PREDICTING MACHINE FAILURE USING MACHINE LEARNING

SDS322E FINAL PROJECT REPORT

Kushagr Bhatia, Rohit Dinesh, Harshit Gupta, Revath Sankar, and Riley Ylagan

The University of Texas at Austin

Fall Semester 2021

Contents

1 Introduction

The problem we are trying to solve is understanding when a machine will fail given that machine's current conditions. To do this we are creating a machine learning model which will help us classify if the machine will fail, and if so, specially which of the 5 different failure types caused it.

The concept that we are working on is called predictive maintenance, and it has several different applications and use cases which make it important. Predictive maintenance helps detect problems within a machine before they occur so they can be quickly addressed. This process helps avoid unexpected failures, which could cause assembly lines to stop, and result in a loss of time and revenue for the factory. The application of predictive maintenance is important for factory owners because it helps them save money and avoid abrupt stops in their operation.

2 Understanding the Data Set

2.1 Data Set Description

The original data set is a "synthetic data set that reflects real predictive maintenance encountered in industry" [1] . It consists of 10,000 observations stored as rows with 14 different feature variables. Each observation represents one single process on a machine. The variables can be split into three different categories: process identifiers, process parameters, and process failure labels. The variables and their descriptions are as follows:

Process Indentifiers-

These variables identify each unique process.

UID	Unique identifier ranging from 1 to 10000
Product ID	Unique ID for each product
Product Type	Product quality of either low(L), medium (M), or high (H)

Process Parameters-

These variables are the condition of the machine when a particular process is run. These variables determine whether or not a machine failure occurs or not.

Air Temperature (K)	Temperature of surrounding environment measured in degrees Kelvin
Process Temperature (K)	Temperature of precision bit measured in degrees Kelvin
Rotational Speed (rpm)	Turning speed of the tool, calculated from a power of 2860 W, overlaid with a normally distributed noise.
Torque (Nm)	The linear force of the tool. The values are normally distributed around 40 Nm with a = 10 Nm and no negative values.
Tool Wear (min)	Time of the current tool has been in use. The quality variants H/M/L add 5/3/2 minutes of tool wear to the used tool in the process.

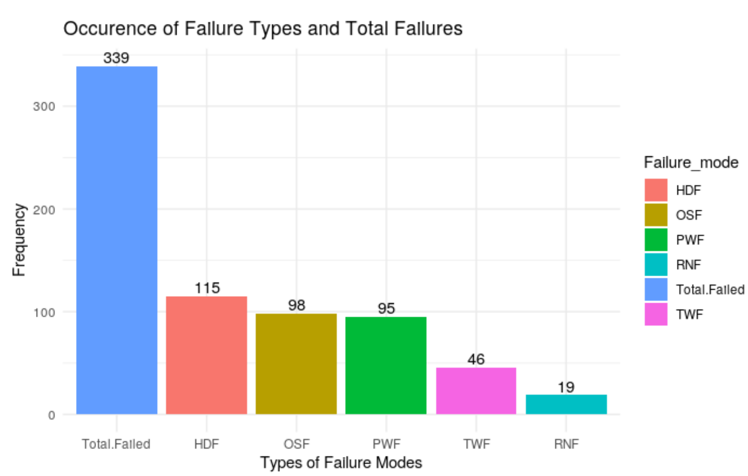
Process Failure Label Failure Modes-

All of these variables are binary variables with 1 meaning the failure mode occurred and 0 meaning the failure mode did not occur. If any of the 5 failure modes occur, then the 'Machine Failure' variable is set to 1 indicating that a failure occurred during the machine process.

Machine Failure	Label that indicates whether the machine has failed in this particular data point if any of the following failure modes are true.
Tool Wear Failure (TWF)	The gradual failure of cutting tools due to regular operation. Tools affected include tipped tools, tool bits, and drill bits that are used with machine tools.
Heat Dissipation Failure (HDF)	Heat dissipation occurs when an object that is hotter than other objects is placed in an environment where the heat of the hotter object is transferred to the colder objects and the surrounding environment.
Power Failure (PWF)	There are many causes of power failures in an electricity network. Examples of these causes include faults at power stations, damage to electric transmission lines, substations or other parts of the distribution system, a short circuit, cascading failure, fuse or circuit breaker operation.
Overstrain Failure (OSF)	Overstraining leads to material transformation, for instance strain hardening, followed by changes in material properties such as toughness loss.
Random Failure (RNF)	A random occurrence that cannot be accounted for as a common preventable failure.

