Machine Failure Prediction

Using Machine Learning Algorithms

Our Agenda

Machine failure prediction refers to the task of using machine learning and data analysis techniques to predict when a machine or equipment is likely to fail or experience a breakdown. By analyzing historical data and identifying patterns and indicators, machine failure prediction models can provide early warnings or alerts, enabling proactive maintenance and minimizing downtime.

About Dataset

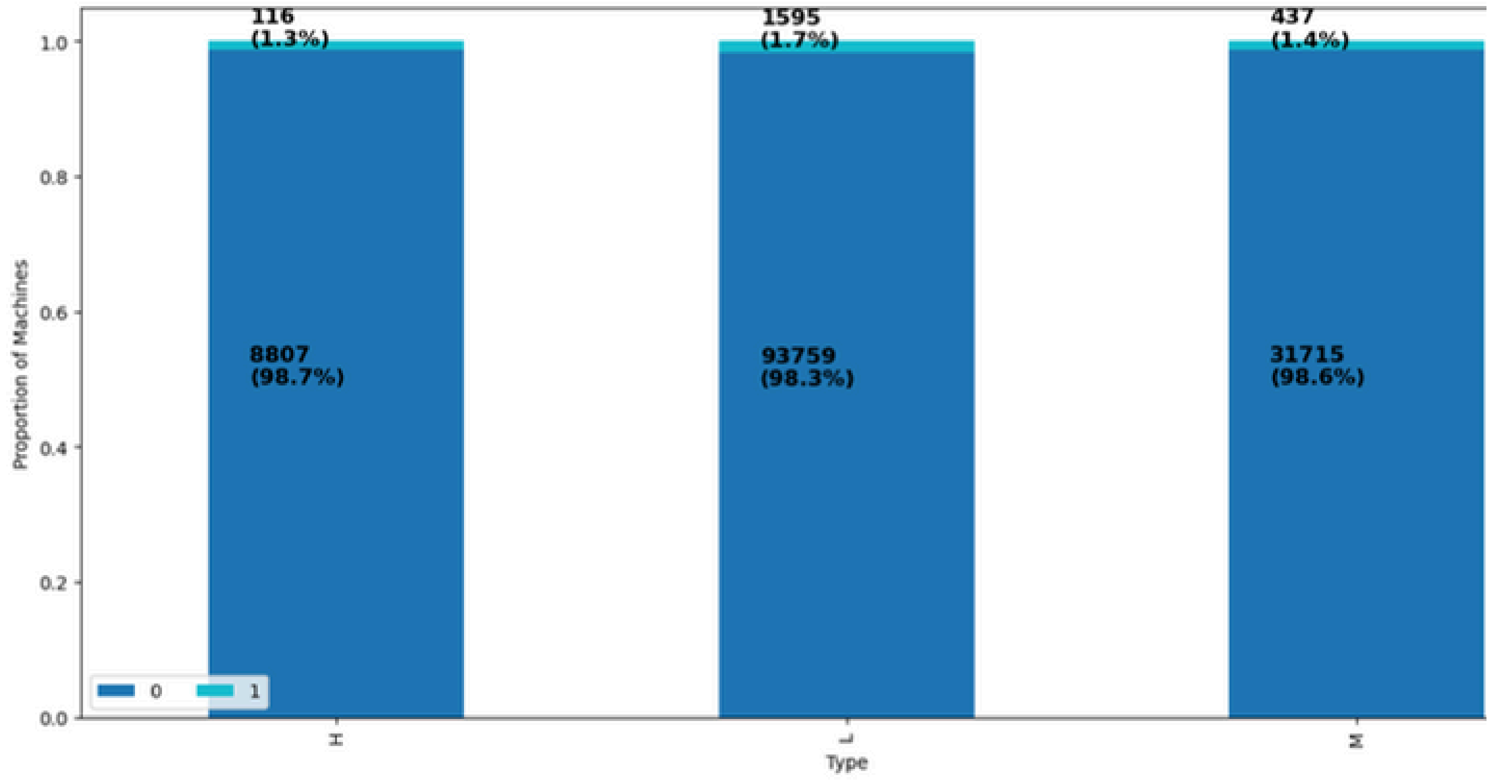
Relevant data is collected from the machines or equipment, such as sensor readings, operational parameters, maintenance records, and historical failure data. This data serves as the basis for training and building the predictive models.
In this project, our task is to predict machine failure based on 13 features:

• **Product ID** is an identification code for each product. This ID is a combination of letters and numbers.
• **Type** is the classification of the product or device into three categories: Low, Medium, and High. The proportions of these types in the dataset are as follows: Low 50%, Medium 30%, and High 20%.
• **Air temperature [K]** represents the temperature of the air, measured in Kelvin units.
• **Process temperature [K]** indicates the temperature during the production process, measured in Kelvin units.
• **Rotational speed [rpm]** refers to the number of revolutions per minute. It is calculated based on a power of 2860 W and is subject to normally distributed noise.
• **Torque [Nm]** measures the force that causes an object to rotate, expressed in Newton-meters (Nm). The torque values are normally distributed around 40 Nm and do not include negative values.
• **Tool wear [min]** represents the time it takes for production tools to erode or become damaged due to regular wear and tear caused by cutting operations.
• **TWF, HDF, PWF, OSF, RNF** are binary classified dummies representing the type of failure that occured.

Exploring the categorical Features

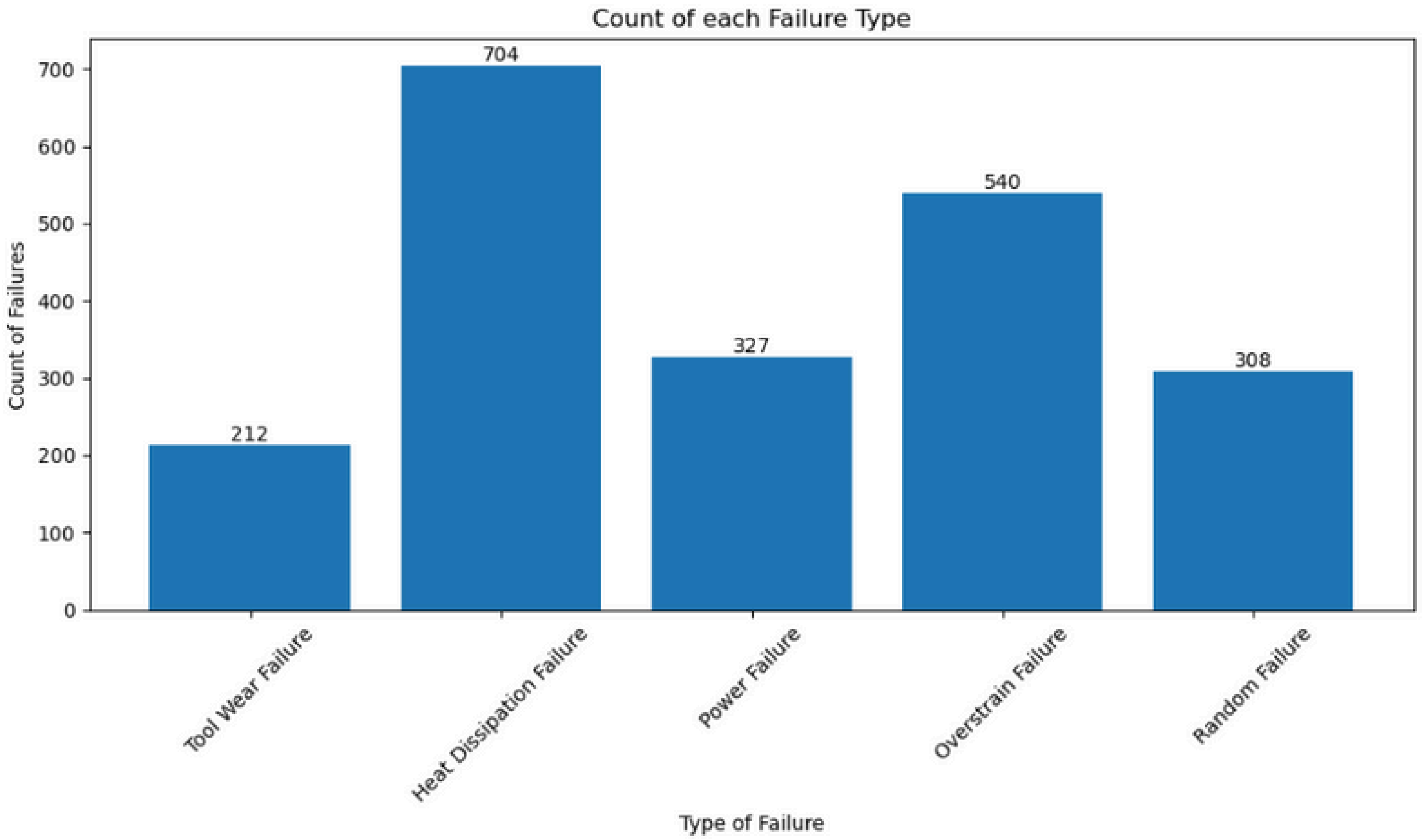
Type Feature

- Type is the classification of the Machine into three categories: Low, Medium, and High.
- The proportions of these types in the dataset are as follows: Low 70%, Medium 23.5%, and High 6.5%.

