DATA ANALYSIS TO REDUCE MAINTENANCE COSTS FOR LAND PLATFORMS

Yang An Yi[1], Cheong Yek Meng[2], Ong Sir Hong[2]

[1] Hwa Chong Institution, 661 Bukit Timah Road, Singapore 269734
[2] Defence Science and Technology Agency, 1 Depot Road, Singapore 109679

ABSTRACT

Land platforms are critical assets for military training and operations. When these platforms break down at suboptimal times, it affects operational objectives and increases maintenance cost due to the downtime. Thus, this project aims to use data analysis of telemetry data taken from land platforms and classification machine learning models to predict if failures would occur using past data trends. The aim is to boost the land platform's availability, reduce maintenance costs by predicting potential platform breakdown using the specific dataset provided. Through tests conducted, a comprehensive model was developed which obtains a prediction accuracy of 97.50%.

TABLE OF CONTENTS

INTRODUCTION

Even with the constant development and improvement of technology over the past few decades, land platforms still play an important role serving as the cornerstone of modern warfare. The maintenance of land platforms, especially legacy ones, still remains a pertinent issue even today. The frequent or seemingly random breakdown of land platforms negatively affects its availability and causes an increase in maintenance costs.

As is frequently said, prevention is better than cure. This project thus aims to use state-of-the-art data analysis methods to process past telemetry data of land platforms and effectively forecast future possible failures in these land platforms in order to enhance availability and reduce maintenance costs, as well as to potentially enhance spares support for land platforms.

In addition, the limited number of individual vehicles results in a smaller data sample and hence may affect the quality of results obtained from data analysis. With this in mind, this project also aims to optimise data analysis methods employed to better extract useful features and conclusions from the dataset.

The parameters given are variant, mileage, engine hours, phase (of production), age till date. Initial intuition indicates that mileage and engine hours may be an important factor when determining breakdown as they indicate usage of the land platforms, and increased usage usually leads to a higher chance of breakdown.

LITERATURE REVIEW

Predictive maintenance to help reduce costs and improve equipment uptime has been a focal point for various industries, especially in sectors where machinery and equipment play a crucial role in operations. For Small and Medium Enterprises (SMEs), the maintenance cost of equipment could be the crucial difference between a profit and a loss. On the other hand, the uptime of crucial equipment matters, for example for land platforms in a military context where the availability is of logistical importance and affects operations and training success.

The utilisation of data analytics has gained significant attention due to its potential in optimising maintenance strategies through predicting failures. Various studies have emphasised its importance through analysing large volumes of telemetry data generated by sensors, monitoring systems and historical records to identify patterns and trends related to equipment performance and failure.

A predominant area of research involves the development of predictive maintenance models using advanced analytics techniques, for example machine learning. These models aim to forecast potential failures by analysing data. Research by Sezer et al. (2018) demonstrated the effectiveness of a low cost predictive maintenance approach which was applied to a CNC turning centre. Research by Mallock et al. (2021) showed that the effectiveness can be further extended to multiple aspects of a transport system, including “predicting upcoming failures”. The model used in the paper (Mallock et al., 2021) utilised the Random Forest model for predicting upcoming failures, which is one of the models that was tested for this project.

