SDS 322E Project: Applying Data Science to Predict Machine Failure

Team: Allysa Dallmann, Michael Montez, Shreya Nallaparaju, Jonathan Saldeen, Sebastian Utama, Kayla White, and Emily Zhou

Introduction

The problem we are trying to solve in this project is how to best predict machine failure. In manufacturing, maintenance is crucial to keep the equipment running optimally. Being unprepared to handle machine faults or failure can be extremely costly to the business in manufacturing efficiency, and throughout the entire production process. The ability to predict machine failure will better allow stakeholders to maintain manufacturing equipment, prevent failure, and reduce revenue loss. This project uses machine learning methods Random Forest, Support Vector Machine (SVM), and k-Nearest Neighbors (kNN) to better predict machine failure in order to inform preventative maintenance.

Data

Description

The UCI Machine Learning Repository provides several datasets for use in machine learning tasks. We downloaded the *AI4I 2020 Predictive Maintenance Dataset* csv file from the repository for our use. The file presents a synthetic dataset generated to reflect real predictive maintenance data in industry.

The dataset contained the following variables: machine-part type (high, medium, low quality), air temperature, process temperature, rotational speed, torque, and tool wear. Machine failure is reached through five different failure modes: tool wear failure (TWF), heat dissipation failure (HDF, when the difference between air and process temperature is less than 8.6 K and rotational speed is less than 1380 rpm), power failure (PWF, when the product of torque and rotational speed is less than 3500 W or greater than 9000 W), overstrain failure (OSF, the product of tool wear and torque exceeds 11,000 minNm for a low quality product), or random failure (RNF). If the conditions of any of these failure modes are reached then there will be machine failure.

Cleaning

In order to provide us with a clear overview of our data and better optimize our models, it was important to clean our data. Fortunately, our raw data had no missing values nor any duplicate entries. We renamed each feature for easier access and encoded categorical variables, like “Type”, for better use by our machine learning algorithms. A normalized version of the data was used for the machine learning models to ensure that no one variable was driving the predictions. Our final dataset contained 10,000 unique entries.

Exploratory Data Analysis

Hypothesis 1: HDF is the most common type of failure

For the first part of the exploratory analysis, we decided to visualize the total probability of each type of failure across the data. Figure 1, shows a comparison of all of the different failure types and the probability of that failure occurring. We suspected that out of all of the five failure types heat dissipation failure (HDF) would be the most common. Figure 1 shows that HDF is the most common type of failure to occur with it occurring approximately 1.15% of the time. The least common failure to occur is random failure (RNF) as this occurred 0.19% of the time. All of the other failures occurred less than 1.2% of the time, which implies that the occurrence of a failure is minimal. This is significant because for maintenance and efficiency of mechanical operation it is essential to understand the occurrence of failures.

