feedback that the instructor received as well as actionable insights. The model presents each instructor with Strengths and Areas for Improvement, as well as specific ways to improve in each of those areas (such as suggesting that an instructor records lectures if students mention that technology is not fully utilized). We were also able to gather all of the data and display it on a web application. Since our overall goal was to make the feedback actionable and usable, it required a platform that made it easy to access so that instructors weren't expected to run the code. They can now search for their name and view their feedback analysis.

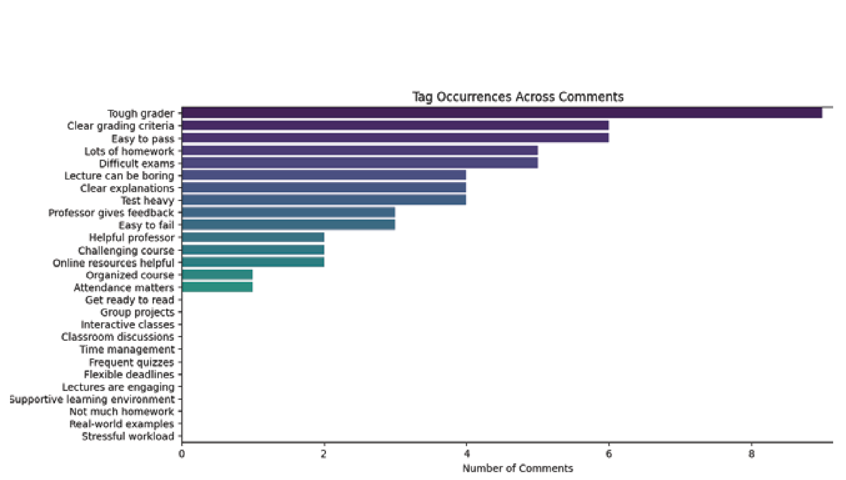

Fig 3. Number of Tag Occurrences (one count per comment)

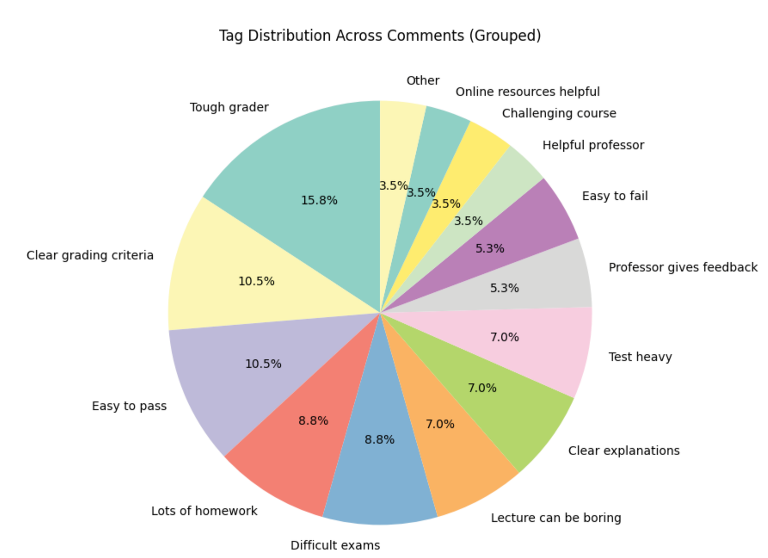

Fig 4. Tag Distributions Across Professor Comments

METHODS

Step 1: Collect student feedback data from RateMyProfessors utilizing a webscraper.

Step 2: Perform NLP Analysis

A basic Sentiment Analysis using VaderSentiment to classify negative, neutral, and positive reviews.

A more robust Emotion Analysis to classify specific tones in the reviews, beyond negative, neutral, and positive. Look for key emotions such as joy, fear, anger, surprise, etc.

Key Word Analysis to classify top tags for each instructor using NTLK n-grams, and Gensim phrasers.

Step 3: Prompt Google Gemini to create feedback summaries and action items based on collected data.

Step 4: Evaluate Results

Complete Human Evaluations using Google Sheets and compare results to the model's Emotion Analysis using accuracy metrics. Verify Gemini outputs through hand coding.

Compare model generated top tags to tags in initial data set from RateMyProfessors, generated and selected by students using accuracy, precision, recall, and F-1 scores.

