Predictive Maintenance for Industrial Equipment using Machine Learning

Tarek Salameh, AlWaleed AlAlami

1. Introduction

Predictive maintenance is a cornerstone necessity in the domain of industrial operations for performance optimization and a reduction of maximum uptime (Kudelina et al., 2023). The project plans to venture in this critical domain by harnessing a spectrum of machine learning models ranging from Random Forest, SVM, Logistic Regression, Deep Learning, Linear Regression, Naive Bayes to MLP benchmarked with RNN, for revolutionizing the predictive maintenance strategies. The complex approach to the problem involves different algorithms and thus should result in a radical rise of the accuracy and efficiency of maintenance practices in case of comparison with traditional methods.

The research work was aimed at comparing those different algorithms in order to find the best algorithm for the prediction of equipment failure. Our results therefore indicate remarkable success with the Random Forest model showing the highest accuracy performance towards a notable rate of 0.9967. This accomplishment underscores the proclivity of our approach in revolutionizing practices of predictive maintenance within industry settings, exposing prospects for unprecedented accuracy and dependability utilizing the robustness of diverse machine learning techniques.

Task	AlWaleed AlAlami	Tarek Salameh
Random Forest	x	✓
SVM	x	✓
Logistic Regression	x	✓
MLP	✓	x
Deep Learning	x	✓
Naïve Bayes	✓	x
Linear Regression	✓	x
MLP with RNN	✓	x
Documentation	✓	✓

Table (1): Division of Labor

2. Problem Definition and Algorithm

2.1 Task Definition

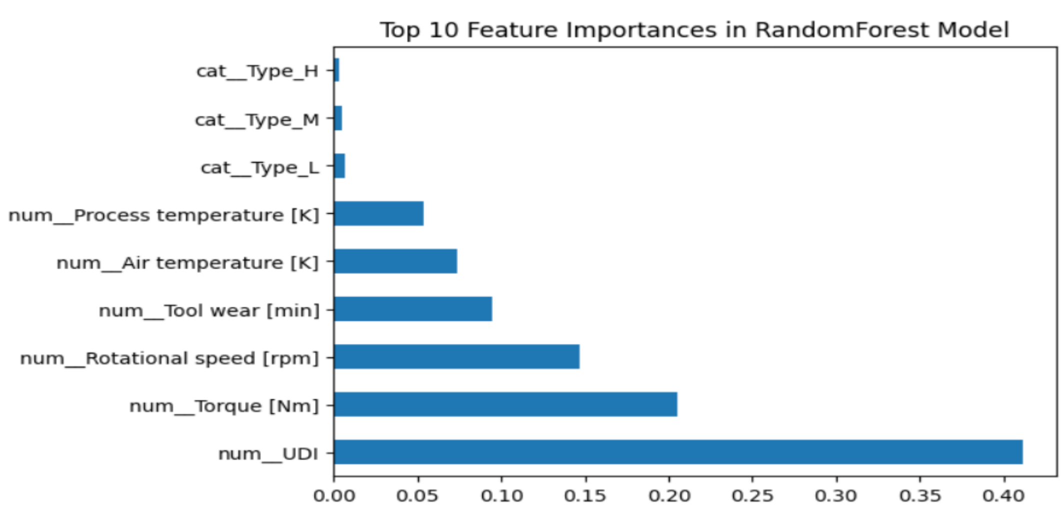
Our project tends to estimate the failure of industrial equipment - whereby the failure prediction helps optimize maintenance schedules and avert unforeseen plant breakdowns (Nofal et al., 2023). The difficulty lies in the multi-faceted aspect of predicting these failures accurately that requires a delicate sophisticated approach. On the contrary, our methodology involves high-dimensional operational dataset with parameters like readings from sensors including temperature and pressure. The aim is the processing of this input data and the production of a predictive score or classification which might indicate that equipment will fail. The significance of this endeavor lies under the premise that it can significantly lead to a reduction in downtime and maintenance costs within industrial environments, which increases not only operational efficiency but also safety.

