BA 476 Team 7

MACHINE PREDICTIVE MAINTENANCE

Classification

Team 7

Yijun Kim, Jaejoong Kim, Woohyeon Her, Seokhoon Shin, Joshua Nahm

Problem Statement

We are living in a world where ...

01

Economies shaped by manufacturing

28% of Korea's GDP is composed of manufacture

Problem Statement

We are living in a world where ...

01

Economies shaped by manufacturing

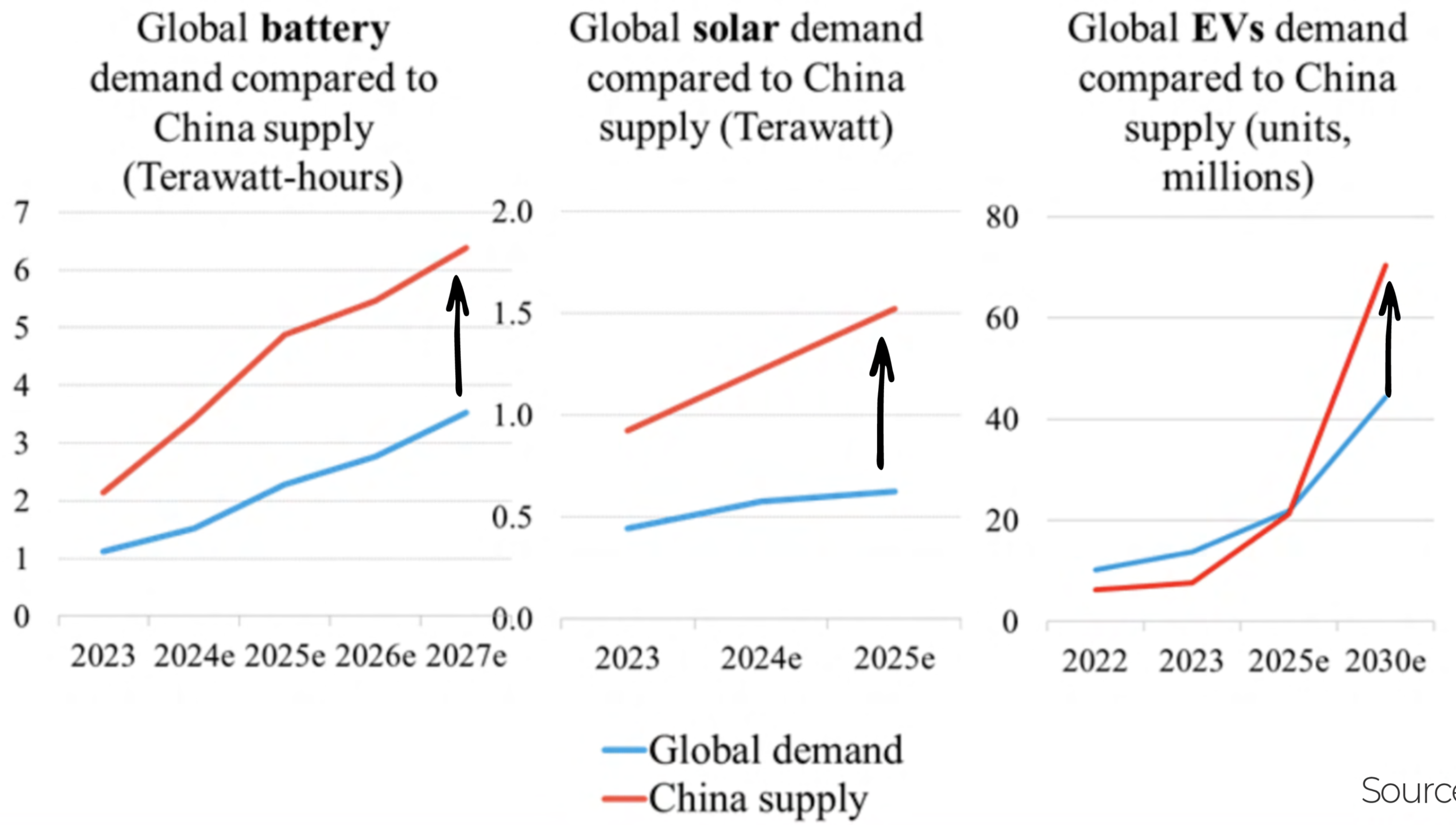
28% of Korea's GDP is composed of manufacture

02

The global manufacturing landscape faces new challenges

- **Overproduction**
- **Rising costs**

Global Demand Compared to Supply



Source: BNEF, IEA, U.S Treasury

Problem Statement

We are living in a world where ...

01

Economies shaped by manufacturing

28% of Korea's GDP is composed of manufacture

02

The global manufacturing landscape faces new challenges

- **Overproduction**
- **Rising costs**

03

So, we believe the key is ...

Predictive maintenance

Problem Statement

By using ML ...

1. Predict equipment failures
2. Reduce downtime
3. Optimize operation

→

Predictive maintenance classification model

Cut costs and improve efficiency

Description of Data - Data Set

Original Dataset from Kaggle

- # of Rows: **10,000**
- # of columns: **10**
- Outcome Variable
  - **Target : No Failure = 0, Failure = 1**
- Predictors: **6 features**
- No Missing values
- Data Imbalance between Failure & No Failure
  - No Failure: **96.7%**
  - Failure: **3.3%**

Variables

Numertical Variables

- Air Pressure [°C] - float64
- Process temperature [°C] - float64
- Rotational speed [rpm] - int64
- Torque [Nm] - float64
- Tool wear [min] - int64
- UDI - int64
- Target - int64

Categorical Variables

- Type - object
- Failure Type - object
- Produce ID - object

07

Description of Data - Data Set

Asked GPT to **RANDOMLY** create 40,000 more instances.

