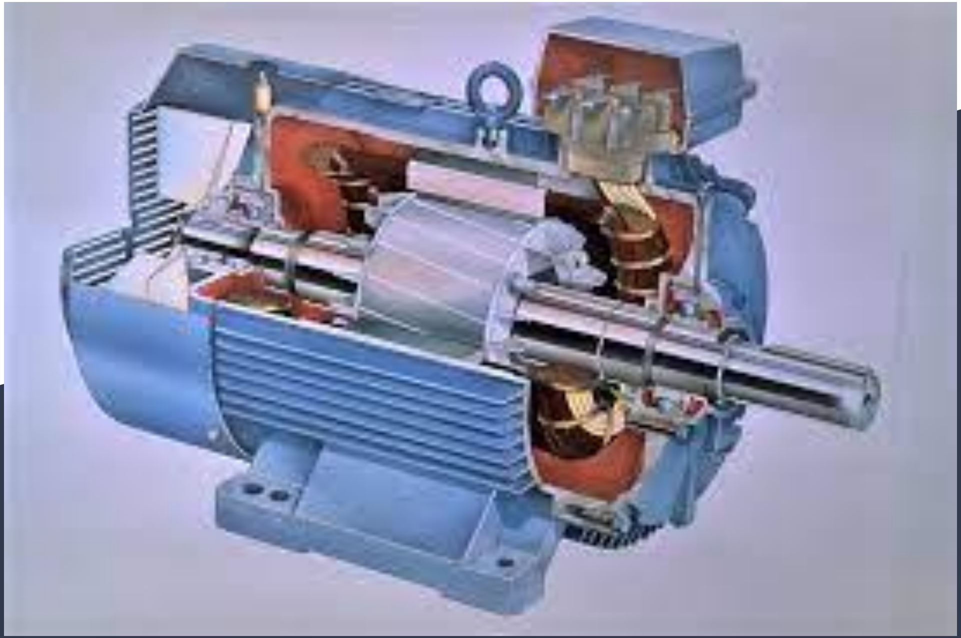
Electric Motor Temperature Prediction.

Data Science ExcelR Project - P189.



Group Members-

1. Ankit Indorkar
2. Shreyas Sadavarte
3. Bhimaraj Tagadagar
4. Kunal Mhaske
5. Meenal Patel
6. Radhika Jadhav
7. Neha Chavan

Project Mentors:-

1. Mr. Madishetti Rajashekar
2. Mrs. Pallavi B

Project Objective :-

Predict Motor Speed based on other attributes available

Details of given Data Set

- The dataset comprises several sensor data collected from a permanent magnet synchronous motor (PMSM) deployed on a test bench.
- All recordings are sampled at 2 Hz. A measurement session can be between one and six hours long.
- Most driving cycles denote random walks in the speed-torque-plane in order to imitate real world driving cycles to a more accurate degree than constant excitations and ramp-ups and -downs would.