2.2 Algorithm Definition

So in order to cover a wide-ranging spectra of predictive maintenance scenario, we very carefully select the range of algorithms because each of it owns distinct strength and capability. These are Robust Random Forests to deal with different type of data, Precise SVM effective in higher dimensions spaces, and Logistic Regression as well as MLP. Each of the chosen algorithms was to deliver in terms of an algorithm that adopts high reliability and dependable predictive maintenance framework catering for different aspects of predictivity data. Further, our approach was marked by openness to multiple scenarios for developing intelligent models further to allow us to scrutiny the effectiveness of each model diligently so as to be able to ensure a best fit solution for purpose of predictive maintenance. (Ehrig et al., 2020)

Pseudocode for Random Forest:

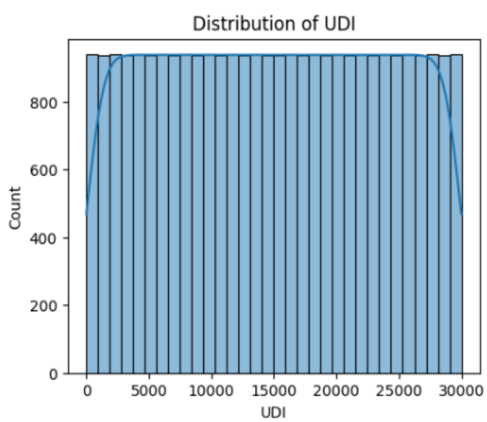
```
Pseudocode for Random Forest Algorithm:

1. Start
2. Define the number of trees in the Random Forest: num_trees
3. For each tree in the Random Forest:
   a. Randomly select a subset of training data.
   b. Randomly select a subset of features.
   c. Build a decision tree using the selected data and features:
      - For each node of the tree:
         i. Choose the best split based on the subset of features.
         ii. Split the dataset into two subsets based on the best split
         iii. Repeat the splitting process recursively until a stopping
   d. Save the decision tree.
4. For making predictions with the Random Forest:
   a. For each decision tree in the forest:
      i. Make a prediction on the input data.
   b. Combine the predictions from all trees (e.g., majority voting for
5. Return the final prediction as the output of the Random Forest.
6. End
```

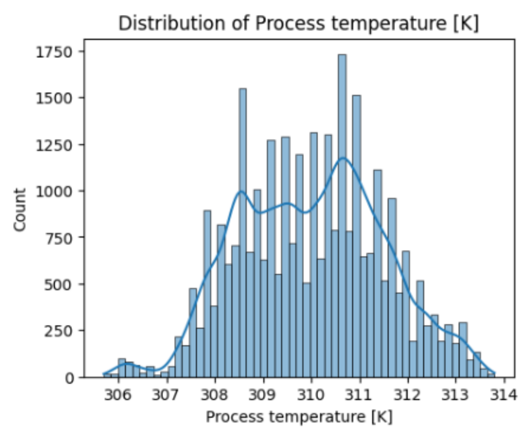


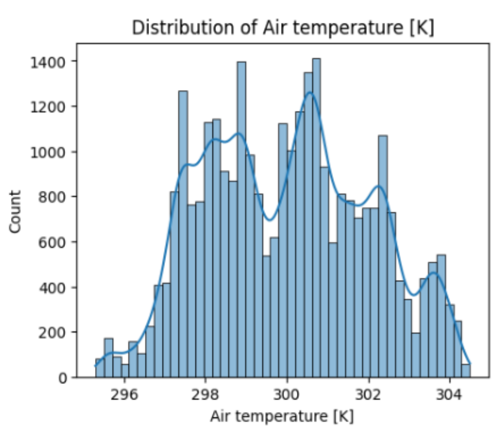
Figure(1): Top 10 Feature Importance's (RF)

