ReneWind

Business Case

Background

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.

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.

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.

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

To 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.

Data Information

The data contains information about the Data

Variable	Type of Variable
V1	Float
V2	Float
V3	Float
V4	Float
V5	Float
V6	Float
V7	Float
V8	Float
V9	Float
V10	Float
V11	Float
V12	Float
V13	Float
V14	Float
V15	Float
V16	Float
V17	Float
V18	Float
V19	Float
V20	Float

Variable	Type of Variable
V21	Float
V22	Float
V23	Float
V24	Float
V25	Float
V26	Float
V27	Float
V28	Float
V29	Float
V30	Float
V31	Float
V32	Float
V33	Float
V34	Float
V35	Float
V36	Float
V37	Float
V38	Float
V39	Float
V40	Float
Target	Float

Observations	Variables
40,000	41

Missing value treatment was applied on the training data via median replacement.

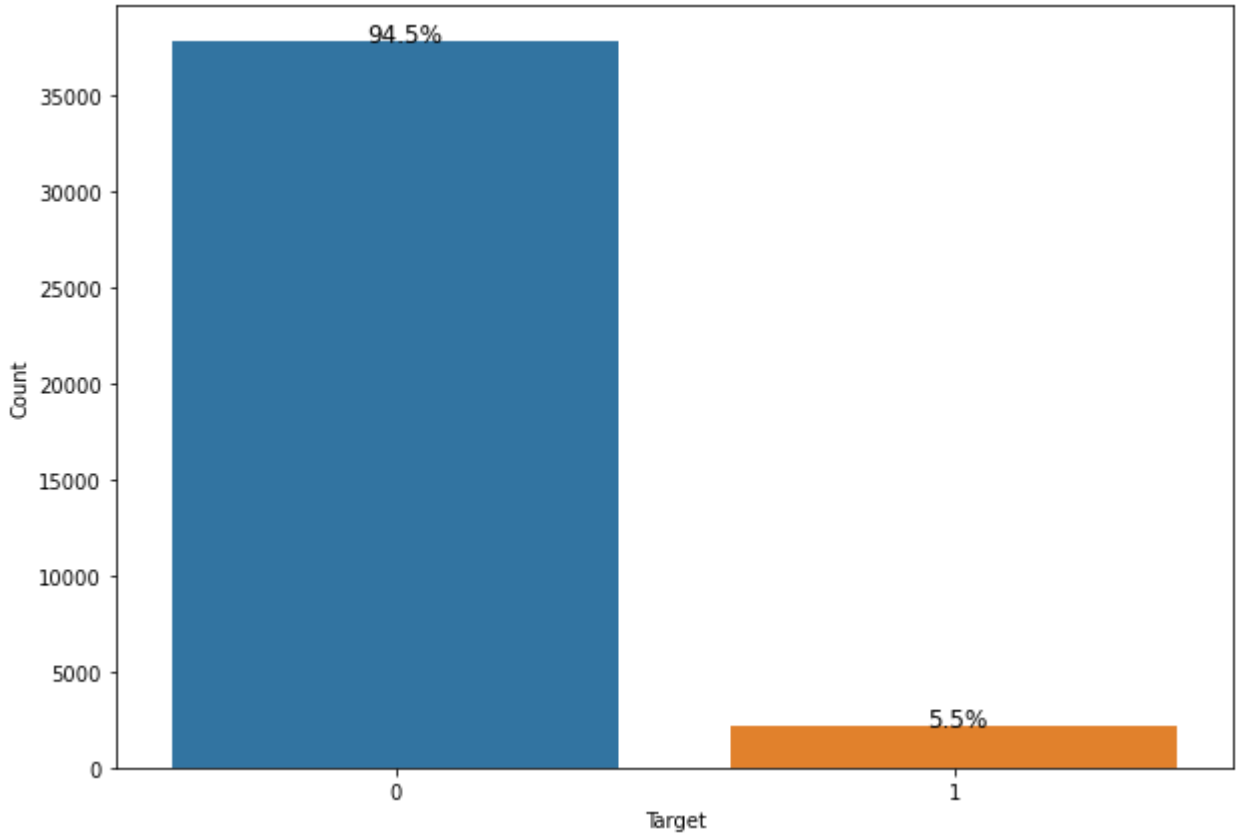
Exploratory Data Analysis – Target

This data contains the time spent on the page of the various customers

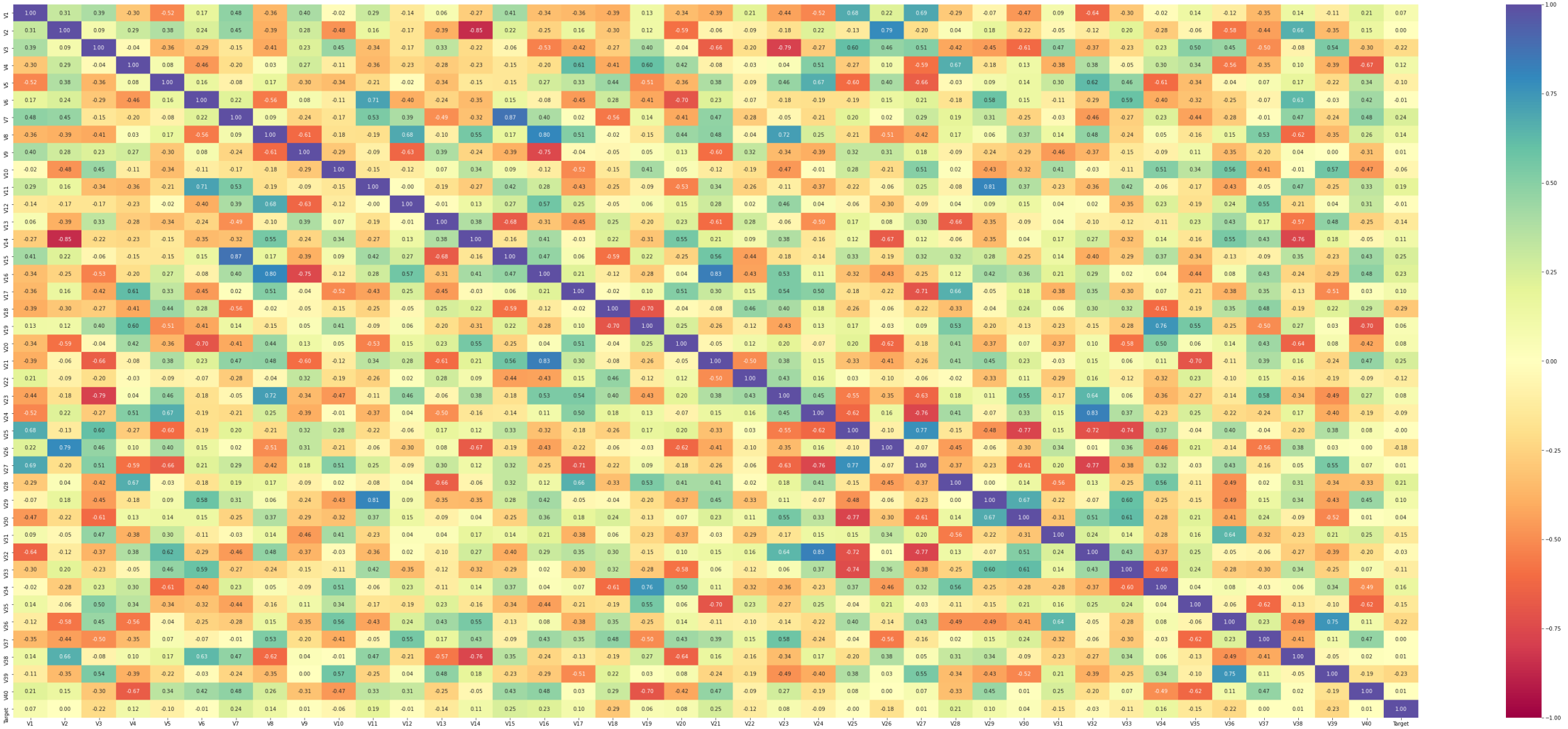
Observations:

