Report of Predictive Maintenance

Dataset :

S. Matzka, "Explainable Artificial Intelligence for Predictive Maintenance Applications," 2020 Third International Conference on Artificial Intelligence for Industries (AI4I), 2020, pp. 69-74, doi: 10.1109/AI4I49448.2020.00023.

Meaning of variables :

1. UID: unique identifier ranging from 1 to 10000
2. product ID: consisting of a letter L, M, or H for low (50% of all products), medium (30%) and high (20%) as product quality variants and a variant-specific serial number
3. type: just the product type L, M or H from column 2
4. air temperature [K]: generated using a random walk process later normalized to a standard deviation of 2 K around 300 K
5. process temperature [K]: generated using a random walk process normalized to a standard deviation of 1 K, added to the air temperature plus 10 K.
6. rotational speed [rpm]: calculated from a power of 2860 W, overlaid with a normally distributed noise
7. torque [Nm]: torque values are normally distributed around 40 Nm with a SD = 10 Nm and no negative values.
8. tool wear [min]: The quality variants H/M/L add 5/3/2 minutes of tool wear to the used tool in the process.
9. a 'machine failure' label that indicates, whether the machine has failed in this particular datapoint for any of the following failure modes are true.

The machine failure consists of five independent failure modes

1. tool wear failure (TWF): the tool will be replaced of fail at a randomly selected tool wear time between 200 - 240 mins (120 times in our dataset). At this point in time, the tool is replaced 69 times, and fails 51 times (randomly assigned).
2. heat dissipation failure (HDF): heat dissipation causes a process failure, if the difference between air- and process temperature is below 8.6 K and the tools rotational speed is below 1380 rpm. This is the case for 115 data points.
3. power failure (PWF): the product of torque and rotational speed (in rad/s) equals the power required for the process. If this power is below 3500 W or above 9000 W, the process fails, which is the case 95 times in our dataset.
4. overstrain failure (OSF): if the product of tool wear and torque exceeds 11,000 minNm for the L product variant (12,000 M, 13,000 H), the process fails due to overstrain. This is true for 98 datapoints.
5. random failures (RNF): each process has a chance of 0,1 % to fail regardless of its process parameters. This is the case for only 5 datapoints, less than could be expected for 10,000 datapoints in our dataset. If at least one of the above failure modes is true, the process fails and the 'machine failure' label is set to 1. It is therefore not transparent to the machine learning method, which of the failure modes has caused the process to fail.

1. Data Overview

1.1 Shape and structure

- **Identification of the target :** Machine failure
- **Numbers of rows and columns :** 10000, 14
- **Variables types :**
  - Discrete :
    - UID,
    - Product ID
    - Type,
    - Machine failure,
    - **TWF,*
    - **HDF,*
    - **PWF,*
    - **OSF,*
    - **RNF*

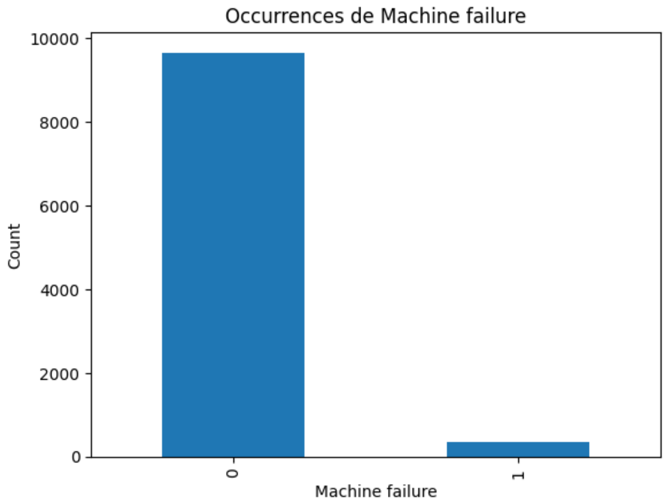
  **These variables are logically correlated with Machine failure, so because I need only one target and to avoid data leak I will drop all these columns.*

  - Continuous :
    - Air temperature [k],
    - Process temperature [K],
    - Rotational speed [rpm],
    - Torque [Nm],
    - Tool wear [min],

1.2 Quality of Data

- **Identifications of the missing values :**
  - No missing values detected in this dataset
- **Identification of duplicate values**
  - No duplicate values

1.3 Target visualization :



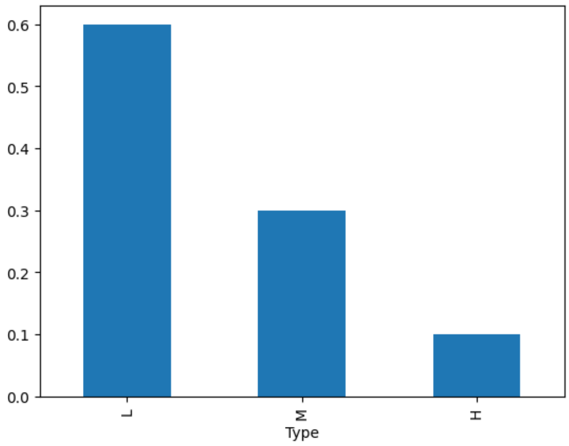
Machine failure

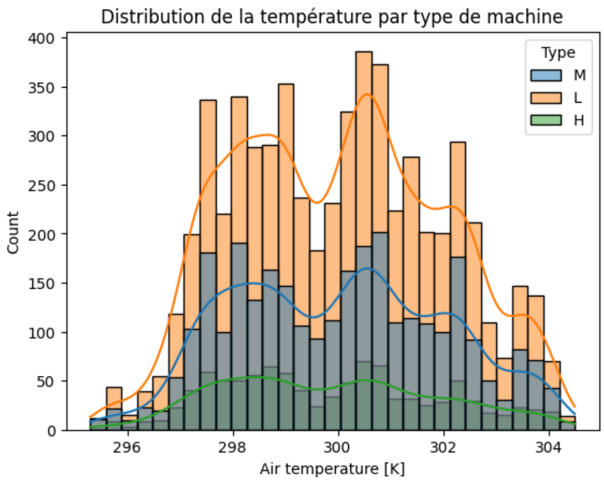
0 - 96.61%

1 - 3.39%

The dataset is unbalanced, this is an important information to choose the criteria of evaluation of the Machine Learning model.