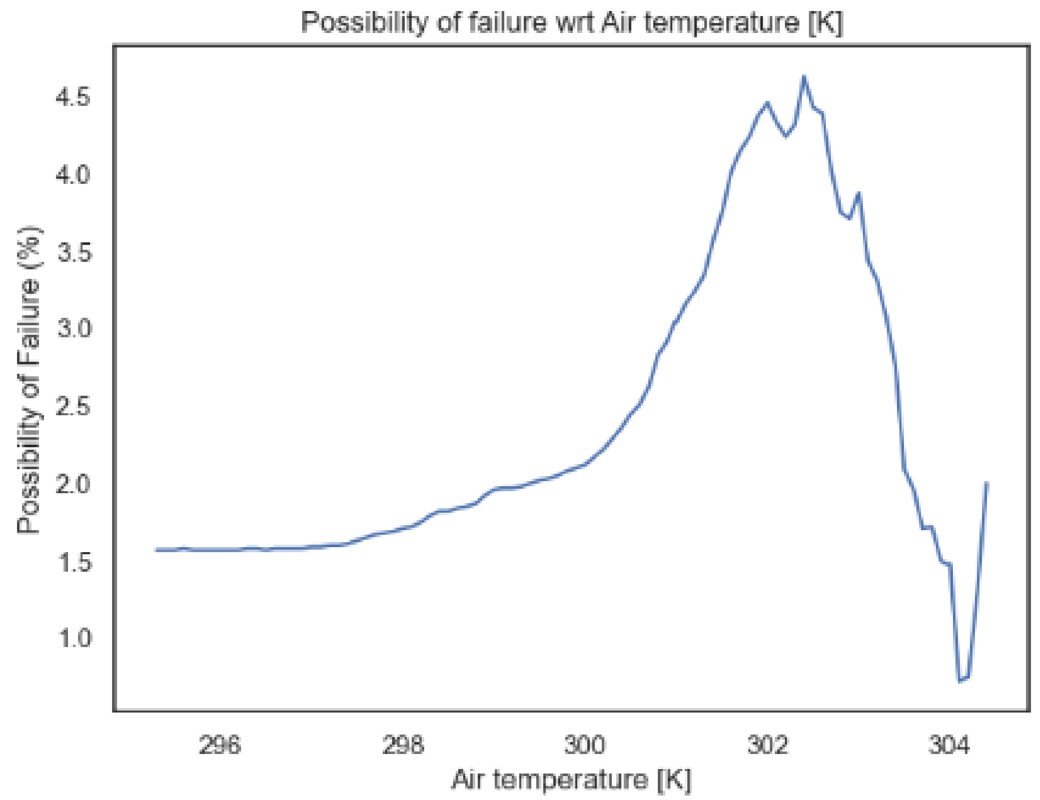


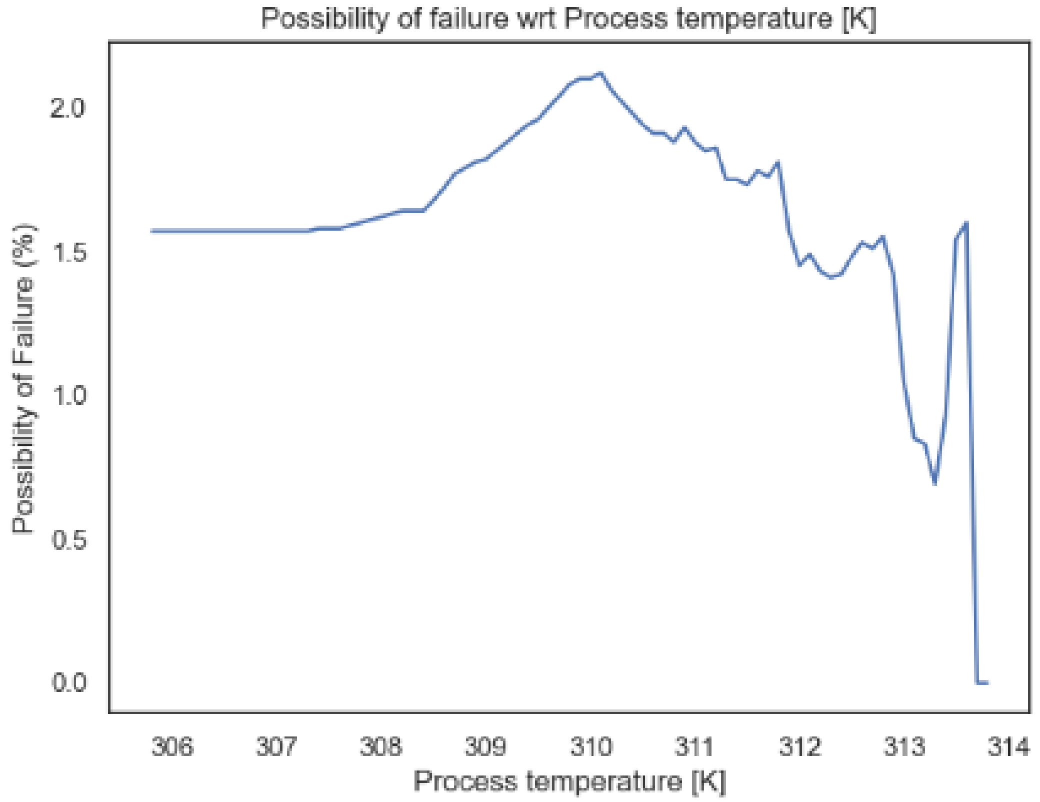
Failure Types

- **TWF (Tool Wear Failure)**: Indicates industrial tool failure, leading to the need for equipment change and defective products.
- **HDF (Heat Dissipation Failure)**: Indicates failure in heat dissipation during the production process.
- **PWF (Power Failure)**: Indicates that the supplied power was not suitable for the production process, resulting in a failure.
- **OSF (Overstain Failure)**: Indicates failure due to product overstains, which may occur as a result of high load and tension during production.
- **RNF (Random Failure)**: Indicates that a random error causes the failure.