2.2 Cleaning the Data Set

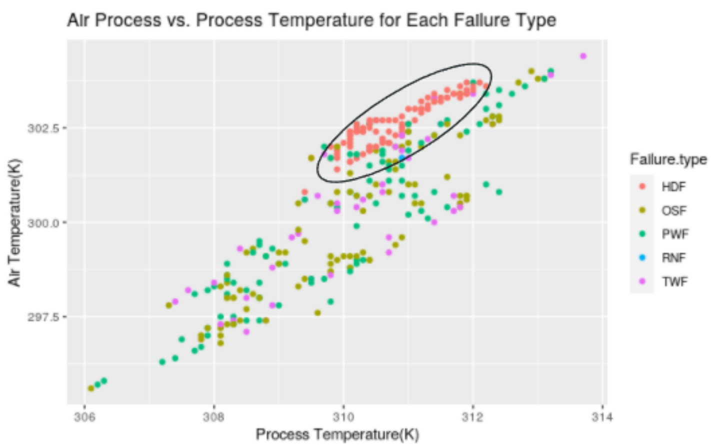
Next the dataset is cleaned and prepared for data analysis by removing NA or null values and duplicate data, and tidying or reshaping as needed. In order to make visualization and classification easier, the last five variables are combined into a singular categorical variable called 'Failure Type' indicating what type of failure occurred for a particular process. This is achieved by pivoting the last 5 columns longer and removing duplicate and unnecessary columns. The resulting dataset now has 11 feature variables and is ready to be used in exploratory analysis. The amount of each failure type can now be easily visualized.



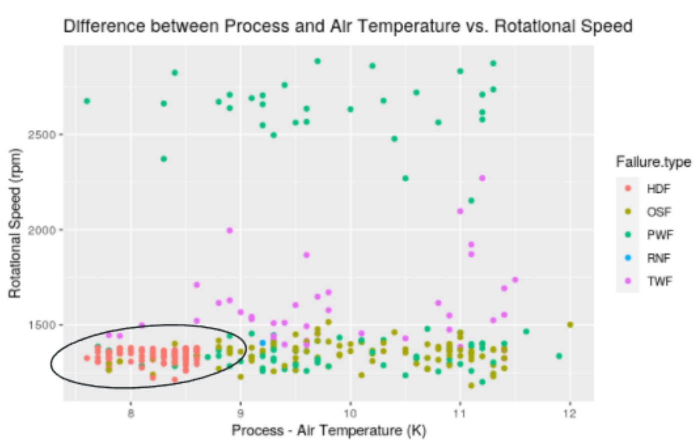
From this bar plot, we can see the amount of times a certain failure mode occurs. We can see that, in this data set, only 3.39 % of the 10,000 processes fail. This imbalance within the data set could prove issues when a classification model is run.

3 Exploratory Analysis-

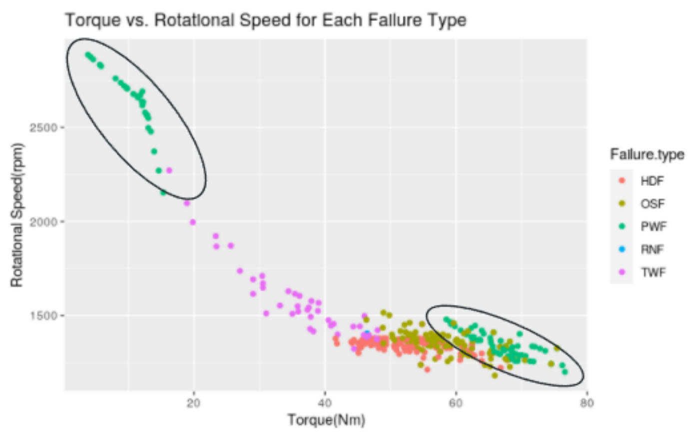
Now that the data set has been prepared for analysis, it can be used to find relationships between the variables in the data. As mentioned before, our main goal is to be able to identify when a failure occurs so we will mainly be looking at the processes that fail. To do this we look for critical values of the process parameters by plotting variables against each other for processes that failed and coloring the points by failure type. In this way, the influence of certain process parameters on the failure type can be identified by looking for tight clusters and sharp cutoff points which might indicate critical values. From this, we can make hypotheses about which parameters influence a particular type of failure mode. Since random failures occur regardless of process parameters, it is excluded from investigation as it cannot be explained from the variables in the data set. Linear regression and correlation matrices would not give relevant information from this data set, so it is not used in analysis. Let's start by looking at a few scatter plots plotting two variables against each other.



