

Progress Report

Nov 14, 2021

Payal Ahuja, Trupthi Hosahalli Chandrappa, Riya Simon

Task	Estimated Hours	Progress
Scrape/collect training data	5	Completed
Data cleaning and preprocessing	5	Completed <ul style="list-style-type: none">• Optimization pending
Design sentiment analysis technique	10	In progress
Design tagging model	10	In progress
Design clustering algorithm	10	Research
Implement model	12	Research
Test and verify	12	
Build interface for user interaction	8	In progress
Documentation and review	5	
	77	

1. Which tasks have been completed?

1. Web Scraping: The code for fetching the comments and rating of professors with more than 50 ratings on “Rate my Professor” has been completed. The code for organizing the data and writing to a csv file has been completed.
2. UI: Research for tools to build a UI for the application has been completed.
3. Data cleaning and preparation for the Coursera review by course - static dataset

```
#Cleanse data for better analysis
stopwords = set(stopwords.words("english"))

def data_cleanse(line):
# Removes all special characters and numerals leaving the alphabets
    line = line.lower()
    line = re.sub('@', '', line)
    line = re.sub('\[.*?\]', '', line)
    line = re.sub('https?://\S+|www\.\S+', '', line)
    line = re.sub('<.*?>+', '', line)
    line = re.sub('[%s]' % re.escape(string.punctuation), '', line)
    line = re.sub('\n', '', line)
    line = re.sub('\w*\d\w*', '', line)
    line = re.sub(r"^[^a-zA-Z ]+", "", line)

    #Tokenize the data
    line = nltk.word_tokenize(line)
    #Remove stopwords
    line = [w for w in line if w not in stopwords]
    return line

data = pd.read_csv('/content/drive/My Drive/Coursera_reviews_subset.csv', engine='python', error_bad_lines=False).dropna()
```

4. Building the tokenized and lemmatized static reviews for model input

	reviews	Cleaned Reviews
0	Pretty dry, but I was able to pass with just t...	[pretty, dry, able, pas, two, complete, watch,...
1	would be a better experience if the video and	[would, better, experience, video, screen, sho...
0	Pretty dry, but I was able to pass with just t...	[pretty, dry, able, pass, two, complete, watch...
1	would be a better experience if the video and	[would, better, experience, video, screen, sho...
2	Information was perfect! The program itself wa...	[information, perfect, program, little, annoyi...
3	A few grammatical mistakes on test made me do ...	[grammatical, mistakes, test, made, double, ta...
4	Excellent course and the training provided was...	[excellent, course, training, provided, detail...
..
154	It was very useful course. This course is very...	[useful, course, course, time, taking, best, w...
155	very tough to understand ...use simple languag...	[tough, understand, use, simple, language, exp...
156	I prefer a course that teaches the fundamental...	[prefer, course, teaches, fundamentals, financ...
157	I can't access the quiz 5 and beyond. I'd real...	[cant, access, quiz, beyond, id, really, like,...
158	This coure seems to completely deviate from th...	[coure, seems, completely, deviate, main, topi...

[159 rows x 6 columns]

	reviews	Cleaned Reviews
0	Pretty dry, but I was able to pass with just t...	[pretty, dry, able, pas, two, complete, watch,...
1	would be a better experience if the video and	[would, better, experience, video, screen, sho...
2	Information was perfect! The program itself wa...	[information, perfect, program, little, annoy,...
3	A few grammatical mistakes on test made me do ...	[grammatical, mistake, test, make, double, tak...
4	Excellent course and the training provided was...	[excellent, course, train, provide, detail, ea...
..
154	It was very useful course. This course is very...	[useful, course, course, time, take, best, wor...
155	very tough to understand ...use simple languag...	[tough, understand, use, simple, language, exp...
156	I prefer a course that teaches the fundamental...	[prefer, course, teach, fundamental, financial...
157	I can't access the quiz 5 and beyond. I'd real...	[cant, access, quiz, beyond, id, really, like,...
158	This coure seems to completely deviate from th...	[coure, seem, completely, deviate, main, topic...