1.4 Distribution of the machines by type :





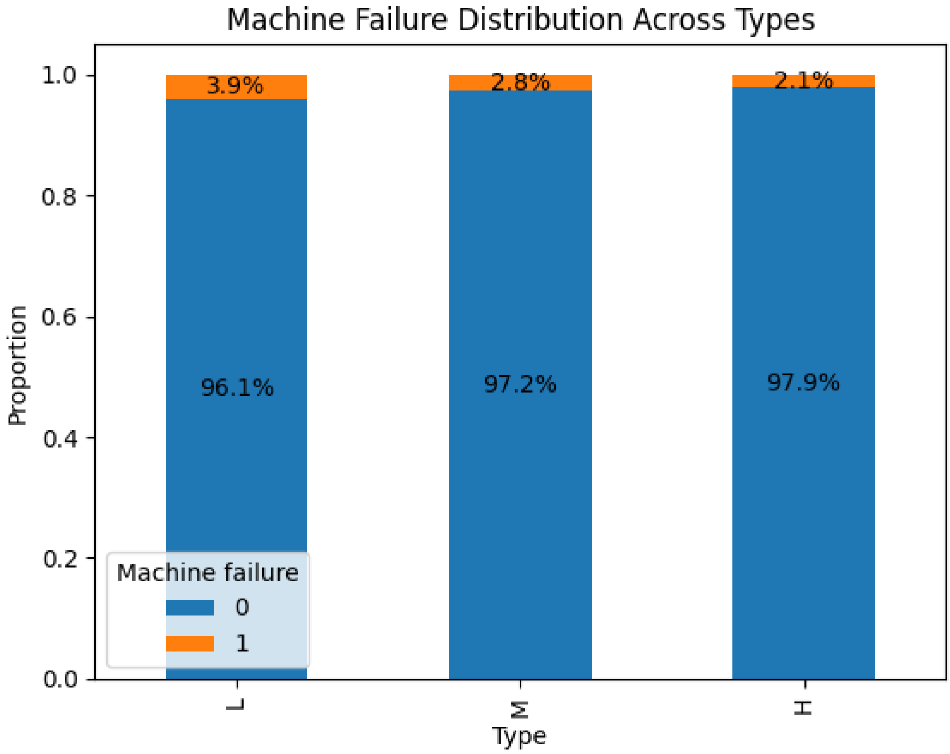
6000 (~60%) machines of TYPE L,
2997 (~30%) machines of type M,
1003 (~10%) machines of Type H.

While all machines operate within the same ranges of Torque and Speed, the **Air Temperature distribution** reveals a fundamental difference in quality.

- **Type L (Low Quality):** Exhibits a jagged, bimodal distribution with sharp peaks, indicating **poor thermal regulation**.
- **Type H (High Quality):** Shows a much flatter and smoother distribution, suggesting **superior heat dissipation capabilities**.
-

This confirms that the 'Quality' of a machine is largely defined by its thermal stability."

1.5 Relation : Type / Machine Failure :

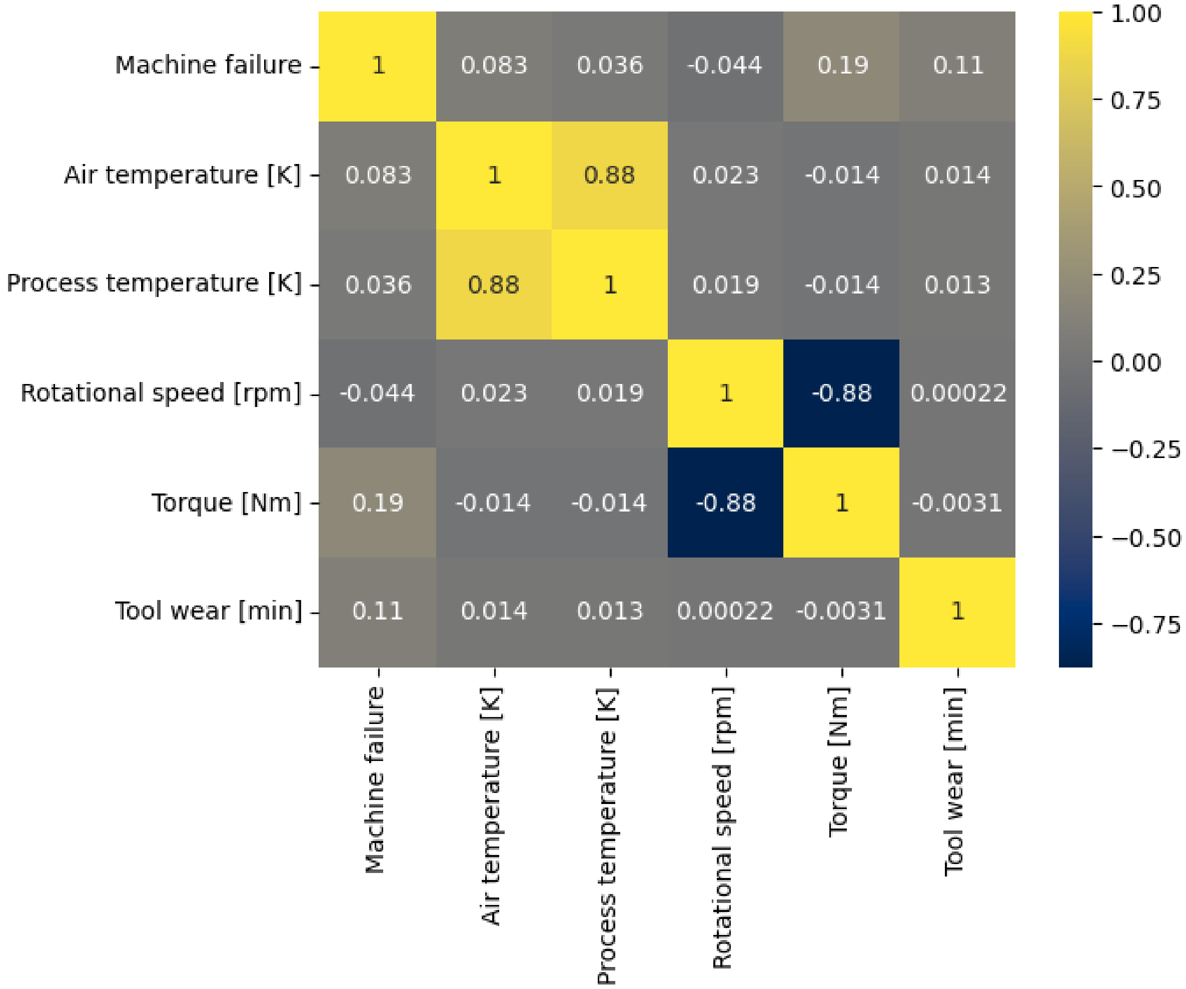


The failure rates are low for all machine types (below 4%). Type L shows a slightly higher failure rate (~3.9%) compared to types H (~2.1%) and M (~2.8%).

