Global CO_2 emission and its relation to climate change

Maitrey Talware, mt3915
Shivam Bawane, sb6818

1. What is the problem ?

Problem:
To predict and analyze global CO_2 level and global temperature level by analysing data from previously recorded data.

Motivation:
To create a system that predicts and analyses global temperatures and global CO2 levels, in order to tackle Global Warming. According to the U.S. Energy officials, in 2017 the US emitted 5.1 billion metric tons of energy-related carbon dioxide, a good number that contributes to increasing wildfires, rising sea levels, catastrophic hurricanes and changes in the climate. According to National Geographic, the average temperature of the Earth is rising at nearly twice the rate it was 50 years ago. This rapid warming trend cannot be explained by natural cycles alone, scientists have concluded. The only way to explain the pattern is to include the effect of greenhouse gases (GHGs) emitted by humans.
This boils down to the day to day activities of humans. We aim to analyze previous as well as current data on the global CO2 levels and create awareness about Global Warming. We aim to show a relation between the amount of CO2 and the overall rise in global temperature. This analysis can be used further to decide national/global climate change policies.

Potential outcome:

We plan to:

1. Predict CO_2 emission for each country according to their historic data
2. Predict temperature for each country according to their historic data

We plan to perform data analysis extensively. Some of which are:

1. Interactive graph to visualize "How CO_2 emission is rising over time" to grab attention of stakeholder
2. Top countries contributing to CO_2 emission
3. Top countries affected by the temperature change

2. How will you learn the background?

We will be referring to these organizations for domain knowledge

Carbon Dioxide Information Analysis Center

It is an organization within the United States Department of Energy that has the primary responsibility for providing the US government and the research community with global warming data and analysis as it pertains to energy issues.

United Nation Climate department

The UN is at the forefront of the effort to save our planet. In 1992, its "Earth Summit" produced the United Nations Framework Convention on Climate Change (UNFCCC) as a first step in addressing the climate change problem. The 197 countries that have ratified the Convention are Parties to the Convention. The ultimate aim of the Convention is to prevent "dangerous" human interference with the climate system.

3. What kind of data will you use?

a. CO_2 :

Annual CO2 emission(mt) across countries and years
http://www.globalcarbonatlas.org/en/CO2-emissions

Attribute Name	Attribute Type	Attribute Details
Country Name	Categorical	Name of the Country
Year	Ordinal	Year
Territorial emissions in $MtCO_2$	Quantitative	CO_2 emission in Metric Ton

b. Temperature :

http://berkeleyearth.org/data/

Attribute Name	Attribute Type	Attribute Detail
Date	Ordinal	Date
AverageTemperature	Quantitative	Average temperature for that month of year
AverageTemperatureUncertainty	Quantitative	Confidence interval around the average
Country	Categorical	Country name

4. **What kind of model will you build?**

   Methodology:

   Dealing with missing values by replacing them with suitable values by doing analysis on data.

   Converting categorical text data into numeric form using methods like one hot encoding.

   Scaling/ Standardising features so that the features are on the same scale and each feature contributes approximately proportionately.

   Data transformation - We also need to add a few additional columns like latitude, longitude, etc.

   Model:

   A. We will use AutoRegressive Integrated Moving Average (ARIMA) to predict global CO2 emission and global temperatures.
      Since the data is time series data, ARIMA can be used to forecast the data. We aim to select best p and q values according to the dataset.
   B. Long Short Term Memory (LSTM) networks can also be used

   As we progress through the project we plan to compare various other models and select the best one.

   Evaluation:

   1. Divide data into train and test set,
   2. Train data on training set and predict data on testing set.
   3. Calculate Mean Absolute Error and Mean Square Error to check the performance of the model

5. Why not any other approaches ?

1. Since, we are using time series data to predict quantitative data, we cannot use classification algorithms.
2. Also, regression methodologies cannot make use of time to predict target variable.