XXVI ENMC
ENCONTRO NACIONAL DE MODELAGEM COMPUTACIONAL
XIV ECTM
ENCONTRO DE CIÊNCIA E TECNOLOGIA DE MATERIAIS
2023

# INFLUENCE OF DATA FILTERS ON MACHINE LEARNING FOR PREDICTING AIRCRAFT COMPONENT FAILURES

**Henrique Mendes Castilho**[1,2] - henrique.castilho@embraer.com.br
**Wallace Hessler Leal Turcio**[1] - wturcio@embaer.com.br
**Luis Carlos de Castro Santos**[1,3] - luis.castro@embraer.com.br
**Rafael Duarte Coelho dos Santos**[2] - rafael.santos@inpe.br
[1]Embraer S.A. - São José dos Campos, SP, Brazil
[2]Instituto Nacional de Pesquisas Espaciais - São José dos Campos, SP, Brazil
[3]MAP-IME-USP - São Paulo, SP, Brazil

***Abstract.*** *Effective maintenance of aircraft systems and components is an important factor for economical operation in the aviation industry. This study explores the impact of data filtering on the performance of machine learning algorithms. The methodology involves the application of feature engineering concepts to derive information from sensors installed in the aircraft and maintenance reports. The effectiveness of each filter is assessed by comparing the performance of a classification tree for each filter. The case study focuses on the pressure control valve of the air management system of an aircraft. This paper addresses challenges that arise when working with field operational data in predicting component failures.*

***Keywords:*** *Prognostic, Exploratory, Failure analysis, Machine learning, Filtering*

## 1. INTRODUCTION

Equipment and component maintenance represents a challenge. For the aviation industry, costs include material, personnel and aircraft downtime, as covered by Pascoal (2013). Aircraft and equipment manufacturers propose guidelines for maintenance, considering an average time when failures might occur.

These recommendations are based on the expected life span, which is estimated based on experiments, specialist knowledge and historical data. Scheduled maintenance actions might occur before a given component reaches its end-of-life, meaning these components would still be able to operate. Unscheduled maintenance actions incur high costs, since there is an unexpected downtime in the aircraft.

A solution is to predict failures and take actions when there is a large probability of potential failure occurring soon. This field of study in named Prognostics and Health Management (PHM), described in Vachtsevanos et al. (2006). One of the goals of PHM is to assess the

health state of a component. It is possible to determine whether a component is degraded using flight data and maintenance records as seen in Moreira (2012).

Various techniques can be used to estimate the failure date, from traditional statistical analysis to machine learning algorithms. The source of the data used in this work are sensors installed in the aircraft and maintenance reports.

However, these data sources tend to have particular challenges. Sensor data may show some noise and also inherent variations from the influence of different operating conditions. Maintenance reports are usually the result of an inspection by a mechanic, bringing to the process variations associated with human factors.

Another major challenge in the PHM field is to define a degradation indicator for a complex system such as air conditioning control, as studied by Graves et al. (2018). Since there are several performance parameters, a technique is needed to combine them into a single health indicator rather than having one indicator for each system component.

The main objective of this work is to evaluate classical filtering techniques applied to the problem of interest and provide some insight for similar application. The original data is very noisy and subjected to operational decisions. This makes it difficult to verify if there is a trend that matches the events that must be predicted. Therefore, if there is a filtering technique that is able to reveal a trend is the data, it will make better features for a machine learning algorithm.

First, a selection of filters are applied to the data, to each feature. Then, a new set of inputs is created using each of the filters. Using each set, some machine learning algorithms are evaluated for each set of inputs. Finally, the analysis of the result of these algorithms will reveal if there is a filtering technique that is better suited for this class of problem.

Using the sensor data paired with the maintenance reports, the aim is to predict the state of degradation of the analyzed component, in order to predict possible replacements.

## 2. METHODOLOGY

Figure 1 illustrates a typical bleed system architecture and valve installation. Air is taken from the engine, an action called “bleeding air”, whose high pressure is reduced and controlled by the PRSOV before being supplied to the clients.

The valve is installed in the aircraft’s engine nacelle, exposed to conditions dominated by the engine’s temperature and subjected to a high level of vibration, factors that contribute to a rapid degradation. It is controlled by a closed-loop control system that keeps the pressure at the reference point. As the component ages, a more severe action by the controller is required, and the pressure becomes more oscillatory.