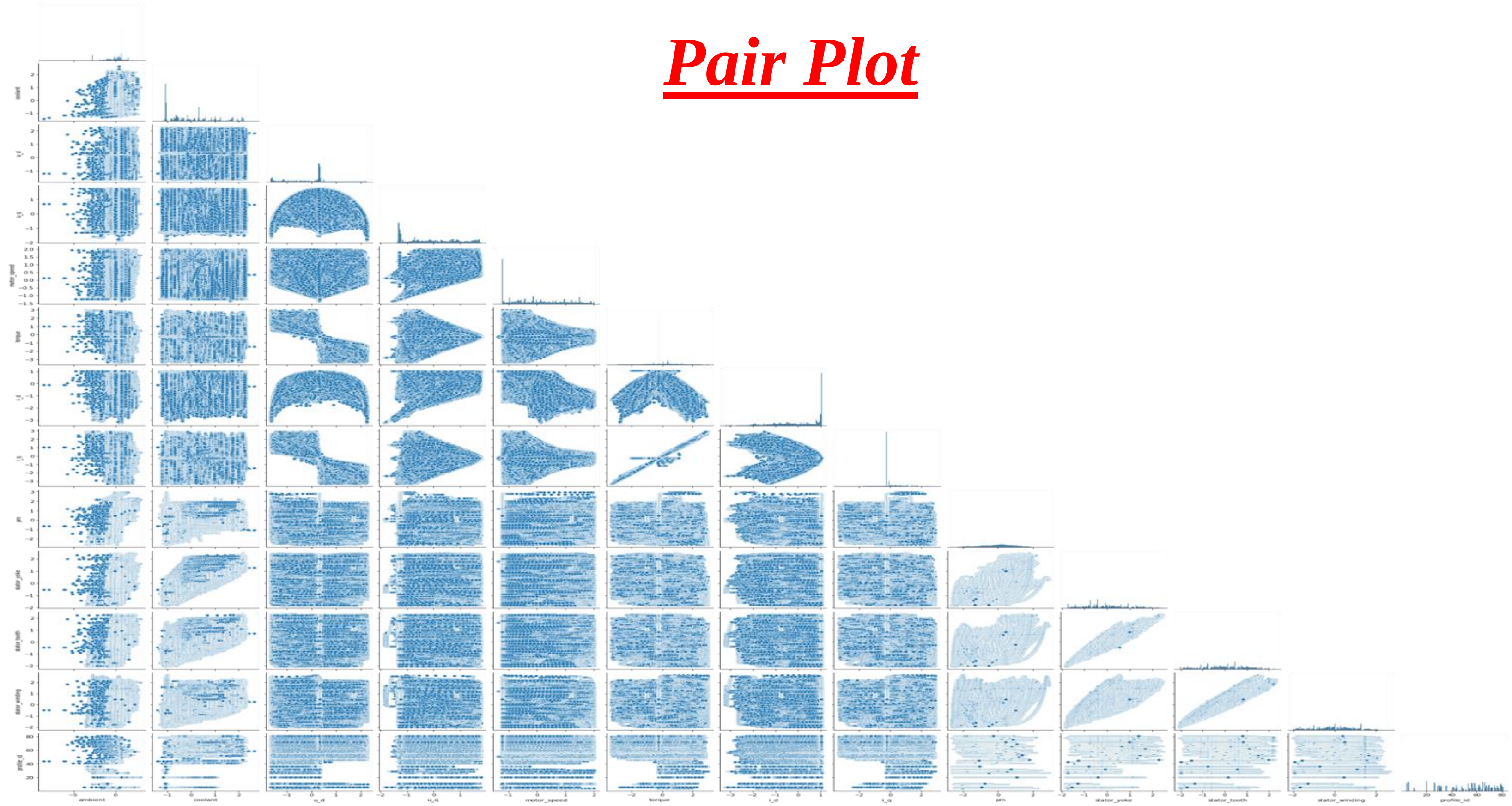
Approach towards the Problem-

- Directly we can’t say they that all of the features are important, but they can have an impact on the motor speed.
- To select the most important features, we can use feature selection techniques like correlation, mutual information, univariate feature selection, and tree-based feature selection methods. These methods will help us to identify the features that have the most impact on the target variable.
- In this particular project, it may be important to consider the features that are most closely related to the motor speed.
- The standard approach for this regression problem would be to use a machine learning model to make predictions about the motor speed based on the other attributes in the dataset.

Exploratory Data Analysis Over given Temperature dataset.

- While Beginning with the EDA, we addressed both Qualitative and Quantitative issues with the data given, which includes Identifying the null values, dropping out the duplicates, handling the missing values and dealing with the Outliers.
- We succeeded in Validating data through EDA over the given dataset which is now ready for model building.
- We did not perform the Label Encoding and One Hot Encoding as data is already in the numeric format.
- Following are the visualisations and Inferences we carried out after performing EDA.

