# Electric Motor Temperature

# Team 14

Angel Fernandes  
Akash Varun

**Introduction**

**Background:**

The dataset under consideration comprises sensor data obtained from a permanent magnet synchronous motor (PMSM) deployed on a test bench, representing a prototype model of a leading German OEM. These measurements were meticulously gathered by the LEA department at Paderborn University, utilizing a range of driving cycles to simulate real-world conditions. Each recording, sampled at 2 Hz, spans one to six hours and is distinguished by a unique "profile\_id." The motor operates under a hand-designed driving cycle, defining reference speed and torque, with resulting quantities such as motor speed and torque derived from standard control strategies.

**Motivation:**

The motivation for this proposal stems from the challenges in measuring crucial parameters like rotor temperature ("pm"), stator temperatures ("stator\_\*"), and torque in a commercial vehicle in a reliable and cost-effective manner. Rotor temperature and torque are particularly difficult to measure economically but are pivotal for efficient motor operation. A robust estimator for rotor temperature not only aids in manufacturing motors with reduced material but also empowers control strategies to optimize the motor's performance. Accurate torque estimates enhance motor control, leading to reduced power losses and minimizing heat build-up, thereby improving overall efficiency.

**Goal:**

The primary goal of this proposal is to develop advanced predictive models for the key features of interest, including rotor temperature, stator temperatures, and torque. Leveraging machine learning techniques, we aim to create robust estimators that can predict these critical motor parameters with high accuracy. Such models have the potential to revolutionize the automotive industry by enabling the manufacturing of motors with optimized materials, improving control strategies, and ultimately enhancing the efficiency of electric vehicles. The proposed predictive models will contribute to a deeper understanding of motor behavior and facilitate advancements in motor design and control strategies.

**Methodology**

1. **Data Preprocessing and Cleaning**

* **Addressing missing or repeated data involves a meticulous approach to fill gaps and eliminate redundancies, ensuring the integrity of the dataset.**
* **Standardization of the dataset is conducted to establish consistency, with particular attention to identifying anomalies or unexpected values.**
* **The data is formatted appropriately for analysis, considering factors such as scale, units, and data types.**

1. **Exploratory Data Analysis (EDA)**

* **EDA is employed to unravel insights into the distribution of the dataset and understand the correlations between various features.**
* **Visualizations, including histograms, scatterplots, and correlation matrices, serve as tools to delve into patterns related to rotor temperature, stator temperatures, torque, and their interrelationships with motor speed and other variables.**

1. **Feature Engineering:**

* **Relevant features, such as power magnitude derived from dq-coordinates (u\_q, u\_d, i\_d, i\_q), are computed to enrich the dataset.**
* **Insights into power consumption patterns are gained, shedding light on their implications for motor performance.**

1. **Model Selection and Evaluation**
2. **Linear Regression**

* **Identification of features and target variables for predicting continuous outcomes, such as motor power.**
* **Leveraging linear regression for its simplicity, effectiveness, and scalability, especially in the context of large datasets.**

1. **Random Forest Classifier**

* **Utilizing the ensemble learning approach of random forests to effectively handle imperfect data.**
* **Focusing on specific temperature variables, the classifier's suitability for the dataset is explored.**

1. **Decision Trees**

* **Implementation of decision trees to classify and predict motor speed, capitalizing on the interpretability afforded by their tree-like structure.**

1. **Model Comparison**

* **A comprehensive evaluation and comparison of the predictive performance of linear regression, random forest classifier, and decision trees.**
* **Metrics such as accuracy, precision, and recall are considered to determine the most effective model for achieving the project's objectives.**

1. **Results Visualization**

* **Creation of visualizations to articulate project findings, encompassing:**
* **Performance metrics of selected models, providing a quantitative understanding of their efficacy.**
* **Insights gleaned from data exploration, emphasizing the influence of different variables on motor behavior.**
* **Comparative analysis of model predictions, highlighting strengths and weaknesses and aiding in informed decision-making.**

# **Description of the Dataset**

All recordings are sampled at a rate of 2 Hz, with the dataset comprising multiple measurement sessions differentiated by the "profile\_id" column. These sessions, identifiable by unique profile identifiers, can vary in duration from one to six hours. The motor is stimulated by manually crafted driving cycles representing a target motor speed and torque. The currents in d/q-coordinates ("i\_d" and "i\_q") and voltages in d/q-coordinates ("u\_d" and "u\_q") result from a standard control strategy aiming to track the specified speed and torque references. The "motor\_speed" and "torque" columns represent the outcomes of this strategy, derived from the set currents and voltages. Many driving cycles simulate random walks in the speed-torque plane, providing a more accurate emulation of real-world driving conditions than constant excitations and gradual changes. The dataset contains 13,30,816 rows.