This difference remains moderate and would require a statistical to assess whether it is statistically significant.

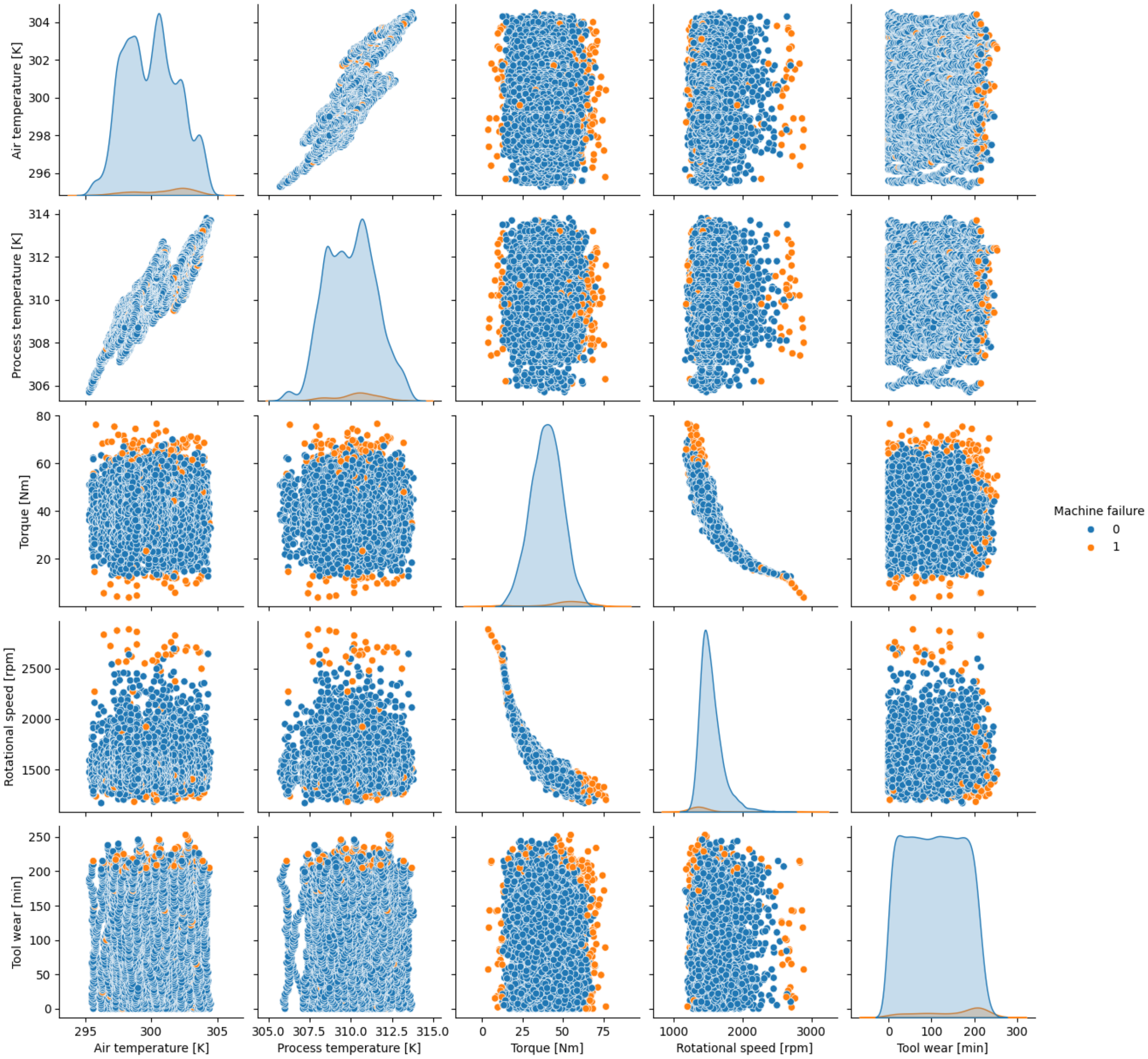
2. Global Exploration of interactions

2.1 Correlation matrix (heatmap)



- The Air temperature and Process temperature have a strong correlation (0.88)
- Torque and rotational speed are strongly correlated (-0.88)

2.2 Pairplot for Feature Relationships and correlation



This graph confirms a linear correlation between :

- Air temperature [k] and Process temperature [K]
- Torque [Nm] and Rotational speed [rpm]

In the following analysis, I will combine these variables in feature engineering.

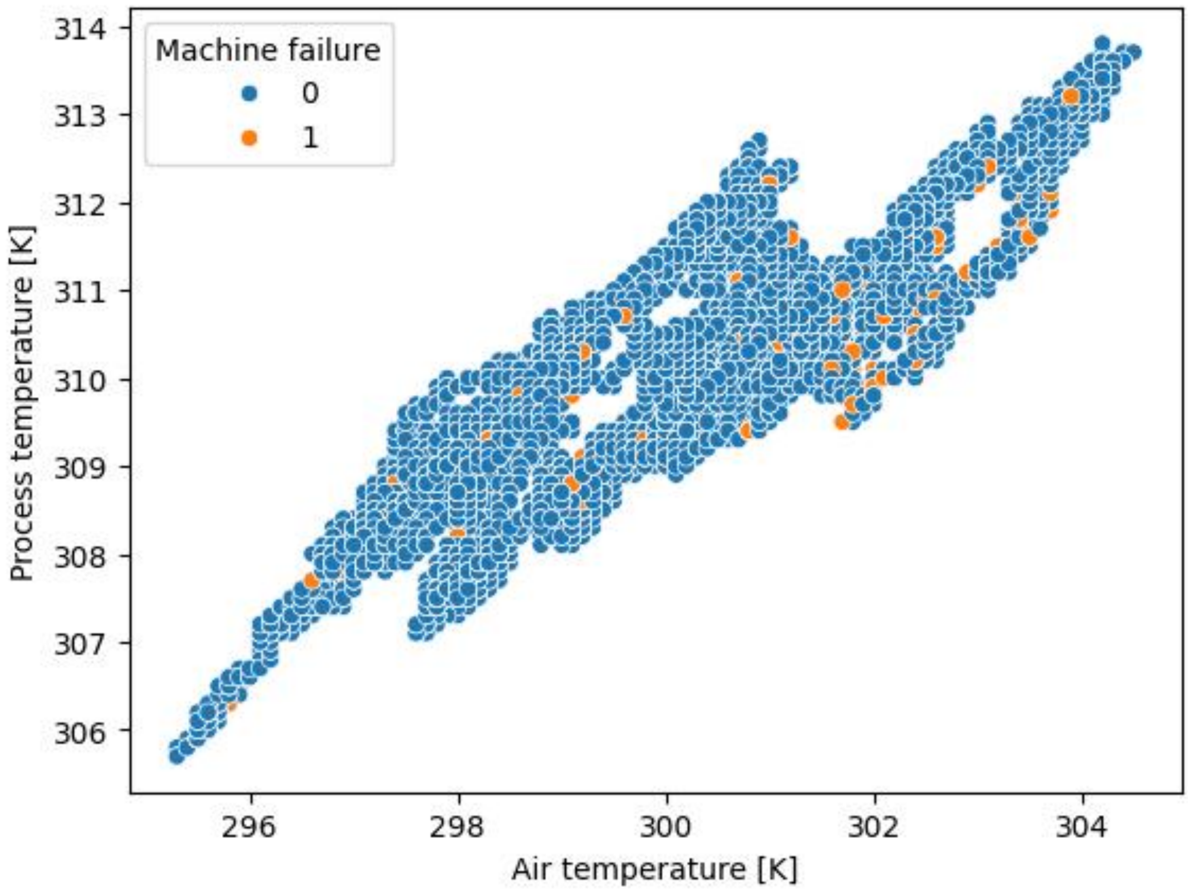
Temp_diff = Process temperature [K] - Air temperature [k]

