Course Review Sentiment Tagging

NLP Trio

Payal Ahuja (payalda2), Trupthi Hosahalli Chandrappa (trupthi2), Riya Simon (riya5)

Topic

The topic that our group chose to explore is Course Review Sentiment Tagging using NLP Techniques. We were interested in exploring how we can gain more insight into students' feedback and understand the underlying sentiment from the colloquial language they use. Additionally, we also want to understand what factors students consider when reviewing a course or a professor (ex. nature of the professor, course usefulness, etc.) At the same time, we want to help the reviewed entity (professors, course designers) make sense of the large set of feedback and extract meaningful data to report and use for further decision-making or performance evaluation. By using NLP techniques to extract key phrases, and building a machine learning model to generate sentiment tags based on those phrases, we hope to gain a better understanding of the underlying sentiment in the feedback process.

Dataset

We plan to scrape data from ratemyprofessors.com, a platform where students can leave reviews on their course instructors. We will use the review text and accompanying tags on the most rated professors at UIUC to train our model.

Another dataset we are planning to use in conjunction is the Kaggle dataset on Coursera reviews, which includes reviews on over 600 different courses. We are planning to perform the necessary preprocessing to combine these datasets so that we have the following fields:

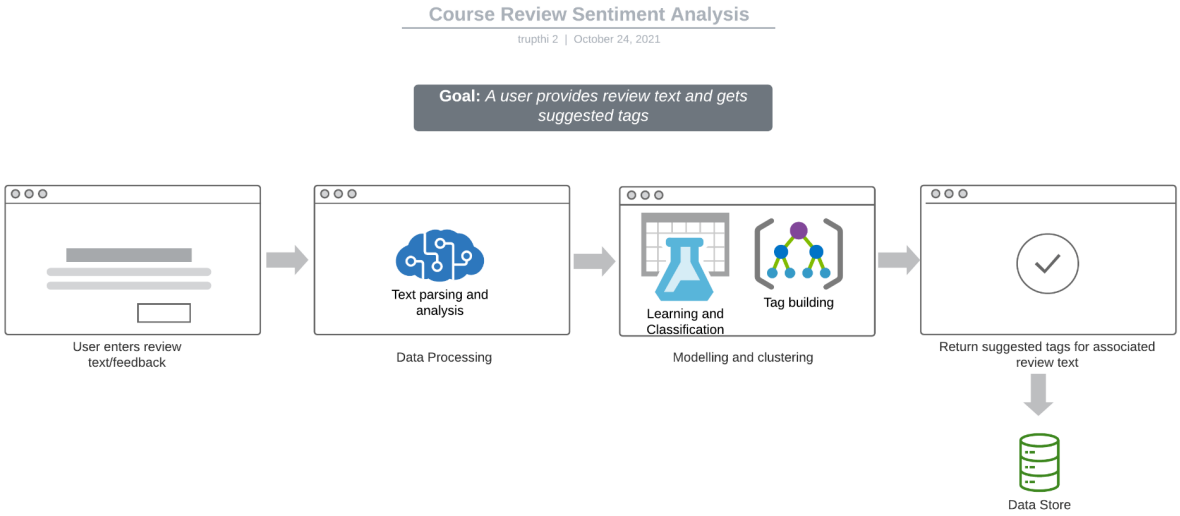
Professor Name/ID	Course Name/ID	Review Text	Numerical Rating	Tags/Additional Info

Implementation

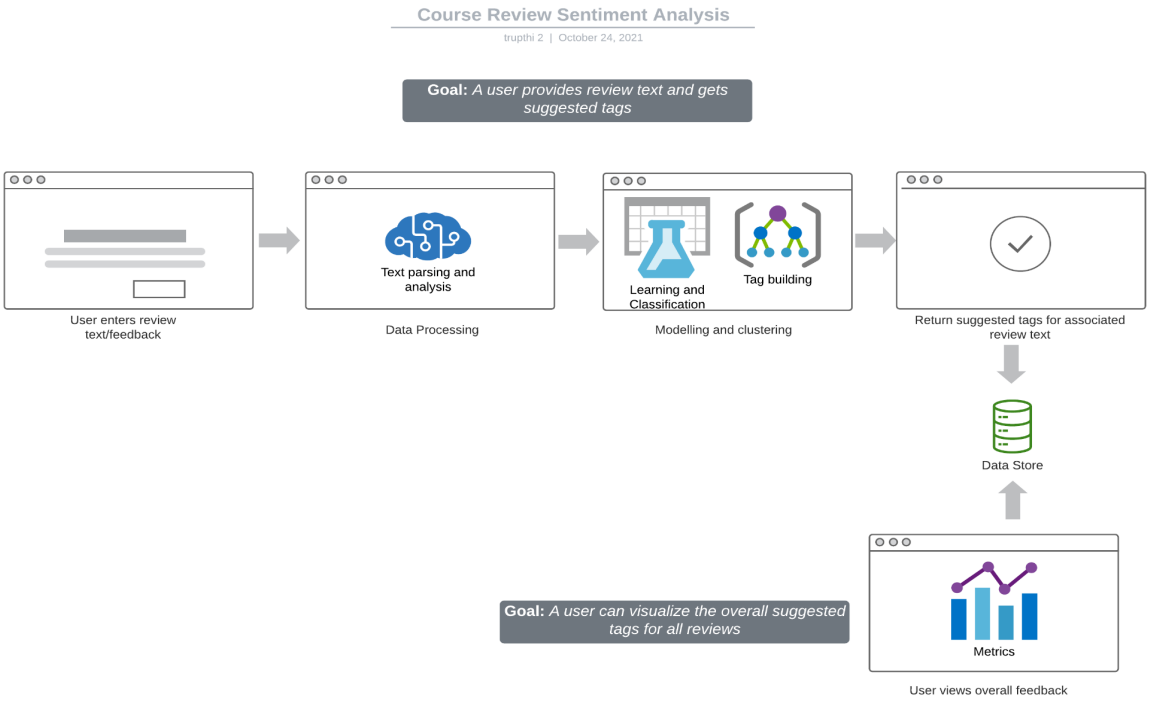
We are planning on creating an interactive interface where the student can type in reviews and receive recommendations on tags that match their review sentiment. Our interface will also include a platform on the professor-side, where course instructors can view reports and visualizations of the feedback they have received.
The tools we are planning to use are:

- **Environment**: JupyterLab/Google Colab, AWS
- **Languages**: Python, Javascript
- **Libraries**: plotly, matplotlib, scikit-learn, tensorflow, keras, nltk
- **Methods**: POS-tagging, TF-IDF, K-means clustering, Markov chains

MVP:



Further Goal:



Verification

Our goal is to ask for explicit human feedback to verify that the tag sentiment that our model generates matches the sentiment of the student review. This will be a binary match, i.e. if the train and test are both of the same sentiment, we classify the match as a "1". Otherwise, we classify it as "0."

In addition to this, we can use the numerical ratings for each course to compare with the sentiment ratings we generate. For example, if the course rating is around 2.0, yet we are generating many positive reviews, we will need to revisit our algorithm.

Workload:

Task	Estimated Hours
Scrape/collect training data	5
Data cleaning and preprocessing	5
Design sentiment analysis technique	10
Design tagging model	10
Design clustering algorithm	10
Implement model	12
Test and verify	12
Build interface for user interaction	8
Documentation and review	5
	77