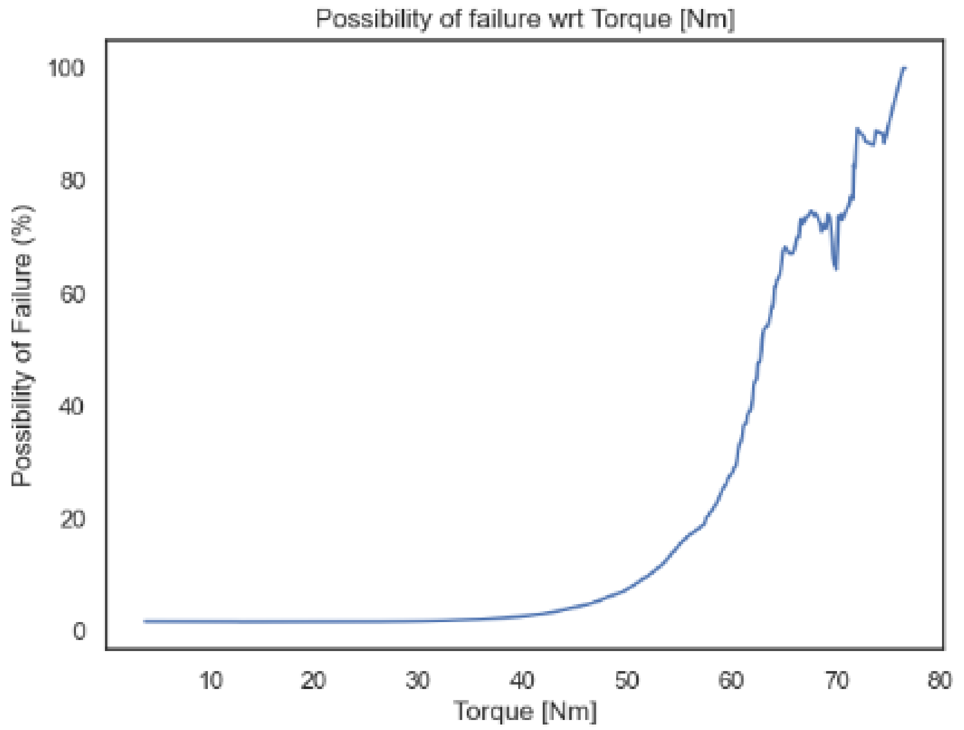


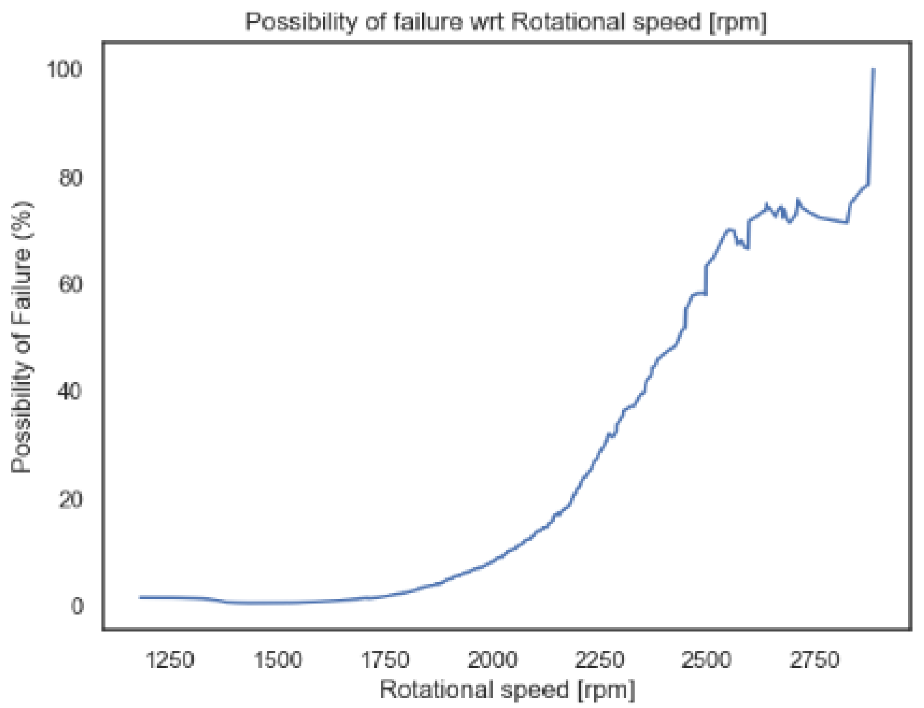
Exploring the continuous Features

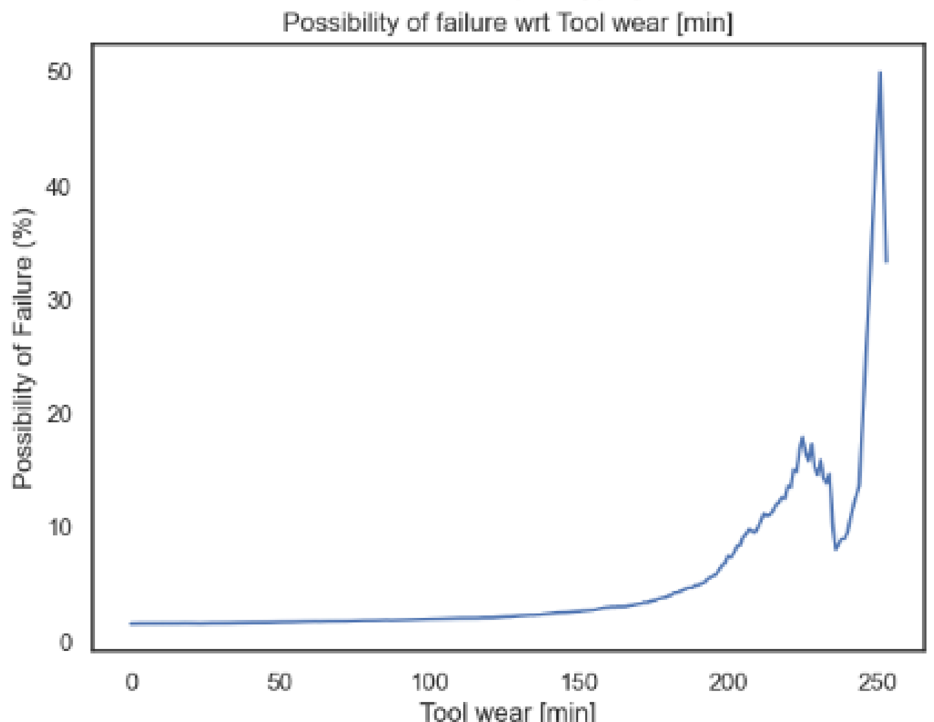




- **The air temperature:** varies between approximately 295.3 K (Kelvin) and 304.4 K.
- **The process temperature:** ranges from about 305.8 K to 313.8 K. Both air and process temperatures seem to have relatively low standard deviations, indicating that they are not highly variable.
- **Rotational Speed:** The rotational speed varies from 1,181 rpm to 2,886 rpm.
- **Torque**: Torque values range from 3.8 Nm (Newton-meters) to 76.6 Nm.
- **Tool Wear:** Tool wear values range from 0 minutes to 253 minutes, with an average of approximately 104.41 minutes.