The sensor data from the aircraft is measured during the entire flight, but presents some noise and variations associated with operation particularities and constraints. Feature engineering aims to reduce the sensor data to a single point that represents the flight for each feature, as presented in França et al. (2022) and Shah et al. (2020). The features range from statistical functions, such as mean and standard deviation to derivatives and overshoots for various sensors. As the information used in the work is proprietary, it had to be mischaracterized to meet Intellectual Property standards, however its essence and content were preserved.

The cruise flight phase provides benefit in describing the component’s degradation, because the pressure is the most stable and the demand is continuous. Figure 2 illustrates the valves’ controlled pressure for two different flights, one with a low degradation and the other with a a high degradation of the PRSOV. These flights are for the same aircraft and valve, but are spaced

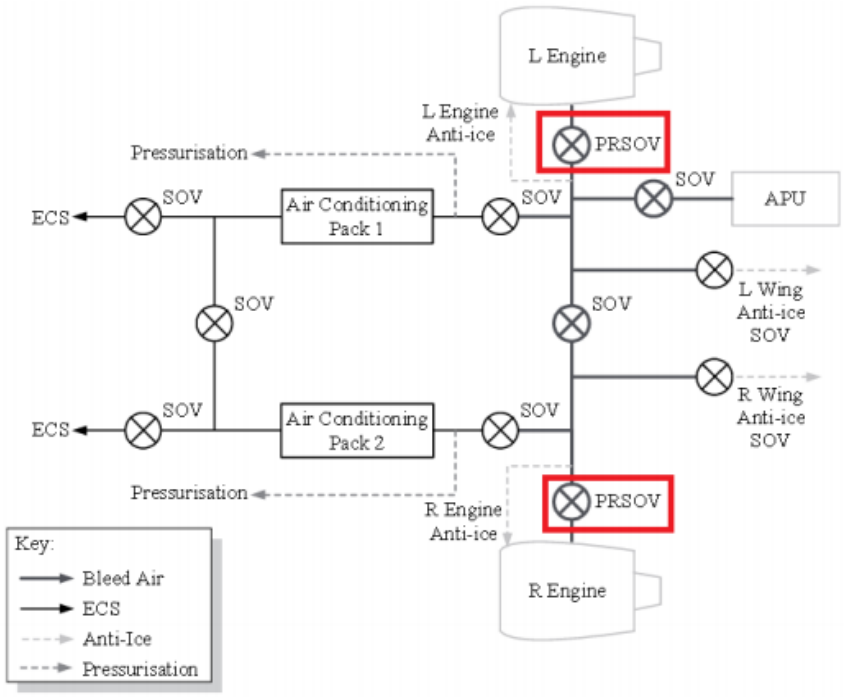


Figure 1: Notional diagram of a typical air bleed system

months apart. The level of degradation could indicate a maintenance action.

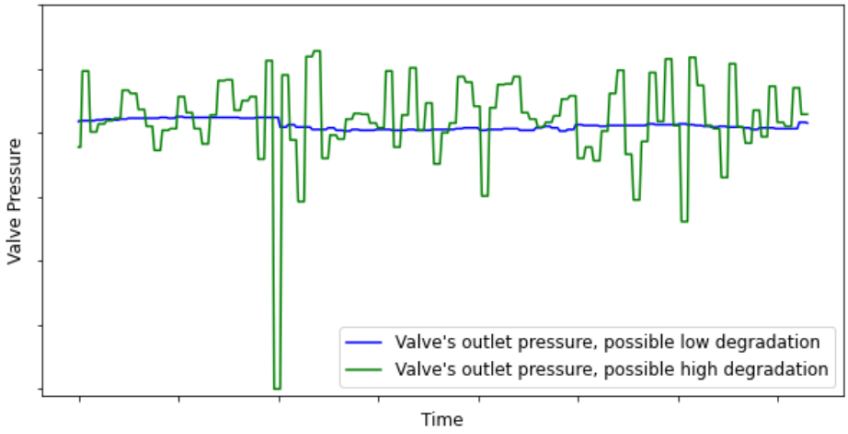


Figure 2: Data during flight - comparison of the degradation

The flights considered “far from the failure event” are that flight that occurred soon after the failed valve replacement, meaning the valve is new. The other flight happened months later and presents a more oscillatory pressure, indicating degradation. A feature can be constructed by calculating the standard deviation of the pressure, since as it increases, the degradation may also be rising.

