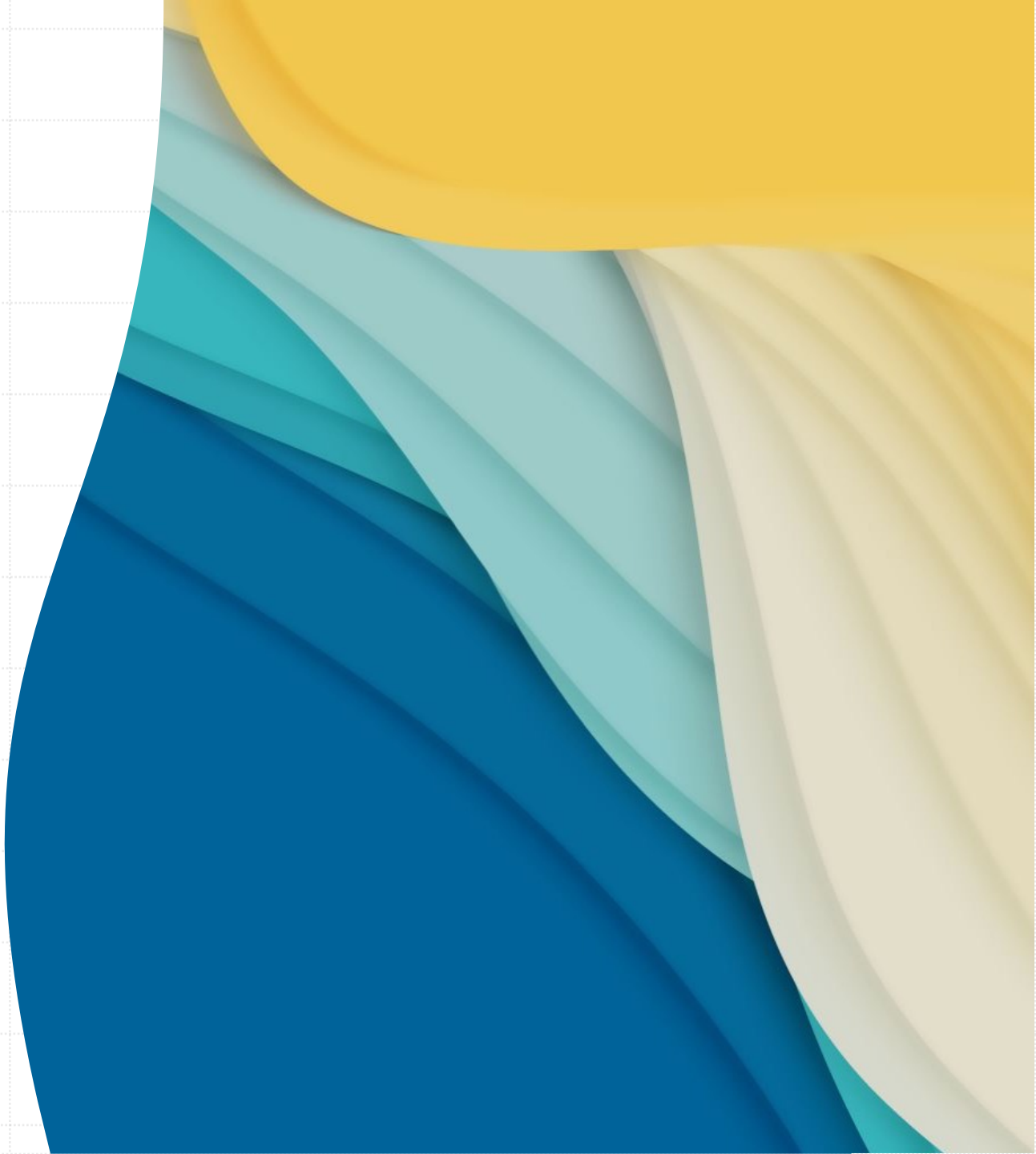
ReneWind

Renewable Energy



Contents

Business Overview

Renewable energy sources play an increasingly important role in the global energy mix, as the effort to reduce the environmental impact of energy production increases.