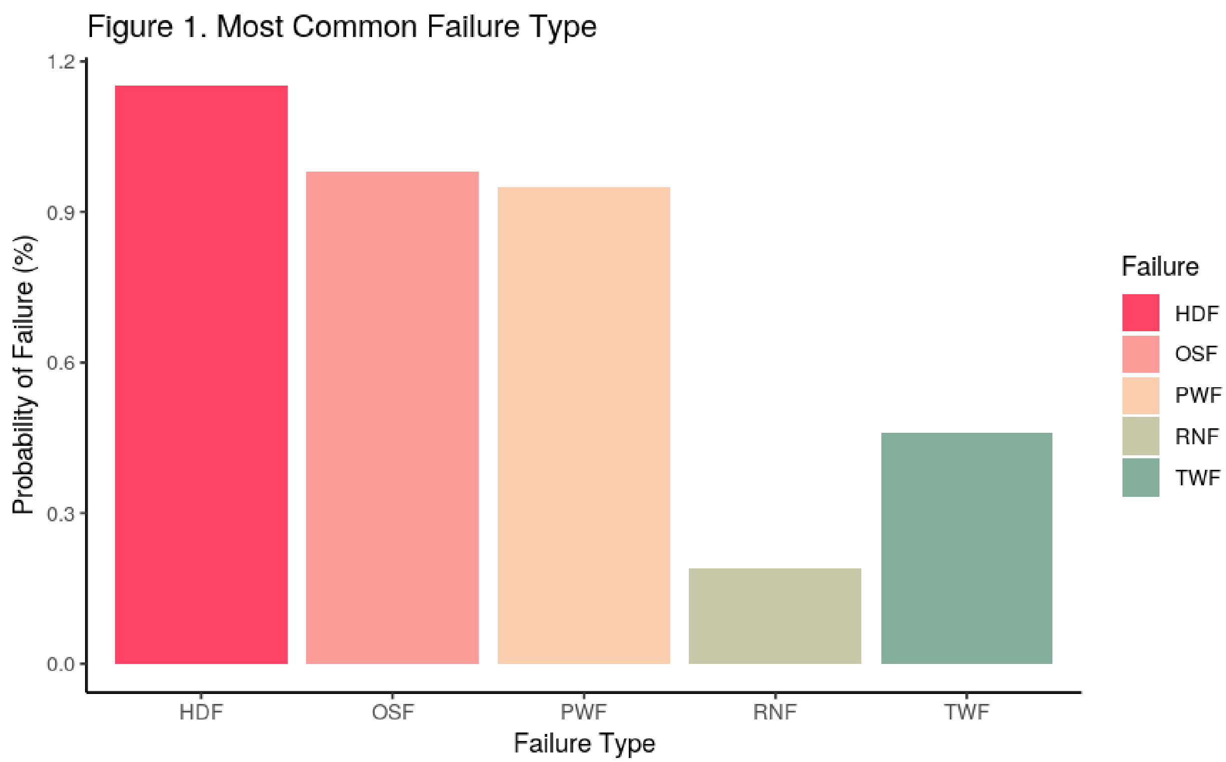


Figure 1. Probability of failure (%) relative to failure type. Heat dissipation failure (pink), overstrain failure (salmon), power failure (peach), random failure (lime), tool wear failure (green).

Hypothesis 2: More failures occur in low quality products

The second part of the exploratory analysis focused on a comparison between the three different quality types and the association to the number of failures that occur. We hypothesize that more failures will occur with lower quality products. The analysis was carried out by analyzing the count of each quality type to failure (Figure 2) along with the proportion of each quality type associated with the count of failures (Figures 3, 4, 5). For the occurrence of one failure, low quality products occurred the most with a probability of 70%, medium is 23%, and high 7%. When focusing on two failures, low quality products remained to have the highest failure rate as the proportion of two failures with a low quality product was 74%, and medium 26%. There were no failures associated with high quality products at this failure count. However, when looking at three failures occurring, there was only one failure reported and it was with a high quality product. After further data analysis (not shown) in the plot, the failures are TWF, PWF, and OSF which are not the most common HDF. Overall, when one or two failures occur, there is a higher proportion associated with a lower quality product, and the least likely with a higher quality product, with exception for when three failures occur.

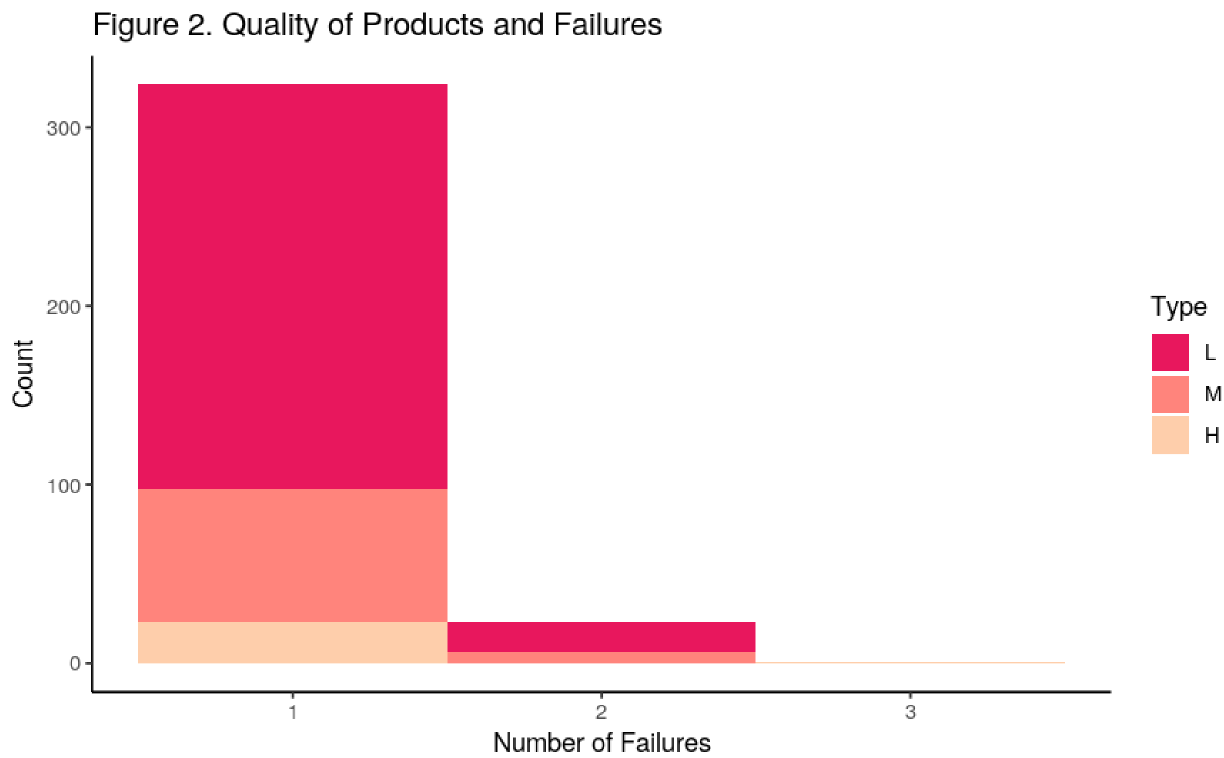


Figure 2. Count of product quality types Low (pink), Medium (salmon), and High (peach) relative to the number of failures that occur.