- # of Rows: **50,000**
- Dropped **'UDI', 'Product ID', 'Failure Type'**
- # of columns: **7**
- Outcome Variable
  - **Target : No Failure = 0, Failure = 1**
- Predictors: **6 features**
- No Missing values
- Distribution of No Failure & Failure
  - No Failure: **68%**
  - Failure: **32%**

Variables

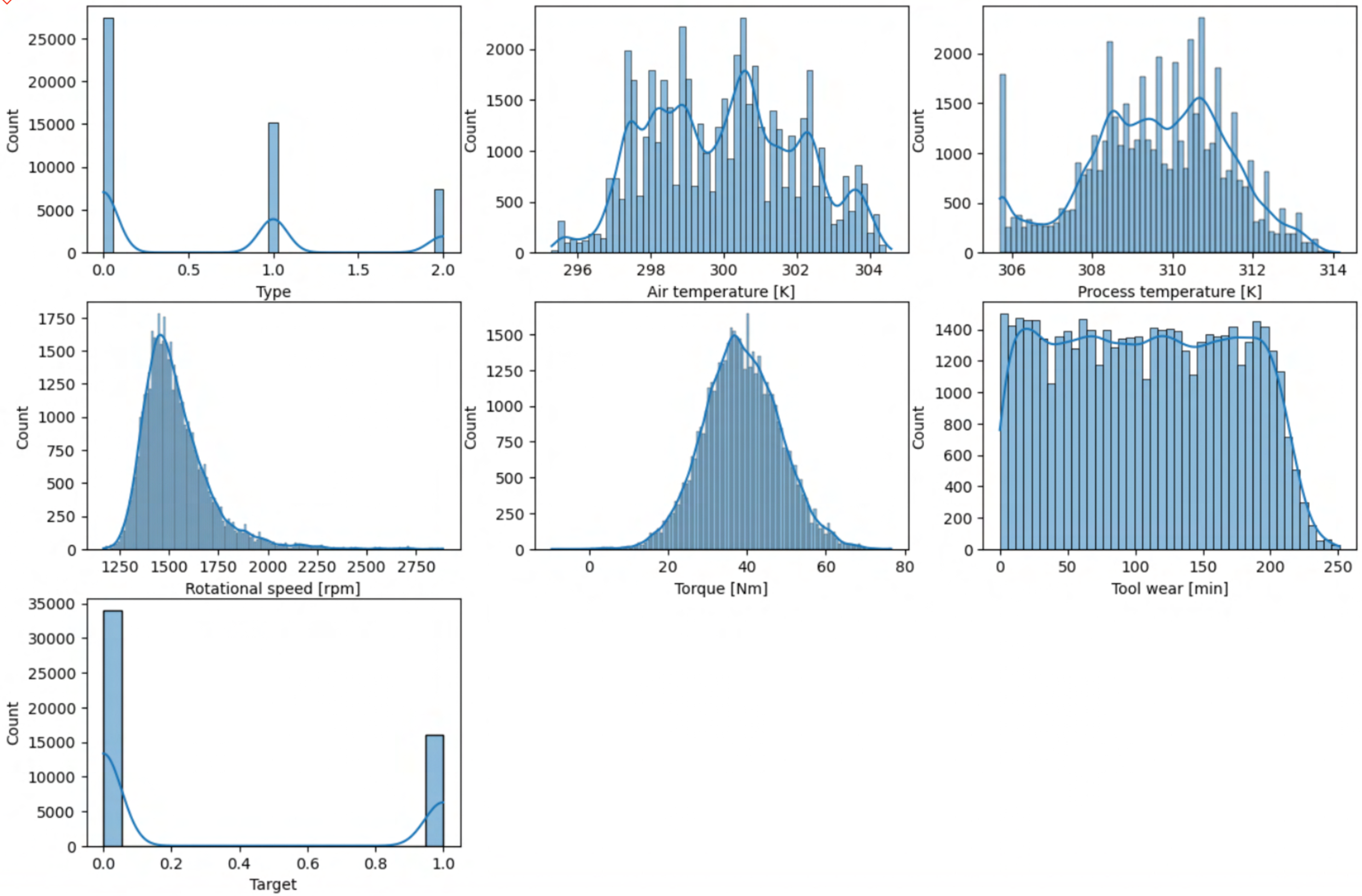
Numertical Variables

- Air Pressure [°C] - float64
- Process temperature [°C] - float64
- Rotational speed [rpm] - int64
- Torque [Nm] - float64
- Tool wear [min] - int64
- ~~UDI - int64~~
- Target - int64