Only 5.5% of the Target variable are below the minimum cost

Time Spent on Page

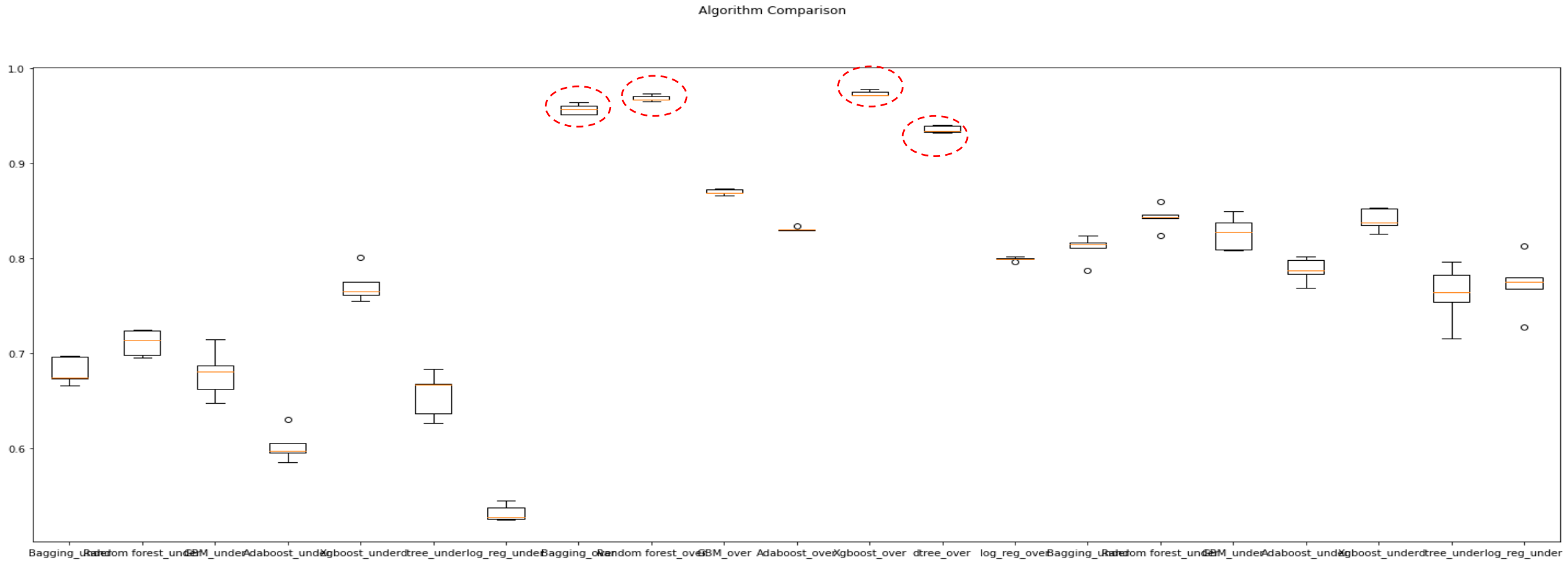


Exploratory Data Analysis – Data Correlation



There are lots of strong correlation between the 40 predictors but none will be dropped due to the nature of the problem

Model Comparison and Choice



Observations

- OverSampling Models have the highest Cross-validation scores followed by UnderSampling Models
- Models for HyperparameterTuning are:

OverSampled Data of Xgboost, RandomForest, Bagging and DecisionTree because they have highest Cross-validation scores and there is no major difference in the training scores and the cross validation scores