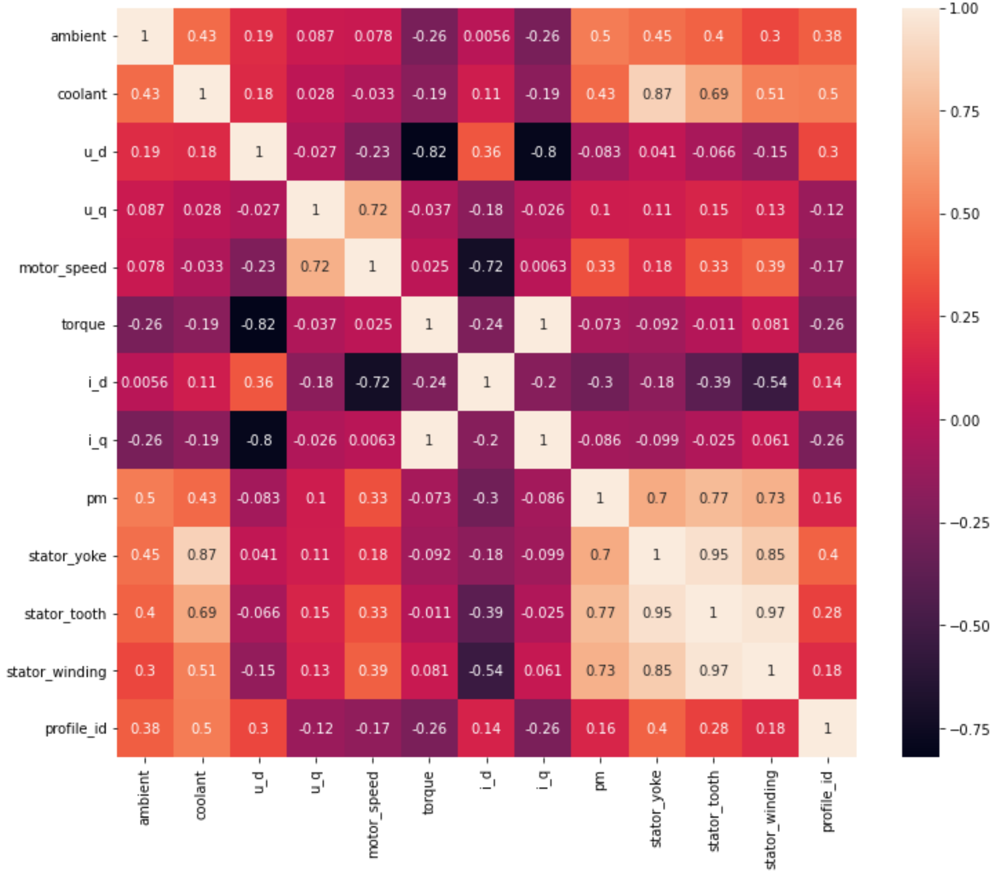
Pair Plot



Pair Plot Over the complete dataset.

Inferences based on the pair plot-

- There is a linear relationship between i_q (current d-component) and torque (torque induced by current).
- We can also find Distance correlation in between motor speed, u_q (voltage q-component), u_d (voltage d-component), i_q (Current q-component), i_d(current d-component).
- There is also a linear correlation between, stator_yoke, stator_tooth, stator_winding.

