Predicting The Energy Output Of Wind Turbine Based On Weather Condition

Milestone 1: Project Initialization and Planning Phase

The "Project Initialization and Planning Phase" marks the project's outset, defining goals, scope, and stakeholders. This crucial phase establishes project parameters, identifies key team members, allocates resources, and outlines a realistic timeline. It also involves risk assessment and mitigation planning. Successful initiation sets the foundation for a well-organized and efficiently executed machine learning project, ensuring clarity, alignment, and proactive measures for potential challenges.

Activity 1: Define Problem Statement

Problem Statement: The project aims to develop a machine learning model to predict the energy output of wind turbines based on current weather conditions. This predictive capability is crucial for energy companies to forecast production levels, enabling optimized energy distribution and pricing strategies. Additionally, it supports wind farm operators in scheduling maintenance to minimize downtime and maximize productivity, while aiding grid operators in efficiently integrating wind energy into the electrical grid.

SmartLender Problem Statement Report:[Click Here]

Activity 2: Project Proposal (Proposed Solution)

The proposed solution involves leveraging historical data on weather conditions (such as wind speed, temperature, humidity) and corresponding energy output from wind turbines. Using machine learning algorithms like regression models or neural networks, we will develop a predictive model. This model will be trained to accurately forecast the energy output of wind turbines based on real-time or forecasted weather data. By continuously refining and updating the model with new data, it will provide reliable predictions essential for energy production forecasting, maintenance planning, and efficient grid integration of wind energy.

SmartLender Project Proposal Report: [Click here]

Activity 3: Initial Project Planning

The project aims to build a machine learning model for predicting wind turbine energy output based on historical weather data. Initial steps include data collection, cleaning, and feature engineering to enhance model accuracy. Model selection will involve testing various algorithms like regression and neural networks, optimizing parameters for

robust performance. Ultimately, this solution will empower stakeholders in energy production forecasting, maintenance scheduling, and grid integration planning.
Here we have divided our roles on jira for better project management adn working together on the project.

SmartLender Project Planning Report: [Click here]

Milestone 2: Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase involves executing a plan to gather relevant loan application data from Kaggle, ensuring data quality through verification and addressing missing values. Preprocessing tasks include cleaning, encoding, and organizing the dataset for subsequent exploratory analysis and machine learning model development

Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report

The data collection plan involves sourcing historical weather data and real-time energy output data from wind farms from kaggle. Quality assurance measures include thorough cleaning for missing values, normalization of units, and validation through cross-referencing multiple sources to ensure data accuracy and consistency for model training.

SmartLender Data Collection Report: [Click here]

Activity 2: Data Quality Report

The data used for predicting wind turbine energy output based on weather conditions exhibits high quality with comprehensive coverage of weather variables (e.g., wind speed, temperature, humidity) and corresponding energy output records, ensuring minimal missing data. Rigorous accuracy checks against multiple reliable sources and internal consistency validations contribute to precise weather measurements and energy production data. Consistent normalization methods further enhance data uniformity, supporting robust machine learning model training for accurate energy production forecasting, optimized maintenance scheduling, and efficient grid integration strategies for energy stakeholders.

SmartLender Data Quality Report: [Click here]

Activity 3: Data Exploration and Preprocessing

The project conducted thorough analysis of weather variables (wind speed, temperature, humidity) to understand distributions and correlations. Preprocessing involved handling missing data with interpolation, ensuring consistency across sources, and engineering features like wind chill factor for improved model accuracy. Data normalization and splitting into training/validation sets prepared the dataset for effective machine learning model training to predict wind turbine energy output based on weather conditions.

SmartLender Data Exploration and Preprocessing Report: [Click Here]

Milestone 3: Model Development Phase

The Model Development Phase entails crafting a predictive model for loan approval. It encompasses strategic feature selection, evaluating and selecting models (Random Forest, Decision Tree, KNN, XGB), initiating training with code, and rigorously validating and assessing model performance for informed decision-making in the lending process.

Activity 1: Feature Selection Report

For predicting wind turbine energy output based on weather conditions, feature selection focused on identifying key variables crucial for model accuracy. Initial analysis revealed wind speed, temperature, and humidity as primary predictors due to their significant impact on energy production. Additional features such as wind direction and atmospheric pressure were considered for their potential influence. Selection criteria included correlation analysis, feature importance from initial model runs, and domain expertise inputs from energy sector stakeholders. The final set of features chosen aims to optimize model performance and enhance predictions under varying weather scenarios for effective energy management and grid integration.

SmartLender Feature Selection Report: [Click Here]

Activity 2: Model Selection Report

The model selection process aimed to identify the most suitable algorithm for predicting wind turbine energy output based on weather conditions. Initial evaluations included regression models (linear, polynomial, and ridge regression) and ensemble methods like random forests and gradient boosting. Criteria for selection considered predictive accuracy metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), computational efficiency, and scalability. Random forests emerged as the preferred choice due to their robust performance in handling non-linear relationships and feature interactions, ensuring reliable energy production forecasts for operational planning and grid integration strategies.

SmartLender Model Selection Report: click here

Activity 3: Initial Model Training Code, Model Validation and Evaluation Report

The initial model training commenced with data preprocessing, including feature scaling and splitting into training and validation sets. The chosen model, a Random Forest Regressor, was implemented using Python's Scikit-Learn library. Training involved fitting the model to the training data and tuning hyperparameters through grid search with cross-validation. Evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were computed on the validation set to assess model performance, achieving promising results indicative of accurate predictions of wind turbine energy output based on weather conditions.

SmartLender Model Development Phase Template: Click Here

Milestone 4: Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Activity 1: Hyperparameter Tuning Document

We have applied different hyper parameters like the n_estimators, minimum leaf samples, random state variations and other to do a better preprocessing of data while training the model and obtained different results after observing their performance metrics.

Activity 2: Performance Metrics Comparison Report

using different functions under the metrics package, we have applied r3 score and rmse values to find the changes in the accuracy after the hyperparameter tuning.

Activity 3: Final Model Selection Justification

We have selected random forest regressor with a random state of 42 as our final model since it was giving a great r2 score of 97% as compared to other models. Also after the hyper parameter tuning its accuracy reduced to 94%, hence we selected the model without tuning

SmartLender Model Optimization and Tuning Phase Report: Click Here

Milestone 5: Project Files Submission and Documentation

We have uploaded the local deployment flask files and the python notebook in the link. **Click here**

For the documentation, Kindly refer to the link. **Click Here**

Milestone 6: Project Demonstration

In the module Project Demonstration, We have recorded ourselves giving a live demonstration of the project on the google meet.

Video link:Click here for video