- XGBoost will not be run due to the time constraints involved

Model Comparison Post Tuning

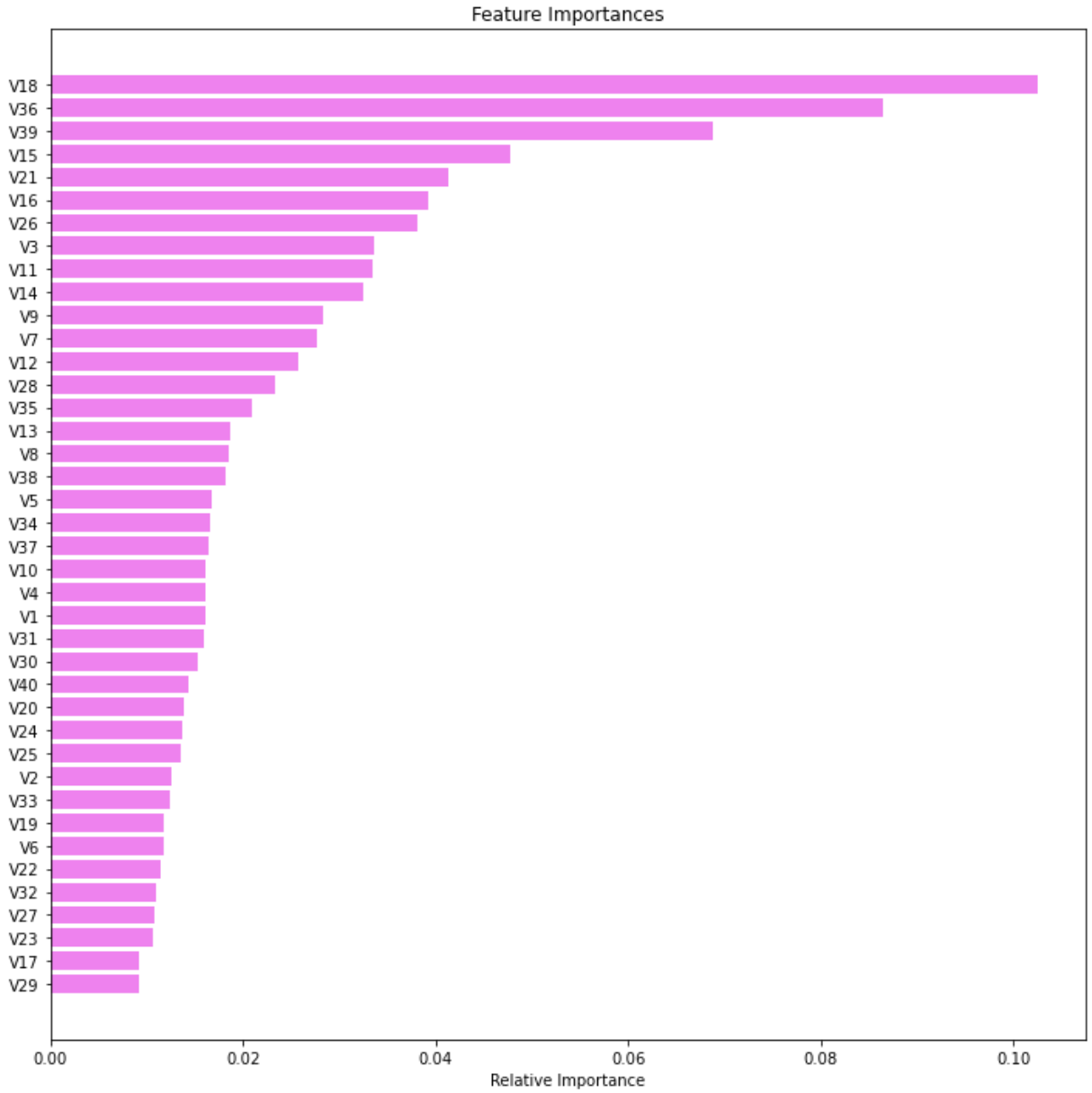
Training Set

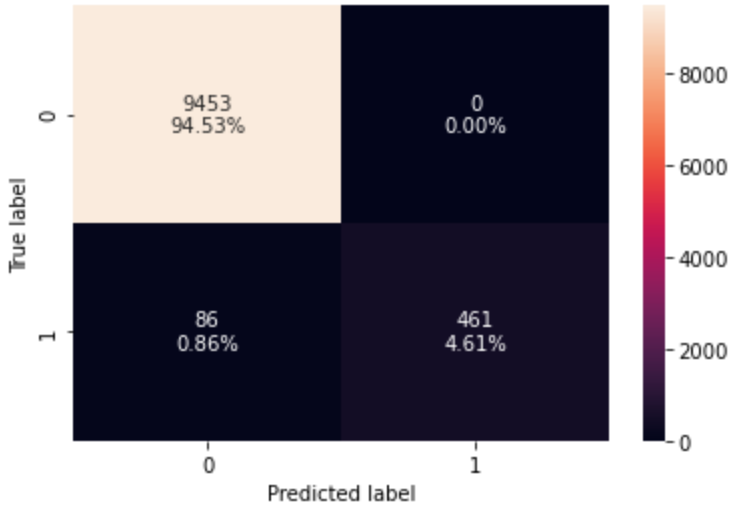
	RandomForest Tuned with Grid search	RandomForest Tuned with Random search	DecisionTree Tuned with Grid search	DecisionTree Tuned with Random Search	Bagging Tuned with Grid search	Bagging Tuned with Random Search
Accuracy	0.9989	0.9991	0.8940	0.8940	0.9999	0.9999
Recall	0.9981	0.9986	0.8542	0.8542	0.9999	1.0000
Precision	0.9997	0.9996	0.9280	0.9280	1.0000	0.9999
F1	0.9989	0.9991	0.8896	0.8896	0.9999	0.9999
Minimum_Vs_Model_cost	0.9967	0.9975	0.7905	0.7905	0.9999	0.9999

Validation Set

	RandomForest Tuned with Grid search	RandomForest Tuned with Random search	DecisionTree Tuned with Grid search	DecisionTree Tuned with Random Search	Bagging Tuned with Grid search	Bagging Tuned with Random Search
Accuracy	0.9896	0.9903	0.9239	0.9239	0.9886	0.9883
Recall	0.8689	0.8720	0.8125	0.8125	0.8567	0.8613
Precision	0.9360	0.9455	0.4029	0.4029	0.9289	0.9202
F1	0.9012	0.9072	0.5387	0.5387	0.8914	0.8898
Minimum_Vs_Model_cost	0.8076	0.8129	0.5835	0.5835	0.7932	0.7961

Final Test Model Information





Accuracy	Recall	Precision	F1	Minimum_Vs_Model_cost
0.9912	0.840951	0.997831	0.912698	0.790082

The performance on the test set is generalised

Conclusion

Despite not knowing the names of the variables being used, it can be seen that not all variables carry a similar amount of weighting in determining the maintenance cost with the most important factors being V18, V36,v39,V15,V21

Recommendations

1.Special focus should be given to the parts that have a higher rate of decomposition
2.Costs of equipment needs to be amortized in order to reduce the effect of the costs on the books