Categorical Variables

- Type - object
- ~~Failure Type - object~~
- ~~Produce ID - object~~

Distribution for each Features



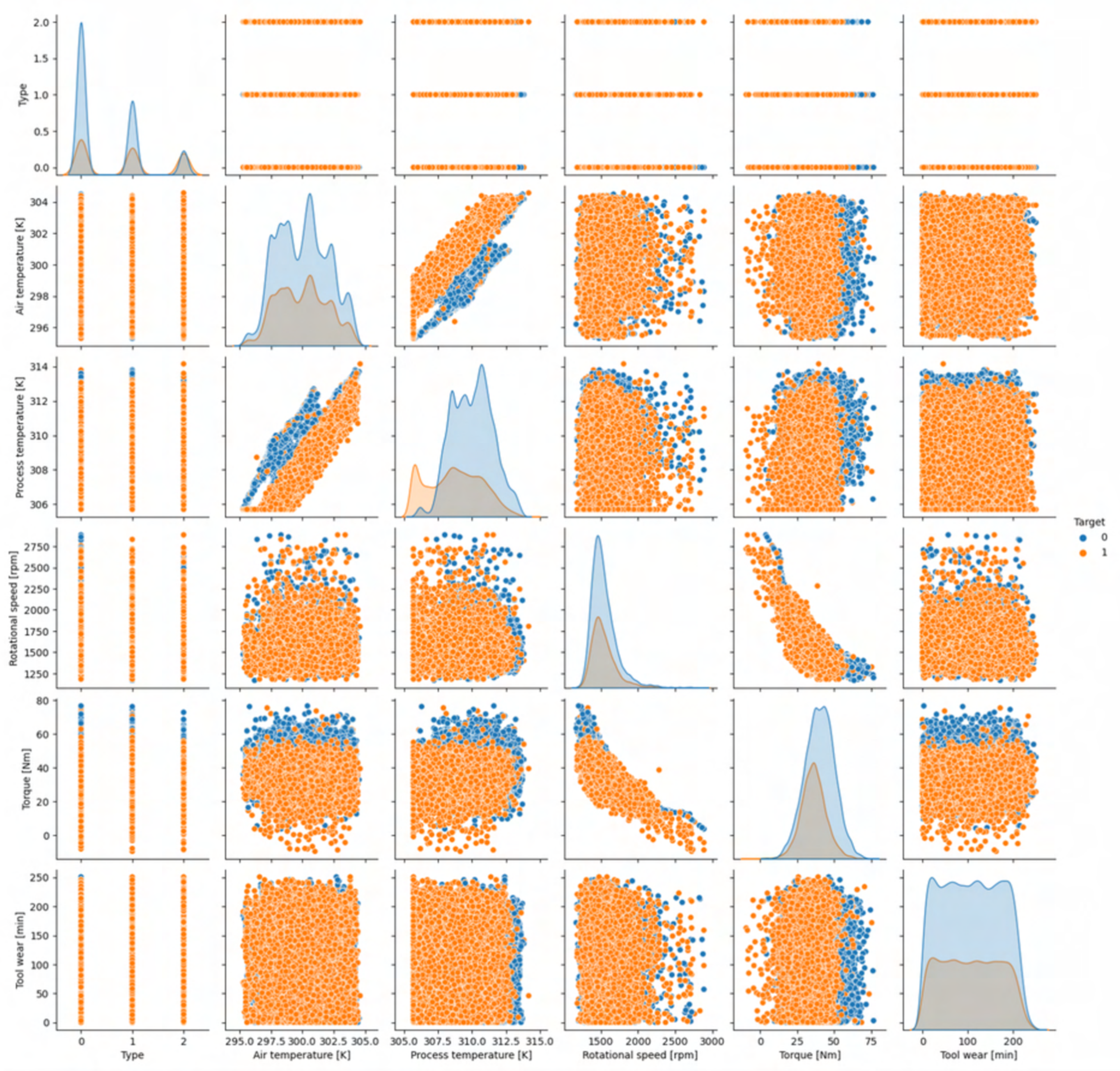
Air temperature

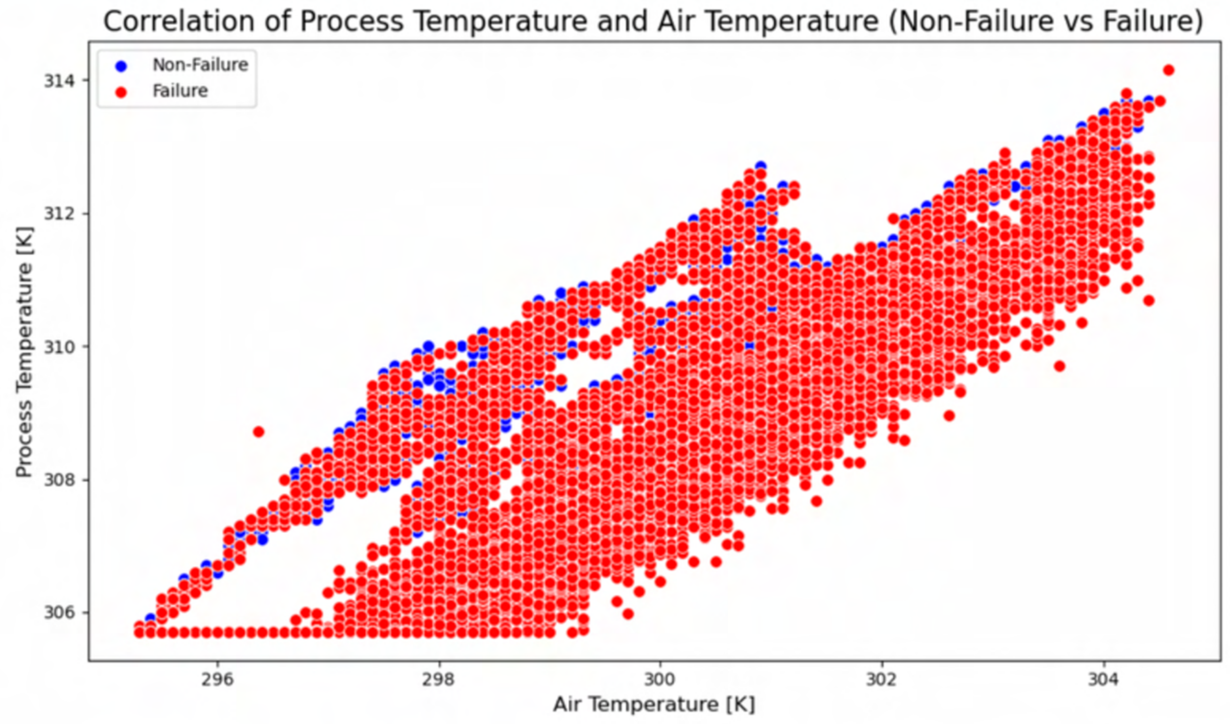
Measures the surrounding temperature affecting the machine's operational conditions.

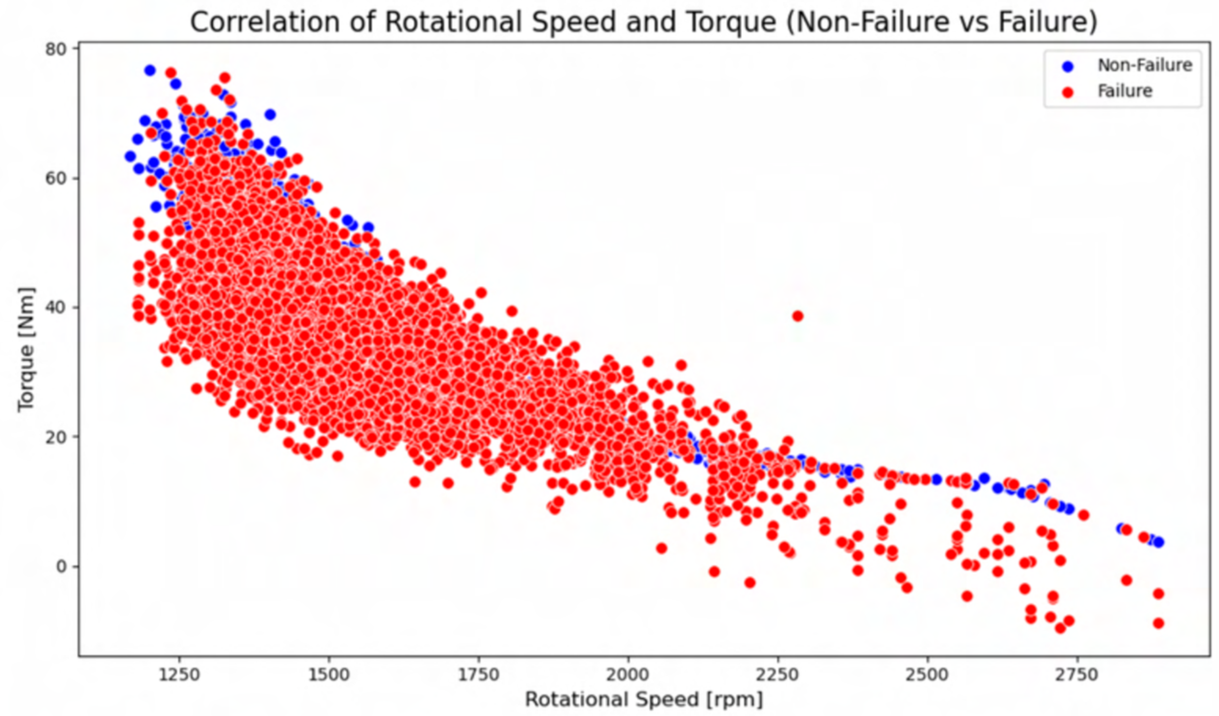
Process temperature

Represents the temperature of the production process.

