Brian Kilburn

CSC 546 Artificial Intelligence

Dr. Wooky

In the python program you will need a too install a few libraries to successfully run the program, the import commands are already included in the code. The main library of concern is to make sure that pandas and NLTK (Natural Language Tool Kit) packages are installed.

If pandas are not installed, you can run pip install:

1. In the command console navigate to scripts folder and execute the following command: **pip install pandas**, if you have a new pip install you can use **pip3 install pandas**.

Fortunately, NLTK is easier to install. If you do not have the package.

1. Open the SMS Filter python code (**detection\_filter.py**) in an IDE.
2. Uncomment line 10 where it should say: **lang.download(‘all’)**. Then wait for the packages to install and try running the program again.

Installing the csv and arff libraries are equally as important. To install these packages, the same process will need to be followed.

**CSV:**

1. Navigate to the scripts folder if you are running in an anaconda environment and execute the following: **pip install csv**

**ARFF:**

1. Navigate to the scripts folder if you are running in an anaconda environment and execute the following: **pip install arff**

Once your library packages are installed you are ready to run the program. Included in the file should be the raw spam csv data file. That will be the file initially used to execute the program. When you run the program just make sure that your data file is in the same directory as the program files. Also make sure that the data file either has the same name in the **read\_csv()** function or change the value to correspond with the name of your data file. After these steps have been completed the program should now execute on your machine. When the program has finished executing two files will be exported, **SpamProcessedData.csv** and **spam.arff**. Next you will be ready to run the R program, ensure that the working directory is set in the same directory as your exported csv file, which will be the relative directory containing the feature program. If you do not set the working directory in R correctly the program will not execute.

To set the working directory in Rstudio, just navigate to the bottom right pane and select the **Files** tab, next click the ellipsis (**…**) in the Files tab window pane and select the same directory relative to the program. After the directory has been selected, click the **More** dropdown option and select **Set As Working Directory,** once the new working directory has been set the Rstudio program **regression.r** file will be ready for execution.