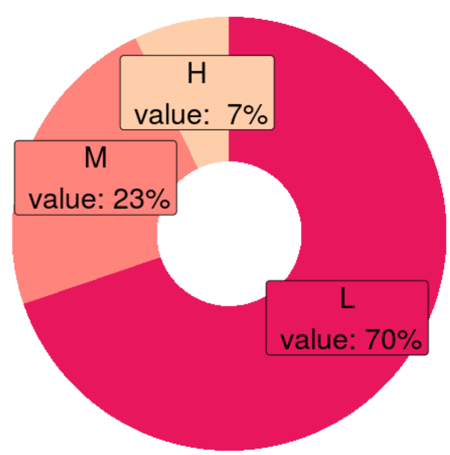


Figure 3. Distribution of product quality types when there is only 1 failure present. Low (pink), Medium (salmon), and High (peach) quality product types.

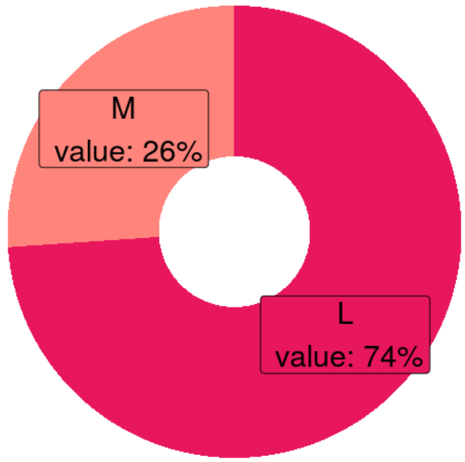


Figure 4. Distribution of product quality types when there are 2 failures present. Low (pink) and Medium (salmon) product quality types. There are no High quality products (peach) present in this distribution.

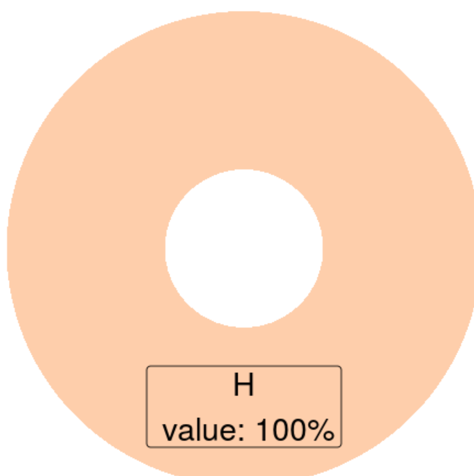


Figure 5. Distribution of product quality types when there are 3 failures present. High (peach) quality types. There are no Low (pink) or Medium (salmon) quality products present in this distribution. There is a single scenario where the high quality product has 3 failures.

Hypothesis 3: When HDF failure is present there will be lower rotational speed across all product quality types

For the third part of the exploratory analysis, we visualized the density of product quality against the rotation speed for those products to see if there was a change in rotational speed with HDF failure present. We hypothesized that the rotational speed would decrease when the HDF failure was present, since low rotational speed is a cause of HDF failure. Figure 6 shows that when HDF equals 0, meaning no failure is present, there is a common distribution of rotational speeds across all product quality types. When HDF equals 1, meaning a failure is present, rotational speeds decrease across all product quality types. In particular, the low quality product becomes more variable in the distribution of rotational speeds. This figure proves our hypothesis correct and shows that the rotational speed decreases when HDF failure is present.