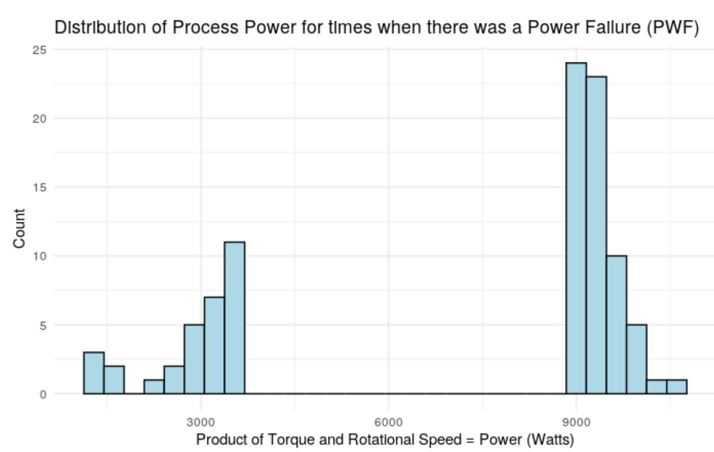
Looking at the scatter plot of failed processes for air temperature vs. process temperature we can see that heat dissipation failure (HDF) type seems to be clustered around a certain area while the other failure types do not appear to be influenced by the temperature. This makes sense in the context of the definition of heat dissipation failure laid out in the previous section. Now, we take a look at the difference between air and process temperature and look for relationships along other variables.



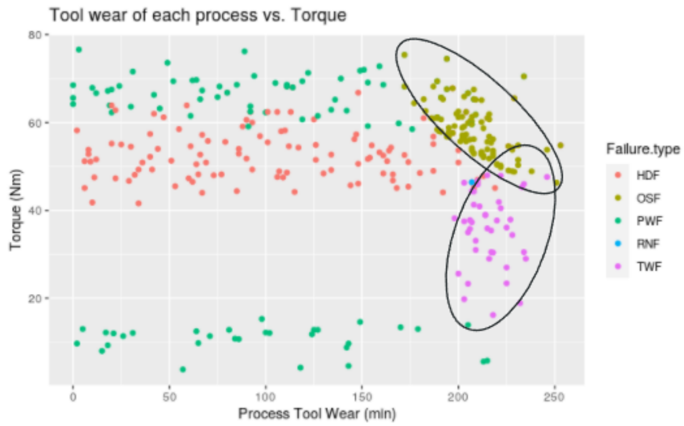
When plotting temperature difference against rotational speed, we can notice further clustering of the HDF type. From this, we can hypothesize that a heat dissipation failure occurs when the temperature difference and rotational speed of the tool is below a certain value. Next, we will try to find which process parameters influence the other failure types.



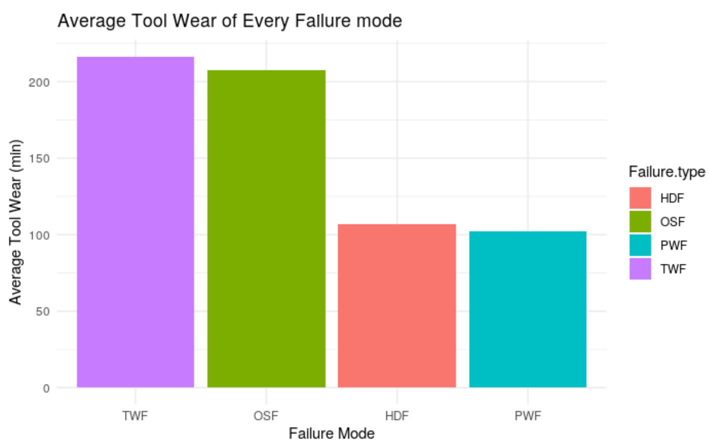
The above scatter plot plots the rotational speed against torque for the processes that fail. We noticed that the power failure (PWF) type appears to be located at the ends of the scatter plot. If we realize that the product of torque (Nm) and rotational speed (in rad/s) is power in watts, it begins to make sense. It appears that power failures might happen when the power is either too high or too low. This is supported by plotting the distribution of the power of each process that has a power failure:



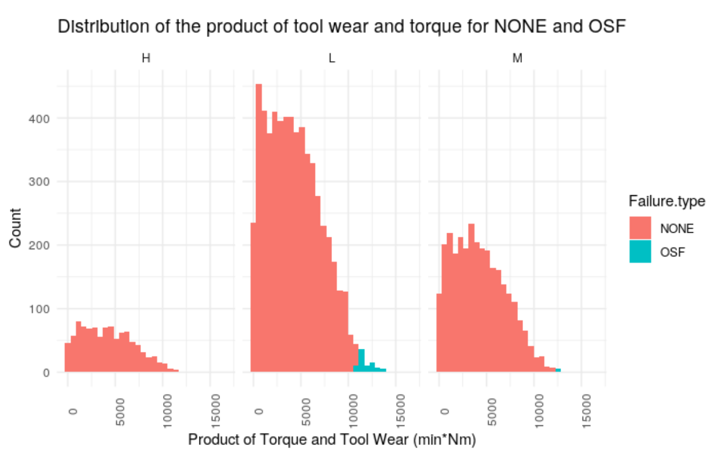
Here, we can see that the distribution of power occurs below about 3500 watts and above about 9000 watts. We can hypothesize that power failures occur when the power of a process is too low or too high. Next, we will look for causes of tool wear failure and overstrain failure.



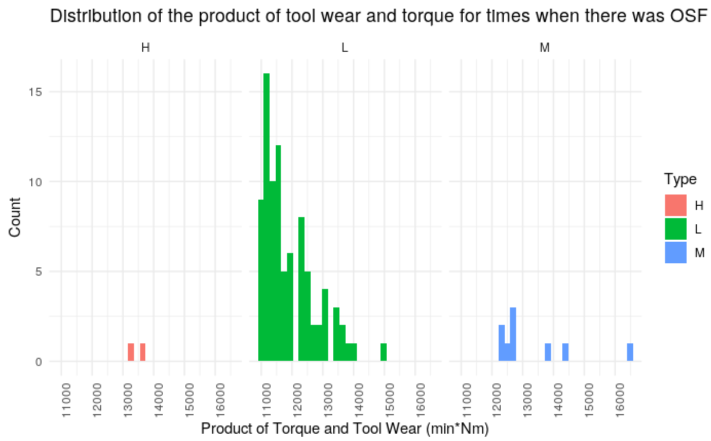
From the plot for tool wear vs. torque, we notice that there are two clusters that appear around the higher values of tool wear. This indicates that tool wear (TWF) and overstrain (OSF) failures might occur when the tool wear at the time of the process is high. Looking at the average tool wear of every failure mode we can see that the average tool wear of TWF and OSF is much higher than the other failure modes:



It makes sense to hypothesize that tool wear failure occurs when the tool wear of the process is too high because of how we defined tool wear failure in the previous section. However, we still want to investigate if other variables might cause an overstrain failure. If we notice that the product of tool wear and torque is the angular momentum of the process, then we can start to see some explanation for when an overstrain failure occurs.



This bar plot shows all of the distribution of the angular momentum for times that either no failure occurs or an overstrain failure occurs stratified by product quality. We can notice that overstrain failures occur at higher values of angular momentum.



Now, looking at only the times when an overstrain failure occurs, we can see that for a product of type L, the an overstrain failure occurs when the angular momentum is at least 11,000; for a product of type M, an overstrain failure occurs when this value is at least 12,000; and for a product of type H, an overstrain failure occurs when this value is at least 13,000.

3.1 Conclusions from Exploratory Analysis

From the exploratory analysis, we can generate a few hypotheses:

- Other than RNF, these five failure modes occur with dependence on the process parameters (i.e air temperature, process temperature, rotational speed, torque, and tool wear).
- From analysis, we can hypothesize that each failure mode is influenced by:
  - Tool wear failure (TWF) occurs only when the tool wear is at a certain value.
  - Heat dissipation failure (HDF) occurs according to the difference between the air and process temperature and also the rotational speed.
  - Power failure (PWF) occurs when a process' power is too high or too low.
  - Overstrain failure (OSF) occurs when the product of torque and tool wear exceeds a certain amount dependent on the processed products' type (L,M, or H).
  - Random failures (RNF) can occur regardless of process parameters.