1. **u\_q:** Voltage q-component measurement in dq-coordinates (in V)
2. **coolant:** Coolant temperature (in °C)
3. **stator\_winding:** Stator winding temperature (in °C) measured with thermocouples
4. **u\_d:** Voltage d-component measurement in dq-coordinates
5. **stator\_tooth:** Stator tooth temperature (in °C) measured with thermocouples
6. **motor\_speed:** Motor speed (in rpm)
7. **i\_d:** Current d-component measurement in dq-coordinates
8. **i\_q:** Current q-component measurement in dq-coordinates
9. **pm:** Permanent magnet temperature (in °C) measured with thermocouples and transmitted wirelessly via a thermography unit.
10. **stator\_yoke:** Stator yoke temperature (in °C) measured with thermocouples
11. **ambient:** Ambient temperature (in °C)
12. **torque:** Motor torque (in Nm)
13. **profile\_id:** Unique measurement session id

**Data Source**  
<https://www.kaggle.com/datasets/wkirgsn/electric-motor-temperature/data>

# **Results and Analysis**

**Data Preprocessing and Cleaning**

The dataset underwent a comprehensive quality check, examining both missing values and duplicates, resulting in the identification of zero missing values. To gain a preliminary understanding, basic descriptive statistics were generated for all numerical variables. Subsequently, an assessment of outliers was conducted by creating a boxplot for all columns, excluding 'profile\_id'. Z-scores were computed for each data point in the numerical columns, employing a threshold set at 3. Rows containing outliers were discerned and subsequently excluded, resulting in the removal of 7,252 outliers and a refined dataset comprising 1,323,564 rows.

A graph with a bar

Description automatically generated

Figure 1. Outliers in the dataset

**Exploratory Data Analysis**

A correlation matrix for the cleaned data was generated.

A screenshot of a graph

Description automatically generated

Figure 2. Correlation Matrix

Analyzing the correlations, we see that we have strong correlations in some data, but there is no need to cancel out these variables since it is a direct illumination, as they are technical variations that are naturally correlated, and not a spurious strong brightness.

A heatmap using the Seaborn library was also generated to visualize the correlation matrix of numeric columns in the DataFrame. The heatmap visually represents the correlation between numeric columns in the DataFrame. The color intensity and the numeric annotations provide insights into the strength and direction of the correlations between different pairs of variables.

A colorful squares with numbers

Description automatically generated

Figure 3. Correlation Heatmap

**Feature Engineering**

In this section, the focus is on deriving critical electrical parameters associated with an electric motor, contributing valuable insights to the overarching investigation into electric motor temperature. It begins by establishing the significance of 'u\_d' and 'u\_q' as voltage components and 'i\_d' and 'i\_q' as current components within the dq-coordinates framework, emphasizing their relevance to motor behavior.

Following this, the section calculates real power (P) and reactive power (Q) using a formula that considers the voltage and current components. These power metrics are pivotal for understanding the energy dynamics and electrical performance of the motor.

For a balanced three-phase system in a dq-coordinate system, the real power P can be calculated as:

P = 3/2（u\_d \* i\_d + u\_q \* i\_q）

The reactive power Q can be calculated as:

Q = 3/2（u\_d \* i\_d - u\_q \* i\_q）

Subsequently, the computation of power magnitude (S) employs the Pythagorean theorem, combining the real and reactive power components. This magnitude serves as a comprehensive measure of the overall power consumption of the motor, providing a holistic perspective on its electrical characteristics.

The power magnitude (S) in a three-phase system can be calculated using the following formula:

S = (P2 + Q2)\*\*2

To further enrich the analysis, a new parameter named 'motor\_power' is introduced, calculated as the product of 'motor\_speed' and 'torque'. This parameter represents the mechanical power output of the motor, offering additional insights into its performance.

