Katie Lunceford

Final Write Up

To start the program, run ProjectGUI.py, and the interface will pop up. Across the top, there is a place to type in files, as well as to add the author, genre, and year for those files. The options for genre are short story, fiction, science fiction, horror, poem, or other. Press “Add Document” when you have typed everything in. You can add as many documents as you would like before applying filters and building trees.

To apply filters, you must first select what filters you want to apply. Click on the button for each filter that you want to apply, one by one. After you have clicked on each filter that you want to use, press “APPLY FILTERS”, and every filter that you have clicked will be applied to all of the documents that you read in.

For training, you must first click on one of the nine training methods. Then, click train, and it will train based off of the most recently clicked method using all of the files that are listed at the bottom. Similarly, predict will make a prediction based off of the most recently clicked on method. That method must’ve been trained in order for a prediction to be made. Type in the file that you want to predict, above the prediction button.. The year and the genre are optional if you are not predicting based off a tree that was built using year or genre. All 9 of the methods will print out the prediction in the python shell, as well as add the document and predicted author to the list of documents in the bottom. Additionally, SKPCA will pop-up with a scatter plot with the trained files in blue and the predicted file in red, so that you can see the distance it is from other files.

This project was very helpful at understanding the uses behind object-oriented programming. It would have been virtually impossible to build the program without having numerous different classes with various methods. Additionally, I had very little knowledge of both GUI interfaces and authorship attribution methods before this program, and I learned a ton about them. I am interested in going into data science in the future, so these skills will be incredibly helpful in my future.

The thing I struggled with the most is conceptually understanding how all of the pieces were going to fit together in the end. I eventually got through this by breaking everything down into smaller functions or methods and keeping my work organized and well labeled in all of the different classes. Additionally, if I was ever struggling with algorithms or technical issues, I learned a lot about using online resources to find answers to my questions by googling or use stackexchange. In conclusion, the project was helpful in applying course material such as heaps or linked lists to a project that is related to a field that I am interested in entering into in the future.