4 Modelling

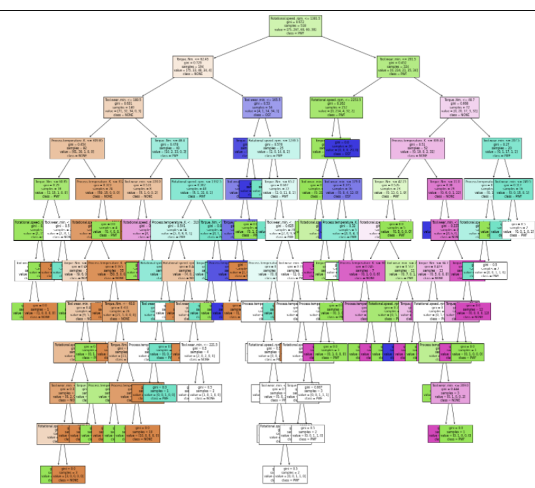
The data presented to us throughout this project dignified a predictive classification task. There were a few things that we as a team needed to look into before deciding which model would best fit our needs. An initial examination of the key features of the data set we were using indicated a strong non-linearity to each feature; indicating that the classification method we would choose needed to account for that. Additionally, in considering the type of data and goal of creating a simplistic feature and cut-off process model that an employee in a manufacturing environment could use as a diagnostic tool in the event of machine failure or act as a preventative measure, allowing them to optimize each performance feature of the machine to reduce the likelihood of machine failure. Thus we needed to account for another factor - interpretability. Finally, what eliminated models like K-Nearest Neighbours was the fact that several factors needed to be accounted for when predicting binary classification of "Failure/No Failure". So, it was ultimately decided that we would use a Decision Tree Classifier for interpretability and clarity alongside the multi feature classification abilities.

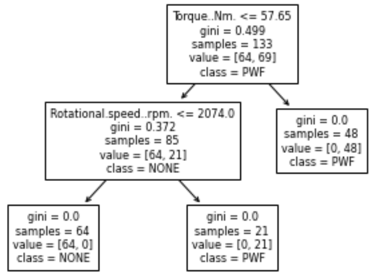
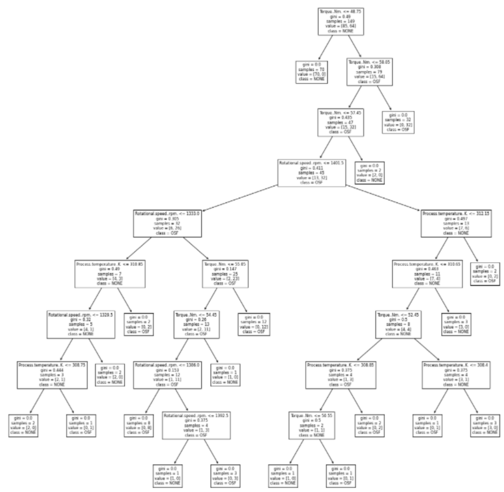
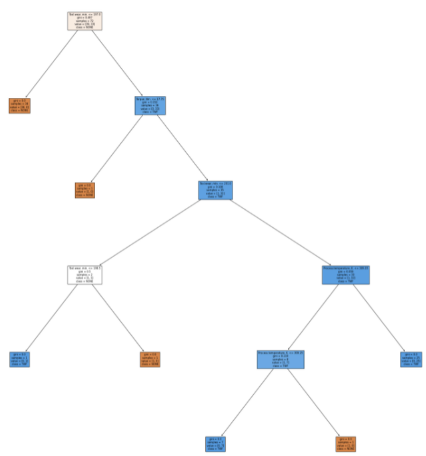
5 Discussion

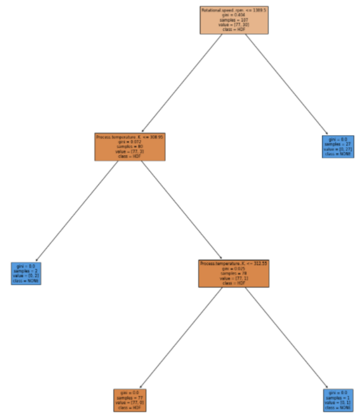
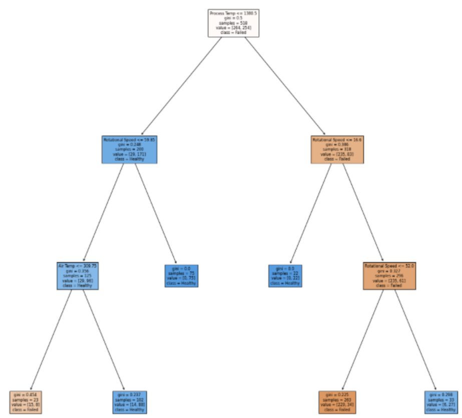
The model had several things to be corrected before actually beginning training. First, as mentioned before, the dataset was heavily imbalanced; with 339 total Failure to 9661 No Failure points. As such we needed to randomly sample about 97% of No Failure points and drop them to achieve as close to a 1:1 ratio of Fail:No Fail ratio to train the model as well as possible. As our team built six models, we found that we needed to repeat this task for each model. In order to best understand the problem of predicting machine failure, we addressed each failure mode head on.

