jcmcdow2
CS 410

Final 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.*

   My NetID is jcmcdow2. I will be working on the project alone because I will have limited availability to meet.

2. *What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?*

   What is your free topic? Please give a detailed description.
   For my topic, I will create a chrome extension that allows users to perform sentiment analysis on video game reviews. When the user submits a review, the extension will send a request to a backend REST API. The API will use a sentiment analysis model to return the overall sentiment of the text.

   What is the task?
   The main task of this project will be creating the sentiment analysis model. The model needs to be accurate enough that users can trust the results but also fast enough that users get results back quickly.

   Why is it important or interesting?
   This project is important because sentiment analysis is a growing aspect of data mining. Due to the large amount of text review data available, users need a way to quickly determine the sentiment of a review. This API can easily be extended to accept a batch of reviews and return a file to the user.

   What is your planned approach?
   The first step will be to create the machine learning model. I will be using a data set containing Amazon video game reviews to create the model. The dataset contains the review and the number of stars given to the product. The model will take the text review and attempt to predict the stars. For 1- and 2-star reviews it will return negative, 3-star reviews will return neutral, and 4- and 5-star reviews will return positive.

   Once the model has been trained it will be written to a file. The next step will be to build the API. I'm planning to create the backend in AWS. The model will be loaded in a Lambda function that will run predictions on request data. The final step will be to

create the extension. The extension will be a simple form that allows users to input the data and send a request to the API.

What tools, systems or datasets are involved?
I'm planning is to use PyTorch for the sentiment analysis model and predictions. For the API I will use an AWS API Gateway connected to a Lambda function. The extension will be created using HTML, CSS, and JavaScript.

I will be using the video game dataset found here https://nijianmo.github.io/amazon/index.

What is the expected outcome
The expected outcome is that the user can submit a text review and receive the accurate sentiment in a timely manner.

How are you going to evaluate your work?
I will evaluate the sentiment analysis model by withholding some of the data and running the model against this unseen data. I will then measure the overall accuracy, precision and recall. Once the model is performing satisfactorily, I will find video game reviews on the internet and compare the model output to the number of stars given.

3. Which programming language do you plan to use?
   I will be using Python, HTML, CSS and JavaScript for my project.

4. Please justify that the workload of your topic is at least 20*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.
   - Create/tune sentiment analysis model – 10 hours
   - Create AWS backend – 4 hours
   - Create extension UI – 3
   - Test extension – 1
   - Create documentation – 2
   - Create demo video – 1