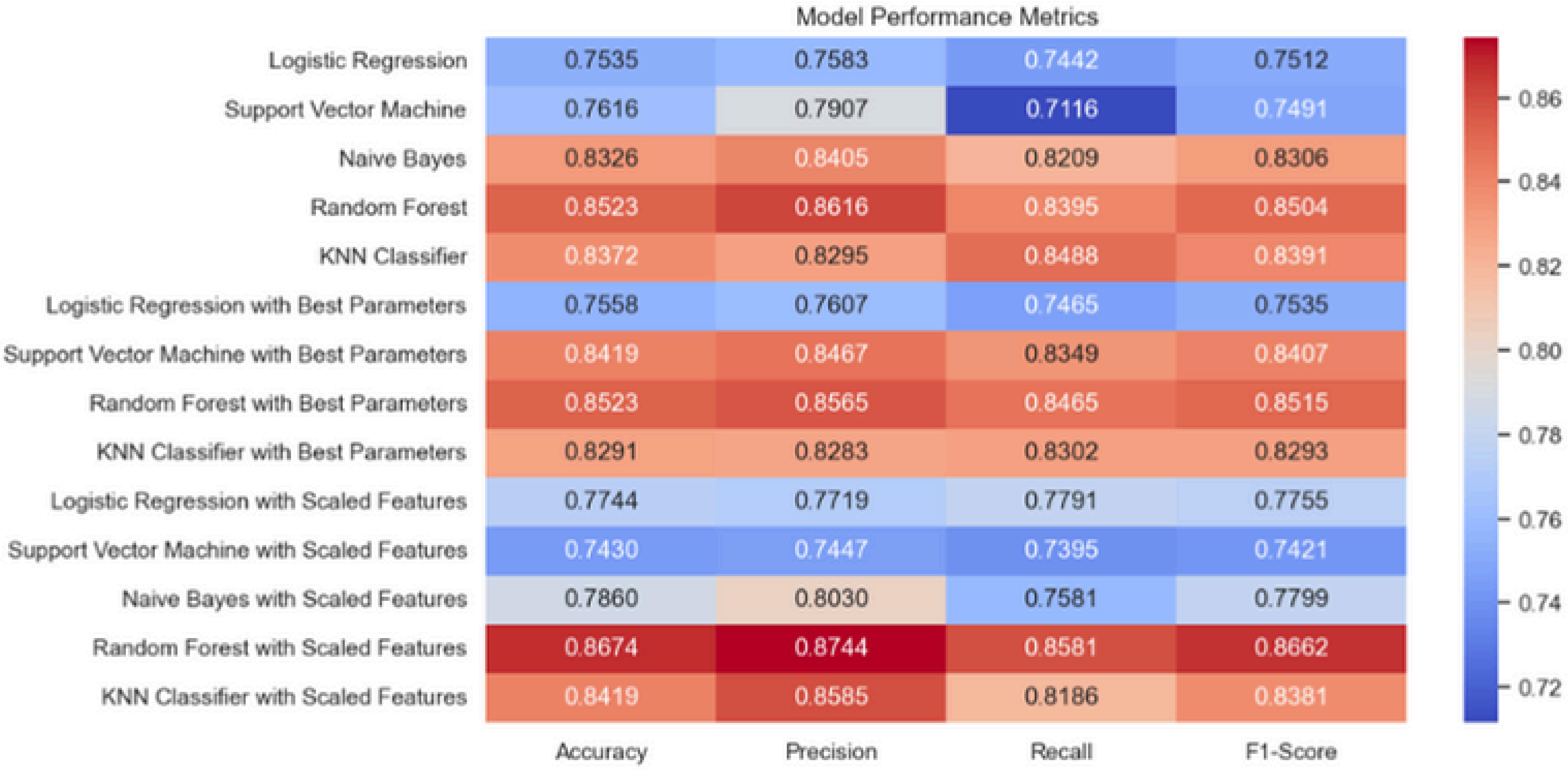




Different Machine Learning Algorithms used in this Prediction:

- **Balancing the Dataset:** With 1.57% of positive Machine failures, it is important to balance the data to avoid biased models and misleading evaluation metrics. The base line accuracy is 98.42%.
- **Train Test Split:** With 20% share of test data and 80% share of train data we are measuring the performance of the models.
- **Logistic Regression.**
- **Support Vector Machine.**
- **Naïve Bayes.**
- **Random Forest.**
- **KNN Classification**
- **GridSearchCV and RandomizedSearchCV:** Including 5-fold cross validation, these algorithms are used for hyperparameter tuning.
- **MinMaxScaler:** To scale the features to a standard scale.
- **Confusion Matrix:** To see the predictions and ROC curve.
- **Accuracy, Precision, Recall and f1 scores:** To measure and compare the performance metrics of the different models.

Performance Metrics



	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.7535	0.7583	0.7442	0.7512
Support Vector Machine	0.7616	0.7907	0.7116	0.7491
Naive Bayes	0.8326	0.8405	0.8209	0.8306
Random Forest	0.8523	0.8616	0.8395	0.8504
KNN Classifier	0.8372	0.8295	0.8488	0.8391
Logistic Regression with Best Parameters	0.7558	0.7607	0.7465	0.7535
Support Vector Machine with Best Parameters	0.8419	0.8467	0.8349	0.8407
