**Title Page – SRS Project Name**

Team Members: William Johnson

Description of project: Analyze a dataset of data on various movies that are available on at least one of four different streaming services to determine which streaming service would be best based on the parameters given.

**Section 1:**

Introduction: I remember when I decided to buy a subscription to a streaming service, I had to rely almost entirely on advice and input from other people for help to figure out which one to get. Something like this would have really helped me out.

Purpose: To help potential users determine which streaming service would be best for them to get a subscription for.

Scope: My project will not interact with any actual websites in any way after the dataset is downloaded. This project will use the existing data in the dataset to help potential users determine which of four streaming services would be best for them to get a subscription for based on anything from movie reviews of different websites to age groups of the target audiences for each of the movies.

Technologies Used: Visual Studio Code, Python

Legal & Security Risks: None

**Section 2a:**

Must Have Requirements: “shall”

* The program shall prepare .txt documents of data based on a number of different parameters available in the dataset.
  + The various parameters used to make the .txt documents shall also be used to create the targets for machine learning classification (no targets for classification currently exist in the dataset and must be made from scratch).
  + These text documents shall all contain at a minimum: the names of movies it pertains to, the movies’ rating on at least one of the two available movie review sites, the countries the movies are available in, and the age group of the target audience.
  + This shall be done before the user attempts to search for shows, streaming services, or anything else they may search for.
* The program shall filter out unnecessary data before preparing the .txt documents.
  + i.e. duplicate data, data that is the same for every movie in the dataset, etc.
* The program shall create the .txt documents with machine learning using a classification technique.
* The program shall have a separate code file for the preparation of each .txt document and for the code to allow the user to search each document.
* The .txt documents shall all include the streaming services the various movies are found on.

**Section 2b:**

Stretch Requirements:

* The user may be able to create their own .txt document based on their own given parameters.
  + If this stretch requirement is done, then the code for it shall be in its own code file.
* The user’s search results that are displayed in the terminal and the search terms used may be written to a .txt file under a name given by the user.

**Section 2c:**   
Weekly schedule:

|  |  |  |  |
| --- | --- | --- | --- |
| Task | Approximate Hours Required | Estimated completion date | Criteria for Completion |
| Research and project proposal | 32 | 2/6/21 | Project Proposal written, submitted, and approved |
| SRS Documentation | 10 | 2/15/21  (during week 7) | SRS Documentation written, submitted, and approved |
| Read and write csv file of dataset | 5 | 2/17/21  (during week 7) | Program can read the data from the csv file and write any given data to a .txt file |
| Determine exactly how many .txt files will be needed and the parameters needed for each one (determine the conditions for each classification target) | 2 | 2/19/21  (during week 7) | Parameters and count of expected .txt files determined and recorded |
| Create the classification algorithm for the data to be written to the .txt files | 40 | 3/4/21  (during week 9) | Classification algorithm successfully writes the correct data into the correct .txt files |
| Design the program to allow the user to search based on possible search parameters | 20 | 3/18/21  (during week 11) | User can search using a number of search parameters and get accurate results back |
| Do the stretch requirements | 17 | 4/7/21  (project delivery date) | Stretch requirements are met and fully functioning |
| Project complete | 126 hours total | 4/7/21  (project delivery date) | Minimum of section 2a (“shall”) requirements met and fully functioning and submitted for grading |

**Section 3:** Design Overview of the Product.

Workflow: The user begins by opening up Visual Studio Code and opening the python code file to search the .txt documents in a tab of Visual Studio Code. Upon running the code file for the search, the user will be given a prompt in the terminal. The prompt will ask if the user wishes to search existing documents or do their own manual search (note: the manual search prompt shall only be given if that stretch requirement is met). If the user chooses to use existing documents, they will receive a new prompt for inputs for the various variables that go into each document. Once they’ve entered data for the prompts, the data from the corresponding .txt document will be displayed in the terminal. If the stretch requirement is met and the user selects the option for their own custom search, they will receive prompts for each of the possible variables. Each of these prompts will ask if they want to search by the given variable and if so which possible value they want to search by. The results will be displayed in the terminal. They will then be given a prompt that asks if they want to write the results to a new .txt document. If they say yes, they will be given a new prompt asking what name they want the .txt document to be given.

Resources:

* Visual Studio Code – a text editor that also runs code of various coding languages.
* Python – The user will need the most up to date version of python installed on their machine for the code to run.
* Github – Version control and program/file download.

Data at Rest: Sorted data from the dataset will be stored in separate .txt files. If the corresponding stretch requirement is fulfilled, the user’s search terms and results will be stored in .txt files at the user’s discretion.

**Section 4:** Verification:

**Demo:** To ensure that the data is properly sorted, the developer will enter search terms to get data from the .txt files and ensure that the results are accurate.

**Testing:**

**Sources/Citation/Resources** Links:

<https://www.kaggle.com/ruchi798/movies-on-netflix-prime-video-hulu-and-disney> (link to the website where the dataset can be found and downloaded)