Predictive maintenance uses sensor information and analysis methods to measure and predict degradation and future component capability. The idea behind predictive maintenance is that failure patterns are predictable and if component failure can be predicted accurately and the component is replaced before it fails, the costs of operation and maintenance will be much lower.

Out of all the renewable energy alternatives, wind energy is one of the most developed technologies worldwide. The U.S Department of Energy has put together a guide to achieving operational efficiency using predictive maintenance practices.

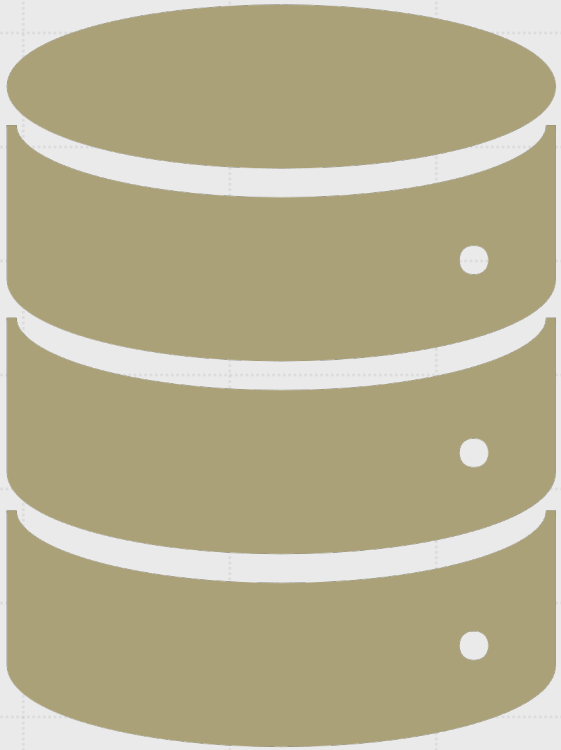
The sensors fitted across different machines involved in the process of energy generation collect data related to various environmental factors (temperature, humidity, wind speed, etc.) and additional features related to various parts of the wind turbine (gearbox, tower, blades, break, etc.).

Objective

- Build various classification models, tune them and find the best one that will help identify failures so that the generator could be repaired before failing/breaking and the overall maintenance cost of the generators can be brought down.
- Reduce the maintenance cost
  - Minimum possible maintenance cost = Actual failures * (Repair cost) = (True Positive + False Negative) * (Repair cost)
  - Maintenance cost associated with model = True Positive * (Repair cost) + False Negative * (Replacement cost) + False Positive * (Inspection cost)

Data Overview

- The data contains 40 features and 1 target column
- Train data contains 40000 observations and test data contains 10000 observations
- All the columns have float data type
- The target variable has value 0 or 1
- There are no duplicate values in the data set
- Column V1 and V2 have missing values
- The data for all the columns ranges between negative and positive values



Exploratory Data Analysis

- All the columns have almost perfect bell curve.
- Normalized data for al the columns
- There are outliers present for for all the columns
- There are few slightly right skewed and slightly left skewed columns
- The skewness for all columns fall between -0.3 to 0.5