Machine_power = Torque [Nm] * [pi() * Rotational speed [rpm] / 30]

3. In-depth Thematic Analysis

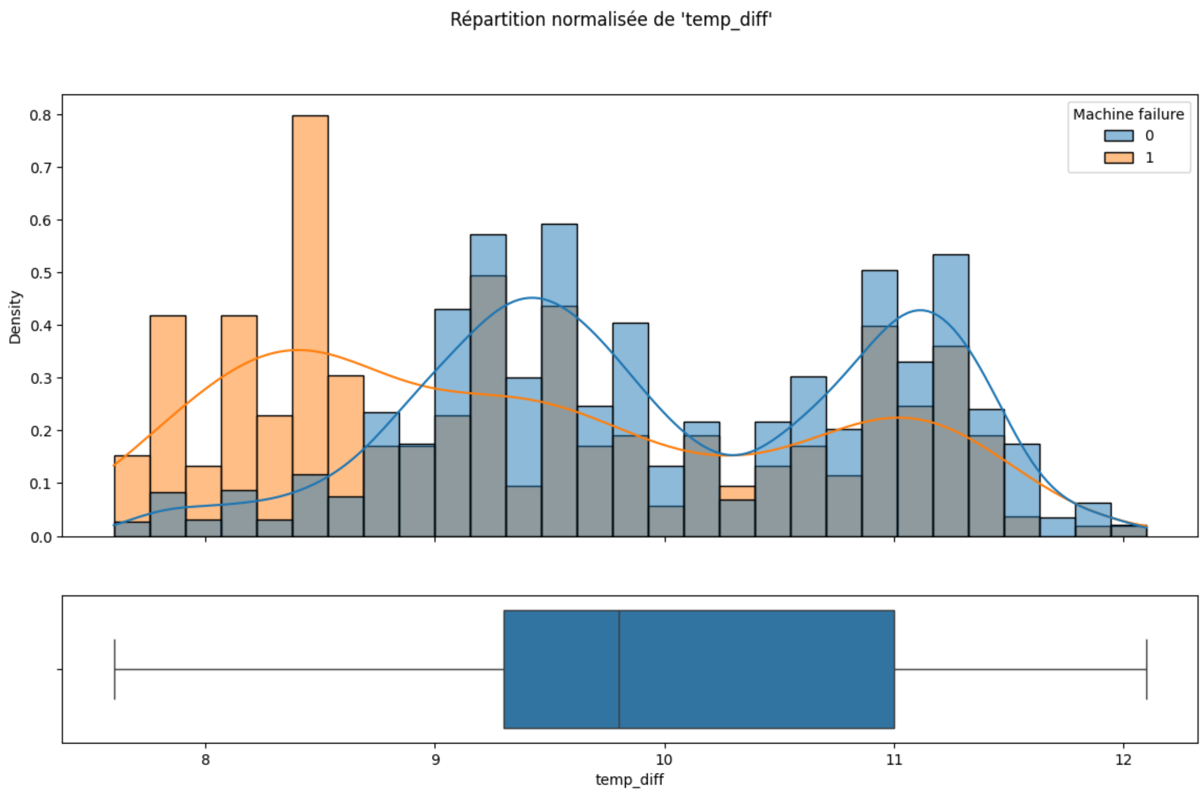
3.1 Thermal Analysis

A preliminary analysis of the raw variables Air temperature and Process temperature was conducted (details in **Appendix A**). While these variables follow stable distributions, they exhibit a significant overlap between normal operations and failures, limiting their individual predictive power.



As suggested by the correlation matrix, the scatter plot above confirms a strong linear relationship (0.88) between the two temperatures. This redundancy implies that monitoring both variables independently adds little value.

Based on the physical properties of the system (Heat Dissipation Failure - HDF), the critical risk factor is not the absolute temperature, but the machine's **inability to dissipate heat**. Therefore, we engineered a new feature: temp_diff (Process - Air).

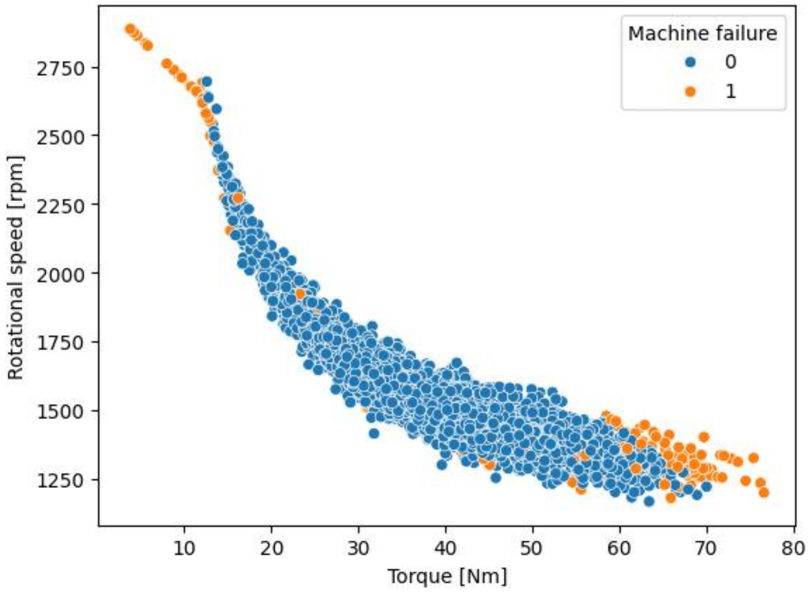


The 8.6 K threshold is critical. As the graph demonstrates, a temperature difference below this value drastically increases the risk of failure.