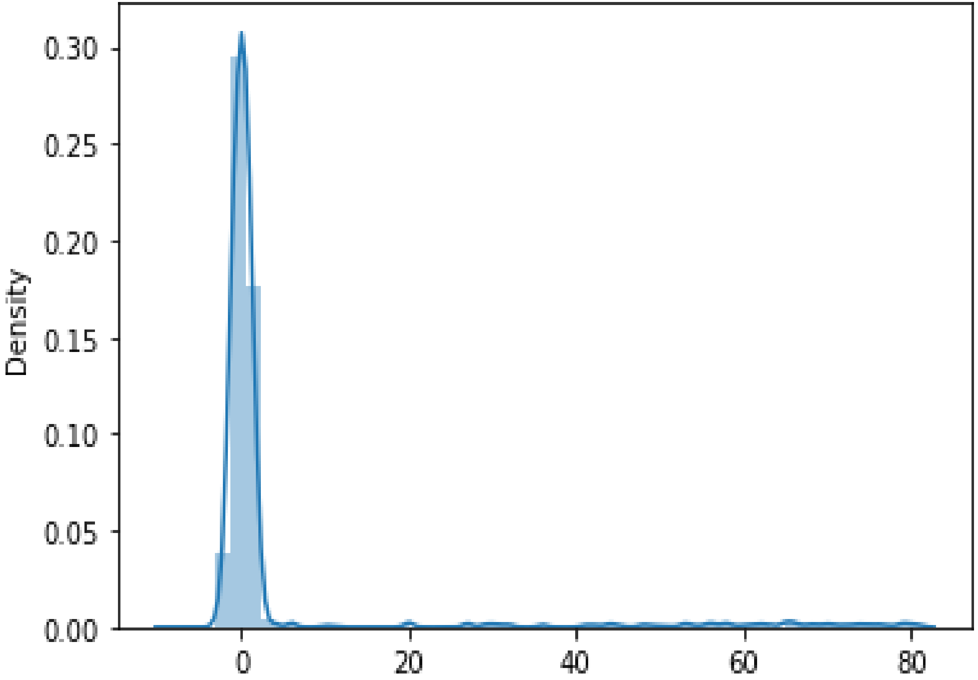


Heat Map for the given data set to understand the correlation.

- From the Heat Map we came to know that there is collinearity in the features of the given dataset.

