Team Mates

- Neha Nooka
- Harshali Narkhede
- Chaturya Gajula

Project Idea

1) **Climate Change and Crop Yield Prediction**

Dataset: faostat.fao.org

Suggested Dataset : 1.) Climate Change Data:
https://www.fao.org/faostat/en/#data/QCL

2.) Crop yield Data
https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data

- Basic Info: FAO Statistical Databases and Kaggle's Earth Surface Temperature Data contains data on climate and crop yield. We plan on merging both the datasets and using the collected data.

- Data Size:
Climate Change Data: 1743-2015 (>1M rows)
Crop Yield Data: 1961-2022 (~79607 rows)

- Attributes:
Climate Change Data:
1)Date
2) Land Average Temperature
3) Land Average Temperature Uncertainty
4) City
5) Country
6) Latitude
7) Longitude

Crop Yield Data:

1) Country/Region
2) Year
3) Crop(TBD)
4) Area Harvested
5) Yield
6) Production Quantity

- Analysis Type: Predictive Modeling (Regression)

- Objective: Build a model to predict crop yields (e.g., corn, wheat, rice) based on historical climate data and climate change projections. By analyzing the relationship between climate variables and crop yields, the model aims to forecast the impact of climate change on agricultural productivity.

- Tasks: Data preprocessing (including feature engineering and handling missing values), model selection (e.g., linear regression, random forest, neural networks), and model evaluation.

- Performance Evaluation Metric: Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) can be used as performance evaluation metrics to assess the accuracy of crop yield predictions. These metrics quantify the average prediction error between actual and predicted crop yields, providing insights into the model's predictive capability.

2) Forecasting Solar Energy Generation Using Machine Learning

Dataset : https://www.nrel.gov/grid/solar-power-data.html

Dataset Contains

1.) Timestamp
2.) Energy(mV)

Description:

To create forecasting models that can precisely calculate solar farms' energy generation depending on a range of environmental conditions. The quantity of power generated by solar panels over various time intervals will be predicted using machine learning algorithms that can be trained using historical data on solar energy, weather.

Steps:

The project includes Data collection and Data preprocessing, selecting and training models by using a variety of machine learning approaches, conducting exploratory data analysis (EDA) to determine the correlations between variables, and conducting an in-depth analysis to evaluate model performance.

Outcomes:

To maximize energy management and grid stability, as well as to help make decisions that are well-informed about investments in renewable energy. Reliable solar energy generation projections enable policymakers, energy suppliers, and grid operators to better anticipate changes in supply and demand, which improves resource allocation and reduces dependency on non-renewable energy sources.

Conclusion:

The project includes Data collection and Data preprocessing, selecting and training models by using a variety of machine learning approaches, conducting exploratory data analysis (EDA) to determine the correlations between variables, and conducting an in-depth analysis to evaluate model performance. The final result of the created models will be their deployment to offer forecasts of solar energy generation in the future or in real time, therefore advancing renewable energy strategies and facilitating the shift to a more resilient, cleaner energy infrastructure.