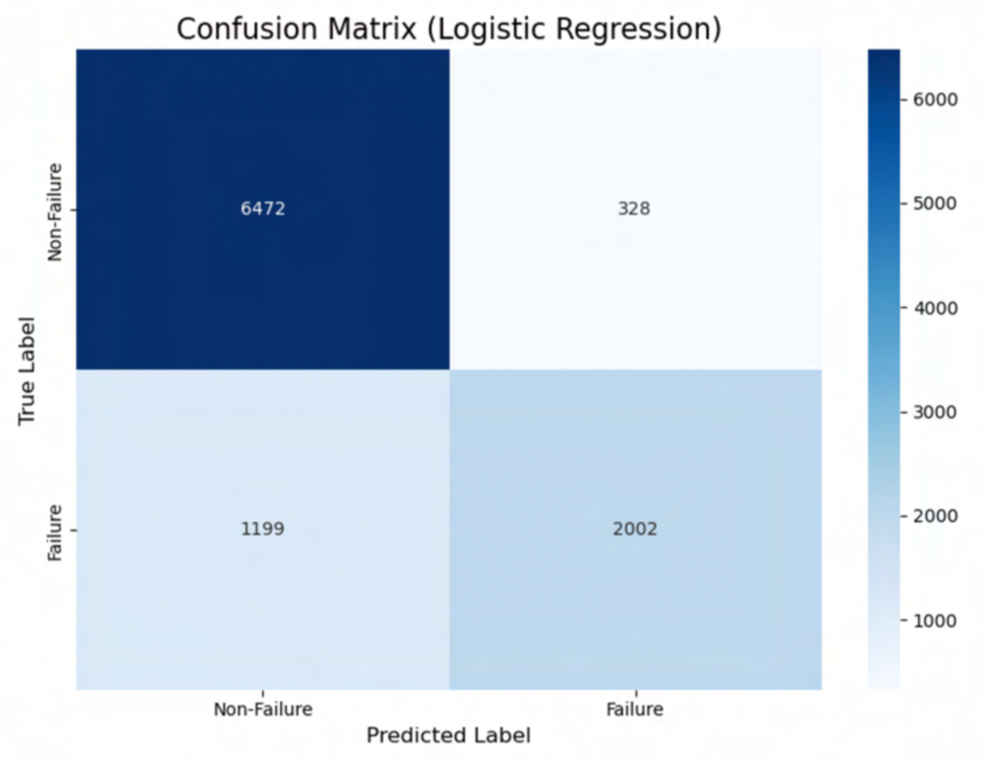
Correlation for each Features





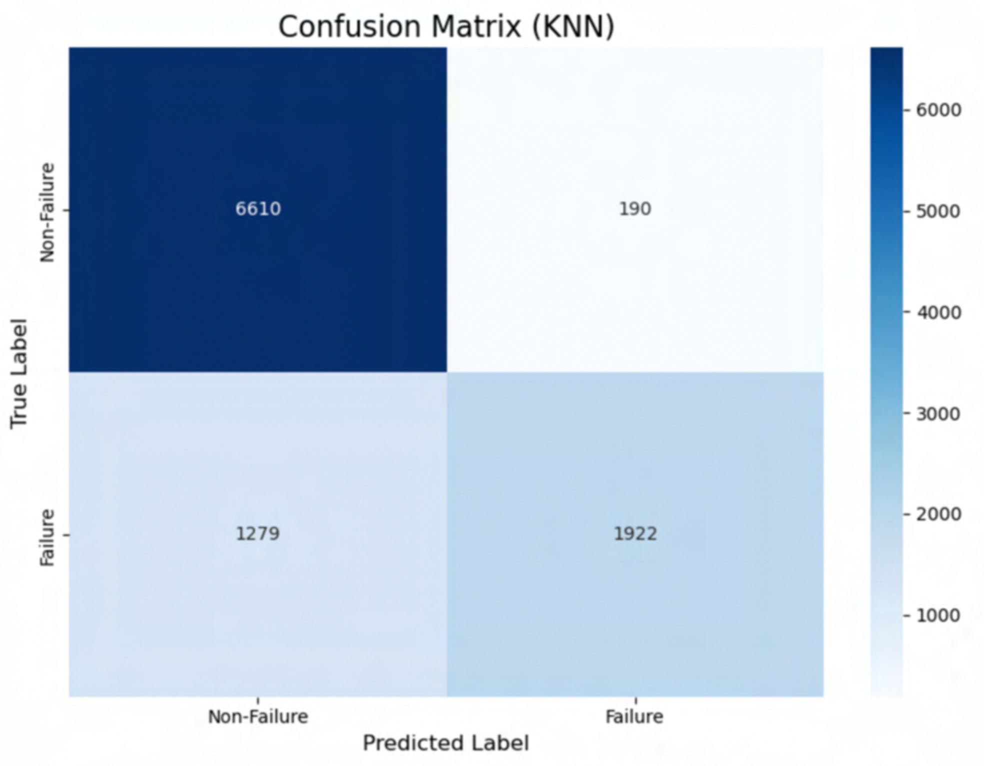


Training different Models



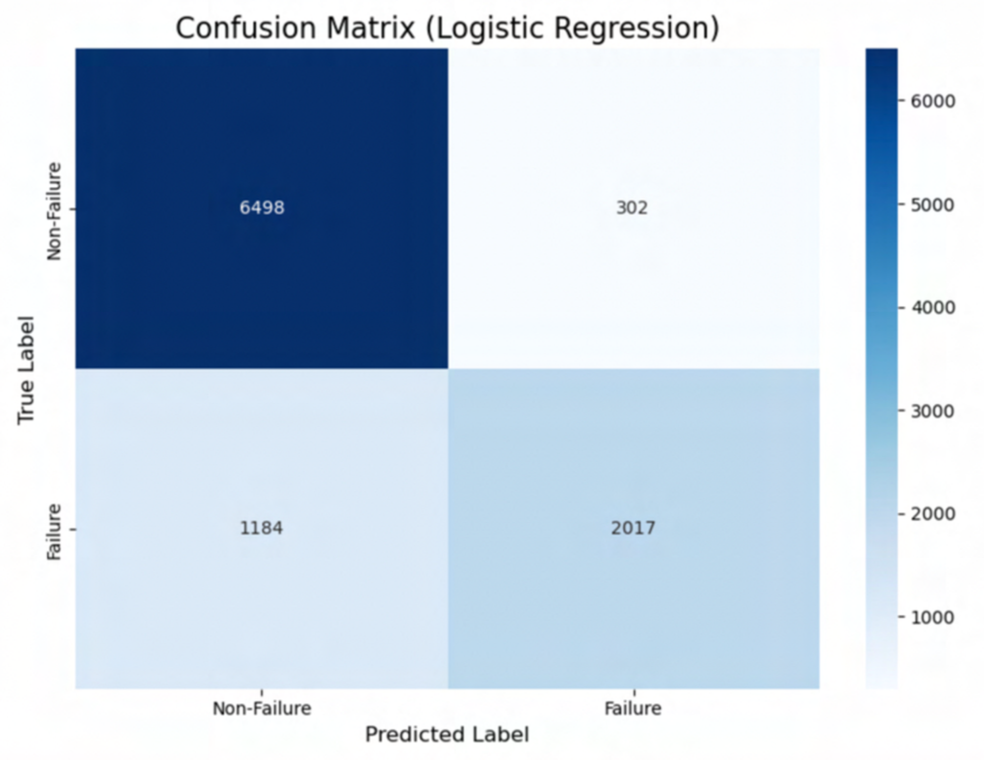
Logistic Regression

Test Accuracy :
0.8514