Model	Model Goal
Type of Machine Failure	The goal of this model is to take all data and predict type of tool wear failure (ie. TWF, PWF, OSF etc.)
PWF Feature Model	The goal of this model is to determine which features are considered most important in power failures.
OSF Feature Model	The goal of this model is to determine which features are considered most important in overstrain failures.
TWF Feature Model	The goal of this model is to determine which features are considered most important in tool wear failures.
HDF Feature Model	The goal of this model is to determine which features are considered most important in heat dissipation failures.
Fail - No Fail Model	The goal of this model was to create a high-level feature based understanding of what makes a machine fail or nor overall regardless of type or quantity of failure.

5.1 Decision Tree Visualization

Model	Visualization
Type of Machine Failure	

Model	**Visualization**
PWF Feature Model	
OSF Feature Model	
TWF Feature Model	

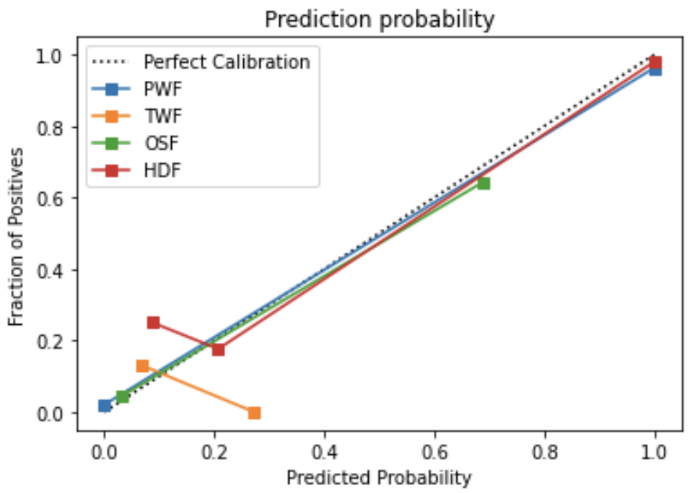
Model	Visualization
HDF Feature Model	
Fail - No Fail Model	

5.2 Model Results

We next needed to see how the models fared as a function of accuracy and precision after the randomized drops took place. This was done using the standard `sklearn` library `accuracy_score` and `precision_score` packages and were used on a test set that was created at a 80% : 20% train to test ratio. The result was overwhelmingly positive, as shown below:

Classification System	Accuracy Score(%)
Prediction of type of failure	75.1677
Prediction of PWF Failure	86.2069
Prediction of TWF Failure	58.5365
Prediction of OSF Failure	89.6907
Prediction of HDF Failure	95.0495
Prediction of overall machine failure	89.9329

We further confirmed this with prediction calibration. Calibration curves compare how well the probabilistic predictions of a binary classifier are calibrated. It plots the true frequency of the positive label against its predicted probability, for binned predictions. The x-axis represents the average predicted probability in each bin. The $y-$axis is the fraction of positives, i.e. the proportion of samples whose class is the positive class for each bin ($n\_bins = 10$, for our purposes). In comparing all the different failure mode classifiers, we see an immediate trend where Tool Wear Failure does markedly poorer compared to the sibling models.



We also created confusion matrices to visualize the classification performance in a cleaner manner than simply precision scores - the same trends were clearly seen - where the precision of all modes are high in precision (> 90%) with the singular exception of the Tool Wear Failure model. This confirmed that there was an additional layer of information that was preventing the model from effectively understanding what led to Tool Wear Failure. Further inspection revealed two key reasons: 1) a low sample size of Tool Wear Failure instances - in a data set of 10,000 points, only 45 were classified as Tool Wear Failure and despite balancing the data set to match the ratio of Tool Wear Failure to No Failures - it was too little information for the model to accurately build off; 2) the nature of Tool Wear Failure was organically random - as a tool would be randomly replaced between 200 - 240 minutes of operating time (denoted as Tool Wear [min] in set) before the tool could fail, which compounded the error rate as some samples would fall within the feature cutoffs established by true failures but suddenly not fail.

