Name: Vankatasaisireesha Bolisetty

Student Id: 23036961

1. Project Overview

1.1 Project Title

Machine Learning-Based Prediction and Optimization of Renewable Energy Production

1.2 Summary of Project Topic and Background

Renewable energy is crucial for reducing carbon emissions and achieving sustainability. However, predicting energy production from solar, wind, and hydro sources is challenging due to weather variations and seasonal patterns. Traditional forecasting models often fail to capture these complexities, whereas machine learning (ML) approaches offer higher accuracy by identifying patterns in historical data. This project applies ARIMA, LSTM, and XGBoost to develop predictive models for renewable energy production and optimize distribution. Using data from the U.S. Energy Information Administration (EIA), this study will preprocess energy records by handling missing values, normalization, and trend analysis. Model accuracy will be evaluated using RMSE and R^2 scores to ensure reliable predictions. This research will provide valuable insights for policymakers and energy providers, aiding in better energy distribution planning. By integrating ML-based forecasting, the project aims to enhance efficiency and reliability in renewable energy integration.

1.3 Research Question

How can machine learning techniques be applied to improve the accuracy of renewable energy production forecasting and optimize energy distribution strategies?

1.4 Project Objectives

- **Data Collection & Preprocessing:** Gather and clean historical renewable energy data from the EIA, handling missing values and normalizing features.
- **Exploratory Data Analysis (EDA):** Analyze trends, seasonal variations, and correlations among energy sources.
- **Machine Learning Model Development:** Implement and compare models (ARIMA, LSTM, XGBoost, Prophet) for energy production forecasting.
- **Optimization & Validation:** Apply hyperparameter tuning, cross-validation, and ensemble learning to improve accuracy.
- **Data Management & Security:** Use GitHub for version control and OneDrive for secure backups.
- **Insights for Policymakers:** Provide recommendations for optimizing renewable energy integration based on model results.
- **Final Report & Presentation:** Summarize findings and present key insights for stakeholders.

1.5 References

1. Hyndman, R.J. and Athanasopoulos, G., 2018. Forecasting: Principles and Practice. 2nd ed. Melbourne: OTexts. Available at: https://otexts.com/fpp2/ [Accessed 9 February 2025].

2. Makridakis, S., Spiliotis, E. and Assimakopoulos, V., 2020. The M4 Competition: Results, findings, conclusion and way forward. International Journal of Forecasting, 36(1), pp.54-74. Available at: https://www.sciencedirect.com/science/article/pii/S0169207019301128 [Accessed 9 February 2025].

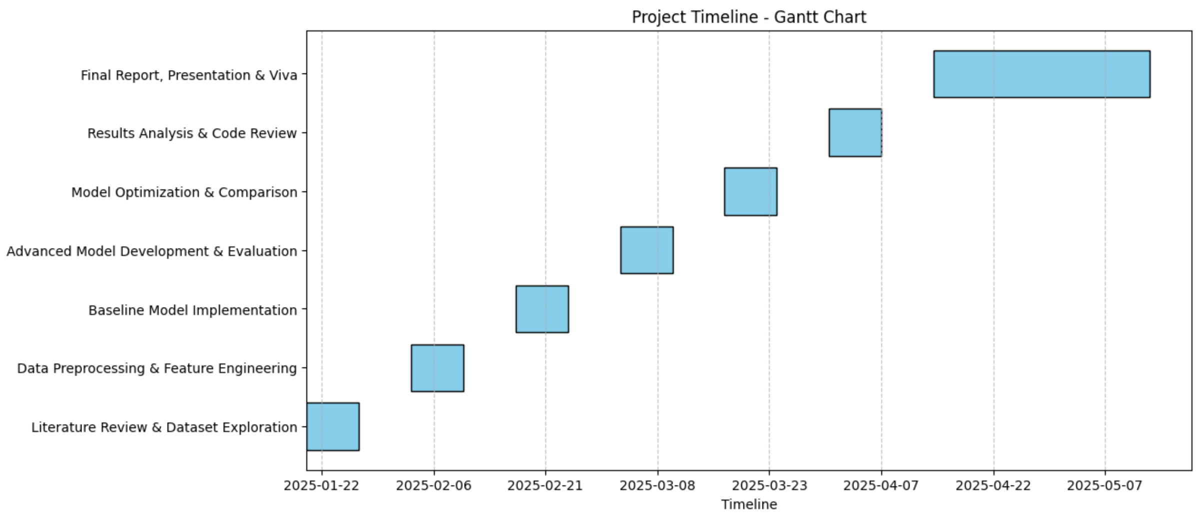
3. Rahman, M.M. and Saha, T.K., 2022. Machine learning-based forecasting of renewable energy production: A review. Renewable Energy Reports, 8, pp.229-248. Available at: https://www.sciencedirect.com/science/article/pii/S2352484722000051 [Accessed 9 February 2025].

