Interim Design Presentation – Group 9

Simon Vellandurai, Lucas Alba, Jorge Rodriguez

Introduction

- The goal of this design project is to develop a **renewable energy production forecasting system** that uses real-time weather data and historical energy output data to predict the amount of energy that can be generated by solar and wind farms.

- By integrating these two datasets, the system would provide actionable insights for energy producers, enabling them to:

    - Optimize their operations
    - Forecast power output more accurately
    - Better manage energy supply based on environmental conditions.

- The scope of this project and the direction it takes moving forward is at our discretion; our sponsors have made it clear that they are only acting as a resource for ideas and improvements.

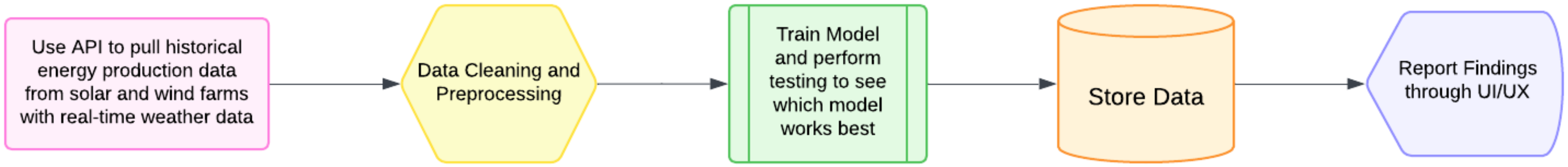
Concept Generation

- The concept that we selected was a centralized data integration model, which focuses on a centralized database that plans to integrate historical energy production data from solar and wind farms with real-time weather data collected via APIs.
- Process:
  - Collect / pull the data from the APIs
  - Preprocessing (Data Cleaning and normalization)
  - Training the model
  - Storage
  - Displaying the resulting to a webpage

Embodiment Design- System Architecture

Centralized Forecasting System:

- Combines real-time weather data and historical energy production records
- Powered by a Random Forest machine learning model for predictive analysis
- Modular design to separate core functionalities: data ingestion, preprocessing, model training, and output visualization



Embodiment Design- Configuration

1. Data Ingestion Subsystem
   - APIs to gather real-time weather data
   - Data import pipeline for historical energy production dataset
2. Data Preprocessing Subsystem
   - Cleaning and normalization of data
   - Feature engineering to extract relevant variables for prediction
3. Forecasting Model Subsystem
   - Random Forest Regressor
4. User Interface
   - Interactive dashboards with filtering and modification capabilities for forecasting based on specific variables
   - Line graphs for predictions, bar charts for comparisons
5. Data Storage and Management
   - Cloud-based storage for real-time and historical datasets