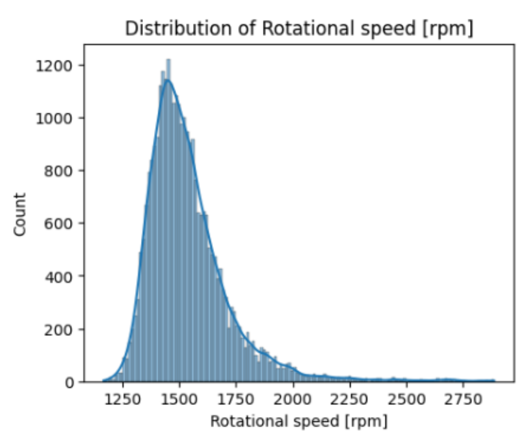
2.3 Dataset

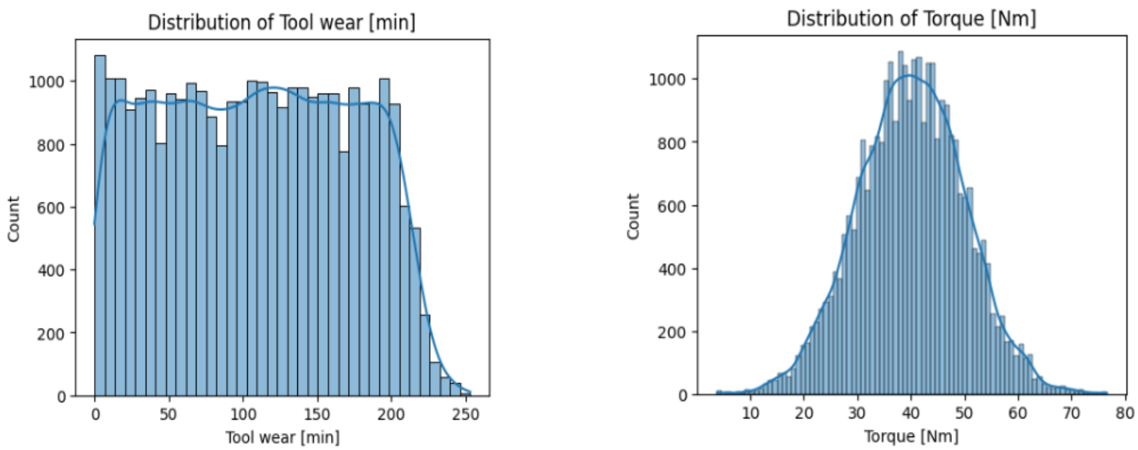
Our predictive maintenance framework leverages on a comprehensive dataset of 30,000 operational records of the industrial equipment with parameters such as temperature, rotational speed, torque and tool wear. The next dataset is a combination of 20,000 synthetic records produced to augment the diversity and representativeness of this dataset and 10,000 historical records sourced from Kaggle at www.kaggle.com providing a range of real-world operational scenarios. The whole hybrid dataset, fully labeled and publicly available has been structured for the purpose of easing the analysis process, meaning that no additional labeling effort from our team or any use of specialized hardware for processing was involved. For creating a more focused seven key attributes that would effectively streamline our data set to be used in our machine learning models, we remove data points like 'product ID', 'failure type', and 'target' which were not significantly relevant to restrict the list of 10 attributes. This approach easily allows a broad user base to access and operate the dataset consequently forming a very strong basis of our study since it gets to augment visibility as well as depth and breadth.











Figure(2): Exploratory Data Analysis (EDA): Visualizing distributions of numeric columns

3. Experimental Evaluation

3.1 Methodology

Our approach consisted of the assessment of several machine learning models as Random Forest, Deep Learning techniques, Support Vector Machine (SVM), Logistic Regression, Linear Regression, Naïve Bayes and a Multi-Layer Perceptron (MLP) benchmarked versus a Recurrent Neural Network (RNN). We focused on Accuracy, Precision, Recall, F1-score, ROC-AUC Score, Mean Squared Error (MSE), and R^2 Score as the main metrics to evaluate how effectively they predicted the failures of equipment accurately.

