Aira Domingo
Data Mining
Final Project Proposal

For the final project, I have chosen to answer the question, “How is the amount of carbon emission affected by a country’s land use and population?”. I chose to ask this question because I think it’s important to see how our consumption behavior is affecting the planet. By answering this question, we will know which factor contributes the most to carbon emissions.

To answer the question, I am using the ‘National Footprint Accounts 2018’ on Kaggle. The information on this dataset was gathered from the Global Footprint Network (www.footprintnetwork.com). I will need to clean the data prior to data mining and evaluation.

I will be using linear regression as my data mining algorithm. I will most likely use the standard algorithm.

To implement the network, I will be using PyCharm because it has an integrated debugger and code completion.

I will be using information from the Global Footprint Network (www.footprintnetwork.com) to get background information to solve the problem.

I will use R-squared, Root mean squared error (RMSE), F-test, and check for multicollinearity to judge the performance of my model.

The following is my rough schedule for completing the project:
April 7th: have the preprocessing done
April 11th: data mining, evaluation, and result exploitation done
12th-13th: GUI
17th-19th: write the report
19th-22nd: presentation and individual report