Ryan Fleury

000758292

Senior Projects 2

Final Project

**Software Purpose and Vision:**

The main purpose of the program that I created was to analyze the emotion of text documents through machine learning. My vision for this project was to create a text editor of my very own that could accomplish this task. I wanted to create this project because I wanted to see how far this could be taken in a topic that I have an interest in.

**Functional and non-functional requirements of the application:**

The following requirements that are below and underlined are the ones that I set out to accomplish this summer:

Functional Requirements:

* The program must analyze text documents text for emotion.
* The program must have a graphical interface that acts as a text editor.
* The program must highlight text based on the emotion grade it receives.
* The program must have a way to recognize the amount of input text in the program. (Impacts accuracy of prediction lower amount of text present).
* The program must pull text from multiple sources to eliminate Bias.
* The program must be able to analyze text regardless of the amount of input text.

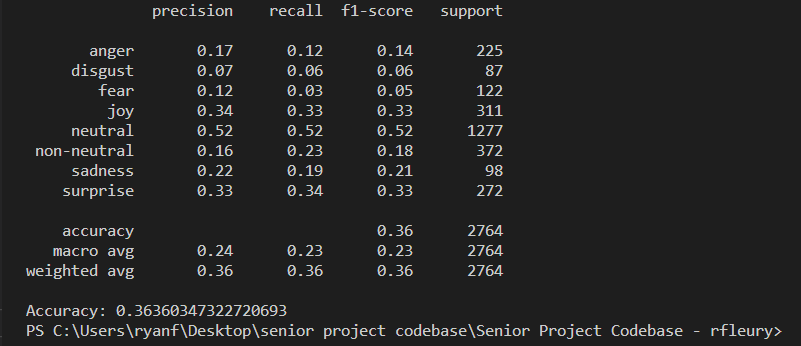
Non-Functional Requirements:

* The program shall be easy to use.
* The program GUI shall look much better than previous iterations of the program.
* The program shall be able to perform more efficiently with large amounts of input text.
* The program shall adapt better to being used on multiple systems.
* The program shall have embedded instructions on how to make it function.
* The program shall accept multiple types of documents. (Word and .txt)
* The program shall have a prediction accuracy higher than 35%.

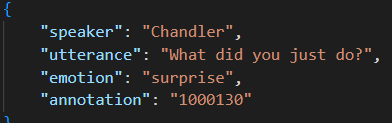
**Analysis and approach to design:**

For my program, my main analysis this summer was to improve the prediction accuracy of my program. I was able to achieve better accuracy overall but had some drawbacks to accomplishing this. The is a large amount of input text that is needed to achieve the prediction accuracy I was able to achieve of 36% overall.

Accuracy data for each emotion:

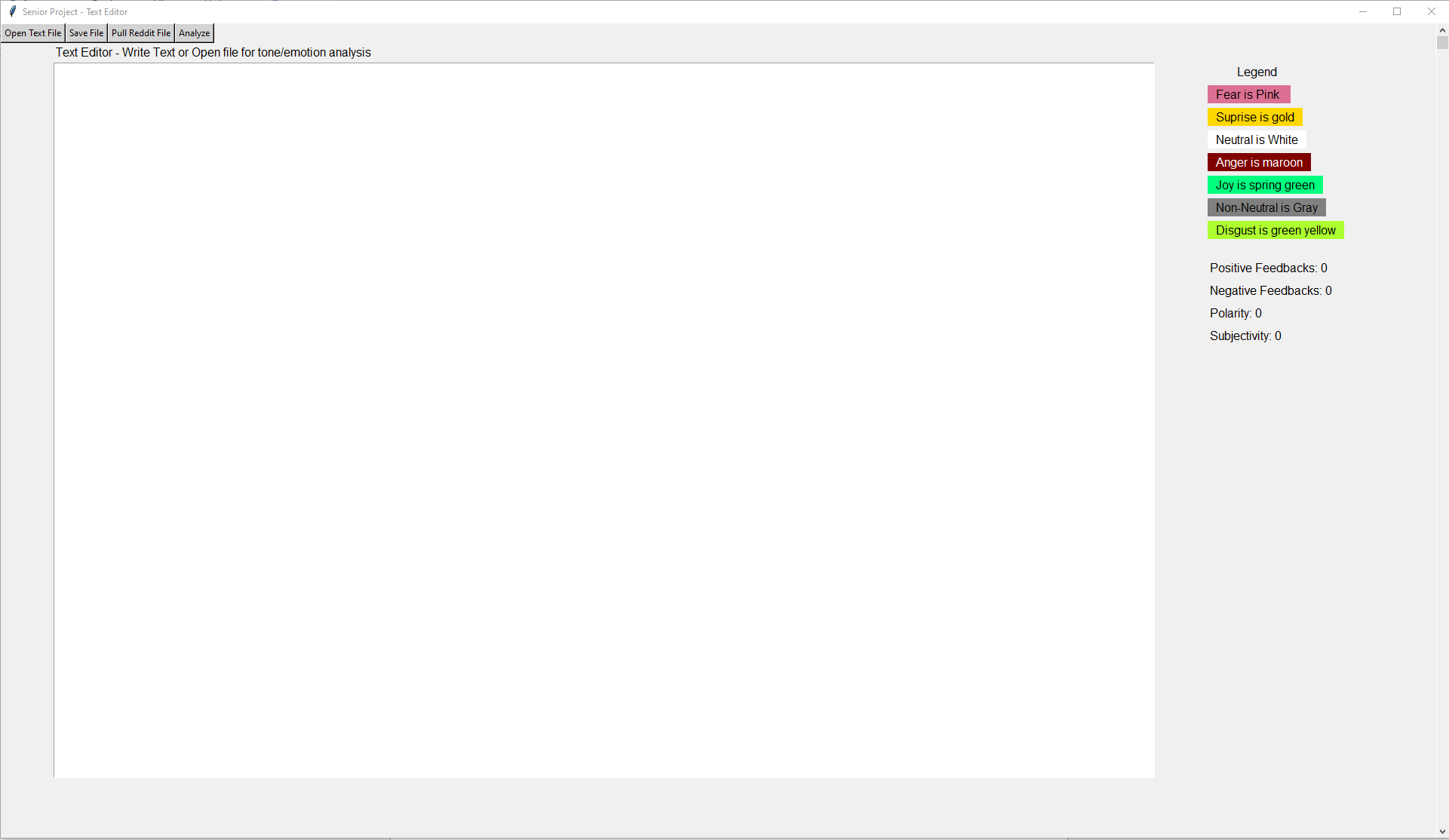


I aimed to improve this accuracy through a better dataset, but I was unable to achieve this goal. The dataset I have is used to train for the emotion of the sample text. An entry from the dataset looks like the following:



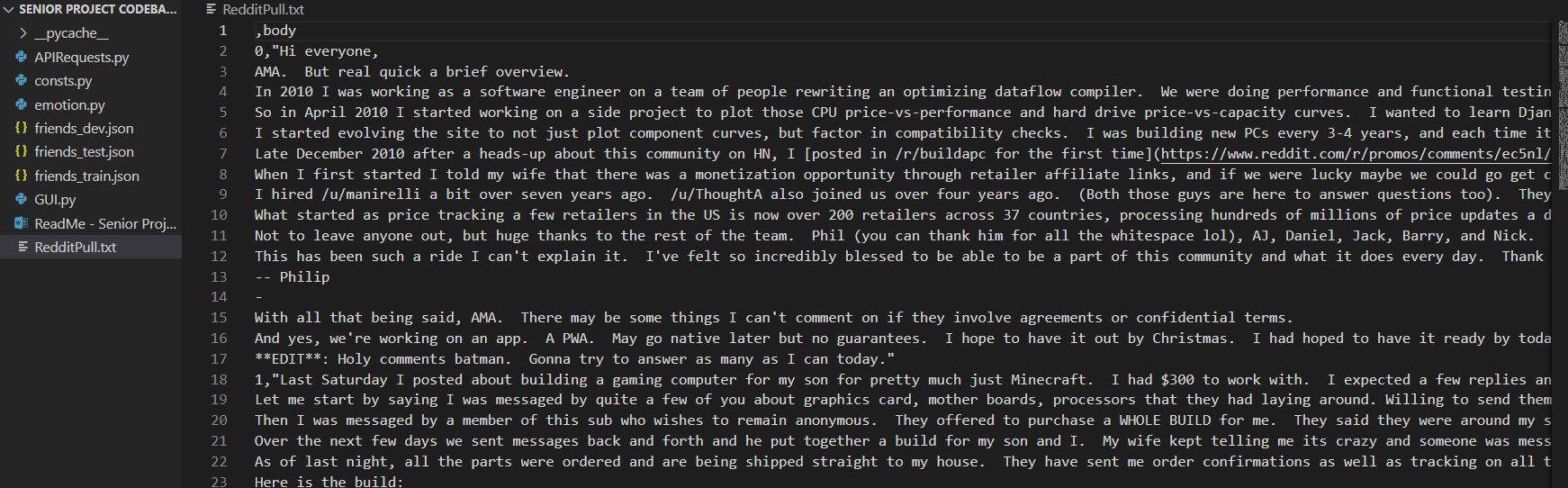
The sentence and emotion are used to train a kernel to predict the emotion of text entries input in the program. The training is done with the library sklearn and uses the support vector machines algorithm to train.