This involved splitting the dataset into a training and testing set in order to ensure a realistic and robust evaluation of the predictive capability for each model. The choice and preprocessing of data were carefully aligned with the specific needs and characteristics of each model making room for an accurate assessment of their predictive ability. Default parameters was used for each model.

3.2 Results

Referring to the performance of performing the evaluated models, it should be said that different systems demonstrated substantially variant results, where the model Random Forests turned out to be a sound method with accuracy equal to 0.9967.

Notably, the accuracy of the Deep Learning and MLP models was averagely high at 0.9928 and 0.9907, respectively.

In addition to the Decision Trees, models such as SVM, Logistic Regression, Linear Regression or Naive Bayes can be no less well applied. All these findings attract attention to the fact that complex machine learning techniques in particular, Random Forest, Deep Learning, MLP models and others can quite efficiently be used for predictive maintenance purposes.

In Figure (3), we present the ROC curves for all evaluated models, showcasing their respective abilities to classify true positives against the backdrop of false positives, with the area under the curve (AUC) as an indicator of overall performance.

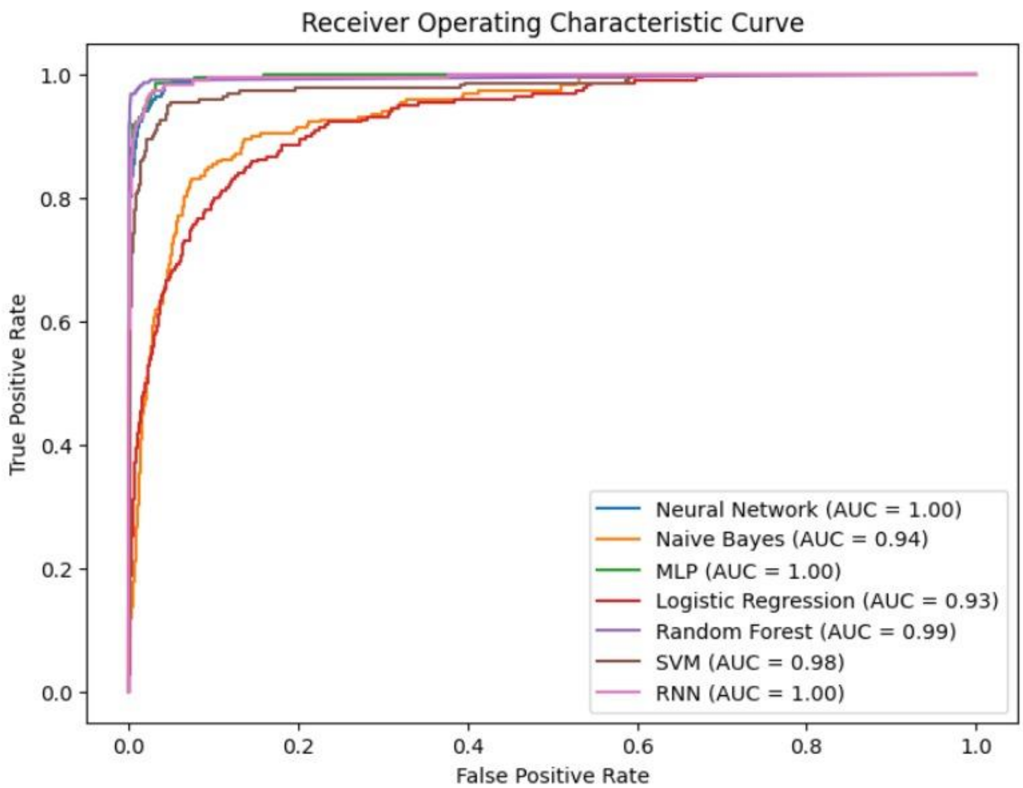


Figure (3): ROC Curve

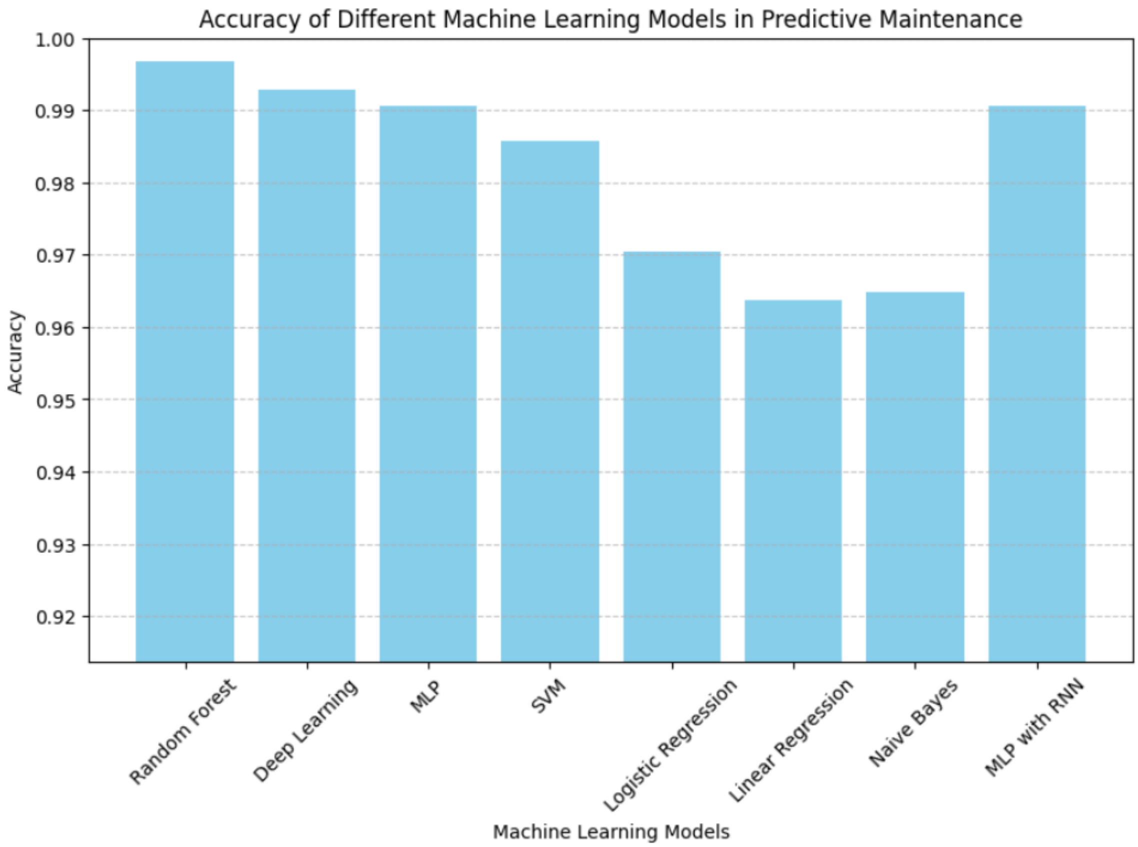


Figure (4): Machine Learning Testing Data Accuracies

	Random Forest	SVM	Logistic Regression	MLP	MLP with RNN	MLP Deep	Naïve Bayes	Linear Regression
Precision Class '0'	100.00%	99.00%	97.00%	99.00%	99.00%	100.00%	98.00%	###
Precision Class '1'	99.00%	91.00%	75.00%	91.00%	91.00%	92.00%	53.00%	###
Recall Class '0'	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	99.00%	###
Recall Class '1'	92.00%	67.00%	28.00%	82.00%	82.00%	88.00%	33.00%	###
F1-score Class '0'	100.00%	99.00%	98.00%	100.00%	100.00%	100.00%	98.00%	###
F1-score Class '1'	95.00%	77.00%	41.00%	86.00%	86.00%	90.00%	41.00%	###
Training Accuracy	100.00%	98.70%	97.10%	99.32%	99.32%	99.21%	96.72%	96.67%
Testing Accuracy	99.67%	98.57%	97.05%	99.07%	99.18%	99.28%	96.48%	96.37%

Table (2): Machine Learning Accuracies

3.3 Discussion

The results of this study thus confirm our working hypothesis on the effectiveness of machine learning in predictive maintenance by way of the very high levels of accuracy and ROC-AUC scores rendered by the developed models, particularly Random Forest and Deep Learning.

