Erdos Data Science Bootcamp Summer 2024 Executive Summary

Team: Md Mehedi Hasan, Kamlesh Sarkar
Github: https://github.com/mmh52/Forecast-Solar-Energy

Overview: Solar energy, a renewable source of power, plays an important role in reducing greenhouse gases, mitigating climate change, and protecting ecosystems. Nowadays, the adoption of solar energy into the power grid has increased, and Direct Normal Irradiance (DNI) is particularly important in forecasting the performance of concentrating solar power (CSP) systems. Photovoltaic panels track the sun to receive more DNI, which accounts for a large portion of solar energy from PV.

Objective: The objective of our project is to select an effective time series model that can forecast a week ahead of Direct Normal Irradiance from solar power. This is crucial for the effective operation and maintenance of power systems, ensuring their ability to harness solar energy effectively.

Stakeholders: Governments and Regulators, Managing Agents, Manufactures, Investors, Local Business, and Universities.

Modeling Approach: we use a dataset from Lowery Power Station, Denver, Colorado, covering the period from June 2008 to December 2013. This dataset contains Direct Normal irradiance, Globa Horizontal, and other important features of Solar Energy

- This dataset contains minute data of direct normal irradiance and has some anomalies. We conducted a thorough data cleaning process, fixing anomalies, filling in missing values, and converting it to an hourly dataset.
- Check the seasonality and trend of the data. Plot autocorrelation and partial autocorrelation to find the seasonal pattern of direct normal irradiance.
- We take DNI's data from the last week of December 2013 as a test set and the remaining data as a train set.
- To forecast a week ahead DNI, we choose some time series forecasting models such as Exponential smoothing, ARIMA, SARIMA, and SARIMAX

Results: Among all the time series forecasting models we applied in our project, the SARIMAX model is much more accurate than others in forecasting DNI. Regarding RMSE, SRIMAX is approximately 39.09% and 59.3% better than SARIMA and Tripple exponential Smoothing, respectively. The exogenous variable significantly contributes to improving the model's performance.

Future Work:

- Incorporate weather and cloud cover data, which can be important features for forecasting direct normal irradiance.
- Use multivariate deep learning models that can incorporate all important features of solar energy and forecast direct normal irradiance more accurately.