The goal is to create a database were each flight is an observation with a number of features that describe the equipment’s health. The database is composed of different aircraft, each with numerous flights and is shown in table 1.

Table 1: Feature database

| Aircraft and Flight | Feature 1 | Feature 2 | Feature 5 | Feature 9 | Feature 10 |
|---|---|---|---|---|---|
| Aircraft 1 Flight 1 | 37.875 | 34.099 | 0.942 | 1.181 | 5.170 |
| Aircraft 1 Flight 2 | 35.375 | 33.956 | 1.062 | 1.645 | 6.274 |
| ... | ... | ... | ... | ... | ... |
| Aircraft 98 Flight 1999 | 40.750 | 34.492 | 5.267 | 6.715 | 7.168 |
| Aircraft 98 Flight 2000 | 38.125 | 41.614 | 5.239 | 4.716 | 11.583 |

Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods. Its objectives are defined in Tukey (1977). EDA can also help assess the need for filters and trending analysis.

Figure 3 illustrates the graph of Feature 5 for one aircraft along time. The vertical lines are replacements. The behavior changes as time progresses and a valve replacement resets the trend. For the feature illustrated the value rises as an event approaches, but not all events behave the same way. This highlights the non-linear aspect of the dataset.

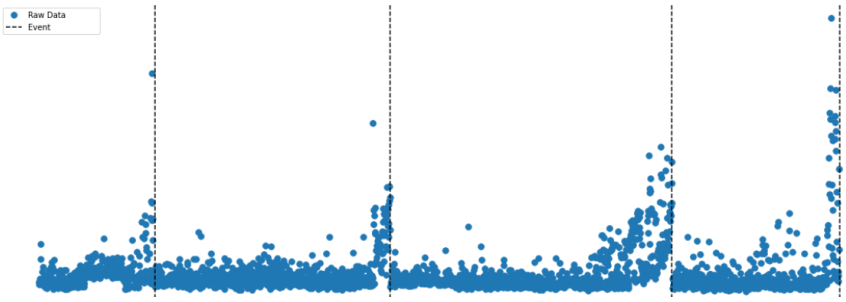


Figure 3: Feature 5 raw data for a single aircraft

The degradation score aims to link the raw data to the degradation, as illustrated by fig. 4, where a single event maintenance event for a single aircraft is represented. The raw data is also Feature 5. The limit was set by the analysis of the same feature for all replacement events for all aircraft. For flights far from the event, most of the points reside under the limit. As the event nears, the points tend to be above the limit.

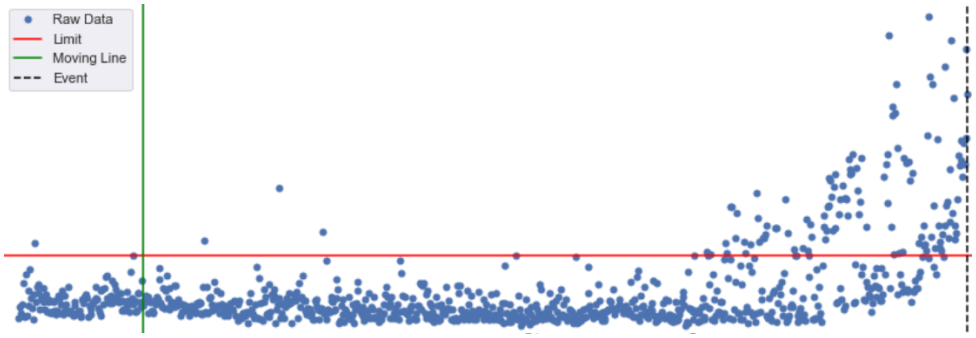


Figure 4: Degradation Score Criteria

The score is defined as the percentage of the number of points above the limit on the right side of the moving vertical line, illustrated by fig. 5. The increase as the event approaches is

because there are more points above the limit relative to the total as the moving line travels to the right.