With a trained kernel the python library tkinter is used heavily to make an interactive GUI have text inputted into. This library allows all of the functionalities to make a text editor GUI that is customizable. I completely overhauled the interface this summer and the screenshot below is a capture of the program:



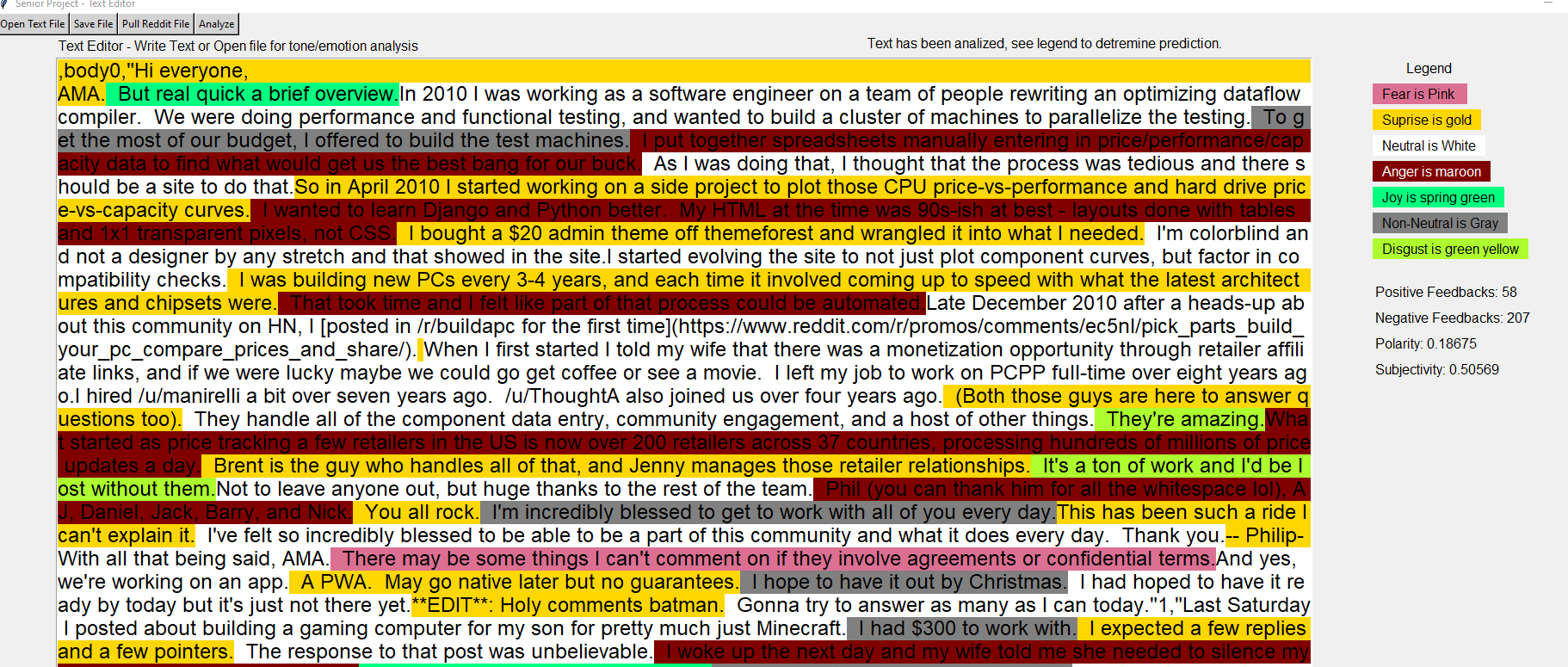
This interface is more adaptive to most computer systems as well as having a much cleaner look than before.