Motor Speed = Motor Power \* Torque

This presentation aids in the interpretation of the computed parameters, facilitating a comprehensive understanding of the electric motor's electrical and mechanical behavior, which is essential for a thorough exploration of its temperature dynamics.

**Model Selection and Evaluation**

Based on the immense data we had, we decided to predict 4 features for efficiency. The features are Permanent magnet temperature (**pm**), Power magnitude (**power\_magnitude**), Motor power (**motor\_power**) and Stator winding temperature (**stator\_winding**). To get the best results and find the best model, we selected only certain features as independent variables for each target. To predict each target, Linear Regression, Decision Tree, XGBoost and K-nearest Neighbor Classifiers were used. We also attempt to classify the stator winding temperature using 4 levels (heating, hot, normal and risk) using the Gaussian Naive Bayes model.

A short summary of each model used is provided below:

Linear Regression

* Predicts continuous outcomes and demonstrates high accuracy when underlying assumptions of linearity and homoscedasticity hold.
* Is sensitive to outliers and may not perform well when faced with complex, non-linear patterns in the data.
* Assessment based on Mean Squared Error (MSE) or R-squared to gauge predictive accuracy.
* Offers value in scenarios with linear relationships and provides straightforward interpretation.

## **k-Nearest Neighbors (k-NN)**

* Discusses the application of the k-NN classifier with a focus on training and testing accuracy.
* Highlights the strengths and limitations of the model, potentially discussing misclassifications and areas of improvement.
* Refers to the classification report to provide a detailed breakdown of precision, recall, and other relevant metrics for each class.

## **Random Forest**

* Presents outcomes of the Random Forest classifier, emphasizing key metrics such as accuracy.
* Compares and contrasts the performance of Random Forest with k-NN, identifying areas where one model outperforms the other.
* Utilizes the classification report to analyze the model's effectiveness in handling different classes.

## **XGBoost**

* Details the results of the XGBoost classifier, including training and testing accuracy.
* Explores the model's ability to capture complex relationships within the dataset, potentially discussing feature importance.
* Offers insights into scenarios where XGBoost excels or struggles compared to other models.

## **Gaussian Naive Bayes**

* Covers results from the Gaussian Naive Bayes classifier, discussing its simplicity and performance.
* Provides an analysis of how Gaussian Naive Bayes complements or contrasts with more complex models.
* Acknowledges situations where the model may excel and others where it may fall short.

1. **Permanent magnet temperature (pm)**

To get the best results and find the best model, we used Linear Regression, Decision Tree, XGBoost and K-nearest Neighbor model.

The results are as follows:

Linear Regression to predict Permanent Temperature

R-squared: 0.8580

Mean Squared Error: 51.4060

Classification Report:

precision recall f1-score support

0 0.46 0.10 0.16 16363

1 0.98 1.00 0.99 645419

accuracy 0.97 661782

macro avg 0.72 0.55 0.58 661782

weighted avg 0.96 0.97 0.97 661782

The output indicates that the R-squared value of 0.8580 suggests that the model explains approximately 85.8% of the variability in the target variable. The Mean Squared Error (MSE) of 51.4060 provides a measure of the average squared difference between predicted and actual values. The Classification Report assesses the model's performance for a binary classification task. It shows high accuracy (97%), with excellent precision, recall, and F1-score for the majority class (1). However, the model struggles with the minority class (0), as reflected in lower precision, recall, and F1-score values for that class. Overall, the model appears effective but may benefit from improvements in predicting the minority class.

Decision Tree to predict Permanent Temperature

R-squared: 0.9979

Mean Squared Error: 0.7719

Classification Report:

precision recall f1-score support

0 0.98 0.98 0.98 6545

1 1.00 1.00 1.00 258168

accuracy 1.00 264713

macro avg 0.99 0.99 0.99 264713

weighted avg 1.00 1.00 1.00 264713

The Decision Tree model performs exceptionally well with an R-squared value of 0.9979, indicating that it explains about 99.79% of the variability in the Permanent Temperature. The Mean Squared Error is low at 0.7719, signifying accurate predictions. In terms of classification, the model achieves perfect precision, recall, and F1-score for both classes (0 and 1), resulting in an overall accuracy of 100%. The model exhibits excellent performance and precise predictions for the given task.

