Project and Data Management (PDM) Plan

Project Overview

PROJECT TITLE: SOLAR ENERGY PRODUCTION FORECASTING USING MACHINE LEARNING AND METEOROLOGICAL DATA

SUMMARY OF THE PROJECT TOPIC AND BACKGROUND: While solar energy is an important part of the transition to renewable energy sources, it is highly variable based on weather changes during the day. Forecasting Solar Power Production has become a popular source of sustainable energy, and accurately forecasting its production is crucial. This project focuses on developing machine learning models for predicting solar energy output based on historical solar power generation data and meteorological parameters. Using historical trends and weather conditions like solar irradiance, temperature, and humidity, the project aims to produce short-term energy forecasts to help solar farm operators and energy planners.

RESEARCH QUESTION: How accurately can machine learning models predict solar energy production based on historical meteorological and solar power generation data?

PROJECT OBJECTIVES:

Data Collection & Preprocessing: Obtain and merge solar power generation and weather data from credible sources (as loaded in the notebook), clean and preprocess the data (handling missing values, formatting datetime, scaling features) for further analysis.

EDA: Explore trends, correlation, and patterns between meteorological variables and solar energy output using visualizations and descriptive statistics.

Feature Engineering: Extract and select relevant features impacting solar power production, potentially including solar irradiance, temperature, humidity, and wind speed, as used in the model inputs.

Model Implementation & Comparison: Implement and train multiple machine-learning models, specifically **LSTM** and **Random Forest (Regressor and Classifier)**, to predict solar energy output.

Model Validation & Optimization: Assess model performance using metrics such as RMSE, MAE, and R^2 (as calculated in the notebook), and potentially optimize hyperparameters to improve accuracy.

Deploy & Interpret: Appropriately deploy the best model (though deployment is not shown in the notebook), interpret results, and deliver insights into how to use these predictions for real energy management in the field.

REFERENCE LIST:

1. Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F. and Antonanzas-Torres, F., 2016. Review of photovoltaic power forecasting. *Solar Energy*, 136, pp. 78-111. Available at: https://doi.org/10.1016/j.solener.2016.06.069 [Accessed 10 February 2025].

2. Voyant, C., Notton, G., Kalogirou, S., Nivet, M.L., Paoli, C., Motte, F. and Fouilloy, A., 2017. Machine learning methods for solar radiation forecasting: A review.