DESCRIPTION OF THE PROBLEM

With a primary focus on greenhouse gas emissions, particularly carbon dioxide (CO2). These emissions result from a variety of human activities, including the use of fossil fuels for energy, transportation, manufacturing, and deforestation. Rising global temperatures, more frequent and severe extreme weather events, the melting of the polar ice caps, an increase in sea level, and disruptions to ecosystems and biodiversity are some of the negative repercussions that follow. To ensure a healthy and habitable planet for both the present and future generations, addressing the emissions problem is a global necessity that requires cooperative efforts to reduce and ultimately eradicate these harmful emissions.

WHAT IS THE MODEL ABOUT

A regression model would be a good option to forecast CO2 emissions from a dataset gathered in Rwanda. Multiple linear regression is a frequently employed regression model for this kind of prediction. By fitting an equation to the data, multiple linear regression enables you to investigate the link between one dependent variable (CO2 emissions) and numerous independent factors (such as carbon monoxide, nitrogen oxide, atmospheric conditions, etc.). You can use this model to generate quantitative forecasts based on the dataset and understand how various factors affect CO2 emissions. To address potential problems like multicollinearity and model complexity, you could also investigate various regression algorithms.

USEFULNESS OF MODEL

There are various reasons why using a regression model to forecast CO2 emissions is important. First off, it is an effective weapon in the struggle against climate change on a worldwide scale. Researchers and policymakers can decide how to lessen the environmental impact of CO2 emissions in Rwanda by having a better understanding of the elements that drive these emissions. Secondly, a wide range of stakeholders, including academics, decision-makers, public figures, and international climate organizations, can benefit from the model. It equips decision-makers to develop efficient policies and regulations targeted at lowering emissions and attaining sustainability goals by supplying insights into the environmental impact.

DATA SOURCES TO USE TO TRAIN MODEL

We obtained the dataset from Kaggle.com | CO2 Emissions in Rwanda

Our project is predicting CO2 emissions using predictive modeling. The factors we've identified encompass a wide range of variables, both atmospheric and pollutant-related, which can significantly influence CO2 emissions. Relevance of each factor is discussed briefly below:

1. **Sulfur Dioxide (SO2)**: SO2 is a precursor to sulfuric acid formation and can contribute to air pollution. Understanding its levels can shed light on industrial emissions and their environmental impact.
2. **Carbon Monoxide (CO)**: CO is a product of incomplete combustion and is a key indicator of combustion-related emissions. Monitoring CO levels helps assess the efficiency of energy and transportation systems.
3. **Nitrogen Dioxide (NO2)**: NO2 is a major component of smog and contributes to respiratory issues. It's an important pollutant to track due to its links with emissions from vehicles and industrial sources.
4. **Formaldehyde**: Formaldehyde is a volatile organic compound (VOC) and is a crucial air quality parameter. It is linked to various health concerns and is used to gauge indoor and outdoor pollution sources.
5. **Aerosol Content**: Aerosols affect climate and air quality. Measuring aerosol content is valuable for understanding their role in emissions and atmospheric processes.
6. **Ozone**: Ozone levels are relevant as they are both a greenhouse gas and an air pollutant. Its presence in the atmosphere is a key factor in air quality and climate change.
7. **Atmospheric Features**: Monitoring atmospheric features such as cloud cover can provide insights into how weather patterns impact emissions. Clouds can affect the amount of solar radiation reaching the Earth's surface, which, in turn, influences CO2 concentrations.