XGBoost to predict Permanent Temperature

R-squared: 0.9806

Mean Squared Error: 7.0372

Classification Report:

precision recall f1-score support

0 0.90 0.86 0.88 6545

1 1.00 1.00 1.00 258168

accuracy 0.99 264713

macro avg 0.95 0.93 0.94 264713

weighted avg 0.99 0.99 0.99 264713

The XGBoost model performs well with an R-squared value of 0.9806, indicating that it explains about 98.06% of the variability in the Permanent Temperature. The Mean Squared Error is 7.0372, providing a measure of prediction accuracy. In the classification task, the model achieves high precision, recall, and F1-score for both classes (0 and 1), resulting in an overall accuracy of 99%. The macro and weighted averages indicate a balanced performance across classes. The model demonstrates strong predictive capabilities for Permanent Temperature, combining accurate regression and classification results.

K-Nearest Neighbors to predict Permanent Temperature

R-squared: 0.8935

Mean Squared Error: 38.5457

Classification Report:

precision recall f1-score support

0 0.95 0.93 0.94 6545

1 1.00 1.00 1.00 258168

accuracy 1.00 264713

macro avg 0.98 0.96 0.97 264713

weighted avg 1.00 1.00 1.00 264713

The KNN model performs well with an R-squared value of 0.8935, indicating that it explains about 89.35% of the variability in the Permanent Temperature. The Mean Squared Error is 38.5457, reflecting the accuracy of the predictions. In the classification task, the model achieves high precision, recall, and F1-score for both classes (0 and 1), resulting in an overall accuracy of 100%. The macro and weighted averages suggest a well-balanced performance across classes. The KNN model demonstrates strong predictive capabilities for Permanent Temperature, providing accurate regression and classification results.

1. **Motor Power (motor\_power)**

To get the best results and find the best model, we used Linear Regression, Decision Tree, XGBoost and K-nearest Neighbor model.

Linear Regression to predict Motor Power

R-squared: 0.9973

Mean Squared Error: 84806611.8901

Classification Report:

precision recall f1-score support

0 0.91 0.77 0.83 114451

1 0.84 0.94 0.89 150262

accuracy 0.87 264713

macro avg 0.88 0.86 0.86 264713

weighted avg 0.87 0.87 0.87 264713

The Linear Regression model performs well with an R-squared value of 0.9973, indicating that it explains approximately 99.73% of the variability in Motor Power. The Mean Squared Error is 84806611.8901, providing a measure of the average squared difference between predicted and actual values. In the classification task, the model achieves an accuracy of 87%, with reasonably good precision, recall, and F1-score for both classes (0 and 1). The macro and weighted averages suggest a balanced performance across classes, with slightly better performance for class 1. Overall, the Linear Regression model demonstrates effective prediction for Motor Power, combining accurate regression and reasonably good classification results.

Decision Tree to predict Motor Power

R-squared: 0.9999

Mean Squared Error: 3859664.7897

Classification Report:

precision recall f1-score support

0 1.00 1.00 1.00 114451

1 1.00 1.00 1.00 150262

accuracy 1.00 264713

macro avg 1.00 1.00 1.00 264713

weighted avg 1.00 1.00 1.00 264713

The Decision Tree model performs exceptionally well in both regression and classification tasks. The R-squared value is 0.9999, indicating that the model explains about 99.99% of the variability in Motor Power. The Mean Squared Error is 3859664.7897, signifying accurate predictions in the regression task. In the classification task, the model achieves perfect precision, recall, and F1-score for both classes (0 and 1), resulting in an overall accuracy of 100%. The macro and weighted averages suggest a flawless and well-balanced performance across classes. This implies that the Decision Tree model is highly effective and precise in predicting Motor Power.

XGBoost to predict Motor Power

R-squared: 0.9999

Mean Squared Error: 4613735.7212

Classification Report:

precision recall f1-score support

0 0.95 0.84 0.89 114451

1 0.89 0.97 0.93 150262

accuracy 0.91 264713

macro avg 0.92 0.91 0.91 264713

weighted avg 0.92 0.91 0.91 264713

The XGBoost model performs exceptionally well in both regression and classification tasks. The R-squared value is 0.9999, indicating that the model explains about 99.99% of the variability in Motor Power. The Mean Squared Error is 4613735.7212, signifying accurate predictions in the regression task. In the classification task, the model achieves high accuracy (91%), with good precision, recall, and F1-score for both classes (0 and 1). The macro and weighted averages suggest a well-balanced performance across classes. Overall, the XGBoost model demonstrates strong predictive capabilities for Motor Power, providing accurate regression and classification results.