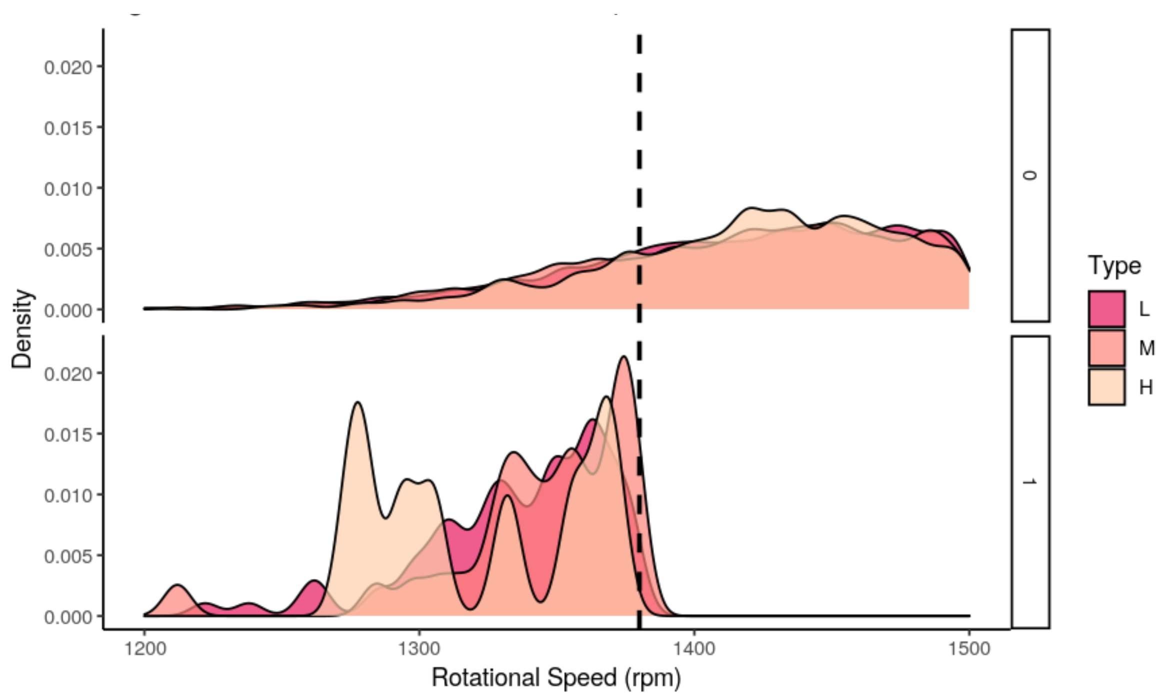


Figure 6. Density of products with rotational speed. This plot shows the Low (pink), Medium (salmon), and High (peach) quality types when a failure is present or not and shows the range of rotational speeds.

Hypothesis 4: The rotational speed and temperature difference (proc. temp. - air temp.) should decrease if an HDF failure is present

For the fourth part of the exploratory analysis, we wanted to compare the temperature difference between the process temperature and air temperature with the rotational speed since these are the causes of HDF failure. We hypothesized that a decrease in temperature difference and a decrease in rotational speed would lead to an HDF failure. Figure 7 shows that when the HDF failure equals 0, there is no failure present. So in the top portion of this figure, as the temperature difference decreases, the rotational speed increases across all product quality types. But when HDF equals 1, meaning a failure is present, then the overall temperature differences and rotational speeds are much lower. The temperature differences drop roughly 2 K and the rotational speeds drop roughly 100 rpm. This is true across the low and medium quality failures. On the contrary, when a high quality product has an HDF failure, the air temperature difference increases as the rotational speed decreases. Although not covered in this hypothesis, this merits further investigation. This figure proves our hypothesis to be correct in the low and medium quality products, but does not hold with high quality products since the temperature difference increases with rotational speed.

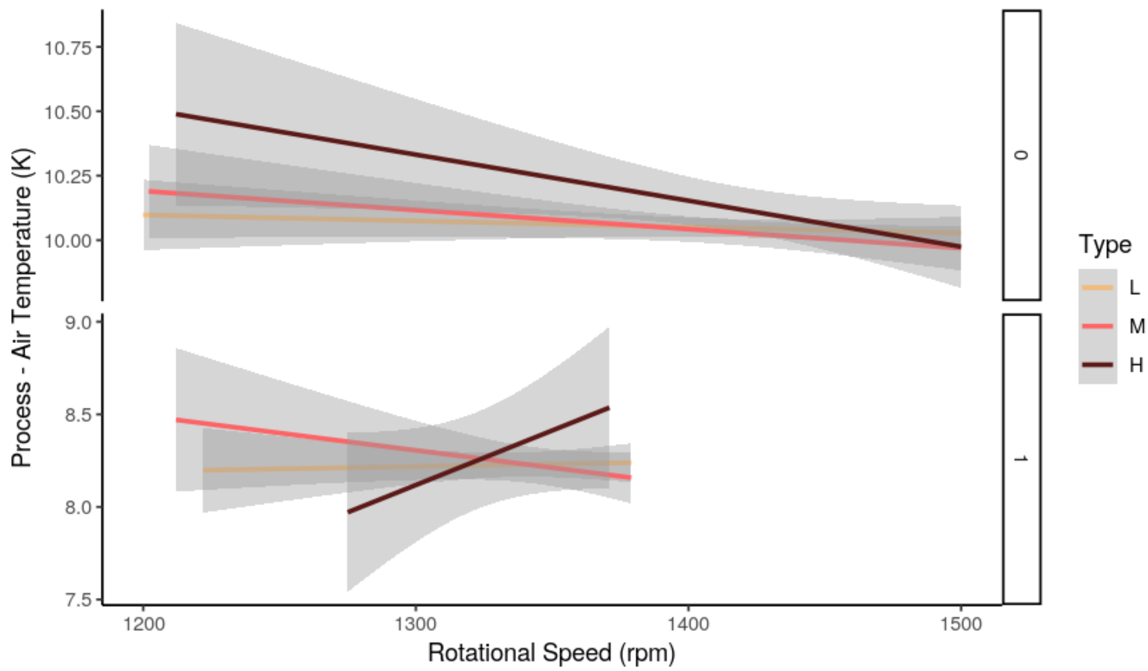


Figure 7. Difference in Process temperature and Air temperature with rotational speed. This plot shows the Low (pink), Medium (salmon), and High (peach) quality types when a failure is present or not and shows the range of temperature differences against rotational speeds.

