

Project: ReneWind

“ReneWind” is a company working on improving the machinery/processes involved in the production of wind energy using machine learning and has collected data of generator failure of wind turbines using sensors. They have shared a ciphered version of the data, as the data collected through sensors is confidential (the type of data collected varies with companies). Data has 40 predictors, 20000 observations in the training set and 5000 in the test set.

Covered Skills and Tools

- Conducted an in-depth Exploratory Data Analysis (EDA) on ciphered data from ReneWind, a company focused on enhancing wind energy production machinery/processes using machine learning. The dataset, containing 40 predictors with 20,000 observations in the training set and 5,000 in the test set, aimed to predict generator failures using sensor data.
- Executed comprehensive data pre-processing, ensuring data readiness for analysis. This involved preparing the data, implementing missing value treatments, and preventing data leakage.
- Employed model-building techniques to create at least 6 classification models, including logistic regression, decision trees, random forest, bagging classifier, and boosting methods. The objective was to identify the most effective model for predicting generator failures and reducing overall maintenance costs.
- Implemented model building with under-sampled training data, constructing an additional set of at least 6 classification models using logistic regression, decision trees, random forest, bagging classifier, and boosting methods.
- Applied Hyperparameter tuning to enhance model performance, selecting at least 3 best-performing models from the entire set. Tuned the chosen models and assessed their improved performance.
- Evaluated the performances of tuned models, comparing and selecting the final model based on its effectiveness in predicting generator failures. Checked the final model's performance on the test dataset for validation.
- Productionized the final model using pipelines, ensuring seamless integration into real-world applications.
- Other Covered Skills: Model Tuning, Up and Down Sampling, Regularization, Hyperparameter Tuning, Exploratory Data Analysis (EDA).
- Used Tools and Python Libraries: Pandas, Numpy, Matplotlib, Seaborn, Sklearn.