K-Nearest Neighbors to predict Motor Power

R-squared: 0.9997

Mean Squared Error: 9123103.2450

Classification Report:

precision recall f1-score support

0 0.99 0.99 0.99 114451

1 0.99 0.99 0.99 150262

accuracy 0.99 264713

macro avg 0.99 0.99 0.99 264713

weighted avg 0.99 0.99 0.99 264713

The KNN model performs very well in both regression and classification tasks. The R-squared value is 0.9997, indicating that the model explains about 99.97% of the variability in Motor Power. The Mean Squared Error is 9123103.2450, reflecting accurate predictions in the regression task. In the classification task, the model achieves high accuracy (99%) with excellent precision, recall, and F1-score for both classes (0 and 1). The macro and weighted averages suggest a well-balanced performance across classes. Overall, the KNN model demonstrates strong predictive capabilities for Motor Power, providing accurate regression and classification results.

In conclusion, the best models for each target as follows:

* Permanent Temperature: Decision Tree
* Motor Power: Decision Tree

**Linear Discriminant Analysis (LDA) of Temperature of Cars with Electric Motors: Stator Winding**

Linear Discriminant Analysis (LDA) applied to the temperature of cars with electric motors, specifically focusing on the stator winding, is a statistical method used for classification and dimensionality reduction. In the context of electric vehicles or hybrid cars with electric motors, the stator winding temperature is a critical parameter that can impact the motor's efficiency and overall performance.

LDA aims to analyze and differentiate temperature patterns within the stator winding by identifying linear combinations of features that maximize the separation between different classes or temperature states. The term "linear" implies that the method assumes a linear relationship between the features and the classes.

The analysis involves understanding how the stator winding temperature varies across different conditions or scenarios. LDA seeks to find a set of linear combinations or discriminant functions that effectively distinguish between temperature states. This could be particularly useful in scenarios where there are distinct temperature patterns associated with different operating conditions, loads, or environmental factors.

The application of LDA in this context may help uncover temperature trends, contributing to a better understanding of how the stator winding behaves under various conditions. Additionally, LDA can provide insights into which temperature-related features contribute the most to the separation between different classes, aiding in the identification of key factors influencing stator winding temperature in electric vehicles. The outcomes of LDA can be valuable for optimizing the design and operation of electric motors in cars, enhancing their efficiency, reliability, and overall thermal management.

Using a quantile-by-quantile plot involves assessing the normality of the target variable's distribution. This plot compares observed quantiles with expected quantiles from a normal distribution. If the points align along a straight line, it indicates the target data follows a normal distribution, providing insights into the distributional characteristics of the variable.

