

CS410 – Fall 2021 – Project Proposal

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

Team Name: Team21

Net Ids:

Subudhi, Pradosh Kumar - [pradosh2](#)

Veera Sikku, Siva Prakasini - spv4

Dunor, Aaron - adunor2

Captain:

Pradosh

2. What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

Topic Chosen:

Email sentiment analysis. We will be creating a process for enterprises to quickly analyze customers' sentiment about their product/ service.

Problem Statement:

Every customer is important to an enterprise for the business growth. To keep the customers loyal to your product/ service, we need to address their issues/ concerns on priority. By building a system which can analyze thousands of emails received every day and categorize emails based on their sentiments (e.g., customers unhappy about a certain product/ service).

Relevance to the Course Subject:

To achieve this goal, we will have to perform Text Mining and Analysis.

3. Briefly describe any datasets, algorithms, or techniques you plan to use

Strategy:

Dataset Selection: We have chosen a dataset available in Kaggle which is a collection of customer complaints from various industries.

Data Source: Evaluating Customer Service

<https://www.kaggle.com/janiobachmann/evaluating-customer-service-sentiment-analysis/data>

Algorithms/ Techniques: We have considered using a Convolutional Neural Network algorithm. We will clean up the data and then convert the data to a text/ CSV format. We will develop a model to perform the sentiment analysis based on some presets.

4. How will you demonstrate that your approach will work as expected? Which programming language do you plan to use?

We come up with step-by-step approach (Choosing the right dataset, clean up/Sort the data/Building the model based on algorithm) by splitting task to achieve the expected outcome. We will provide a demo where our model will correctly classify the sentiment of random customer emails.

Programming Language: Python

5. Please justify that the workload of your topic is at least $20 \times N$ hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

Justification:

For successful completion of the project, we will have to perform the following tasks. We are estimating about seventy five hours is required to complete the project. Considering this we believe; each team member will have a minimum of twenty hours of contribution.

- Analysis and research – 10 hrs.
- Gather and clean data – 5 hrs.
- Build machine learning model – 25 hrs.
- Train and evaluate model – 12 hrs.
- Team collaboration – 15 hrs.
- Project documentation – 4 hrs.
- Demo preparation and presentation – 4 hrs.