Team A: Project Proposal

September 11, 2020

Team Members:

* Goldy Malhotra
* Yuvraj N Singh
* Scotty Williams
* Priyadarshni Suresh Sagar
* Saumya Nagia

Executive Summary:

The current multi-class model to predict the recommendation of a text review is fairly inaccurate and does not perform to acceptable standards. It currently has only an accuracy of 60% and tends to be biased towards a specific class. However, we have a proposal to refine and retrain this model to achieve much better performance. The overarching end goal will be to increase the accuracy to 80-85% (along with removing the class bias).

Requirements:

The major requirement of this proposal is to achieve a machine learning model that is able to better predict recommendations based on a text review than the current model. A major portion of that requirement is to improve the overall accuracy of the model. The target goal there is 80-85% accuracy, up from the current 60% accuracy. The current model also has a bias towards a specific class which should also be removed in the updated implementation. To this end, the implementation will require a new architecture and design, a better machine learning model with finer data preprocessing and superior training, and an improved implementation of said model. Another portion of this implementation is better documentation for the entire system. Each major component of this design will be documented for future reference and client needs.

Solution:

The solution to this problem and these requirements is to redesign the machine learning model used and then to retrain the model and to further test its performance. We are going to follow the below approach:

1. Data Preparation: It includes EDA, and preparation of training, testing data set
2. Feature Extraction: We will extract useful features using various feature extraction algorithms
3. Model Selection : Test different models in Azure ML studio, Model giving accuracy more than 80-85% will be selected to build the final model
4. Model Building : We will build the final model, train it on 60-70% of data
5. Model Testing : Test the model on remaining 30% of data and check the accuracy

Proposal:

To achieve the end goal of a better machine learning model with adequate performance, the primary step is to review the requirements and validate said requirements with stakeholders. Then, those requirements are converted to technical specifications and an initial design and architecture are created. These designs are iterated and refined, including a technical plan for remodeling the existing system. The new model is then created and implemented into this new system. The model is then trained, after which the model is tested for performance and requirement matching. Meanwhile, this entire process is well documented for future needs.

Deliverables:

After the project requirements have been scoped out, a web page will be created that will outline the problems with the current model, the overall project requirements, proposed solutions to achieve the target requirements, and other documentation. After reviewing the proposed solutions, a new machine learning model will be created to attempt to meet the project requirements. Once a suitable model has been created, a presentation of the model along with how we went about in creating it will be made, discussing obstacles faced while attempting to meet the target requirements. All documentation will be posted on the web page and Trello will be used to manage and view each process step of the project.

Technical Process:

Upon team creation, we began the Inception Phase (sprint 1) during which we discussed the overall project proposal and key requirements. Tools used during this timeframe were Slack and email for basic communication and Google Apps for collaboration. Much of the first sprint (and some parts of the second) included team discussion and planning surrounding overall ideas for the project. The initial phase was used to set the groundwork for the Elaboration Phase (sprint 2) in which administrative tasks such as creating the Trello board and Github Repository will be conducted. Trello is being used as a project management tool to track progress. Team members will be using the tool to create cards for each step or task of the project (and members will be assigned to their respective tasks). Other parts of this phase will include the overall design of the project as well as brainstorming what the project webpage will look like and include, followed by it’s actual creation. The bulk of the next phase (Construction/sprint 3), will be to build and troubleshoot the actual working product. Finally, during the Transition Phase (sprint 4), we will polish the working product and put together the project presentation, report, and documentation. There will also be a team reflection exercise and various wrap-up activities in order to fulfill any outstanding items. It’s important to note that throughout the project lifecycle, team members will use all the aforementioned tools (Google Apps, Trello, etc.) to stay in contact with each other and continue to collaborate.