

Analysis of “Rate My Professor”: Guidance for Both Students and Instructors

Team Members: Yinuo Hu, Yanbo Wang, Xinlin Huang, Shichen Wu, and Jiaming Xiong

1. Introduction of Problem

Students' reviews can be essential for evaluating teaching and learning experiences. Our group aims to investigate the main factors that affect students' learning experiences, provide suggestions for students on whether they should take a particular class, and determine which part of the lectures students like the most to help instructors improve their teaching quality.

2. Dataset

The dataset we will use was shared by Jibo He on Mendeley Data (<https://data.mendeley.com/datasets/fvtfjyvw7d/2>). It was originally crawled from [RateMyProfessor.com](https://www.ratemyprofessor.com), where students can rate and comment on their experiences of taking classes with different professors. The dataset contains 20,000 observations with 51 variables.

3. Techniques

As the dataset involves some noises and null values, in our analysis, we will first conduct data cleaning and feature engineering to select appropriate variables for analysis. Since our dataset contains quantitative and qualitative attributes and open-ended comments, we will investigate the dataset from two angles: student's vs. instructor's.

1. Student's Angle:

- We integrate all students' reviews in the dataset for classification, including features like 'Student_star,' 'student_difficult,' 'attendance,' etc.
- Consider the 'take_again' attribute as the dependent variable and use the selected features as the independent variables to build models. We plan to explore different models such as logistic regression or tree-based algorithms and possible ensembling methods to improve the model's efficiency.

2. Instructor's Angle:

- For the data, we will use students' comments.
- We will conduct several NLP techniques to explore the patterns of student comments, which may include 1) A word cloud describing the most frequently used words in students' comments, 2) Sentiment analysis of whether student comments are positive or negative, 3) Topic modeling to identify the main topics in student comments (e.g., LDA).

4. Impact

The project studies data from “RateMyProfessor” to provide suggestions for both students and professors. Students can use the model with up-to-date ratings for the professor and get an overall idea of whether or not to take this course from the binary predicted value. With sentiment analysis, professors can use the model to identify the qualities that are essential for students' learning experience. In conclusion, this project aims to build models to quantitatively evaluate the features affecting students' ratings for professors and provide suggestions for students' course selection and lectures' teaching quality improvement.