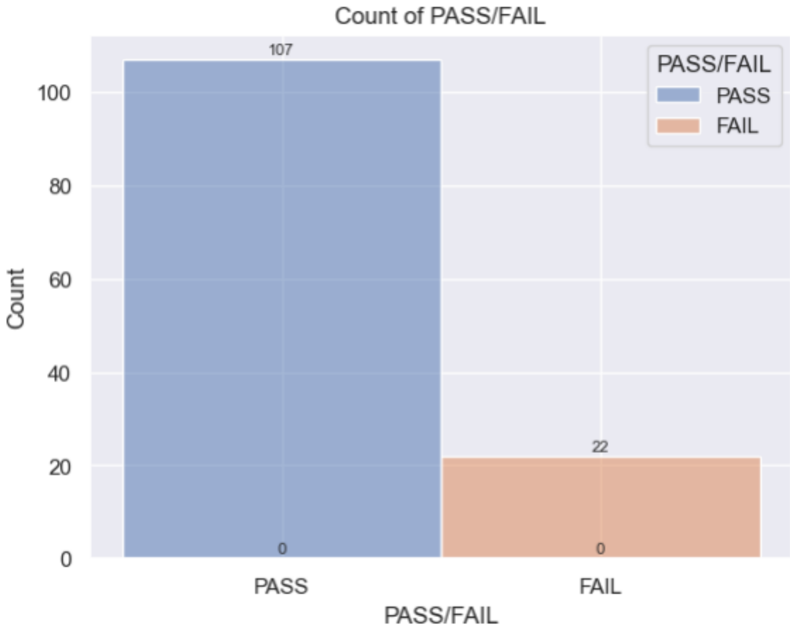
METHODS

Preliminary Observations

There are in total 129 individual data points, split into 107 non-breakdowns and 22 breakdowns as shown in Figure 1. Hence, the minimum target to achieve is 107 ÷ 129 = 82.9% accuracy and an ideal final accuracy is >90%.

Figure 1

Number of Pass and Fails



The five parameters from the dataset provided are "variant" (of the land platform), "mileage", "engine hours", "phase" (of production) and "age till date". "Variant" and "phase" are interpreted as categorical data, whereas "mileage" and "engine hours" are interpreted as numerical (continuous) data, with age till date interpreted as numerical (discrete) data.

Initial intuition indicates that the main parameters that matter are "mileage", "engine hours" and "age till date" since these relate to how much the land platforms are used. The hypothesis would be that "mileage" and "engine hours" show a strong correlation with land platforms failure rate.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a method to study, visualise and analyse datasets to identify their predominant traits. It also allows for the discovery of patterns and identification of correlations between variables. EDA will be employed on this dataset in an attempt to uncover any interesting relationships and to possibly prove the initial hypothesis erroneous.

For EDA, the packages Pandas, Matplotlib and Seaborn will be used in Python as statistical tools for data visualisation.

```
df[“phase”].value_counts()
```

Table 1

Number of Data Points per Phase

Phase	Count
1	36
2	31
3	62

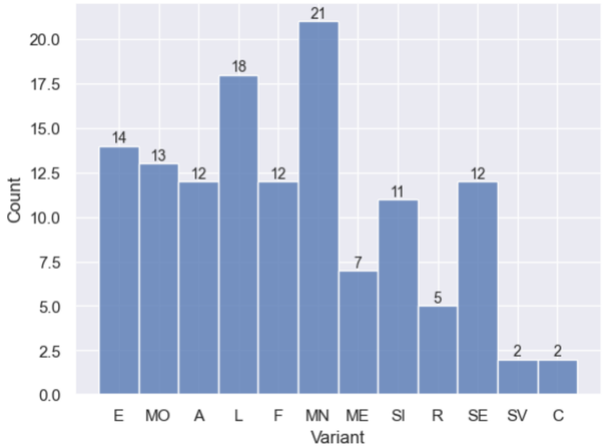
Note. In the original data, the phases were labelled 1, 3 and 4. They have been replaced by 1, 2 and 3 respectively for readability.

In Table 1, there is a notable preponderance of data points in Phase 3 as shown above. It is important to take this into account as in the event that the "phase" has a huge impact on the pass/fail rate, Phase 3 will be heavily weighted and the results would tend to be biased towards Phase 3.

Figure 2 shows the number of data points per variant in the dataset. Each variant does not have a lot of data since the data is split into many variants. Hence trying to train individual models on each variant can be ruled out. However, as seen in Table 1, there is enough data for each individual phase, which allows model training on each individual "phase" as a possible method for consideration.

Figure 2

Number of Data Points per Variant.



