Nate Smolczyk (nds6)
CS410
23 October 2022

<p style="text-align:center">Course Project Proposal – Video Game Review Classifier</p>

For my final project I would like to create a site which can suggest a star number based on a review. (i.e. 1 star for very negative review, 5 stars for a very positive review, and somewhere in the middle for the rest.) I will be working alone. I plan to utilize this dataset as a basis for the project, which has a multitude of Amazon reviews for a variety of products. I am particularly interested in the video game review data. I am interested to analyze what type of language people tend to use for each star-rating of a video game. While ratings are subjective, I think it will be interesting to see how well I can train a model to guess how many stars a user would give their review based on how the Amazon community as a whole tends to assign ratings.

My final product will have two components: one page that visualizes and explains my analysis of the dataset, and another page for an application that takes as input a review and give the user a suggested star-rating. I have a couple of use cases in mind for this kind of tool. Since ratings are subjective, users who are unsure how to rate their review could utilize this tool to get a suggested star rating, based on how the community as a whole has assigned ratings. Alternatively, a user may have their rating in mind and want to use the tool to see how their ratings compares to users on average. For example, I may have really enjoyed a game and had 5 stars in mind, but after plugging my review into the tool, could find that similarly worded reviews tend to only give 4 stars.

To accomplish this, I plan to use the data to train a machine-learning model that can classify a review based on its content. With over 2 million reviews provided in the dataset, there is plenty of data to both train and test such a model. I plan to do NLP analysis to determine what kinds of words are most indicative of each star rating. Once trained, this model will support a program which can take in an unseen review and classify it as one of 5 options: 1 star, 2 stars, 3 stars, 4 stars, or 5 stars. I will then create a UI which will be used to document my analysis and support a site which can interact with the model to give the user a suggested rating for their review. I will be using Python for all of the backend tasks: training the model, analyzing the dataset, and setting up a server which will support the UI. I will be using JavaScript/React to build the UI itself. To host my website, I plan to use either AWS, GitHub Pages, or a combination of the two. Here is a breakdown of the approximate workload for each task:

NLP Analysis and Model training (Python)	~10 hours
Server Setup (Python)	~3 hours
UI Creation (JavaScript/React)	~3 hours
Hosting (AWS, GitHub pages)	~3 hours
Connecting UI to Server	~1 hour
Testing	~2 hours
Total	~22 hours

The model training and NLP analysis will be the core task for my project and will take the most time, which makes sense since it is the most relevant to course material. Each of the other tasks will be somewhat less time consuming. I'm not terribly familiar with the server setup and hosting steps so these will have some trial and error involved. I put these both at 3 hours since there isn't a ton of code to write, but will take time to get the pieces all communicating together correctly. I also assigned 3 hours to the UI creation since while this will involve more code to write, I am very familiar with React. Finally, I allotted about 2 hours for testing my results and analyzing my work. I will test my work by running the model on testing data and seeing how accurately I can predict star ratings for the Amazon dataset. I also plan to create some of my own reviews and checking how the model categorizes them.