Hypothesis 5: Some features of the production process are strongly correlated

Since we observe the encouraging positive correlation between the difference in process and air temperatures with rotational speed, we set out to see whether we can find other correlations between the process features. We constructed a correlogram containing all of our process features and divided our data points based on the quality of product manufactured. A strong positive correlation exists between air temperature and process temperature as we have expected. This positive correlation might suggest that creating a sufficiently large temperature difference to prevent heat dissipation failure could be a challenge. The only other significant relationship we find is the negative correlation between rotational speed and torque. Other features only exhibit a weak correlation between each other. An interesting note is that torque seems to be the feature that has the largest correlation with whether we would find machine failure or not. Additionally, torque and tool wear seems to have a larger influence on machine failure during the production of low quality products. On the other hand, air temperature and rotational speed influences machine failure more during the production of high quality products.

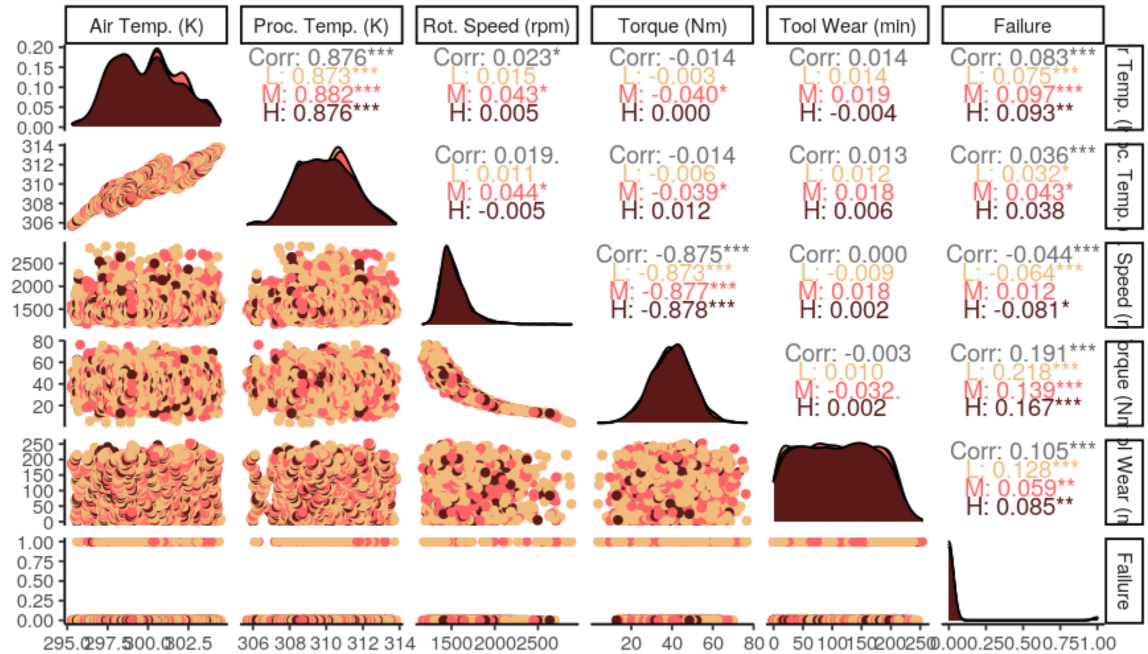


Figure 8. Correlogram of production features with datapoints divided between low, medium, and high quality products.

Hypothesis 6: Distribution of feature values differ depending whether or not we find machine failure

We constructed a second correlogram to see the distribution of feature values. We divided our data points depending on whether they are associated with a machine failure or not. From the correlogram we find that machine failures have feature values that are on the edge or even outside of the distribution for data points that do not have machine failure. This suggests that machine failure is associated with extreme production conditions. High air and process temperatures are more common during machine failures. This might explain the difficulty in creating a sufficiently large temperature difference needed to avoid heat dissipation failure. Higher torque is associated with machine failure. The negative correlation between torque and rotational speed also explains why lower rotational speeds tend to produce machine failure. As we can expect, when we have more tool wear we find more machine failures during production.