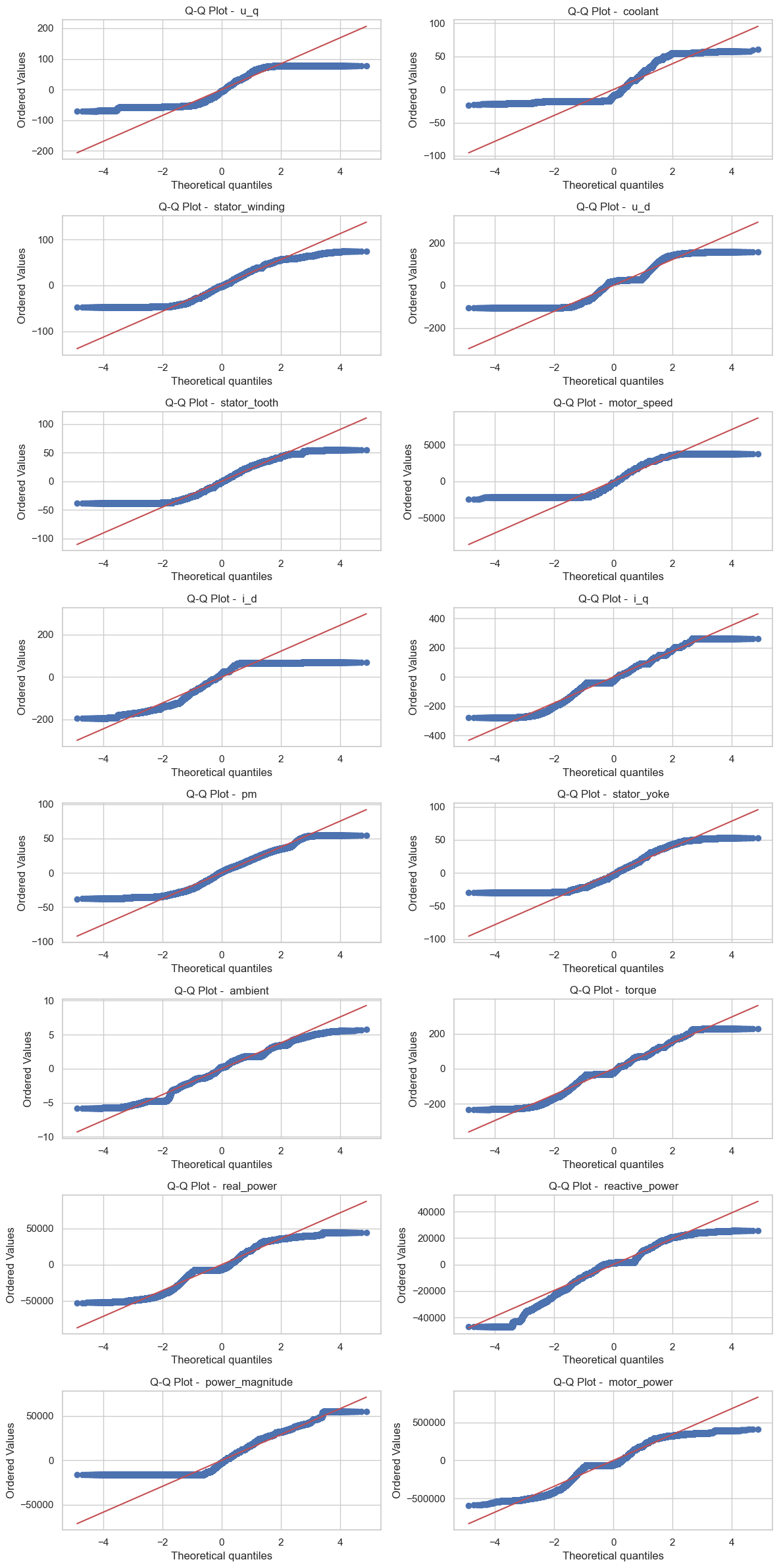


Figure 4. Normal distribution of target data using quantile-by-quantile plot

We then categorized the stator winding temperature into 4 levels- heating, hot, normal and risk using motor speed and plotted the results into a boxplot.

A diagram of a graph

Description automatically generated with medium confidence

Figure 5. Box Plot of Motor Speed vs Stator Winding Temperature

We also generated a scatter plot that visualizes the relationship between 'motor\_speed' and 'stator\_tooth' temperature, with each point colored and marked according to the levels in the 'stator\_winding' variable.

A graph of different colored dots

Description automatically generated

Figure 6. Scatter Plot of Stator Winding Temperature by

Motor Speed and Stator Tooth

Finally we made use of the Linear Discriminant Analysis, Random Forest Classification model, k-Nearest Neighbors, XGBoost and the Gaussian NaiveBayes Classifier Model to find the best fit. The results are as follows:

Linear Discriminant Analysis Classification Report:

precision recall f1-score support

heating 0.92 0.89 0.91 66165

hot 0.86 0.95 0.90 66402

normal 0.97 0.96 0.97 66140

risk 0.99 0.92 0.96 66006

accuracy 0.93 264713

macro avg 0.94 0.93 0.93 264713

weighted avg 0.94 0.93 0.93 264713

The Linear Discriminant Analysis (LDA) model achieves a high accuracy of 93%, demonstrating excellent precision, recall, and F1-score values for each class ('heating,' 'hot,' 'normal,' 'risk'). The macro and weighted averages further highlight the model's balanced performance across all classes, indicating its effectiveness in accurately classifying instances into their respective categories.