Random Forest with Best Parameters	0.8523	0.8565	0.8465	0.8515
KNN Classifier with Best Parameters	0.8291	0.8283	0.8302	0.8293
Logistic Regression with Scaled Features	0.7744	0.7719	0.7791	0.7755
Support Vector Machine with Scaled Features	0.7430	0.7447	0.7395	0.7421
Naive Bayes with Scaled Features	0.7860	0.8030	0.7581	0.7799
Random Forest with Scaled Features	0.8674	0.8744	0.8581	0.8662
KNN Classifier with Scaled Features	0.8419	0.8585	0.8186	0.8381

Conclusion:

- The Random Forest model with scaled features consistently performs well across all metrics and has the highest accuracy, precision, recall, and F1-Score among all the models. This indicates that it's a strong candidate for a machine failure prediction problem.
- Additionally, Random Forest models are known for their ability to handle complex data, capture non-linear relationships, and provide good generalization. They are also robust against overfitting and can handle both numerical and categorical features, making them suitable for various types of machine failure prediction tasks.

What’s Next?

- More complex models like Gradient Boosting and Neural Networks can be used improve the performance of the model.
- Imbalanced Data Algorithms like SMOTE (Synthetic Minority Over-sampling Technique) can be used to see the performance of the model.

Thank you!