Looking at the number of pass and fail for each variable (variant, mileage, engine hours, phase and age) gives us the figures 3-7 (in Appendix). From those figures, it is obvious that the strongest correlation exists between phase and the number of passes and fails as per Figure 6. In Figure 6, Phase 2 shows a high failure rate as compared to Phase 1 and 3. This relationship is to be observed further after preliminary EDA is completed. From Figure 7, it can be observed that age matters as well, ignoring the deviation at age 6 which could potentially be explained by the smaller dataset.

As such, the EDA portion can be concluded.
The main findings are:
(1) Phase 2 is strongly correlated with high failure rates (38.7% failure rate)
(2) Age is correlated with failure rates
(3) Variant seemingly influences failure rates, although it is not yet clear
(4) Mileage and engine hours show seemingly no correlation with failure rates

Method Selection

The aim of the model is to predict whether a breakdown would occur, given the parameters mentioned above. Since this is a classification task and due to the relatively smaller dataset size of 129 land platforms, the following methods will be evaluated: Decision Tree, Gradient Boosting, Random Forest and K-Nearest Neighbours. The model will be first tested on the dataset in its entirety excluding the "variant" variable since there are not many data points per variant.

Explanation of models

Decision Tree is a supervised hierarchical model used in decision support that depicts decisions and their potential outcomes using a tree-like model. It classifies complex objects by recursively breaking them down into smaller groups based on their features, quantifying the values of outcomes and the probabilities of achieving them. The first Classification and Regression Tree (CART) was developed by Breiman et al. (1984).

Boosting is an ensemble modelling technique that aims to improve predictive accuracy of models by combining multiple weaker learners into a strong one. Boosting is based on the question posed by Kearns and Valiant (1988, 1989): "Can a set of weak learners create a single strong learner?". The affirmative answer was later provided by Schapire (1990).

Random Forest (or random decision forest) is an ensemble learning method that constructs a multitude of decision trees at training time. For this use case (classification), the output of the random forest is the class selected by the most trees. This helps correct for decision trees' habit of overfitting to their training set. The first model was created by Tin Kam Ho (1995) and an extension of the algorithm was done by Breiman (2001).

k-nearest neighbours algorithm (KNN) is a non-parametric supervised learning method first developed by Evelyn Fix and Joseph Hodges in 1951 (a technical report, which was never published) which is used for classification. A commentary for the paper was later published (B.W. Silverman & M. C. Jones, 1989). For knn classification, the output is a class membership which is determined by a plurality vote of its neighbours. k is usually to be determined (usually, $k = \sqrt{n}$)

Training and Assessment of Models

The Decision Tree, Random Forest and KNN models were implemented using the Sklearn package, whereas boosting was done with the XGBoost package. The performance of the models were evaluated using accuracy measures from the Sklearn package (sklearn.metrics.accuracy_score) and if the performance is good, a Receiver Operating Characteristic (ROC) curve will be further used to benchmark the models against each other using the area under the graph. F1 scores will also be used to double check the accuracy of the models.

Based on our EDA, the models will be run with the following variables: mileage, engine hours and phase. For all models, the classifier was used to predict the final result – a pass or fail on inspection. For model accuracy, the model with the highest accuracy (obtained from training) was selected. Whereas for precision, it analysed the reliability in reproducing the same accuracy with 1000 iterations of the model with the same parameters. This will be expressed as a "reproducibility" percentage in the results.

Preliminary observations suggest that anything below 82.9% accuracy is equivalent to ineffective as it means that the model can only achieve the baseline for the pass/fail probability through assuming that all land platforms pass. Ideally, the optimal model is one that can predict pass or fail both accurately and precisely.

PERFORMANCE EVALUATION OF MODELS

As part of the performance evaluation, there will be two main indicators: accuracy and reproducibility. The datasets are split into training (70%) and testing (30%) to mitigate overfitting since the small dataset is at a higher risk of overfitting occurring during the training phase.

Dataset as a Whole

When age is included as a feature inputted into the models, all four models showed significantly lower accuracy rates. As such, the finalised models do not consider age and only have the following features: “mileage”, “engine hours” and “phase”.