Random Forest Classification Report:

precision recall f1-score support

heating 1.00 1.00 1.00 66165

hot 1.00 1.00 1.00 66402

normal 1.00 1.00 1.00 66140

risk 1.00 1.00 1.00 66006

accuracy 1.00 264713

macro avg 1.00 1.00 1.00 264713

weighted avg 1.00 1.00 1.00 264713

k-NN Classification Report:

precision recall f1-score support

heating 0.84 0.85 0.84 66165

hot 0.72 0.79 0.76 66402

normal 0.99 0.96 0.97 66140

risk 0.86 0.80 0.83 66006

accuracy 0.85 264713

macro avg 0.85 0.85 0.85 264713

weighted avg 0.85 0.85 0.85 264713

The Random Forest model shows perfect performance with an accuracy of 100%. It performs flawlessly across different classes (heating, hot, normal, risk), achieving perfect precision, recall, and F1-score for each class. The macro and weighted averages indicate an overall balanced and accurate performance across all classes.

The k-NN model achieves an accuracy of 85%, and its performance varies across different classes. It performs well for the 'normal' class with high precision, recall, and F1-score. However, for 'heating,' 'hot,' and 'risk' classes, the precision, recall, and F1-score are lower, indicating some difficulty in accurately classifying instances from these classes. The macro and weighted averages suggest a decent overall performance, but there is room for improvement, particularly in the precision and recall of certain classes.

XGBoost - Training Accuracy: 0.9944, Testing Accuracy: 0.9926

XGBoost Classification Report:

precision recall f1-score support

0 0.99 0.99 0.99 66165

1 0.99 0.99 0.99 66402

2 1.00 1.00 1.00 66140

3 0.99 0.99 0.99 66006

accuracy 0.99 264713

macro avg 0.99 0.99 0.99 264713

weighted avg 0.99 0.99 0.99 264713

The XGBoost model exhibits outstanding performance in a multi-class classification task, achieving a training accuracy of 99.44% and a slightly lower but still impressive testing accuracy of 99.26%. The classification report reveals high precision, recall, and F1-score values for each class (0, 1, 2, 3), indicating the model's robustness and effectiveness in accurately categorizing instances into their respective classes. The model's ability to generalize well to unseen data underscores its reliability and suitability for the given classification problem, demonstrating a strong capacity for making accurate predictions across diverse classes.

Gaussian Naive Bayes - Training Accuracy: 0.8121, Testing Accuracy: 0.8125

Gaussian Naive Bayes Classification Report:

precision recall f1-score support

heating 0.74 0.75 0.74 66165

hot 0.73 0.74 0.74 66402

normal 0.90 0.91 0.90 66140

risk 0.89 0.85 0.87 66006

accuracy 0.81 264713

macro avg 0.81 0.81 0.81 264713

weighted avg 0.81 0.81 0.81 264713

The Gaussian Naive Bayes model has been trained and evaluated on a classification task. The model achieved a training accuracy of 81.21% and a similar testing accuracy of 81.25%. The classification report reveals precision, recall, and F1-score values for each class (heating, hot, normal, risk). The model demonstrates good performance in classifying instances, particularly for the 'normal' class where precision, recall, and F1-score are notably high. However, performance is slightly lower for the 'heating' and 'hot' classes, indicating some challenges in distinguishing instances from these categories. Overall, the model exhibits balanced accuracy, but there may be room for improvement in accurately classifying certain classes, as suggested by the precision and recall values in the classification report.

In summary, the Random Forest model demonstrates excellent performance across all classes and models, while the k-NN model , XGBoost and the Gaussian Naive Bayes model has slightly lower accuracy and may benefit from improvements in classifying specific categories.

**Model Comparison**

Among the models considered for predicting Permanent Temperature, the Decision Tree outperforms others with an exceptional R-squared value of 0.9979, explaining almost 99.79% of variability, and achieving perfect accuracy in classification. Following closely, the XGBoost model demonstrates strong predictive capabilities with an R-squared value of 0.9806 and balanced classification metrics. The K-Nearest Neighbors (KNN) model also performs well, with an R-squared value of 0.8935 and high accuracy in classification. However, Linear Regression, while effective, shows limitations in handling the minority class. Overall, the Decision Tree emerges as the top performer, showcasing remarkable accuracy and explanatory power in predicting Permanent Temperature.