Heat Map which represents Collinearity in Numbers-

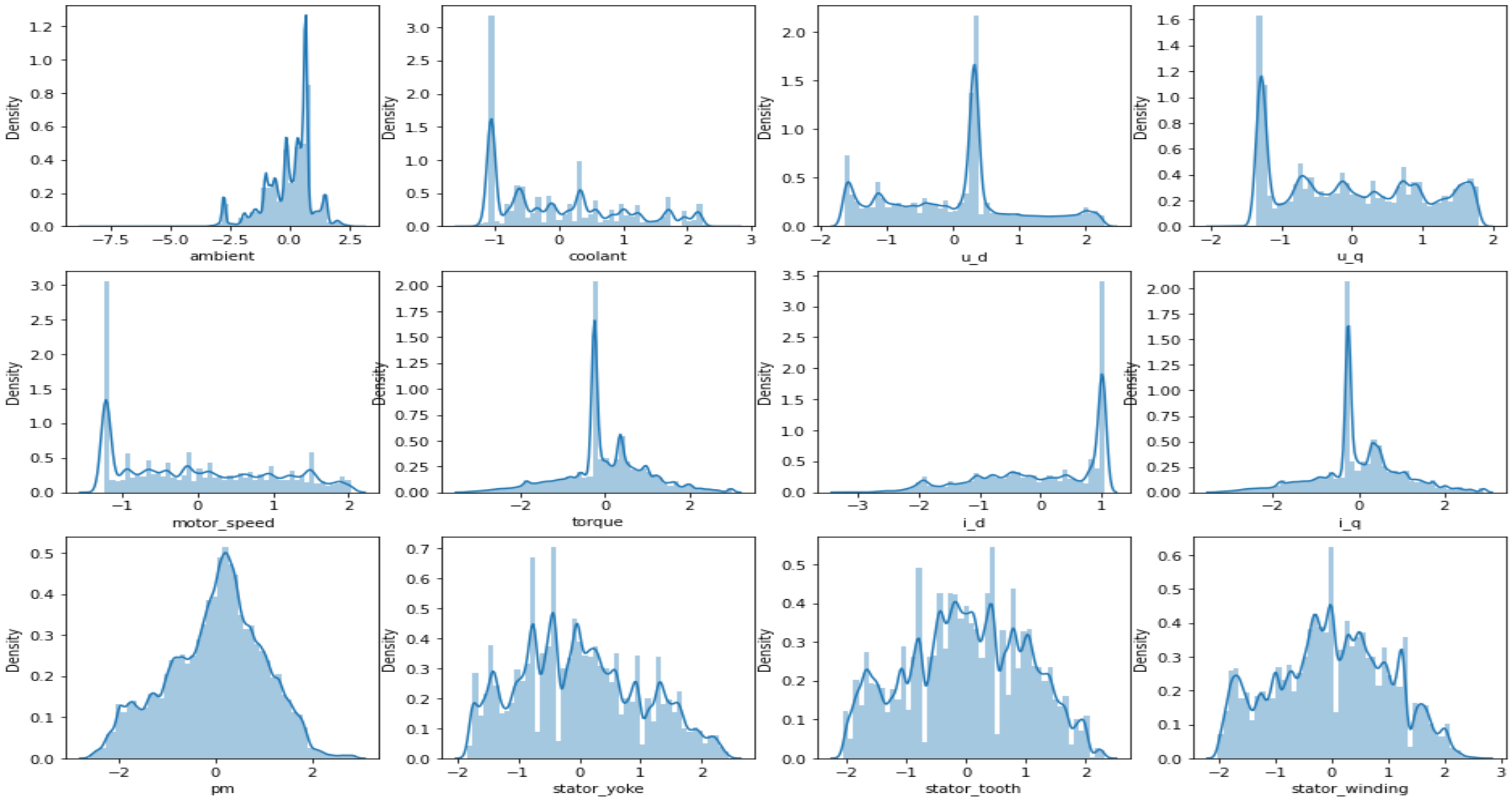
Density Plot



- From the Density plot we came to know that the given dataset is rightly skewed.

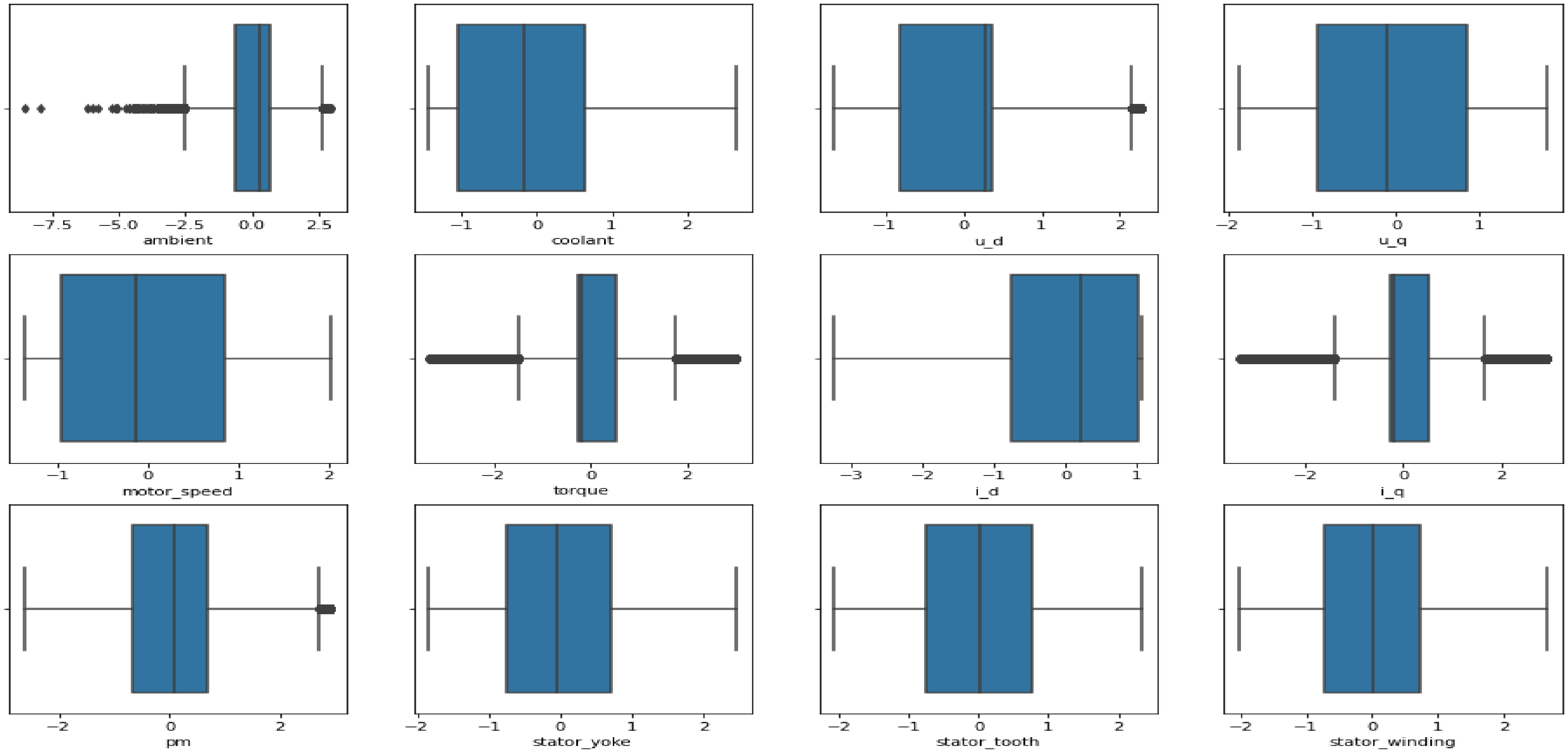
Above is the Density plot of the dataset. Here we identify that the data is Rightly Skewed.