As a team of engineers on the floor we were most interested in what the driving factors were in each case of failure (ie. What feature was most important in PWF, OSF, TWF etc.?). As such our team began to conduct a SHAP analysis in order to better understand what of the pure features were causing each failure type. A simple explainer plot from the standard SHAP package (`https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html`) indicated some interesting trends. First, we learned that Tool Wear time was the driving factor for Tool Wear and Overstrain Failures - which actually contradicted our teams' hypothesis that Torque would be the driving feature given the nature of overstrain in mechanical systems. Heat Dissipation Failure was primarily a function of Rotational Speed, contrary to our previous prediction that Process Temperature would be the dominant feature. Finally, in our examination of Power Failure, we saw that Torque was the driving factor in that particular failure mode. As mentioned before, the goal of our project was to understand the driving factors that influenced overall machine failure from this data set as such we ran a similar SHAP Explainer analysis on the Fail/No Fail model to determine what the overall driving factors were for machine failure - this time including Random Failures - and found that the overall environmental variables (Air Temperature and Process Temperature) played very little role in a machine's failure - rather the applied Torque and Rotational Speed of operation were the two most dominant factors that determined failure status.

5.3 Limitations

No study in the field of data science is without its limitations and our model is no exception. Firstly, there is something to be said about the data itself, in that the set is purely synthetic - as such we have no real datum to quantify the quality of the synthesized data and thus have to rely on low deviations from raw data. Secondly, Tool Wear Failure had both a low number of samples and a strong element of randomness that despite our best efforts in balancing, could not be totally accounted for by our Decision Tree Classification model. Thirdly, in cases of multiple failures, it is impossible for us or the machine to know which failures occurred first, and thus may have compounded the effect into another failure type - and as such there might be some minor skewing to feature cutoffs for the different failure modes. Fourthly, it should be noted that many randomized drops were required to achieve the 1:1 ratio required for each model - when repeatedly tested, there were elements of variability in the accuracy and precision scores for each model - were the data set not so wholly imbalanced - the results of the tests might provide more stability to the classification model. Finally, we did not account for Random Failures in our model - as the data set indicated an element of true randomness in their failures that failed to respond to models as there were elements that impacted its failure outside of the provided process variables.

5.4 What's Next?

In the future, we will try to account for these by testing the quality of the synthetic against the real data it was based on to better gauge the quality of the data and in turn, our model; in regards to the Tool Wear Failure variance in accuracy, the goal is to capture more data points such that not so many drastic cuts will need to be made and we as a team can gather which features are truly most important and their respective cutoff points. Thirdly, it would be wise to record the order of failure moving forward - such that we can further determine if one failure mode might influence the confluence of another as the process continues. Fourth, the gathering of a balanced data set would eliminate the need for so many drops to be made going forward - so the goal would be to gather data that would reflect as close to a 1:1 ratio of Failure : No Failure moving forward while having reasonable number of each failure type reflected in the data set. As for the Random Failures, there are likely additional variables that could be attributed to this mode of failure such as product material, product size, quality of processing (refinement of product) and time since last failure. This is an ongoing area of research within data science known as predictive maintenance - and is likely to play a critical role in process optimization and quality management in manufacturing processes going forward.

6 Conclusion

Now that you have read through our code, seen our visualizations, and got to learn how we predicted machine failure, here's a quick summary of what we learned. Our purpose in pursuing this project was to determine which factors are most important in predicting machine failure and determining their relative importance in determination. Essentially, we wanted to figure what causes machines to fail and which contribute to failures the most and which factors contribute the least.

Through our analysis we found the following:

- Tool Wear Failure is poorly classified which can be attributed to two factors: Low number of TWF cases and the Random nature of tool replacement or failure.
- Torque and Rotational Speed dominate the classification process. Interestingly, the dominant factor in Heat Dissipation Failure Cases was due to Rotational Speed rather than Process Temperature.
- Overstrain Failure was more heavily influenced by Tool Wear time rather than applied Torque

Our findings from the synthetic data set were promising as we were able to find not only the type of failure but also the factors that caused it. If we were to continue working on predictive maintenance, we would carry forward our learnings to go from a classification task to a time-series associated predictive task. One variable from the raw data set that we did not wish to explore was the 'time of failure' for a certain quality ('L', 'M', 'H') of equipment. Supplementing our classification task with the time series will help provide a life cycle estimate for our products. This means that equipment could be replaced before catastrophic failure and and not hamper the production line till replacement.