Table 2

Models Effectiveness Ranked Treating Dataset as a Whole

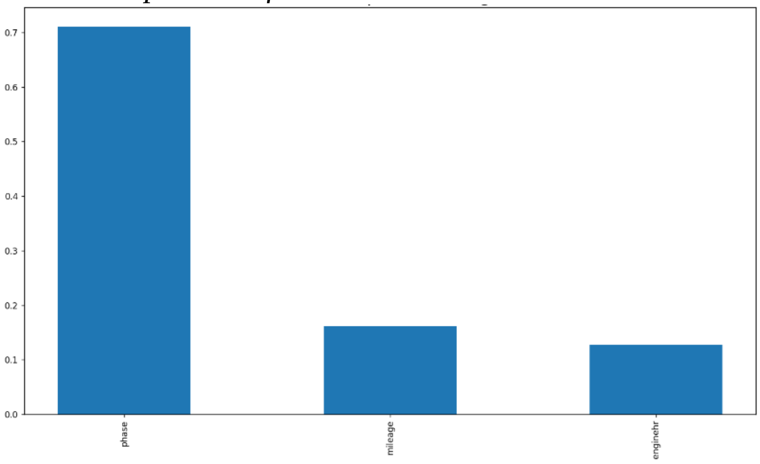
Rank	Model	Accuracy (%)	Reproducibility (%)
1	K-Nearest Neighbours	97.44	100.00
2	Boosting	92.31	100.00
3	Random Forest	94.87	79.50
4	Decision Tree	92.31	14.80

From Table 2 (no adjustment of hyperparameters), it can be concluded that knn is the most reliable model for the whole dataset. The same accuracy (97.44%) can be obtained on xgb while retaining the reproducibility if the *scale_pos_weight* parameter is modified (to 15000).

From the feature importance graphs (Figure 8 – 10), it can be observed that the different models have different feature importance yet all can reach >90% accuracy. This tells us that all features are important in determining whether a failure of the land system will occur. Initially in the EDA stage, there was no visible correlation for “mileage” and “engine hour”, yet it is visible here that there is some pattern, albeit not visible to the human eye.