Data Pre-processing

Split Training data into training and validation

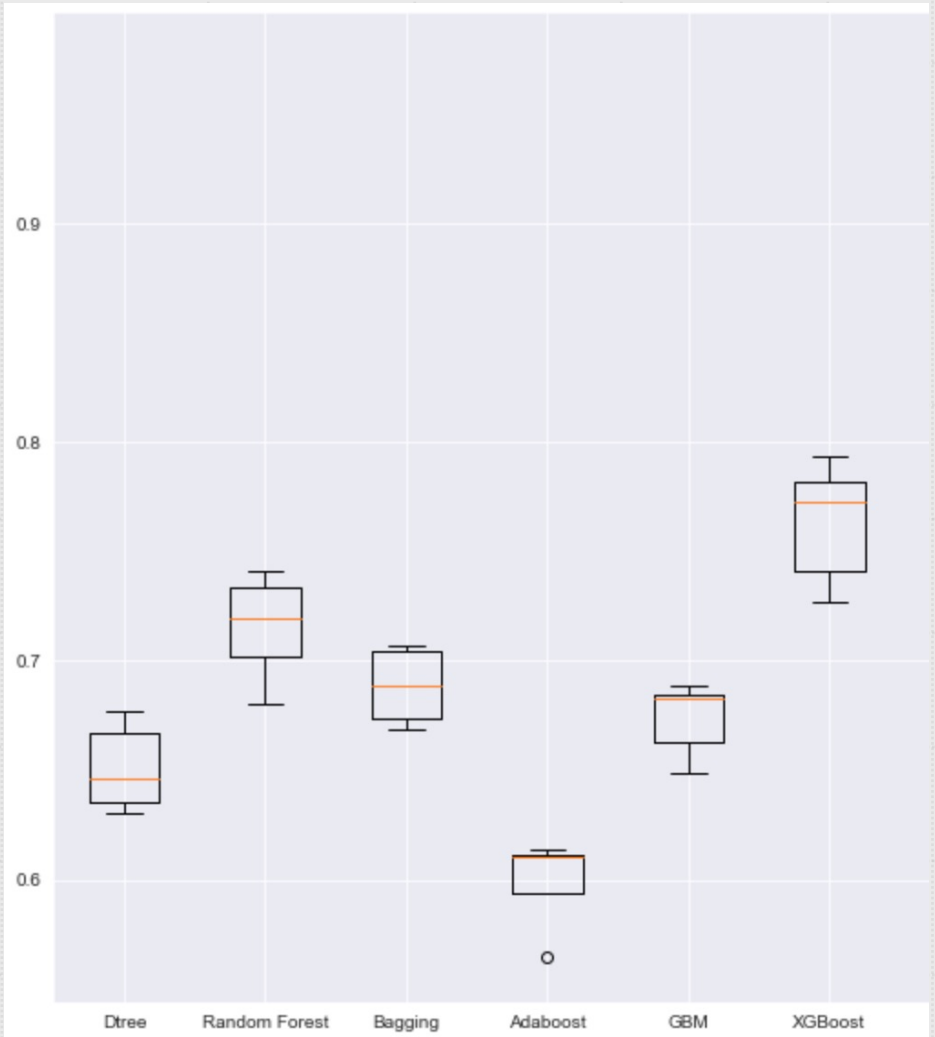
Missing value treatment

Model Building

- Models built with Original data, Oversampled and Under sampled data
- List of models
  - Decision Tree
  - Random Forest
  - Bagging
  - AdaBoost
  - Gradient Boost
  - XGBoost

Model Selection

- Models selected based on cross validation score
  - XGBoost
  - Random Forest
  - Bagging



Model Performance

Training performance comparison:

	XGBoost Tuned with Random search	Random Forest Tuned with Random search	Bagging Classifier Tuned with Random search
Accuracy	99.797	99.347	99.920
Recall	100.000	88.232	98.537
Precision	96.414	99.793	100.000
F1	98.174	93.657	99.263
Minimum_Vs_Model_cost	98.775	83.560	97.619

Validation performance comparison:

	XGBoost Tuned with Random search	Random Forest Tuned with Random search	Bagging Classifier Tuned with Random search
Accuracy	99.050	98.580	98.530
Recall	87.751	74.771	75.686
Precision	94.488	99.031	96.729
F1	90.995	85.208	84.923
Minimum_Vs_Model_cost	81.886	70.278	70.733

Final Model Selection

- Tuned XGBoost with performed best on the validation dataset
- In the test performance, the model shows minimum cost of 80%

Test performance:

	Accuracy	Recall	Precision	F1	Minimum_Vs_Model_cost
0	98.870	86.472	92.383	89.330	80.049

Business Insights Recommendations

- The tuned XGBoost model gives 80% minimum model cost
- The accuracy, recall, precision and f1 scores for the selected model are good
- Oversampling and Under sampling techniques are improving the model performance
- Random Forest and Bagging Classifiers overall performed well
- As the features were masked, outlier treatment was not performed
- Outlier treatment and feature enhancement can help performance of the model