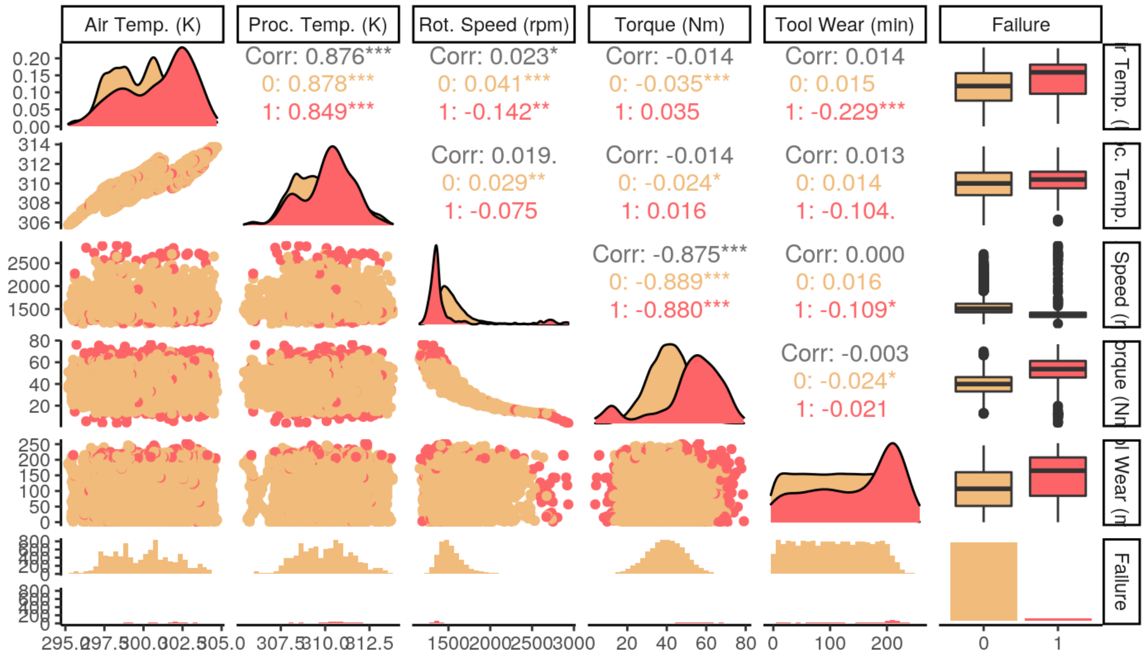


Figure 9. Correlogram of production features with datapoints divided between whether we find machine failure or not.

Modeling

Classification

In this project, we aimed to develop a machine learning model to inform predictive maintenance decisions based on synthetic data of measurements taken from operating machines. Our model performed binary classification, predicting machine parts as a failure or a non-failure. For this task, we tried several different types of machine learning models, including the Random Forest, Support Vector Machine (SVM), and k-Nearest Neighbors (kNN) models, and determined which model was more effective. We fit each of the three models to training data and evaluated their performance on separate testing data.