Figure 8

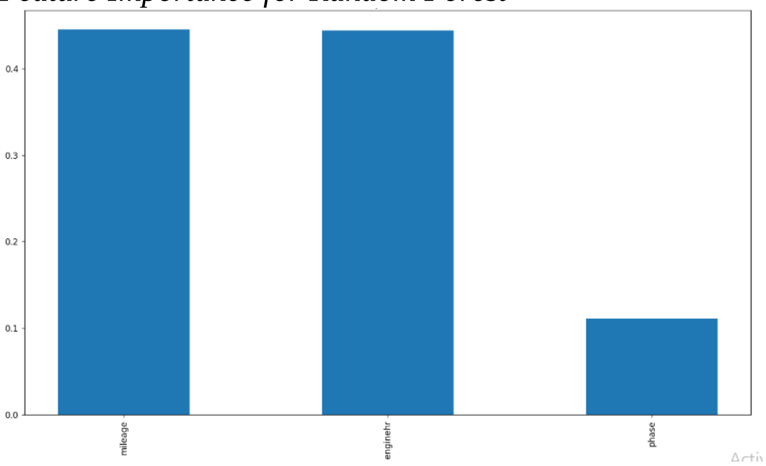
Feature Importance for XGBoost



Note. From left to right: phase, mileage, enginehr

Figure 9

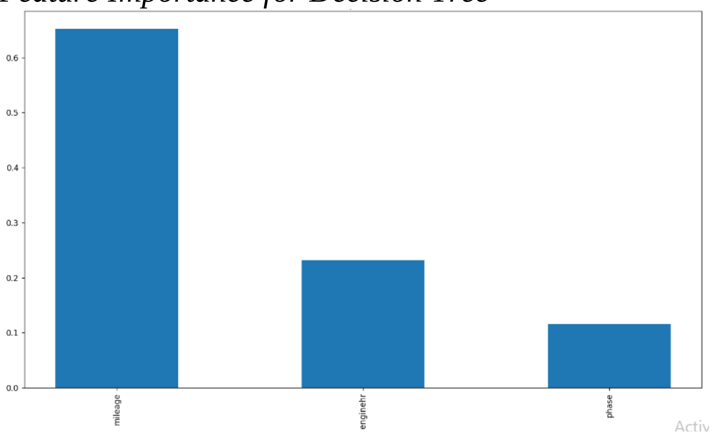
Feature Importance for Random Forest



Note. From left to right: mileage, enginehr, phase

Figure 10

Feature Importance for Decision Tree



Note. From left to right: mileage, enginehr, phase

Further Feature Engineering: Dataset as its Parts

As there was a notable disparity in failure rate between the different phases, as seen from Figure 6, the dataset was spliced into its different phases to train the machine learning model on each one of them separately in an attempt to improve accuracy score.

Table 3

Most Effective Models for Each Phase

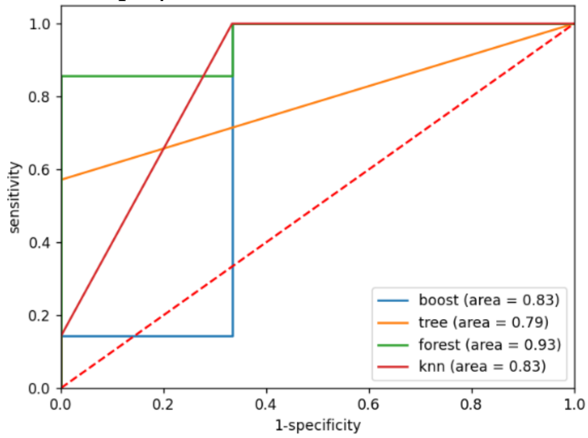
Phase	Model	Accuracy (%)	Reproducibility (%)
1	K-Nearest Neighbours	99.92	100.00
2	Random Forest	89.75	100.00
3	K-Nearest Neighbours	99.96	100.00

From Table 3, it is observable for Phase 1 and 3, knn is the most effective model whereas for Phase 2, random forest is the most effective model. Random forest shows itself to be rather effective in Phase 1 and 3 as well, compared to the other models (decision tree and boosting).

Phase 2 is particularly interesting and as such the ROC graph is as follows (Figure 11). Random forest indeed does its job (as a model that is designed to counter overfitting) and is not overfitted to the training dataset.

Figure 11

ROC Graph for Phase 2



Results

Treating the dataset as a whole, a 97.44% accuracy can be obtained, which is extremely desirable for the small dataset that was available. Treating it as three separate parts based on the phases, a $(99.92 \times 36 + 89.75 \times 31 + 99.96 \times 62) \div 129 = 97.50\%$ accuracy can be obtained with the F1 score being 0.95. Overall, the combination of EDA, Feature Engineering and Machine Learning techniques resulted in an additional 17.6% increase in efficiency over the baseline of 82.9%.

CONCLUSION AND RECOMMENDATIONS

Real-Life Application

Table 4

Averaged Confusion Matrix of Model for Dataset as a Whole

Predicted	Actual	
	Pass	**Fail**
Pass	32.34	0.98
Fail	0.00	5.68

The initial method is to send in all Phase 2 vehicles for repair and deal with the additional cost of checking 19 vehicles (the passes in phase 2) which are actually not broken down. The model is accurate enough and there are no False Passes, which means no vehicles were sent for maintenance when they do not actually need it (maintenance cost avoidance). As shown in the confusion matrix in Table 4, there is close to 1 False Fail, where it predicted one vehicle would pass when it actually would fail. While this error in prediction would incur unexpected downtime, this is a relatively low percentage (2.5% - 1 out of 40) and furthermore does not incur unnecessary maintenance costs. Trading-off the low False Fail and the ability to identify 5.68 vehicles (True Fail) as fails when they would have failed, would allow targeted allocation of resources to maintain the vehicle, saving maintenance cost and effort while reducing the downtime of each vehicle.