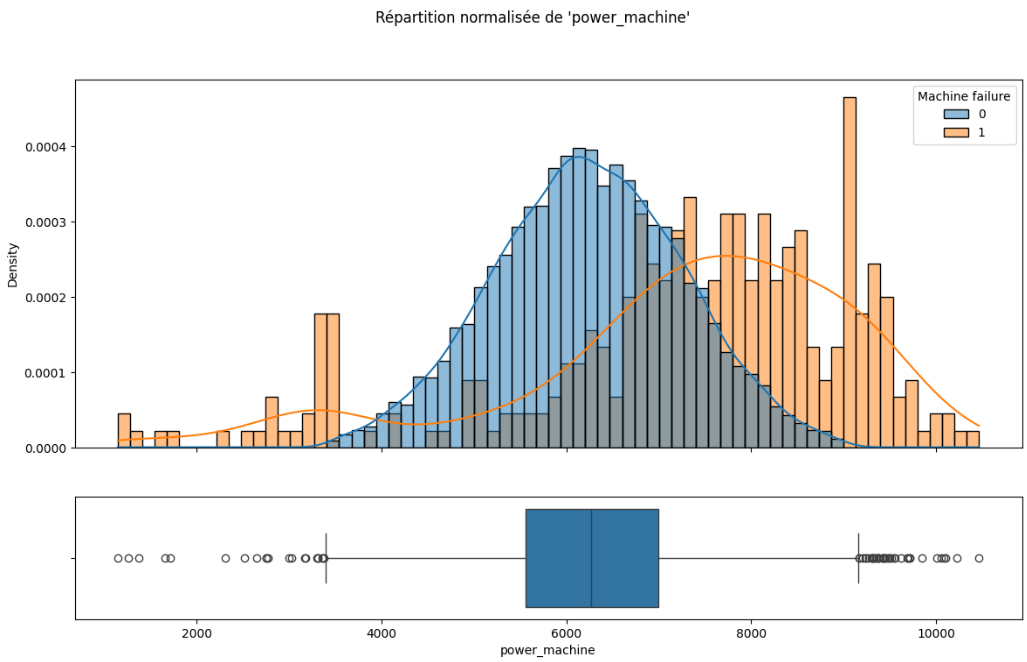
The distribution above visually confirms this theory. To rigorously validate it, we will now verify if the failures strictly fall below the 8.6 K threshold mentioned in the domain documentation.

3.2 Mecanical and power analysis

A preliminary analysis of the raw variables 'Torque [Nm]' and 'Rotational speed [rpm]' was conducted (details in **Appendix B**).



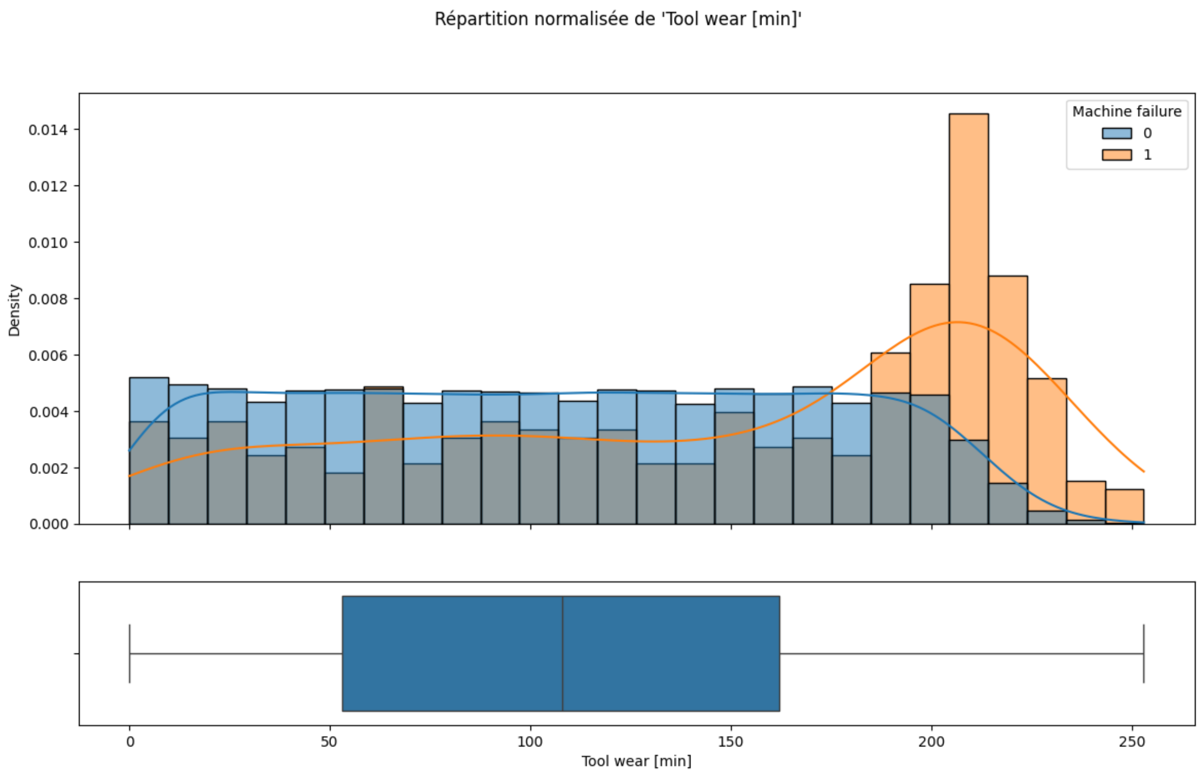
As suggested by the correlation matrix, the scatter plot above confirms a strong linear relationship (-0.88) between the Torque and RPM. This redundancy implies that monitoring both variables independently adds little value.



This graphic is normalized to make easier the analysis. Clearly it appears that the outlier are outside of the normal range of the machines and correspond to failure points

3.3 Maintenance and Wear Analysis

- A preliminary analysis of the raw variable 'Tool wear [min]' was conducted (details in **Appendix C**). Tool Wear follows a **Uniform Distribution**, the Mean (107.9 min) and Median (108.0 min) are identical, confirming the symmetry of the uniform distribution. The maximum value is of 253 minutes.