Density Plot for all indusial Features



Density Plot over the Dataset- Here we can say that Permanent Magnet is normally distributed.

Box Plot for all indusial Features



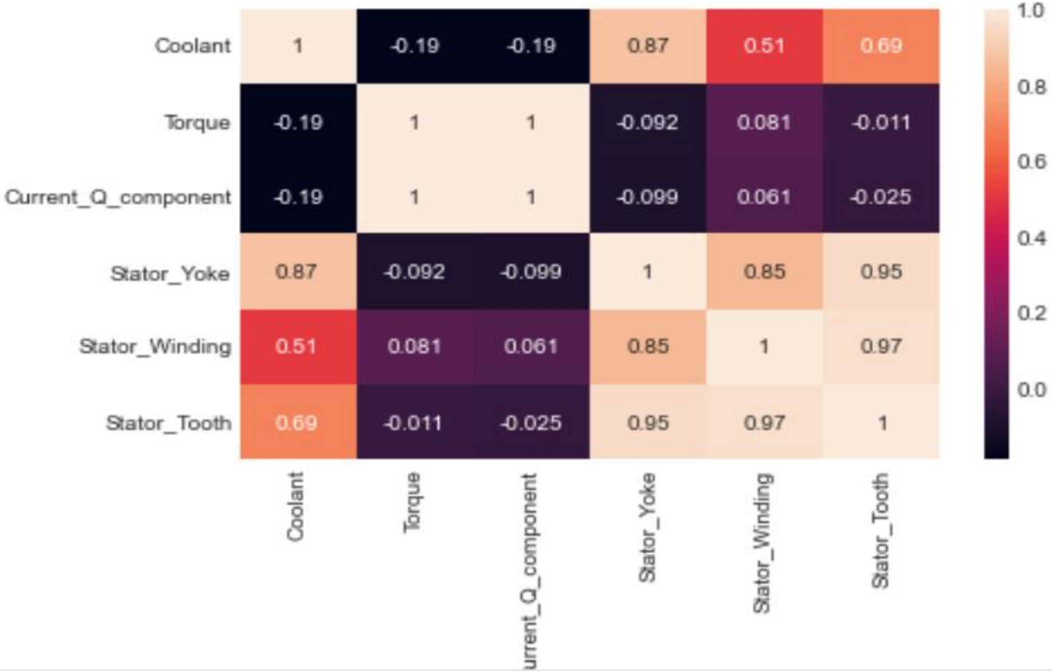
From above box plot we came to know that, some of the features have the outlier in the given data set.

