**Pragna Goura**

**Predicting Carbon Emission by Country**

1. **Introduction:**

A hot topic in our political conversations is the environmental crisis of our planet, specifically CO2 emissions. This paper aims to create a machine learning algorithm that predicts the CO2 emissions country by country, analyzes which country or countries emit the most and least CO2 emissions as well as predicts the amount that will be produced over the coming years.

1. **Data Set**

This data set is a CSV file with all the countries and their year-by-year carbon emission and what fossil fuels they are using. There are 11 variables and over 60000 instances with a 2.96 MB file.

1. **Data Preparation**

The first step I will take is data cleaning because the data sets give us carbon emissions from 1750, there were no technological advancements at the time that would cause any extensive emissions. I will look through the data for missing values and outliers and take care of them as I see fit.

1. **Exploratory Data Analysis**

Next, I will start data visualization, plotting each country’s emissions and what fossil fuel contributes most to the total pollution created. EDA is a very important tool in this process because it helps the program visualize trends and patterns that may not be picked up by the human eye. Making sure we find a trend for every fossil fuel used, for every year produced, for every country will help create a well-balanced analysis of how modern-day fossil fuels affect our global environment.

1. **Modeling**

The machine learning modeling algorithms I will use are regression, decision trees, and neural networks. The goal of using these models is to find what country and what specific fossil fuel will cause the most carbon emission in the next 50+ years. Even after data cleaning, I will have a lot of variables and over 100 countries to analyze for 200 years, neural networks will help with this large data set while decision trees may give us a less accurate result but will give us a general idea as to how our data will be presented.

1. **Evaluation**

The goal of this project is to predict how much CO2 will be produced 50+ years into the future, although there is no true telling of what could happen in this scenario after all this prediction relies on all factors staying the same in our future, we can use the F1 score to determine how well our data has been analyzed.