This normalized visualization highlights a clear pattern: while normal operations are evenly distributed, failures drastically spike after 200 minutes. This confirms the existence of a specific 'Risk Zone' for tool wear.

4. Synthesis and features selection

4.1 Deleted variables

We remove variables that do not contribute to prediction or introduce data leakage.

- **Identifiers (Noise):** UDI, Product ID
- **Failure Modes (Data Leakage):** TWF, HDF, PWF, OSF, RNF
    - *Justification:* These variables directly indicate the cause of the failure. Including them would prevent the model from learning the underlying patterns leading to these failures.
- **Not useful (multicollinearity)** : Air temperature [K], Process temperature [K]

4.2 Useful variables

We select features that describe the physical state and properties of the machine.

- **Categorical:** Type (Encoded)
- **Physical (Engineered):**
    - Diff_temp (Captures heat dissipation issues)
    - Power_machine (Captures power/strain anomalies)
- **Process History:** Tool wear [min]
- **Useful:** Torque [Nm], Rotational speed [Rpm]
- **Target:** Machine failure

4.3 Conclusion

This Exploratory Data Analysis allowed us to thoroughly understand the dataset and prepare it for modeling.

The dataset is now cleaned, structured, and enriched with physics-based features. We are ready to proceed to the **Pre-processing** stage.

5. Pre-processing

5.1 Divide the dataset

- **Trainset:** 80% of the dataset size
- **Testset:** 20% of the dataset size

The original dataset has 3.39% of machine failure. So to the trainset and testset have to be the most similar.

- **Trainset:** 3.3% of machine failure
- **Testset:** 3.75% of machine failure

5.2 Features Engineering

We hadded the 2 features we analysed previously.

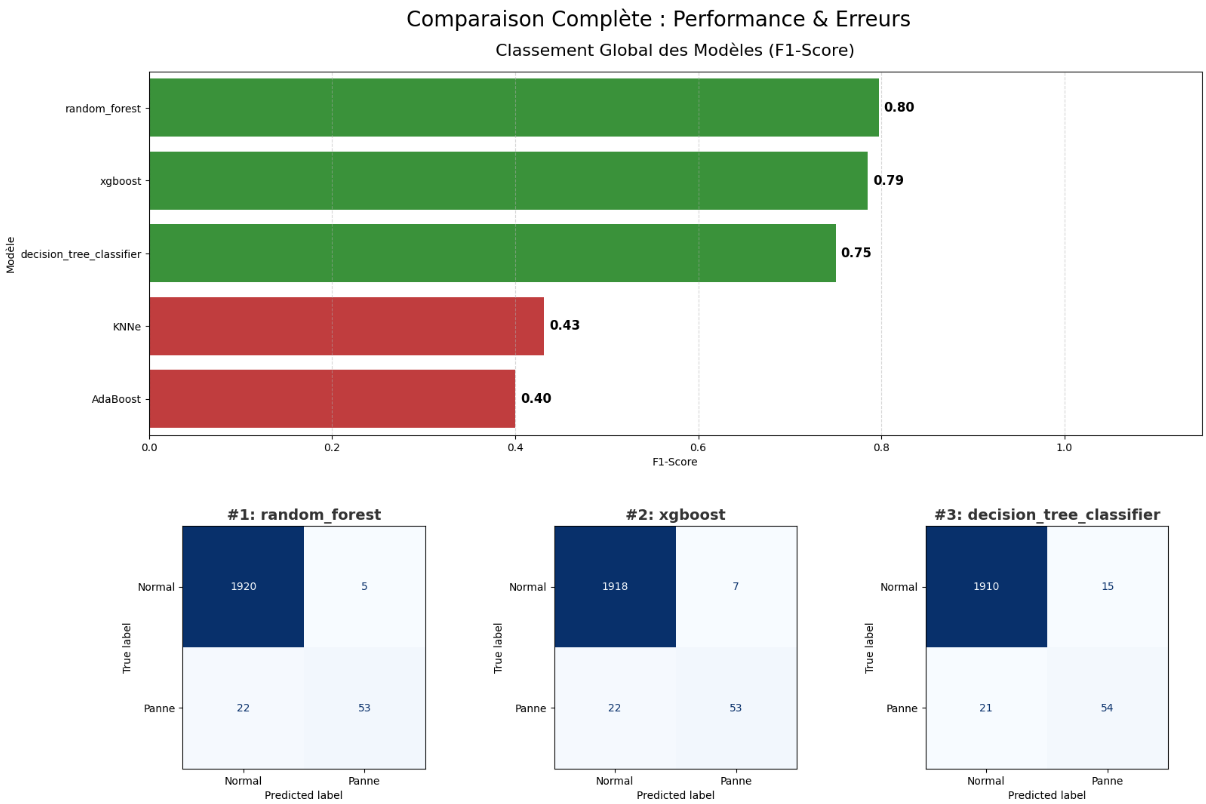
```
def features_engineering(df):
    df['power_machine'] = df['Torque [Nm]'] * (  np.pi * df['Rotational speed [rpm]'] / 30 )
    df['temp_diff'] = df['Process temperature [K]'] - df['Air temperature [K]']
    return df
```

5.3 Modeling

5.3.1 Metric Selection

I selected the **F1-Score** as the main metric because the dataset is unbalanced. That ensure the best choice between the recall and the precision.

5.3.2 Model comparison



Analysis:
We benchmarked 5 algorithms ranging from simple (KNN, Decision Tree) to complex ensemble methods (Random Forest, XGBoost, AdaBoost).

- **The Leaders:** Random Forest and XGBoost clearly outperform the others, achieving F1-scores above 0.80. They effectively handle the complex interactions between features (Power, Temp_diff).
- **The Laggards:** Linear models and simple boosters (AdaBoost) failed to capture the minority class

Decision: We retain **Random Forest** and **XGBoost** for the optimization phase.

5.4 Optimization & Final Choice

5.4.1 Hyperparameter Tuning Strategy

I performed a **RandomizedSearchCV** to optimize both models with the f1-score. Hyperparameter optimization gives the advantage to the Random Forest Classifier. It achieves a higher F1-score (0.83) and, crucially, a significantly better Recall (0.77).

	precision	recall	F1-score
Random Forest	0.89	0.77	0.83
XGBoost	0.93	0.72	0.81

The learning curves (Fig. 1 & 2) show a narrowing gap between training and validation scores, indicating that the models are generalizing well without severe overfitting.