Train Accuracy :
0.8549



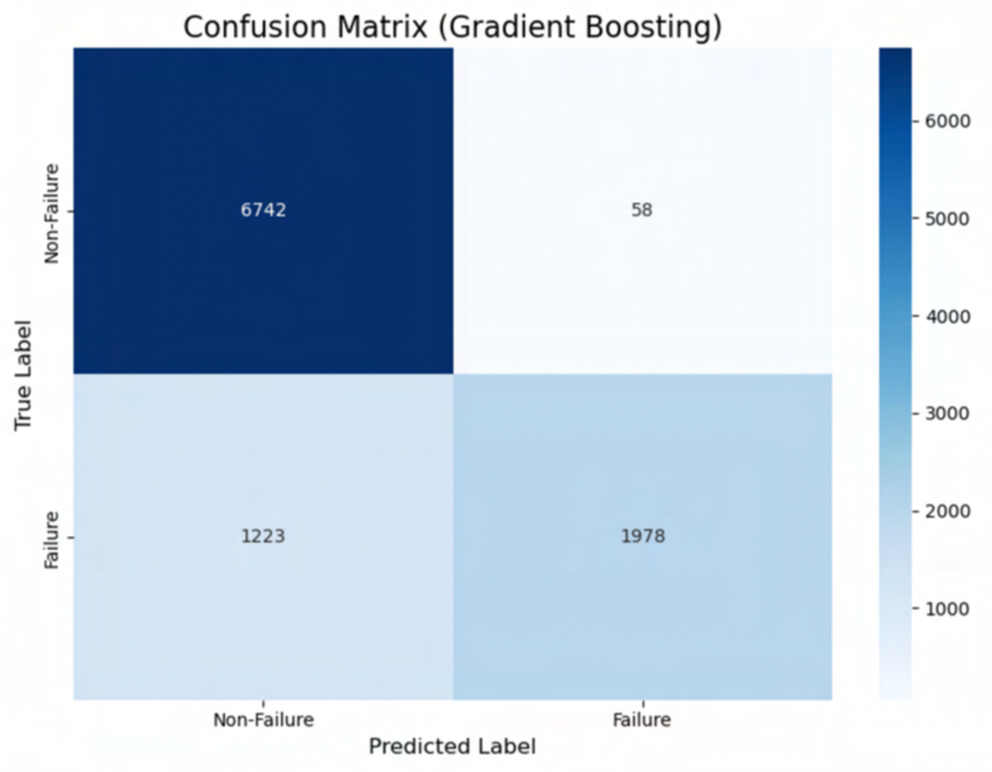
KNN

Test Accuracy :
0.8519
Train Accuracy :
0.8790



Random Forest

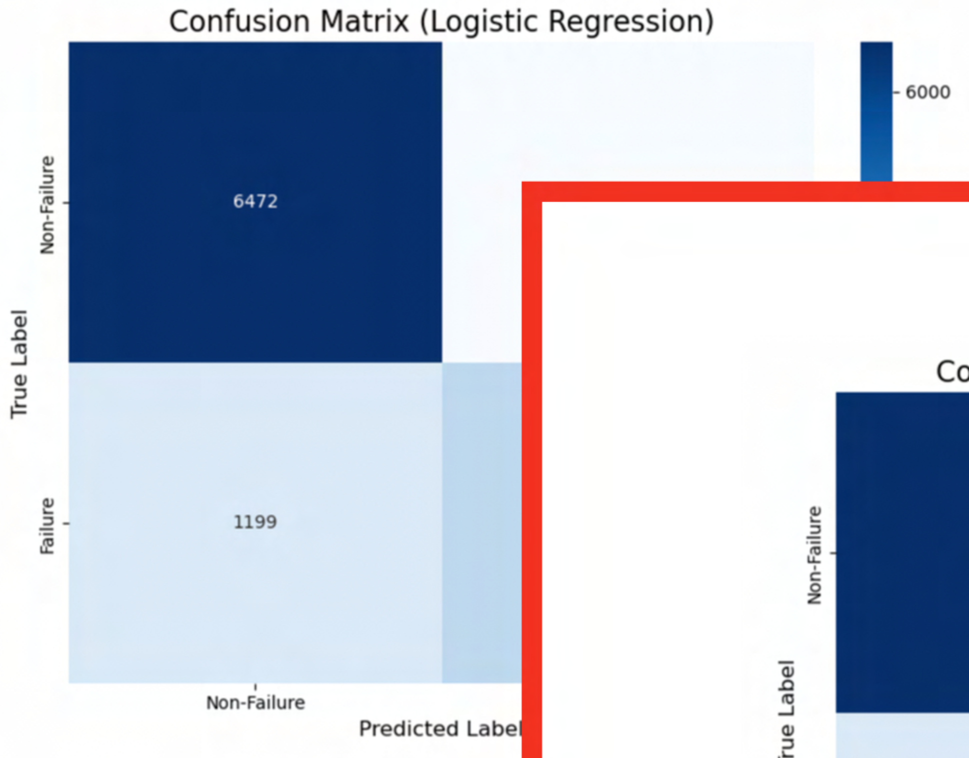
Test Accuracy :
0.8638
Train Accuracy :
0.9076