Embodiment Design- Component Information

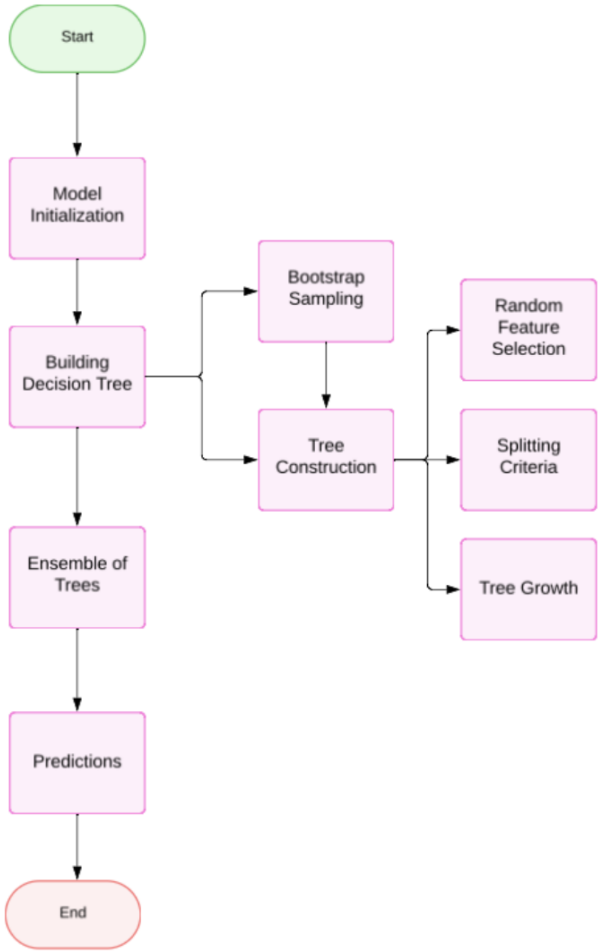


Figure 6. Random Forest Model

- While our project does not contain any dimensions or materials, the most important component within our system is the machine learning model.
- **Initialization**:
  - Developer defines model parameters, such as the number of trees and maximum depth.
- **Building Decision Trees**:
  - Bootstrap sampling generates multiple samples of the training data.
  - Subsets of features are randomly selected at each split.
  - Splitting criteria (e.g., Gini impurity, entropy) are used to grow the trees.
  - Trees grow until they no longer meet the splitting criteria.
- **Aggregation and Random Forest Creation**:
  - Predictions from all decision trees are combined to form the random forest.
  - The most popular prediction (majority vote) is selected as the final output.
- **Model Readiness**:
  - Once aggregated, the model is ready to make predictions.

Design Improvements

Topic	Issue	Solution
Data Ingestion from APIs	• Missing or outdated data can damage prediction accuracy Issues can arise from: Network problems, server outages, changes within the API version and modifications to data formats	• Robust error-handling mechanisms – System attempts reconnection a specified number of times before triggering an alerts • Fallback data source – ensures the system remains operational even if the main APIs are down
Real-Time Prediction and Forecasting (Delay in predictions)	• Caused by high data volumes or environmental changes / cause model drift • Decreased system accuracy and reliability, outdated forecasts affecting decision-making.	• Modified design to support adaptive model retraining. • Monitoring the environment to detect changes. • Periodic retraining on the latest data to maintain relevance.
Ergonomics	• Primary Load: Understanding user strengths and weaknesses. • Experience/Usability: Ensuring product accessibility for a general audience.	• Simplified symbolism to aid comprehension. • Interactive annotations for deeper understanding of complex data. • Developing seamless integration of historical and real-time data.
Data Storage and Management	• Lack of regular backups. • Software bugs causing database corruption. • Inadequate storage solutions for large data volumes.	• Regular and automated backup procedures. • Data safely stored in multiple locations to avoid total loss. • Integration of regular data checks: Verifies stored data integrity and Counters potential corruption issues.

Test Planning

Testing will focus on the model's predictive performance and its robustness under varying conditions

- The testing process will begin with a thorough evaluation of the model using historical data, split into training, validation, and testing sets
  - We will use standard evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score to quantify the model's accuracy.
  - This step will also help identify any underfitting or overfitting issues during initial deployment.
- Stress testing will be conducted to evaluate the model's performance under extreme weather scenarios
  - Unusually high wind speeds, prolonged cloud cover, etc.
  - This will be done by incorporating artificial data that represent such conditions.
  - Doing so will help determine the model's ability to handle outliers in the data or edge cases without a significant drop in predictive accuracy.

Test Planning

- Assessing real-time functionality
  - Compare the model's predictions against live weather data collected over several weeks
  - Analyze the alignment between predicted and observed energy outputs
- Latency Testing
  - Ensure the system processes real-time data inputs without delays that could impact usability

All testing results will be documented to refine the model's hyperparameters, optimize performance, and improve integration with the system's overall architecture. This process ensures that the model meets both technical and user-centric requirements before deployment.

Summary

- Stakeholder Needs
  - Stakeholder input from Two Six Technologies prioritized practical, dual-purpose design for academia and industry.
  - Scalable machine learning models (e.g., Random Forests) ensure adaptability to evolving data.
  - User-friendly features include simplified visualizations and interactive annotations for accessibility by non-technical audiences, such as farmers and renewable energy users.
- Our Project
  - Combines historical solar/wind data with real-time weather via APIs.
  - Delivers predictive insights through a user-friendly web interface.
  - Offers accessible visualizations for non-technical users (e.g., farmers, energy consumers).
  - Scalable for academic and industrial applications.

Questions?