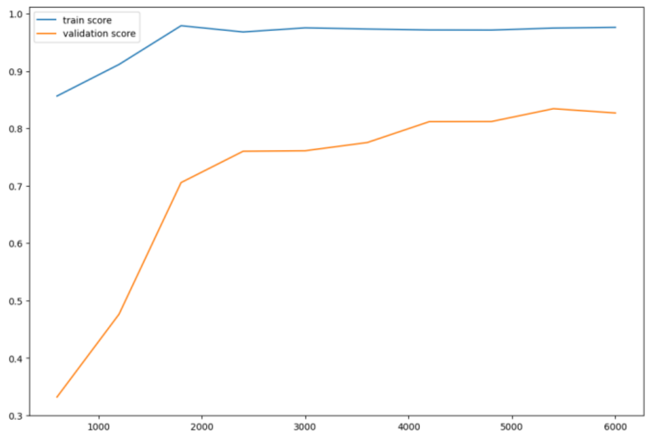


Figure 1: Random Forest learning curve

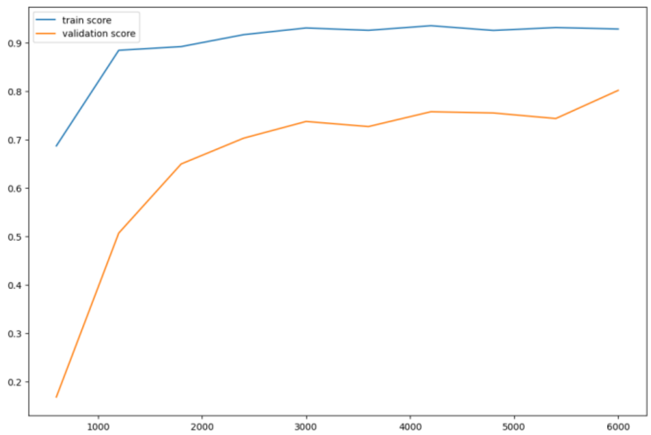


Figure 2: XGBoost learning curve

Conclusion:

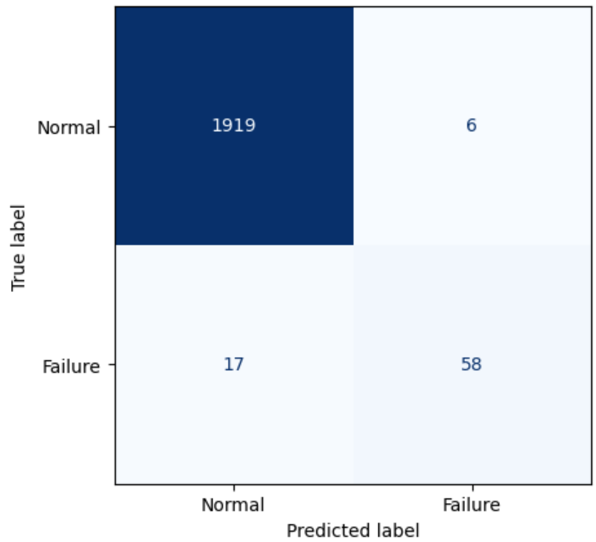
The Random Forest is the superior choice because it captures more actual failures than XGBoost (Recall 77% vs 72%), which is the priority in predictive maintenance.

Then in tuning the threshold decision to 0.531 of the Random Forest, the f1-score increase to 0.834 vs 0.828 and the precision to 91% vs 89%. The recall remains at 77%.

6. Conclusion :

6.1 Final Model Performance

Our final model achieves a **Recall of 77%** and a **Precision of 91%**, resulting in an **F1-score of 0.83**. The confusion matrix shows that out of 75 actual failures, the model successfully **detected 58** (True Positives). It **missed 17 failures** (False Negatives) and generated only **6 False Alarms** (False Positives).



6.2 Robustness & Confidence Interval

To assess the model's reliability on future unseen data, we calculated a **95% Confidence Interval** based on the Recall. Given the number of actual failures in the test set (n=75), the interval is: **[67.8% - 86.8%]**

This means that even in a conservative statistical scenario, we can expect the model to detect at least **68%** of the machine failures, ensuring a minimum safety net for operations.

6.3 Conclusion

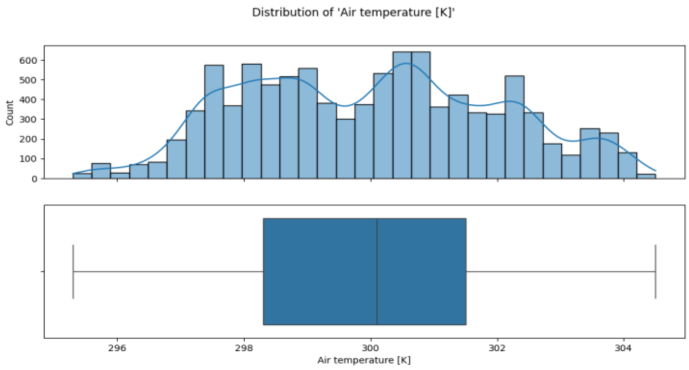
We successfully created a simple yet robust Random Forest model using a limited set of variables. This excellent performance is largely attributed to the physics-based Feature Engineering.

We now have all the necessary information with high precision, solid recall, and statistical guarantees to confidently recommend putting this model into production for these machines.

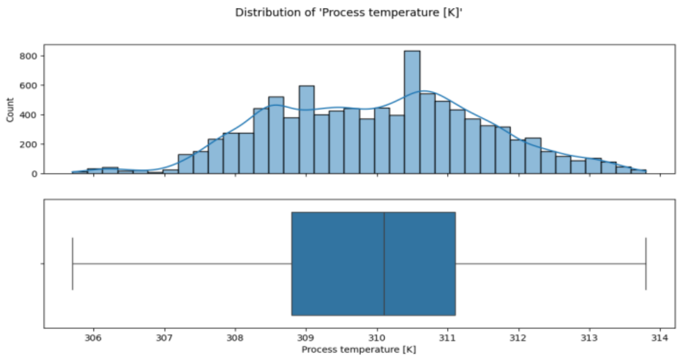
Appendix A :

