

## Using Sentiment Analysis to assess Student Ratings

### 1. Team Members:

- Chris Hickingbottom
- Madeline Udowa
- Polina Nemkova

Github Project Repository: <https://github.com/PoliNemkova/NLP-project>

### 2. Goals and Objectives:

#### a. Motivation

Sentiment Analysis is known as the process of finding positive and negative opinions, also known as sentiment, in a text. Our motivation is to use sentiment analysis to detect sentiment in the ratings of professors by students. There is a website called Rate My Professor, the best source for professor ratings based on the feedback from students. We will be using sentiment analysis to detect the opinions of students from the Rate My Professor website.

#### b. Significance

As earlier stated in the motivation, Rate My Professor is a website where students give feedback on their professors through ratings and comments. It is possible that a comment made might need more analysis to fully understand the reason for the rate given by the student, and this is where the importance of sentiment analysis comes in. Using sentiment analysis to monitor the comments and assess the language or words used by the student will give more reasoning behind the rate.

#### c. Objectives

- i. Create customized NLTK's sentiment analysis to use the most useful features
- ii. Run model with different classifiers such as KNN and Regression to compare the models for best results
- iii. Create landing page to show average emotional score for UNT
- iv. Create search by UNT professor name and page displaying average emotional score for professor

#### d. Features

- . User can see the average emotional rating of UNT
- i. The user can search for a specific UNT professor to see his/her average emotional rating if that professor has comments and ratings within RateMyProfessors.com

### **3. Existing work**

An existing project that makes use of a dataset from Rate My Professor would be a project that was done in 2020 by Tisuela. The purpose of the project was to scrape and search through the Rate My Professor website, hence the name of the project “Rate My Professor Scraper and Search”. In this project, a Python class was created to scrape and search through Rate My Professor data from all professors of a single University. The class created can have multiple uses and can be modified easily to be used in RMP applications. In this project, what will be returned is basic information on any professor searched, like the department, employee ID, the name of the university, the number of ratings for the professor and the overall rating for that professor.

### **4. Limitations and Solutions**

The limitations of the project are in the fine tuning of the existing model for the best performance on our dataset. Additionally, we would like to contribute to the project in regards to the ratings of the professor. We would like to find out the emotions or reasons for the rating given by the students to the professor, using sentiment analysis. We will look at the overall rating comment (good, medium, bad) and the comments specifically made by the students to identify the emotion and reason for the rating given.

### **5. Strength and Weakness**

The state of the art results in the area of sentiment analysis with ST-2 Binary classification task are reported by the SMART-RoBERTa Large mode (with accuracy 97.5%) that is based on a [ACL2020 paper](#). There are other models with a similar strong performance such as [T5-3B](#) model (with accuracy 97.4%) that utilizes transfer learning with a Unified Text-to-Text Transformer ([Journal of Machine Learning Research 21](#)) or [MUPPET Roberta Large](#) which uses Massive Multi-task Representations with Pre-Finetuning (with accuracy 97.4%) and was published by the researchers at Facebook (Meta?).

### **6. Innovation**

Our goal for this project is to perform a sentiment analysis with a high accuracy. As mentioned earlier, the project we are referencing only pulls out the information on a particular professor but we would like to reach high accuracy by using sentiment analysis to classify the emotions and reason for the ratings given to the professor. Since we are just learning NLP, our innovation is minor but we aim to add other features to the projects such as front end for the user.

## 7. Complexity

- Data cleaning & preparing: getting reviews in a suitable format; getting the ratings;
- We will use [VADER Python library](#) to get the sentiment score on each review;
- Average the sentiment score per one professor;
- Use classification to predict RateMyProfessor rating based on our sentiment score;
- Building a simple front end page for the presentation of the results.

## 8. Dataset

- 4295 professors with 12 columns of data
- 43443 ratings with 33 columns of data

## 9. Project Management

- Implementation Status Report
  - Training and Test Data Gathering
    - Utilize the RateMyProfessor.com api to gather UNT professors and their ratings.
    - Christopher Hickingbottom
    - 100%
  - Research and Analysis
    - Research past usage and implementation of RateMyProfessor.com datasets within the scope of NLP
    - Madeline Udowa 50% - Polina Nemkova 50%

## 10. Reference:

Provide more references and link them in the document

- Python API for searching Rate My Professor data from a single university:  
<https://github.com/tisuela/ratemyprof-api>
- VADER: Sentiment Analysis Library for Python <https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/>
- A tutorial for Sentiment Analysis: First Steps With Python's NLTK Library:  
<https://realpython.com/python-nltk-sentiment-analysis/#compiling-data>

- A tutorial for a basic sentiment analysis: <https://towardsdatascience.com/basic-binary-sentiment-analysis-using-nltk-c94ba17ae386>
- Emotion and Sentiment Analysis: A Practitioner's Guide to NLP: <https://www.kdnuggets.com/2018/08/emotion-sentiment-analysis-practitioners-guide-nlp-5.html>
- Sentiment analysis benchmark papers <https://paperswithcode.com/task/sentiment-analysis>
-