The variance in model performances hammers home the suitability of different machine learning approaches to this domain and not least emphasize primarily on the significance of nuanced feature selection and engineering.

From the results above, better performance from Random Forest and those that are more advanced such as Deep Learning models underlines their robustness of being able to handle complex predictive maintenance data as well as applicability in industrial settings. Insights such as these, coupled with the intention to integrate more complex models in the future, emphasize that machine learning has the potential to disrupt the current situation of predictive maintenance strategies. (Ehrig et al., 2020).

4. Related Work

The difference in our approach is from the traditional ways of accomplishing predictive maintenance since it leans too much on single-model approaches, like Kudelina et al.'s (2023) reliance on signal processing and Nofal et al.'s (2023) on ensemble learning. Instead, we followed this multi-modelling approach combining many other machine learning models like Random Forest, SVM, Deep Learning among others. This multifold algorithmic strategy allowed conducting the comprehensive analysis of equipment failure data making our work standing out on a number of its features:

1. **Enhanced Accuracy:** With the value of comparing multiple models, we identified the most effective solution and succeeded to improve the predictive accuracy as compared to the traditional single-model methods.

2. **Robust Analysis:** The very range of models used in the analysis enabled a more robust and wide-ranging analysis to be done. Each model was able to throw light from its own perspective or viewpoint on various dimensions of the data, and together they helped build up a fuller picture of equipment failure.

3. **Methodological Innovation:** Presented methodology hence is a significant methodological innovation in the domain of predictive maintenance. Not only it integrated various different methodologies of machine learning to bridge up the gaps within the existing methodologies, rather our presented methodology presented versatile and relatively more accurate solution both when compared with others.

Generally, our project highlights the addition to be realized through adoption of multi-model predictive maintenance. This is traceable in realizing that the approach does not only magnify the pinpointing of accuracy of prediction but also realization of an all round understanding of forces propelling equipment failure hence generating new knowledge and strategies for the same.

5. Code

We share the code for our project as well as the merged dataset over GITHUB(LINK) with clear instructions in order to reproduce our results. Such sharing allows better transparency and surely would help in carrying out further research and development into this zone.

6. Conclusion

This project demonstrated how the power of machine learning can transform the offerings in predictive maintenance, creating a new standard in the space. Our ingenious use of a number of algorithms, with Random Forest being the shining star, has not only improved the accuracy in prediction but also opened new horizons towards cost saving and risk mitigation in industrial settings. Which gives prominence to Random Forest as the most promising model, and therefore drives our multi-model approach, significant in machine learning evolution of maintenance strategies.

As such, these findings are significant contributions to the field thus demonstrating how data analysis smart results in improved maintenance strategy. By presenting functioning applications and advantages of variant machine learning models, our project becomes a role model for upcoming study and appliance using the data-driven tailor-made solution for industrial maintenance. This work, therefore, only further extends this current understanding of predictive maintenance more into the field of innovations and improvements in this critical area.

Bibliography

- Kudelina, K., et al. (2023). Signal Processing and Machine Learning Techniques for Predictive Maintenance of Rotor Bars in Induction Machines.
- Nofal, Z. A., et al. (2023). Ensemble Learning for Predictive Maintenance on Wafer Stick Machine Using IoT Sensor Data.
- Ehrig, L., et al. (2020). Customizable Asymmetric Loss Functions for Machine Learning-based Predictive Maintenance.