Boosting

Test Accuracy :
0.8737
Train Accuracy :
0.8767

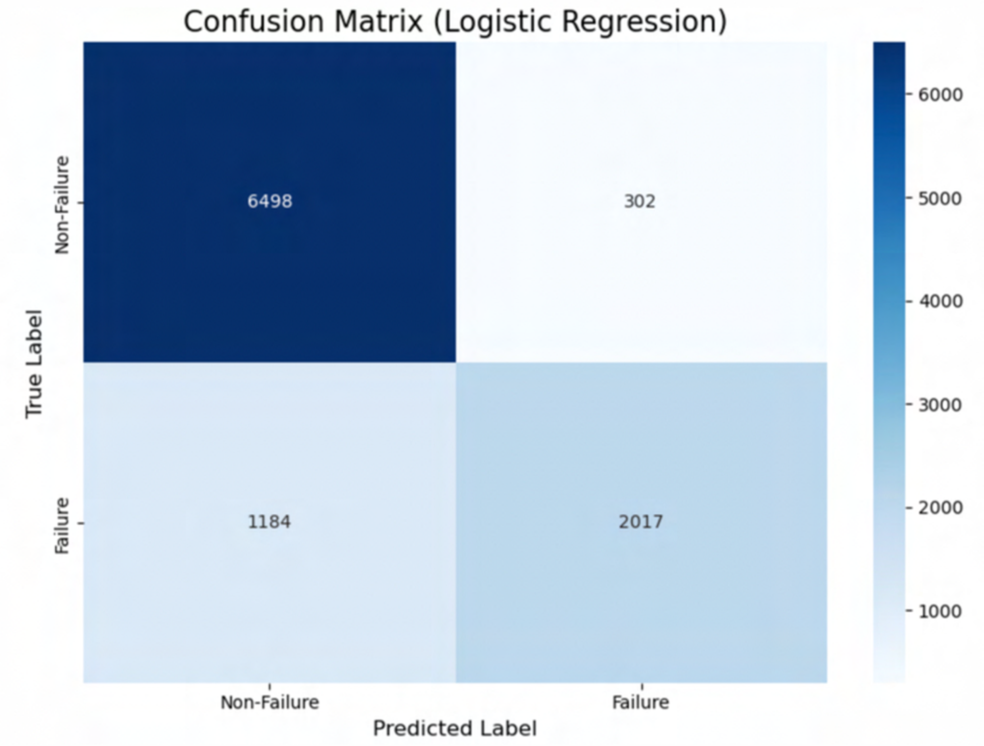
Training different Models



Logistic Regression

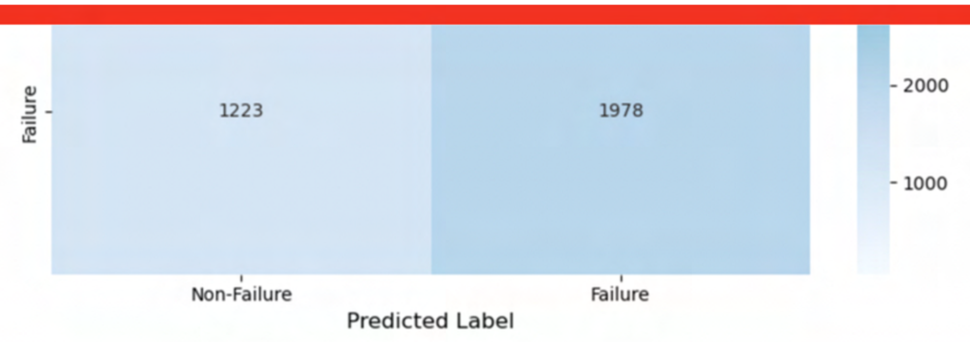
KNN

Test Accuracy :
0.8519
Train Accuracy :
0.8790



Random Forest

Test Accuracy :
0.8638
Train Accuracy :
0.9076



Boosting

Test Accuracy :
0.8737
Train Accuracy :
0.8767

Interpretation of Important Features

Process Temperature

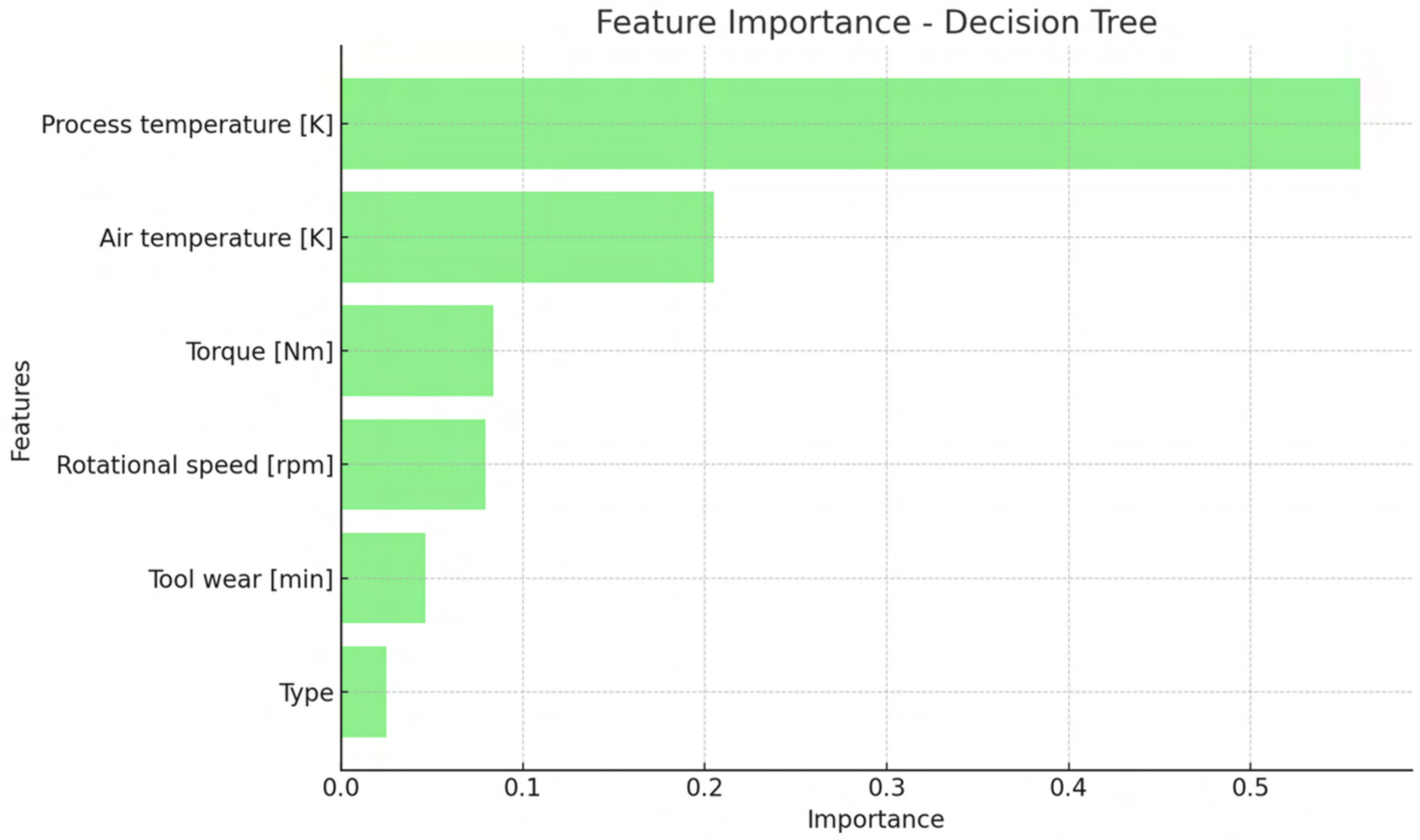
- The most critical feature, strongly influencing predictions.