This program required input text to function so it is integrated with Reddit to make this happen. At the press of a button, it will gather large amounts of Reddit data to be analyzed for emotion. It will also pull from multiple subreddits at random to make sure different types of text are generated. When the “Pull Reddit File” is pressed a text file will be generated in the directory of the project that can be opened into the program. This is what the file looks like:



When this data is analyzed, the program will preprocess and grade the data for emotion. The text does not have to be from Reddit, but it is one way to ensure that random text is generated for the user. Each sentence is parsed and plugged into the support vector machine's kernel to be graded. Based on the grade that is received dictates which emotion highlight the sentence will receive.

When text is graded with the Analyze button the output looks like the following:



**Diagram for program layout:**

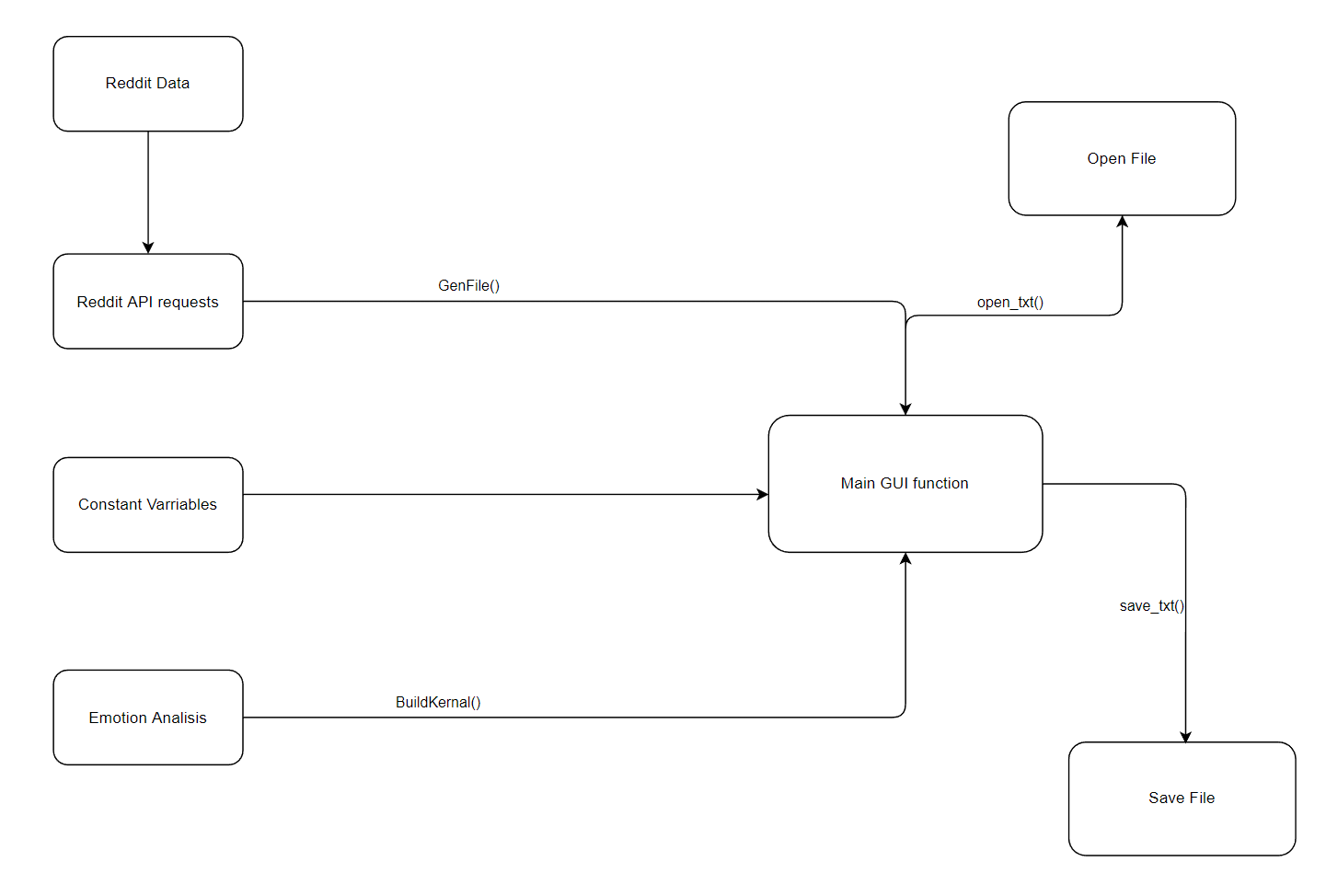
GenFile() -- Generates a Reddit file

BuildKernal() -- Generates a SVM kernel to grade emotion

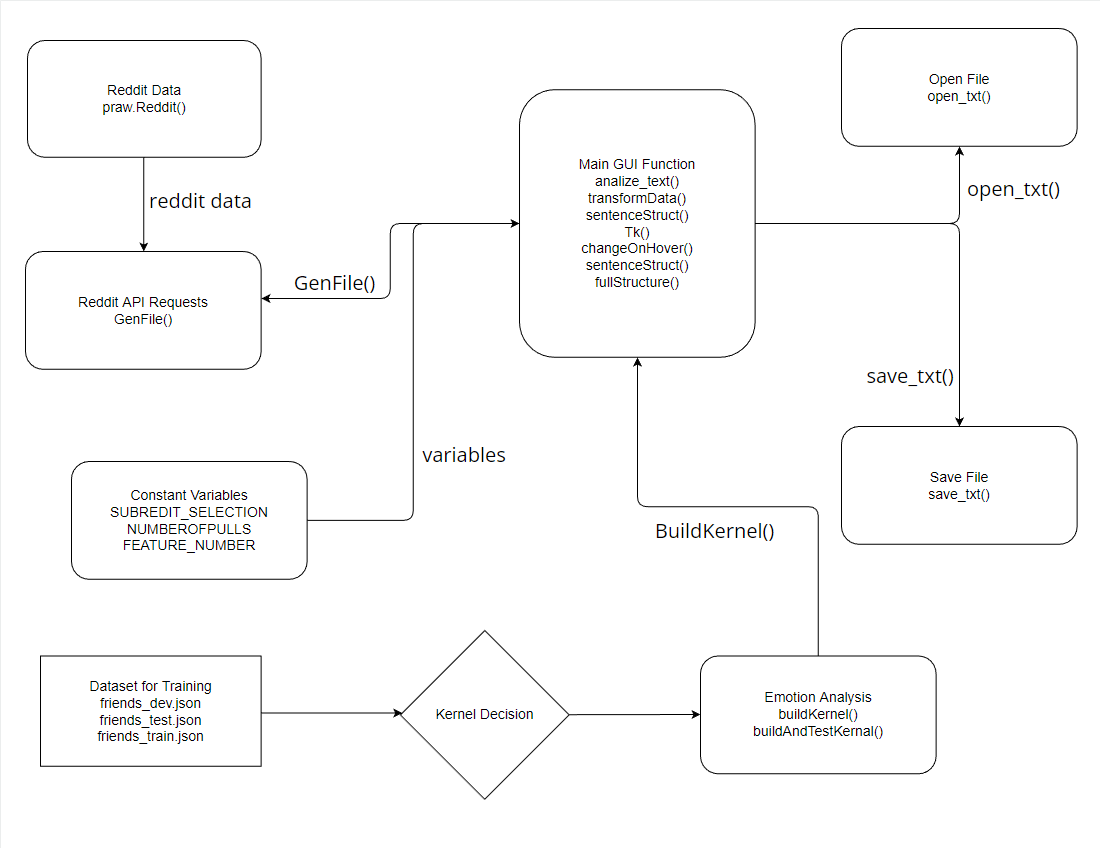
Open\_txt() -- Opens a word or text file

Save\_txt() -- Save to a text file

**Low-Level Design**



**In Depth Design**



**Solution:**

My solution achieves all the original goals that I set out to accomplish. I was unable to reach the prediction accuracy that I had set out to. However, all aspects of the program in other aspects have improved. The functionality, look, and adaptability are all things I was able to improve over this summer.