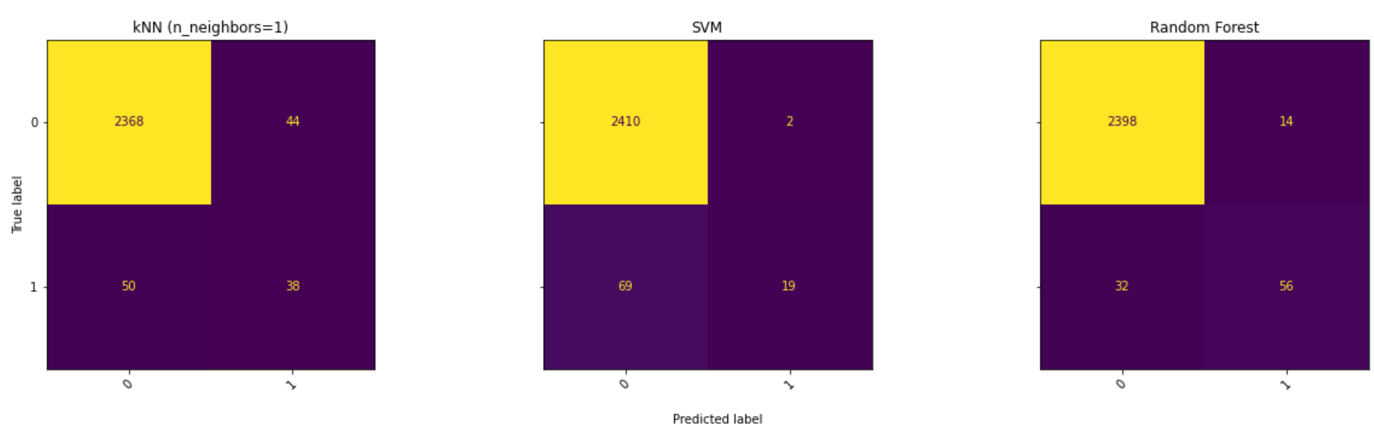


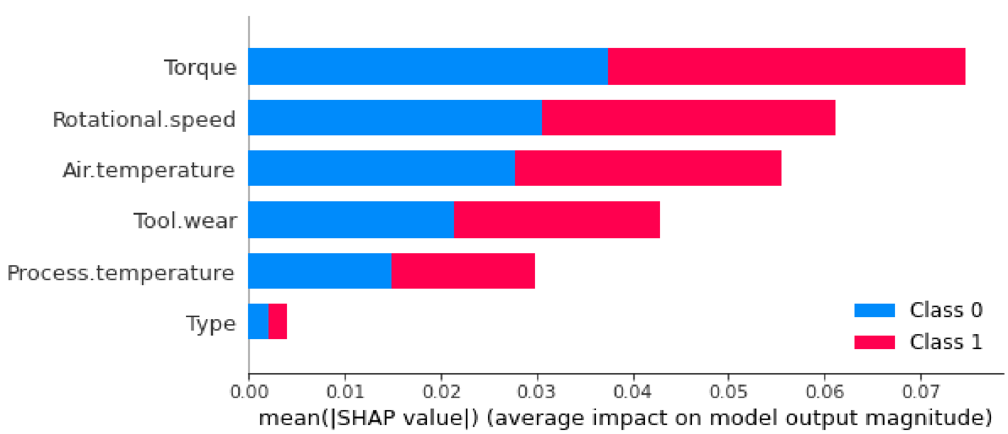
Figure 10. Confusion matrices.

Table 1. Model Accuracy and Recall

Model	**Accuracy**	**Recall**
SVM	0.97	0.22
kNN	0.96	0.43
Random Forest	0.98	0.64

Feature Importance

To evaluate the contribution of each of the features in the best fit model when predicting the binary target variable (whether or not a machine fails), we performed a SHAP (SHapley Additive exPlanations) analysis. Specifically, to obtain global importance, we calculated the mean of the absolute Shapley values per feature across the data — as such, the variables in the plot displaying the largest SHAP values, sorted higher on the plot, most significantly impact the model's prediction of whether a machine failure event occurs or not. Further, we concentrated our analysis to a singular prediction (machine failure occurs) by visualizing feature attributions as "forces" only for instances in the dataset in which machine failure occurs. From the baseline, determined as the actual prediction of machine failure or the average of all the predictions, each feature value represents a force pushing to increase (positive value) or decrease (negative value) the baseline prediction.




```
True Class (if 1, Machine Failure, othewise No Machine Failure): 0
Probability that this data point belong to class 0 = 1.00
Base Probability for class 0 = 0.97
```

Figure 11. Mean SHAP plot for machine failure and no machine failure.

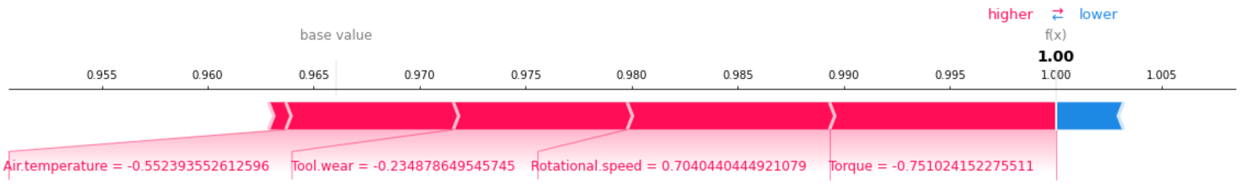
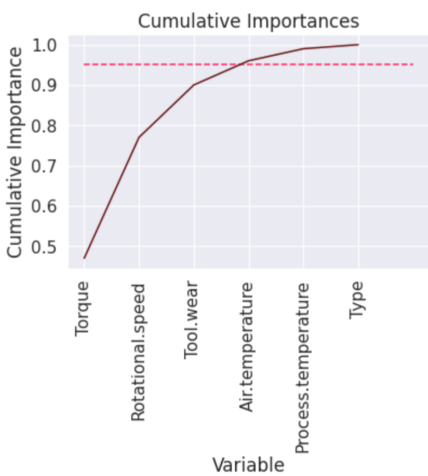


Figure 12. Force SHAP plot for only instances in which machine failure occurs.

We also constructed a cumulative importance graph that represented the contributions to predicting machine failure events from the successive addition of the feature's importance. By using the features' importances extracted from the Scikit-Learn module, we manually derived the cumulative sum values and plotted in ascending order with a dashed line drawn to benchmark that 95% of total importance has been accounted for. From visually assessing the changes in linear slope between the features or referencing the outputted table with the calculated cumulative sum values of the features, one can effectively determine the relative importance of each of the machine attributes in predicting a machine failure event.




```
Variable: Torque                Importance: 0.47
Variable: Rotational.speed      Importance: 0.3
Variable: Tool.wear             Importance: 0.13
Variable: Air.temperature       Importance: 0.06
Variable: Process.temperature   Importance: 0.03
Variable: Type                  Importance: 0.01
```

Figure 13. Cumulative importance plot and table to both visualize and quantitatively determine feature importance relative to machine failure occurring.