Inferences based on Boxplot-

➢ From above box plot we came to know that, some of the features have the outlier in the given data set.

Outliers Are:

1. ambient (ambient temperature)
2. u_d (voltage d-component)
3. Torque (torque induced by current)
4. i_q (current d-component)
5. pm (permanent magnet)



Feature Engineering:

Calculating the VIF Score

Here we are checking corelation between the features with the VIF Calculation.

	Variables	VIF
0	Ambient	1.688646
1	Coolant	39.402307
2	Voltage_D_component	4.759660
3	Voltage_Q_component	1.519728
4	Torque	193.483442
5	Current_Q_component	189.868569
6	Current_D_component	2.939579
7	PMST	4.247227
8	Stator_Yoke	356.752414
9	Stator_Winding	146.668203
10	Stator_Tooth	606.911896
11	Profile_ID	1.522807

- As we know that the standard threshold for the VIF score is 10.
- In the given data set some features are crossing the threshold, that means these features are highly corelated.

VIF Score

	Variables	VIF
1	Coolant	39.402307
4	Torque	193.483442
5	Current_Q_component	189.868569
8	Stator_Yoke	356.752414
9	Stator_Winding	146.668203
10	Stator_Tooth	606.911896

- These are the features which have more collinearity in them.

- We are going to drop the below features, as they have more collinearity.

1. Torque
2. Stator_Yoke
3. Stator_Tooth

Final Data Set after Feature Engineering:

After Feature Engineering we reshaped our given data set as follows:

The Feature are:

1. **Ambient**
2. **Coolant**
3. **Voltage_D_Component**
4. **Voltage_Q_Component**
5. **Current _D_Component**
6. **Current_Q_Component**
7. **PMST**
8. **Stater Wwinding**

The Target Feature

Motor Speed

Model Building:-

We have tested our dataset on the following models:

1. Multi-Linear Regression
2. Decision Tree Regression Model
3. Random Forest Regression Model
4. Ada-Boosting Regression Model
5. Gradient Boost Regression Model
6. Bagging Regression Model

Model Building:-

- All the above listed models has been first built with random parameters.
- After that, as our data set is very large, so we have taken a sample data for performing hyper parameter tuning. And we have used Randomized Search CV.
- Optimized parameters obtained by Hyper parameter tuning is then used to get best result for each individual model.

1. Multi-Linear Regression

OLS Regression Results

Dep. Variable:	Motor_Speed	**R-squared:**	0.919
Model:	OLS	**Adj. R-squared:**	0.919
Method:	Least Squares	**F-statistic:**	1.412e+06
Date:	Thu, 02 Feb 2023	**Prob (F-statistic):**	0.00
Time:	11:23:52	**Log-Likelihood:**	-1.6445e+05
No. Observations:	998070	**AIC:**	3.289e+05
Df Residuals:	998061	**BIC:**	3.290e+05
Df Model:	8		
Covariance Type:	nonrobust		

Intercept	0.000742
Ambient	-0.053668
Coolant	0.049495
Voltage_D_component	-0.181347
Voltage_Q_component	0.595396
Current_D_component	-0.650700
Current_Q_component	-0.232008
PMST	0.189161
Stator_Winding	-0.192270

Above are the Multi-Linear Regression Result

Best Results:

	Model	R_square	Mean_Squared_Error	Mean_Absolute_Error	Root_Mean_Squared_Error	Mean_Absolute_Percentage_Error
0	DecisionTreeRegressor	0.961495	0.038675	0.117855	0.19666	Nan
1	RandomForest	0.999738	0.000263	0.00624	0.016228	0.038014
2	Ada_Boosting	0.989822	0.010223	0.075897	0.101111	0.479641
3	Gradient_Boosting	0.995836	0.004183	0.042206	0.064673	0.467851
4	Bagging	0.999837	0.000164	0.003845	0.012809	0.02891

- Using this data frame we can conclude that, the best results given by Random Forest Regression Model
- We selected Random Forest Regression as our best model.

Deployment:

We have performed our deployment using StreamLIT

Thank You!