In military operations, the percentage availability of vehicles has logistical importance and affects operations and training success. As such, with this dataset, the machine learning models are able to help raise the average availability of the land system with the increased accuracy of 17.6%. This can have significant benefits on resource optimisation and military operations.

Recommendations and Future Work

With an expanded dataset in the future, especially with more sensors and vehicle health utilisation and monitoring systems to collect telemetry data, deep learning models could be considered to process the dataset. The best performing methods should still be considered as they should still produce >95% accuracy. The fairly robust model generated from the relatively small dataset shows promise that this same method can be applied to breakdown and maintenance of different land platforms. Possible extensions include extending the use of these predictive models to other systems, for example naval and air platforms to also help reduce costs in those areas.

ACKNOWLEDGEMENTS

I would like to thank my mentors, Mr Cheong Yek Meng and Mr Ong Sir Hong for their support and technical guidance for the entire duration of this project. Without their guidance and invaluable mentorship, this project would not have been possible.

Additionally, I would also like to thank Ms Esther Chew for her assistance in the administrative and HR portion of this project, as well as DSTA for providing me the opportunity to participate in the 2023 iteration of Research@YDSP.

REFERENCES

[1] Sezer, E., Romero, D., Guedea, F., Macchi, M., Emmanouilidis, C. 2018. An Industry 4.0-Enabled Low Cost Predictive Maintenance Approach for SMEs. *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, Stuttgart, Germany (pp. 1-8) IEEE.
https://doi.org/10.1109/ICE.2018.8436307

[2] Mallouk, I., Sallez, Y., El Majd, B.A. 2021. Machine learning approach for predictive maintenance of transport systems. *2021 Third International Conference on Transportation and Smart Technologies (TST)*, Tangier, Morocco (pp. 96-100) IEEE.
https://doi.org/10.1109/TST52996.2021.00023

[3] Breiman, L., Friedman, J.H., Olshen, R.A., Stone, J. 1984. Classification and Regression Trees. In: *Biometrics Vol.40, No.3* (p. 874). International Biometric Society.
https://doi.org/10.2307/2530946

[4] Kearns, M. 1988. Thoughts on Hypothesis Boosting [Unpublished manuscript (Machine Learning class project)].

[5] Kearns, M., Valiant, L.G. 1989. Cryptographic limitations on learning Boolean formulae and finite automata. *STOC'89: Proceedings of the twenty-first annual ACM symposium on Theory of computing Vol.41* (pp. 433-444) ACM.
https://doi.org/10.1145/73007.73049

[6] Schapire, R.E. 1990. The strength of weak learnability. In: *Machine Learning Vol.5* (pp. 197-227).
https://doi.org/10.1007/BF00116037

[7] Ho, T.K. 1995. Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition Vol.1* (pp. 278-282)
https://doi.org/10.1109/ICDAR.1995.598994

[8] Breiman, L. 2001. Random Forests. In: *Machine Learning Vol.45* (pp. 5-32)
https://doi.org/10.1023/A:1010933404324

[9] Silverman, B.W., Jones, M.C. 1989. E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951). In: *International Statistical Review Vol.57 No.3* (pp. 233-238) International Statistical Institute.
https://doi.org/10.2307/1403796