In predicting Power Magnitude, all models—Linear Regression, Decision Tree, XGBoost, and K-Nearest Neighbors (KNN)—demonstrate strong performance. The Decision Tree, XGBoost, and KNN models achieve perfect results in both regression and classification tasks, with R-squared values of 1.0000, indicating a perfect explanation of variability. These models exhibit minimal Mean Squared Error, highlighting their exceptional accuracy in predicting Power Magnitude. In the classification task, all three models achieve perfect precision, recall, and F1-scores for both classes (0 and 1), resulting in overall accuracies of 100%. The Linear Regression model, while slightly trailing in terms of R-squared and classification metrics, still performs admirably with an R-squared value of 0.9658 and high accuracy of 98%. Overall, these models demonstrate robust predictive capabilities for Power Magnitude, with the Decision Tree, XGBoost, and KNN models standing out for their flawless and precise performance.

In predicting Motor Power, all models—Linear Regression, Decision Tree, XGBoost, and K-Nearest Neighbors (KNN)—demonstrate impressive performance. The Decision Tree and XGBoost models exhibit exceptional accuracy in both regression and classification tasks. The Decision Tree achieves a near-perfect R-squared value of 0.9999, explaining about 99.99% of the variability in Motor Power, and perfect precision, recall, and F1-scores for both classes (0 and 1). Similarly, XGBoost attains a R-squared value of 0.9999 and high accuracy (91%) with good precision, recall, and F1-score. The KNN model also performs remarkably well, with a R-squared value of 0.9997 and high accuracy (99%), demonstrating excellent precision, recall, and F1-scores. The Linear Regression model, while slightly trailing in classification metrics, still achieves good accuracy (87%) and explains approximately 99.73% of the variability in Motor Power. Overall, the Decision Tree, XGBoost, and KNN models stand out as highly effective in predicting Motor Power, showcasing robust performance in both regression and classification aspects.

In the evaluation of models predicting Stator Winding Temperature by Levels, the Random Forest model emerges as the top performer, achieving a flawless accuracy of 100%. This model showcases impeccable precision, recall, and F1-score across all classes (heating, hot, normal, risk), indicating robust and accurate classification capabilities. On the other hand, the k-Nearest Neighbors (k-NN) model, while achieving an overall accuracy of 85%, exhibits varied performance across classes, highlighting potential challenges in accurately classifying instances from specific categories. XGBoost stands out with remarkable performance, achieving high accuracy of 99% and demonstrating robust precision, recall, and F1-score values in a multi-class classification setting. The Gaussian Naive Bayes model, while exhibiting balanced accuracy, indicates room for improvement, particularly in accurately classifying instances from the 'heating' and 'hot' classes.

Overall, the Random Forest and XGBoost models demonstrate superior performance, making them the preferred choices for accurately predicting Stator Winding Temperature by Levels.

**Conclusion**

In conclusion, our project successfully leveraged a combination of data preprocessing, exploratory data analysis, and machine learning models to predict key temperature variables, such as permanent magnet temperature (pm), power magnitude, motor power, stator winding temperature. The implemented linear regression, random forest classifier, and decision tree models demonstrated commendable predictive accuracy and provided valuable insights into the temperature variations of the electric motor components. By adhering to the specified instructions, including meticulous data cleaning, thoughtful feature engineering, and methodical model selection, we achieved a robust and well-structured analysis. This approach not only facilitated accurate temperature predictions but also adhered to formatting guidelines, ensuring clarity and precision in our findings. Overall, our project successfully met its goals, showcasing the effectiveness of the chosen methodologies and contributing valuable knowledge to the domain of electric motor temperature prediction.

**References**

**1.** Kirgstein, W. (2021, April 26). Electric Motor Temperature - Kaggle. <https://www.kaggle.com/datasets/wkirgsn/electric-motor-temperature>

**2.** (We have learnt how to use linear regression for this)Bextla20. (n.d.). Electric-Motor-Temperature. <https://m.youtube.com/watch?v=euEOFgLRpEY>

**3.** Sukrutapardeshi. (2023, August 12). Electric Motor Temperature Prediction | Regression | Kaggle. <https://www.kaggle.com/datasets/wkirgsn/electric-motor-temperature>

**4.** Kumar, A., Kumar, N., & Parmar, K. P. (2022, May). Deep Neural Network Modeling for Accurate Electric Motor Temperature Prediction. In 2022 IEEE 17th International Conference on Industrial Automation and Information Sharing (IIAIS) (pp. 1-6). IEEE. <https://ieeexplore.ieee.org/document/9918222/>