WHY? Process temperature is directly linked to operational conditions and material stress.

Air Temperature

- The second most important feature, impacting cooling efficiency and thermal stability.

WHY? Variations in air temperature affect equipment overheating risks.



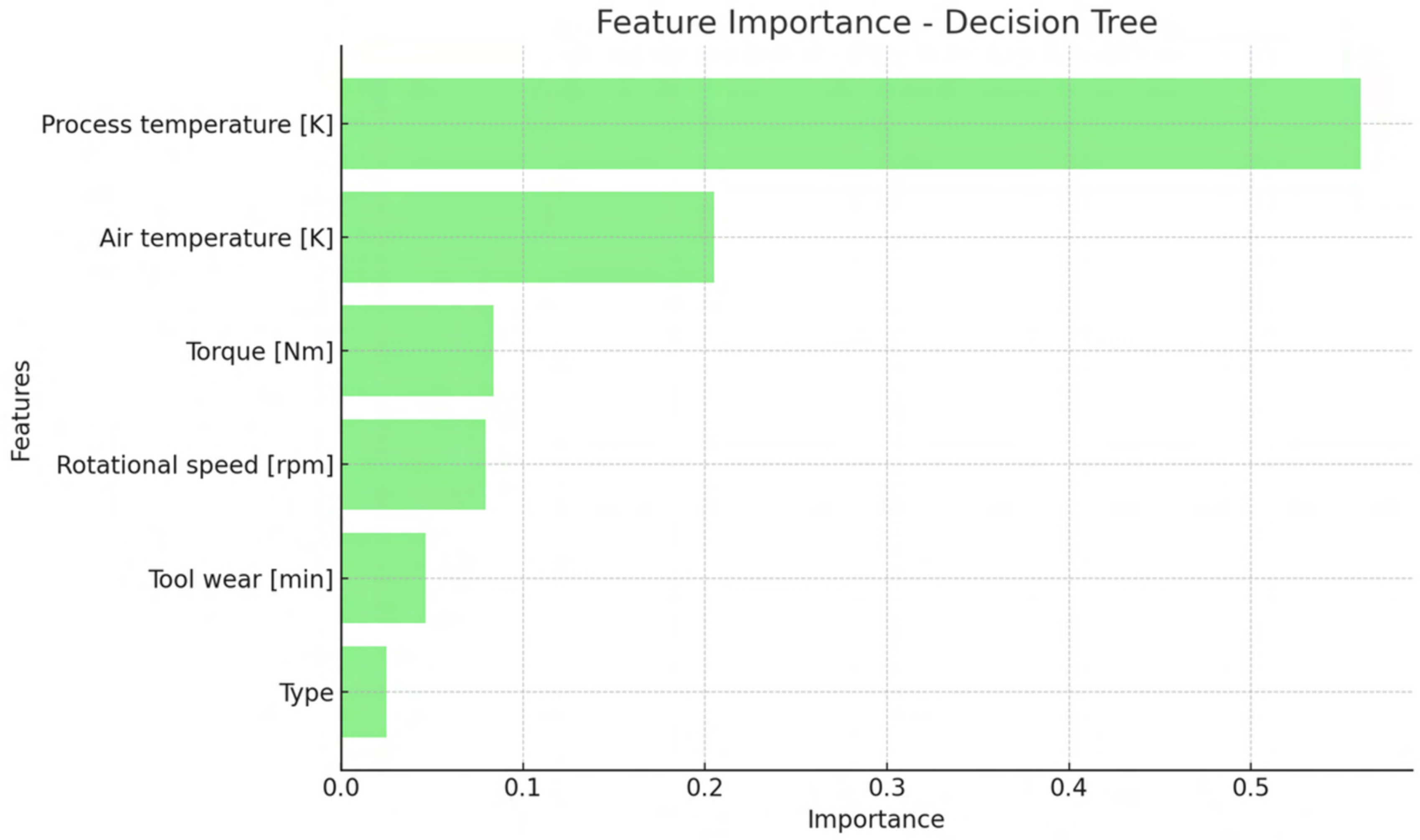
Interpretation of Important Features

Torque & Rotational Speed

- Moderate influence, reflecting mechanical load and stress on equipment.

WHY?

Mechanical performance metrics are critical but secondary to temperature-related factors.



Further Interpretation

Process Temperature

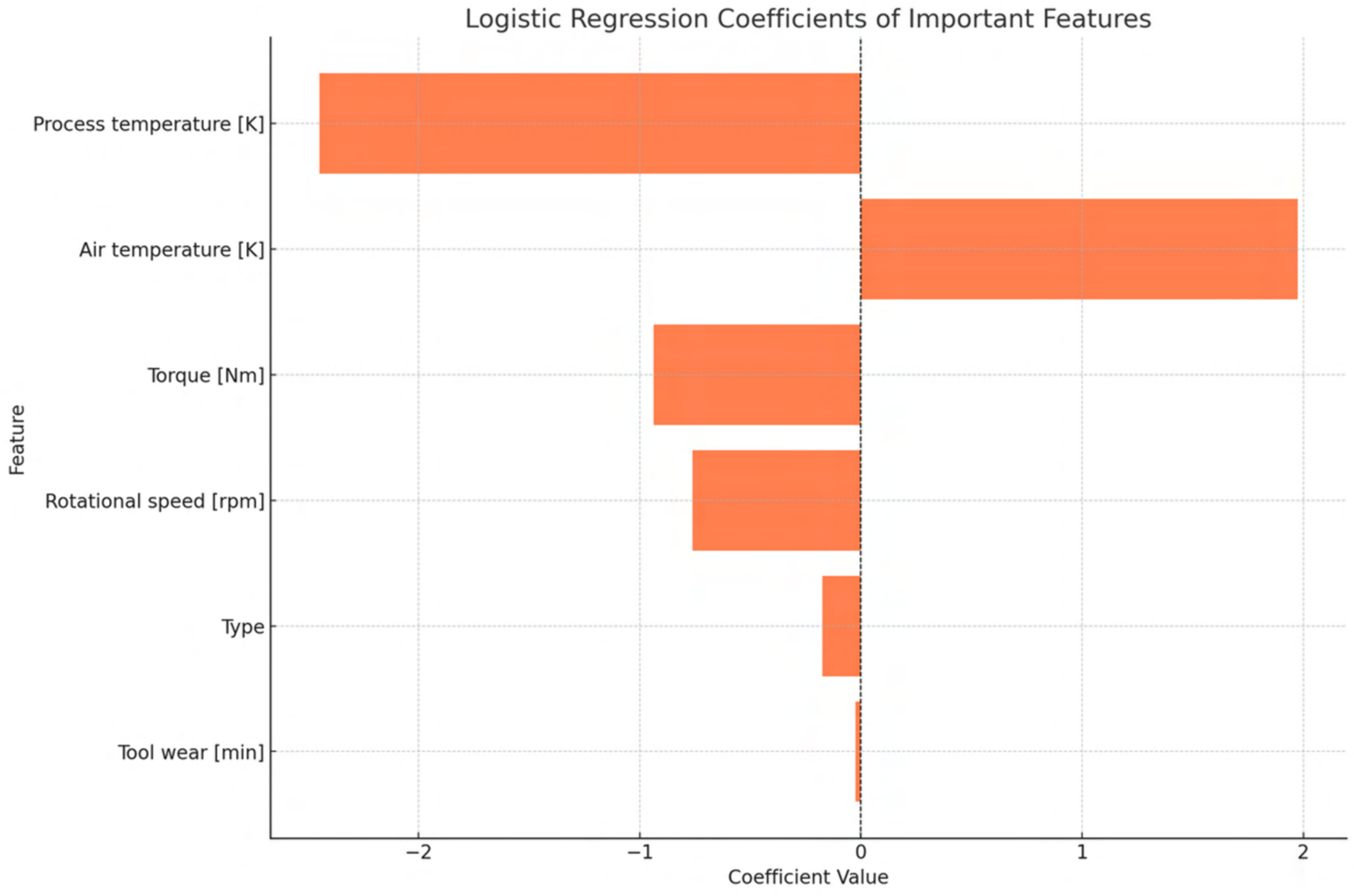
- The largest (negative) coefficient in magnitude.