[159 rows x 6 columns]

5. Research on Multi-label classification and model creation: Reviewed various methods for classification using text to feature conversions with TF-IDF features using BagofWords, GLoVe and ELMO.

6. Sentiment Analysis: Researched various methods, implemented Vader, SVM to get probability distribution of sentiment scores

professor	comment	rating	pred
Gretchen Adams	Gretchen is the best Chem teacher ever created...	5.0	{'neg': 0.228, 'neu': 0.5, 'pos': 0.273, 'comp...
Gretchen Adams	Gretchen is an AMAZING professor. If you take ...	5.0	{'neg': 0.0, 'neu': 0.858, 'pos': 0.142, 'comp...
Gretchen Adams	She is really nice and gives great lectures as...	5.0	{'neg': 0.0, 'neu': 0.518, 'pos': 0.482, 'comp...
Gretchen Adams	Her lectures definitely aren't boring. It's ve...	5.0	{'neg': 0.0, 'neu': 0.602, 'pos': 0.398, 'comp...
Gretchen Adams	Great professor!	4.5	{'neg': 0.0, 'neu': 0.185, 'pos': 0.815, 'comp...
...
Carol Augspurger	Wow...carol is somebody else. She's extremely ...	4.5	{'neg': 0.0, 'neu': 0.792, 'pos': 0.208, 'comp...
Carol Augspurger	Don't get me wrong, this class can be very cha...	5.0	{'neg': 0.038, 'neu': 0.783, 'pos': 0.178, 'co...
Carol Augspurger	Simply amazing professor. She knows everyone's...	4.5	{'neg': 0.0, 'neu': 0.554, 'pos': 0.446, 'comp...
Carol Augspurger	Her classes were the staple of my college care...	5.0	{'neg': 0.082, 'neu': 0.667, 'pos': 0.251, 'co...
Carol Augspurger	She was such a sweet lady and really cared abo...	4.5	{'neg': 0.026, 'neu': 0.819, 'pos': 0.155, 'co...

2. Which tasks are pending?

1. UI: We are currently building more components of the UI and working on deploying the application so it can be available to the public for testing. Below is a screenshot of the web application, built using Streamlit. We are continuing to work on the integration of the web application with the sentiment tagging model functions. After the sentiment model functions are complete, we need to implement a callback method

that calls the function with a given input.

CS410 Project: Course Review Sentiment Tagging

An application that performs sentiment analysis on course/professor reviews and provides a (+/-) rating and relevant key tags.

Using data from Rate My Professor

Done! (using `st.cache`)

☒ Show raw data

Raw data

	professor	comment	rating
0	Gretchen Adams	Gretchen is the best Chem teache...	5.0000
1	Gretchen Adams	Gretchen is an AMAZING professo...	5.0000
2	Gretchen Adams	She is really nice and gives great L...	5.0000
3	Gretchen Adams	Her lectures definitely aren't bori...	5.0000
4	Gretchen Adams	Great professor!	4.5000
5	Gretchen Adams	Gretchen is probably one of the b...	5.0000
6	Gretchen Adams	Gretchen is amazing! She is a fant...	4.0000
7	Gretchen Adams	Gretchen is absolutely amazing! S...	5.0000
8	Gretchen Adams	Gretchen is a front-runner for mo...	4.5000
9	Gretchen Adams	Gretchen is by far the best profes...	4.5000

Try it!

Write review.

The professor was the BEST!

Submit!

2. Define parameters to categorize sentiment as [strongly negative, negative, neutral, positive, strongly positive].
3. Finessing the cleaning of data for better keyword extraction
4. Combine sentiment analysis and keyword extraction functionality.
5. Building the actual model for tagging utilizing the understanding of multi-label classification.
6. Integration of the UI input to the model and providing tag results for the final result display
3. Are you facing any challenges?
 - a. Data cleaning with static Coursera dataset
 - b. Integration of the sentiment analysis results and the multi-label classification approach for tagging the reviews