DISCUSSION

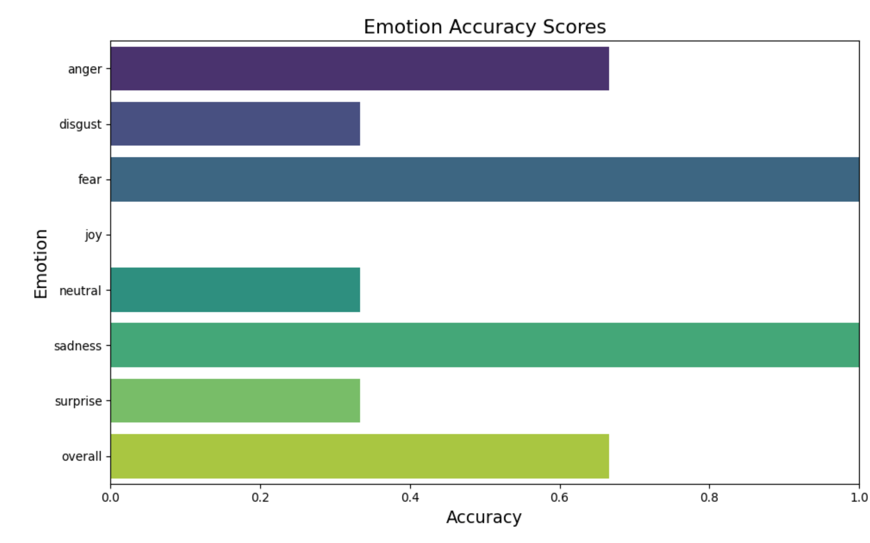

Fig 5. Accuracy of Emotion Analysis based on Human Evaluation Comparison

The accuracy of the emotion analysis seems to perform similarly to the human evaluations that were completed. The model performs the worst on Joy, with 0% accuracy. It also has a less than 50% success rate on Disgust, Neutral, and Surprise. This is important to take into account when considering the overall accuracy.

The evaluation metrics for the Top Tag Analysis show that Precision, Recall, and the F-1 Score are all 74.07% while the Accuracy is 20%. This indicates that the model is doing generally well and is avoiding excessive false positives, but the data set may be imbalanced overall shown by the low accuracy. This is likely because the model's tags go beyond that of the original data set, so there are many more possible tags that could be used.

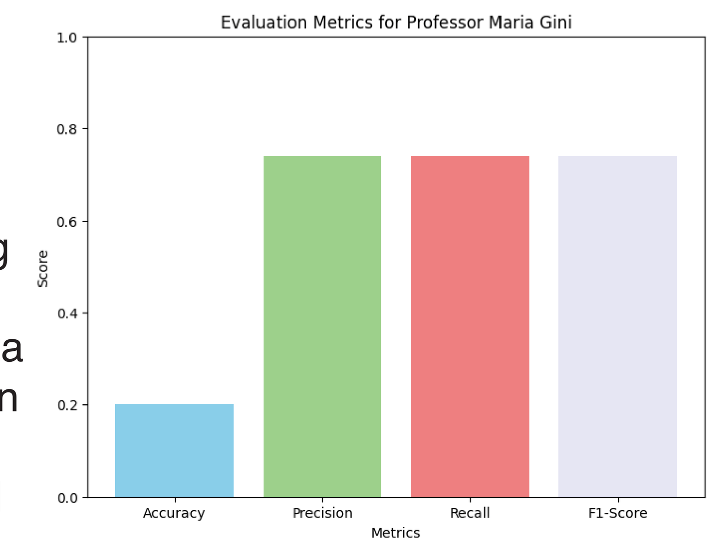

Fig 6. Accuracy, Precision, Recall, and F-1 Score of Top Tag Analysis based on Original Tags in RMP Data

Based on the results, we can conclude that the model works relatively well. In this context, it seems to provide the kind of data that we needed to create a web application for instructors to learn from. We reviewed select results and compared them with the initial data set to see if there were any complete outliers as well, but nothing seemed to be completely wrong, which furthers the promising results. It's important to understand that the models are not perfect, but are able to produce relatively accurate strengths and areas of improvement.

FUTURE WORK

1. Interview Instructors to improve the models and their usability. Implement different metrics based on their reviews and update visualizations to best fit needs.
2. Create a dynamic web application that can pull from and analyze new RMP data as it's submitted to create a self-sufficient application.
3. Gather additional data. For a more robust system and application, allowing the models to access SRT data from the University would be essential, but this was a privacy/security issue that we ran into for this initial project.
4. Improve web application security and user experience as our project relied on practical application of NLP tools.

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