A decrease in the process temperature is strongly associated with an increased likelihood of failure

Air Temperature

- The second most influential feature, with a positive coefficient.

Indicates that higher air temperatures increase the likelihood of failure.

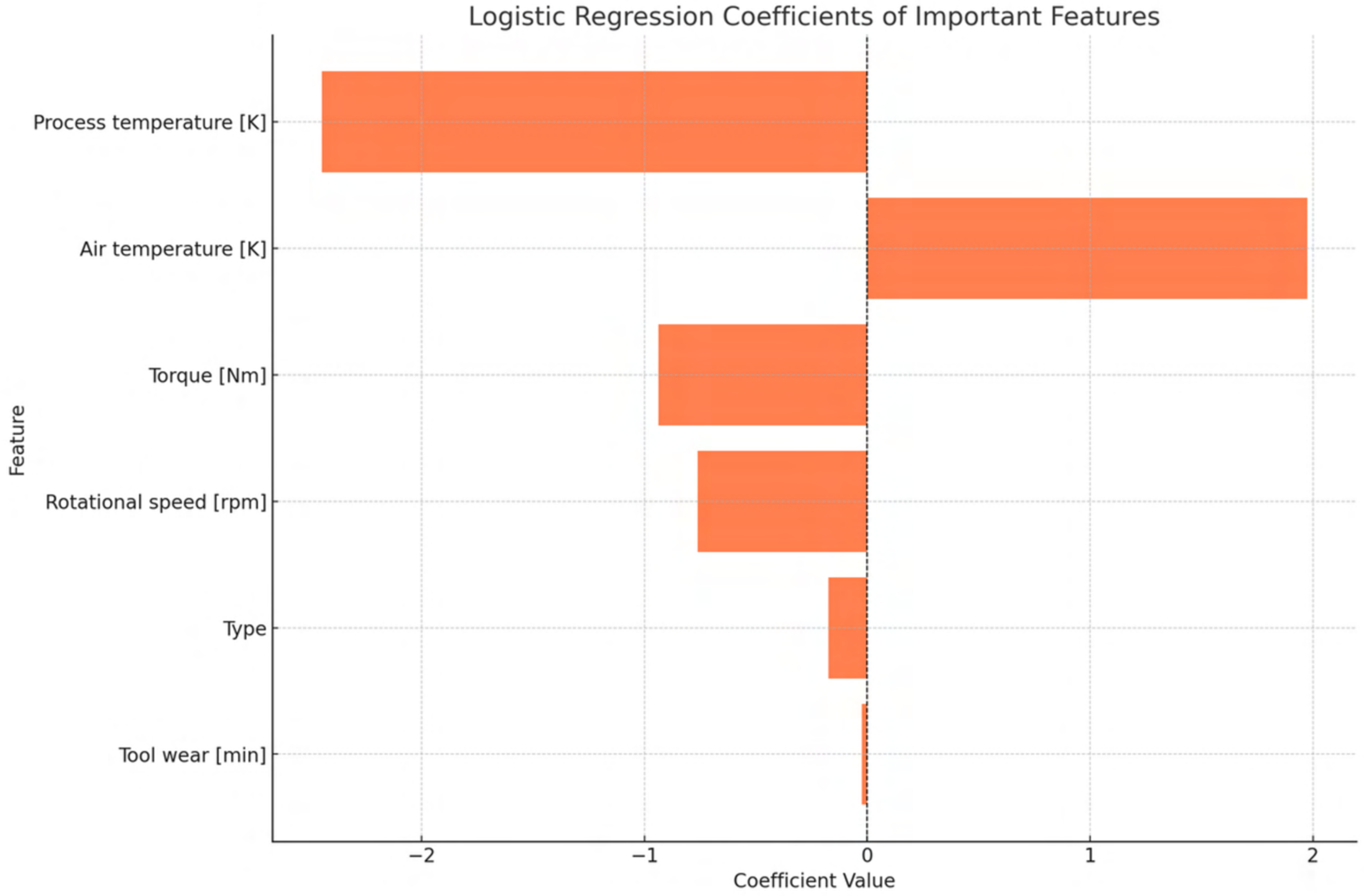


Further Interpretation

Torque [Nm] and Rotational Speed [rpm]

- Moderate negative impacts

Implying that increases in these variables reduce the likelihood of failure.



Biggest Challenge

Creating New Dataset (Adding Rows)

- Created about 50 random datasets and trained individual models.
- Tried different number of rows and different ratio of fail to non-fail.
- Most of them did not have high accuracy (approx. 50% test accuracy).
- Thought it was model’s problem first, so we tried using different parameters for each models. However, did not improve the accuracy.

Biggest Challenge

Creating New Dataset (Adding Rows)

- Created about 50 random datasets and trained individual models.
- Tried different number of rows and different ratio of fail to non-fail.
- Most of them did not have high accuracy (approx. 50% test accuracy).
- Thought it was model's problem first, so we tried using different parameters for each models. However, did not improve the accuracy.

Our dataset was the problem since it was created Randomly

- **Captured too many Noises**

→

We generated a dataset where the values stay within a similar range as the original dataset.