For our analysis and decision-tree classifier model to be put in production, we discovered three main areas of focus:

Reciprocity: translating our model to work with real data as opposed to the synthetic data as in this scenario

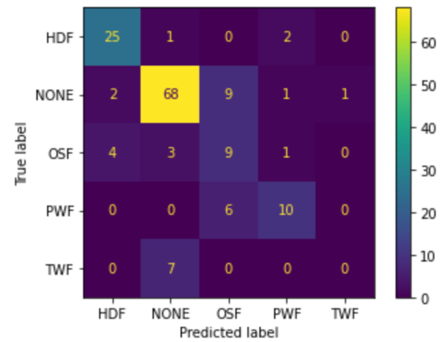
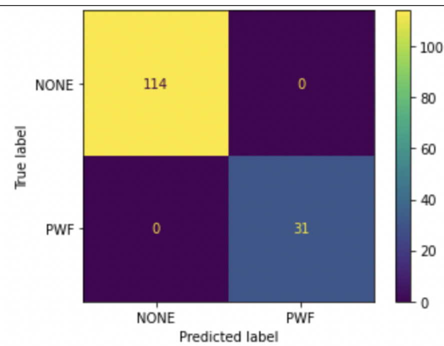
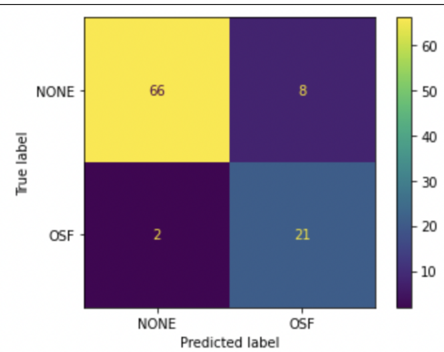
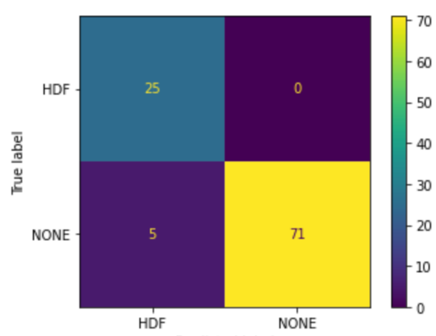
Applicability: figuring out the usability of our model to hold ground in a variety of industry specific applications, i.e. healthcare equipment, production lines of various industries, potentially a human resources tool to measure worker productivity

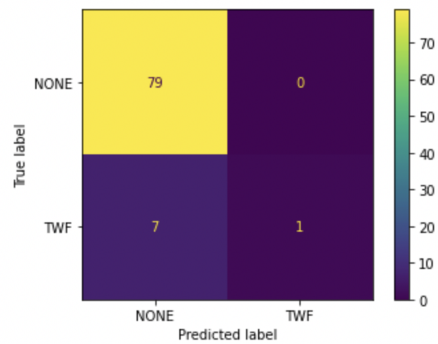
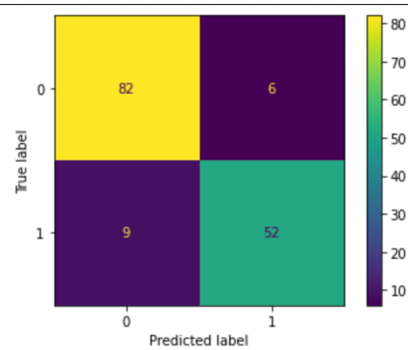
Scalability: bolstering the performance of our model using feed-forward DNN Classifiers to work with more sensor collected data variables

We hope you learned as much as we did through this process! Thanks for reading our findings!

7 Appendix

A. Confusion Matrices generated for each model

<table>
<tr><th>Failure Model</th><th>Confusion Matrix</th></tr>
<tr><td>Overall Type of Failure</td><td>
</td></tr>
<tr><td>Power Failure</td><td>
</td></tr>
<tr><td>Overstrain Failure</td><td>
</td></tr>
<tr><td>Heat Dissipation Failure</td><td>
</td></tr>
</table>

Tool Wear Failure	
Holistic Fail-No Fail	

8 Acknowledgement

Revath Sankar - Model Building, Testing, Slide, Paper and Website Development
Riley Ylagan - Data Cleaning, Data Analysis, Model Interpretation, and Formatting Report in LaTeX
Kushagr Bhatia - SHAP development, Paper and Slides Development
Rohit Dinesh - Slide, paper, testing DNN classifier
Harshit Gupta - Paper and Slides Development

9 References

[1]. Stephan Matzka, 'Explainable Artificial Intelligence for Predictive Maintenance Applications', Third International Conference on Artificial Intelligence for Industries (AI4I 2020), 2020

[2]. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

[3]. Lundberg, Scott M., and Su-In Lee. "A Unified Approach to Interpreting Model Predictions." Advances in Neural Information Processing Systems, 1 Jan. 1970, https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.