Outline of how to use our program:

When you run the main.py file, our UserGUI will automatically run and show the first window. It’s easy to see that there are 4 parts that users need to give inputs. First, number of documents to be analyzed needs to be the sum of both documents to be trained and those to be predicted. Second, it’s time for the users to choose which statistical method they prefer to utilize to analyze data and thus to predict the authors. Third, for ID3 and SKTree method, users need to enter the attributes they are going to use to train and predict the files. In order to make the prediction accurate, we suggest users to enter at least 2 attributes. All attributes should be entered in form of a string with comma. Finally, for SKPCA method, users can choose multiple text filters they want to use when analyzing the files.

After entering enough information and clicking the confirm button at the bottom of the window, our second window will appear automatically. There are five columns in the window. The first column is for the users to input the name of documents they want to train or predict. If a document users enter should be used to train our program, then check the box in the third column, which is ‘For training?’. If a document users enter should be used for prediction, then check the box in the fourth column. Users should make sure that a document could not be used for training and for predicting at the same time. If users choose ID3 or SKTree method in the first window, they need to enter the value of attributes for training in the form of string with comma in the second column. If users choose SKPCA method, they need to enter the author name for training in the second column. It’s important that commas are necessary here, since our program identify two attributes according to commas. Finally, after entering necessary information and checking boxes, hit the ‘Run’ button and the authors’ names of documents for predicting will be displayed in the ‘RESULT’ column. If users want to run the entire program again, they need to exit the window and rerun the main.py file again.

What we learned:

In general, we learned how to conduct a project from the beginning to the end, including project design, utilities breaking down, sub programs and classes establishing, GUI making, programs combining, debugging and so on. In details, it’s important for us to have an experience of conducting an entire project on ourselves, allowing us to know the procedure of software production. Furthermore, we put the knowledge learned from class into practice, enhancing our learning of course material. Also, we improved the ability of pair programming, which offered us a good opportunity to find our fallacy, and of outside-of-class learning, such as Googling. Besides, we found that pair programming is efficient and effective, especially when debugging, since it’s hard for a coder to find his or her own typing problems. Finally and significantly, we learned that the power of computer science is huge and is able to make our life convenient and comfortable, providing us energy and power to study harder and learn more.