2. Project Plan: Task List and Timeline

2.1 Task List

Weeks	Start Date	End Date	Task	Task Description
01-Feb	20-Jan	27-Jan	Literature Review & Dataset Exploration	Reviewing academic papers on renewable energy forecasting, dataset selection, and feasibility study
03-Apr	03-Feb	10-Feb	Data Preprocessing & Feature Engineering	Handling missing values, normalization, encoding categorical features, and feature selection for model input
05-Jun	17-Feb	24-Feb	Baseline Model Implementation	Developing initial predictive models using ARIMA, Linear Regression, and Decision Trees for benchmarking
07-Aug	03-Mar	10-Mar	Advanced Model Development & Evaluation	Implementing advanced models like LSTM, XGBoost, and Prophet for renewable energy forecasting
09-Oct	17-Mar	24-Mar	Model Optimization & Comparison	Hyperparameter tuning, feature importance analysis, and model selection based on RMSE and R^2 scores
11-Dec	31-Mar	07-Apr	Results Analysis & Code Review	Interpreting model results, comparing performance, debugging issues, and ensuring code quality
13-15	14-Apr	13-May	Final Report, Presentation & Viva	Preparing the final project report, creating PowerPoint slides, and rehearsing for viva assessment

2.2 Timeline



3. Data Management Plan

3.1 Overview of Dataset

The dataset for this study is obtained from the U.S. Energy Information Administration (EIA), a trusted and widely used source for energy data. It contains monthly renewable energy production records from

1973 to 2024, covering energy sources such as solar, wind, hydro, biomass, and geothermal energy. This dataset is structured in CSV format, with attributes including year, month, energy type, and production volume. The EIA collects this data for national energy monitoring and policy analysis, ensuring credibility and consistency. The dataset is regularly updated, making it a valuable resource for time-series analysis and machine learning applications. The availability of long-term historical data allows for analyzing energy trends, identifying seasonality, and forecasting future production. For this project, data preprocessing steps such as handling missing values, normalization, and feature engineering will be applied to enhance the dataset's quality before training machine learning models. This structured approach ensures that the dataset is optimized for accurate renewable energy forecasting and optimization. Dataset Link: EIA Renewable Energy Data

3.2 Data Collection

The dataset used in this study is sourced from the U.S. Energy Information Administration (EIA), a credible source for energy production statistics. It includes monthly energy production data from 1973 to 2024, covering solar, wind, hydro, biomass, and geothermal energy. The data is publicly accessible via the EIA Total Energy Data Browser, ensuring accuracy and transparency. The dataset will be downloaded in CSV format and analysed using Python (Pandas, NumPy) in Jupyter Notebook. Data preprocessing steps will include handling missing values, normalizing variables, and integrating additional datasets to improve model accuracy.

3.3 Metadata

- Format: CSV file
- Size: Approx. 100MB
- Attributes: Year, Month, Energy Type (Solar, Wind, Hydro, Biomass, Geothermal), Production Volume

3.4 Version Control and Code Administration

All project advancements will be monitored effectively by GitHub via methodical weekly commits.
GitHub Link: https://github.com/sireesha1010/Renewable_energy_production

3.5 ReadMe File

A ReadMe file will be included in the GitHub repository to provide: Project description, Instructions for dataset usage, Required dependencies for running the code,Explanation of model implementation

3.6 Security and Storage

- GitHub
- OneDrive (for additional backups)
- Backup Frequency: Weekly backups to ensure data safety.
- Data Sharing: Access will be restricted to project collaborators and markers.

3.7 Ethical Considerations

1. GDPR Compliance: The dataset is publicly available and does not contain personally identifiable information, ensuring GDPR compliance.
2. University Ethics Policy: The project follows ethical research standards as outlined by the University of East London.
3. Data Usage Permission: The dataset is available under open data access policies, allowing unrestricted academic use.
4. Ethical Data Collection: The data was collected by the U.S. government for statistical and research purposes, ensuring ethical integrity.