APPENDIX

Figure 3

Number of Pass and Fails per Variant

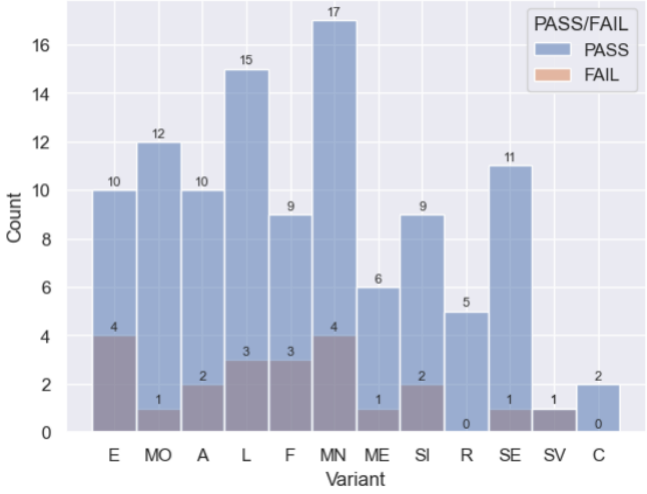


Figure 4

Number of Pass and Fails per Mileage range

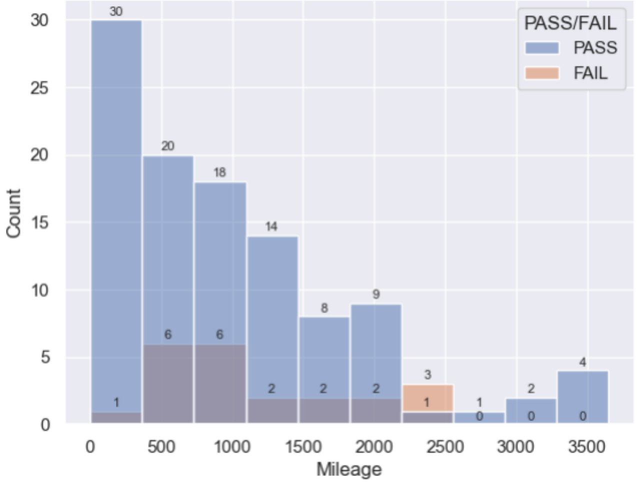


Figure 5

Number of Pass and Fails per Engine Hours range

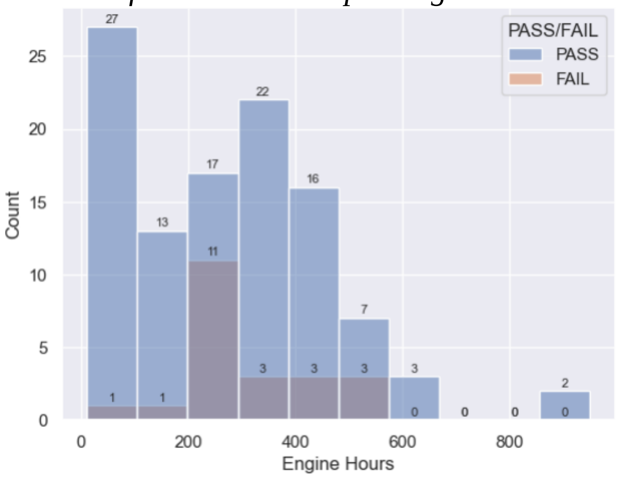


Figure 6

Number of Pass and Fails per Phase

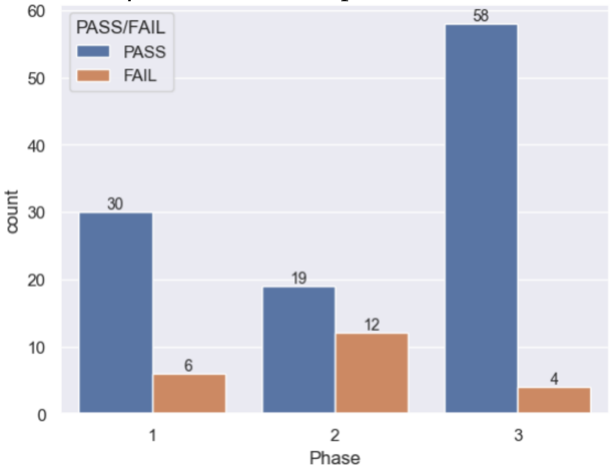


Figure 7

Number of Pass and Fails per Age