Discussion

In the industrial environment where this model would operate, machine failures are very costly, both in terms of direct equipment damage and factory downtime. This high cost of machine failure places an imperative on proper maintenance to ensure that machines never reach their failure point. As a predictive maintenance model, false positives, where our model predicts a failure for machines that do not actually fail, are much more acceptable than false negatives, where our model fails to predict a machine failure.

Reflecting this, we chose recall, which measures the proportion of positives our model correctly predicted out of all actual positive values, as the key metric for model evaluation. The results of our modeling bear out the veracity of this approach. As shown in Table 1, all of the models tested display extremely high accuracy, which could lead one to conclude that these models are all equally good. However, after looking at the recall, we can clearly see that the random forest model is the best model for this dataset.

Further validation of the superior performance of the random forest model can be seen in the confusion matrices presented in Figure 10. All three models correctly predict a similarly high amount of non-failures correctly, but the models clearly differentiate themselves on the number of the failures they predict correctly as the Random Forest model correctly predicts 56, while the kNN and SVM only predict 38 and 19, respectively. Furthermore, the Random Forest model has by far the lowest amount of false negatives, which is the error we most want to avoid in this model.

As shown in Table 1, the random forest model had the highest recall, a recall of .64, meaning that our random forest model successfully identified 64% of the positives in the dataset. This performance is reasonable, but it still misses over a third of all the actual failures in the dataset. This limitation in predictiveness will likely keep this model from being deployed in production.

This model is limited by a couple of important considerations. Due to the nature of binary classification and the random forest model as a black box model, our current model cannot give us information into how exactly the machine is going to fail and exactly what type of maintenance should be performed to prevent this failure. There is also a risk that this model has been fit to the specific physical situation of the dataset and will not be generalizable to future data. For example, if this data were taken at a certain season of the year or with a certain group of operators working, this model may not work for future time periods.

Conclusion

For our exploratory data analysis, we investigate six hypotheses to better understand the failure types in our dataset. Our first hypothesis addresses which failure type occurred the most, which resulted in HDF. Our second hypothesis analyzes the occurrence of failures relative to the quality of the product, which revealed that failures occur mostly in lower quality products. The third hypothesis explores the density of product quality against the rotation speed relative to the various quality of the products, which showed that the rotational speed decreases when HDF failure is present. The fourth hypothesis addresses the rotational speed and temperature difference when an HDF failure occurs. We discover that in the low and medium quality products both rotational speed and temperature difference decrease, but this does not hold with high quality products since the temperature difference increases with rotational speed. The fifth hypothesis investigates which features of the production process are strongly correlated. We discovered that some features did have a strong correlation such as air temperature and process temperature, and rotational speed with torque. The last hypothesis focuses on the distribution of feature values that differ depending whether or not we find machine failure. This analysis reveals that machine failure may be associated with extreme production conditions.

Based on our modeling, we found that the most useful method to predict machine failure was the random forest model. Our model was able to successfully predict 64% of true machine failures with a recall score of 0.64. Possible next steps in the development of our project would be to model and predict each possible type of machine failure. As stated above, one of the limitations of our model is that machine failure is a binary variable, “0” for no failure, and “1” for failure. With this classification, the type of machine failure is disregarded making it more difficult for manufacturers to address the failure. In the future we would like to eliminate this limitation by making the model more detailed by predicting which of the five types of failure is reached. This would be more useful to stakeholders as they would not have to try to figure out why the machine failed. Therefore, they would have a much better idea about what maintenance to perform on the failing equipment.

Acknowledgement

Contribution Table:

Group Member	Contribution	Percentage
Allysa Dallmann	Exploratory data analysis: data cleaning & visualization, hypotheses, helped make hypotheses slides, wrote exploratory data analysis section	100
Michael Montez	Made SVM model and github website, wrote Discussion and part of Modeling report sections and helped make Modeling presentation slides	100
Shreya Nallaparaju	Random Forest model, model evaluation (confusion matrix and feature importance), modeling presentation slides, designed slides/intro visualizations	100
Jonathan Saldeen	Introduction, conclusion, helped make the same slides on the presentation, part of the data description in report.	100
Sebastian Utama	Exploratory data analysis: data cleaning & visualization, hypotheses, wrote exploratory data analysis section	100
Kayla White	Exploratory data analysis: data cleaning & visualization, hypotheses, wrote exploratory data analysis section	100
Emily Zhou	Data cleaning, kNN model, model evaluation, hypothesis, slides: modeling, report: modeling	100

Bibliography

Stephan Matzka, 'Explainable Artificial Intelligence for Predictive Maintenance Applications', Third International Conference on Artificial Intelligence for Industries (AI4I 2020), 2020 (in press)