Detailed Analysis of Raw Temperature Variables

1. Global Distribution & Statistics (Univariate)

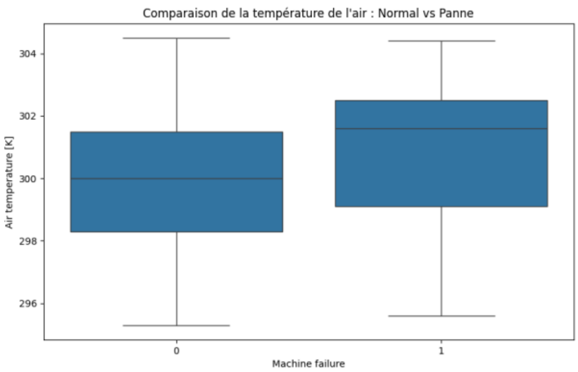


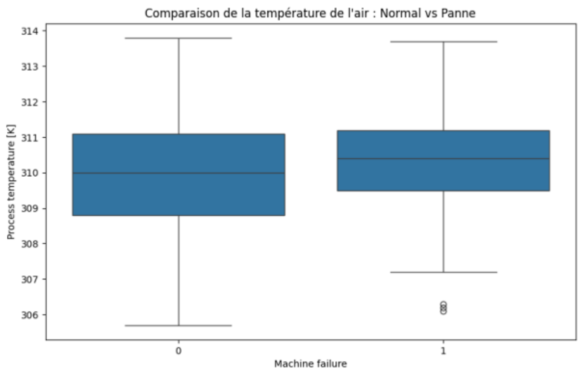
count	10000.000000
mean	300.004930
std	2.000259
min	295.300000
25%	298.300000
50%	300.100000
75%	301.500000
max	304.500000



count	10000.000000
mean	310.005560
std	1.483734
min	305.700000
25%	308.800000
50%	310.100000
75%	311.100000
max	313.800000

2. Discrimination Capability vs. Machine Failure (Feature-Target)





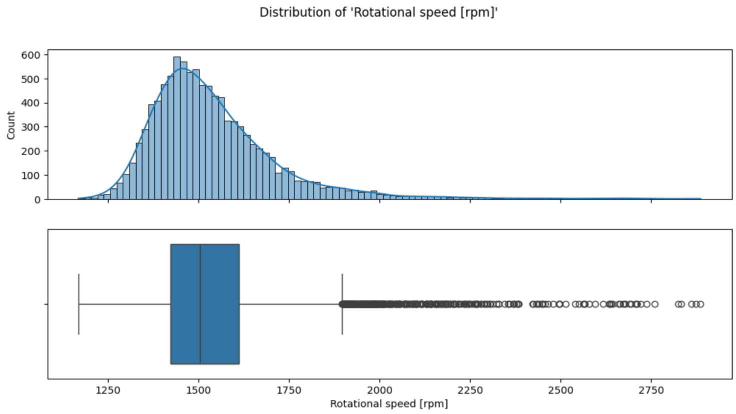
Observation: The univariate analysis (top) confirms that both Air and Process temperatures are stable and follow a multimodal distribution. However, the Feature-Target analysis (bottom) reveals a **significant overlap** between the interquartile ranges of normal operations and machine failures.

Conclusion: This confirms that raw temperature values lack sufficient discriminative power to predict failures on their own. This observation **justifies the Feature Engineering strategy** (creation of the temp_diff variable) presented in the main body of the report.

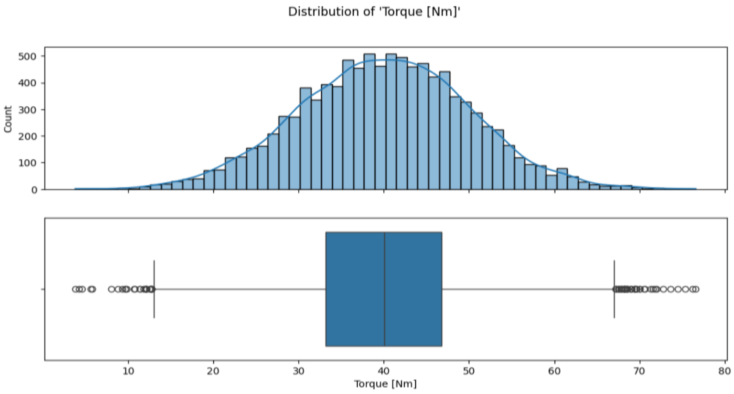
Appendix B :

Detailed Analysis of Raw Torque and RPM Variables

1. Global Distribution & Statistics (Univariate)



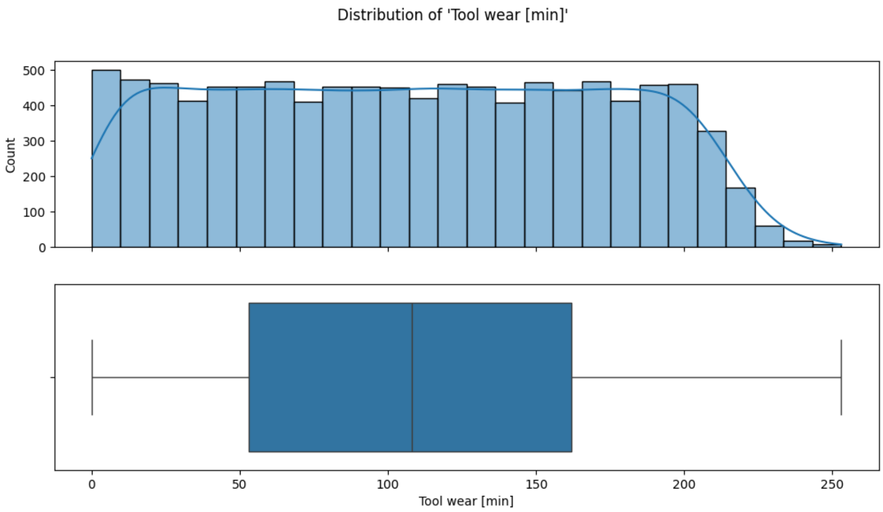
count	10000.000000
mean	1538.776100
std	179.284096
min	1168.000000
25%	1423.000000
50%	1503.000000
75%	1612.000000
max	2886.000000



count	10000.000000
mean	39.986910
std	9.968934
min	3.800000
25%	33.200000
50%	40.100000
75%	46.800000
max	76.600000

Appendix C :

Detailed Analysis of Raw Tool wear Variable



count	10000.000000
mean	107.951000
std	63.654147
min	0.000000
25%	53.000000
50%	108.000000
75%	162.000000
max	253.000000

- **Distribution:** Tool Wear follows a **Uniform Distribution**. The frequency is roughly constant from 0 to 200 minutes. The Mean (107.9 min) and Median (108.0 min) are identical, confirming the symmetry of the uniform distribution. The maximum value is of 253 miutes.