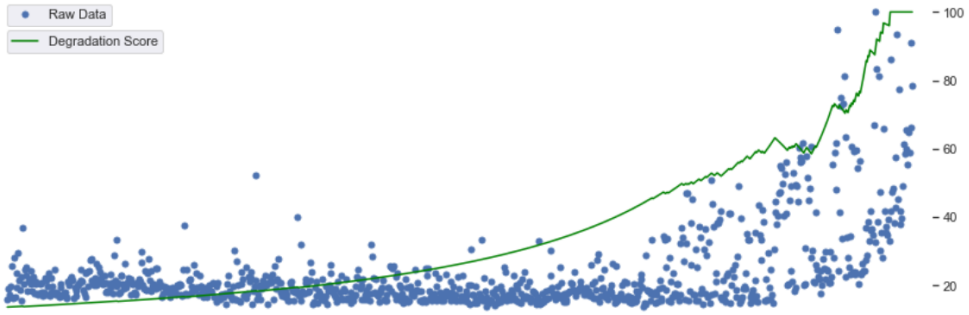


Figure 5: Degradation Index - Single Event

A threshold is chosen for the *degradation score*, where a new indicator is considered *False* when less than it and *True* when greater or equal, creating a binary output called *degradation index*. Figure 6 illustrates the result, which is then calculated for all events of all aircraft and will be used as the classification target.

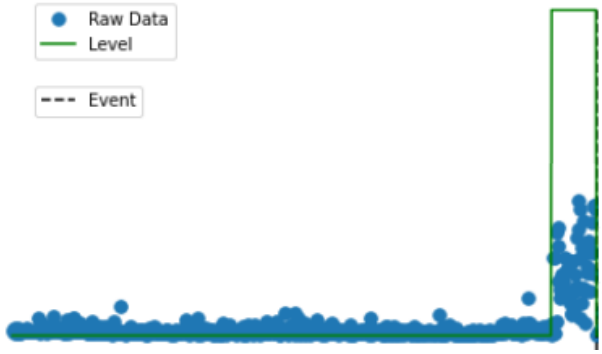


Figure 6: Degradation Index

The next step is selecting the most valuable features. The *degradation index* was constructed based on a feature that correlates very well with the events, but a more thorough analysis must be conducted in order to select the features that will compose the data for the machine learning algorithms. A possible method is a bivariate analysis performed with respect to the target class. Figure 7 illustrates the boxplot for two different features. The distributions are similar for Feature 6, which may indicate that there is no difference in behavior for the target classes for this feature. There is a clear difference for Feature 5, which means that it may be helpful in identifying the target class.

## 2.1 Filtering

Filtering is applied to extract a trend and reduce the noise in the data. The comparison of the results for all filters is the focus of this work, with the aim to find a filter that produces good results for this problem class. The chosen filters were:

- Moving average/maximum/median
- Exponentially weighted average
- Kalman filter

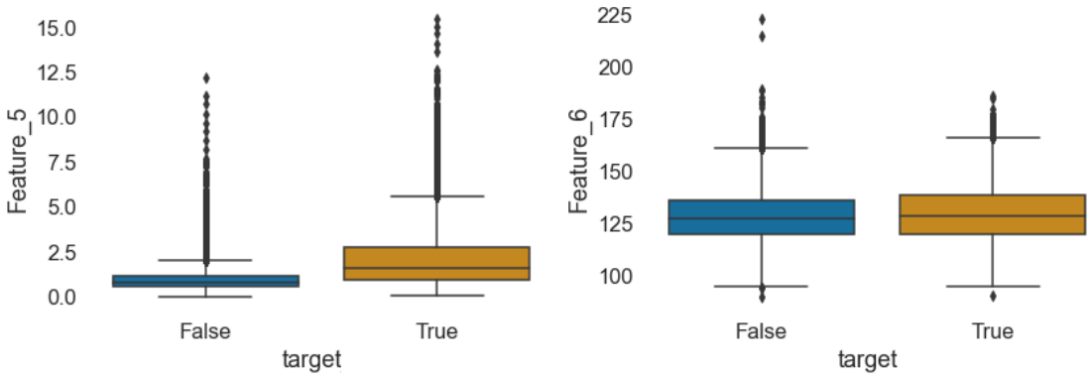


Figure 7: Boxplot of the raw data - Features 5 and 6

The filters are presented in fig. 8 for a single event.

The moving average is a basic filter. It can be observed in fig. 8 that as the event approaches, the maximum of the raw data rises, so the moving maximum was chosen. The median was selected because the data does not have a symmetrical distribution.

The exponentially weighted average presented good results in detecting process variability as studied by Huwang et al. (2009). The Kalman filter proved useful in filtering sensor data for remaining useful life estimation as seen in Baptista (2018). These filters were also chosen to verify whether they are necessary or a simpler approach is enough.

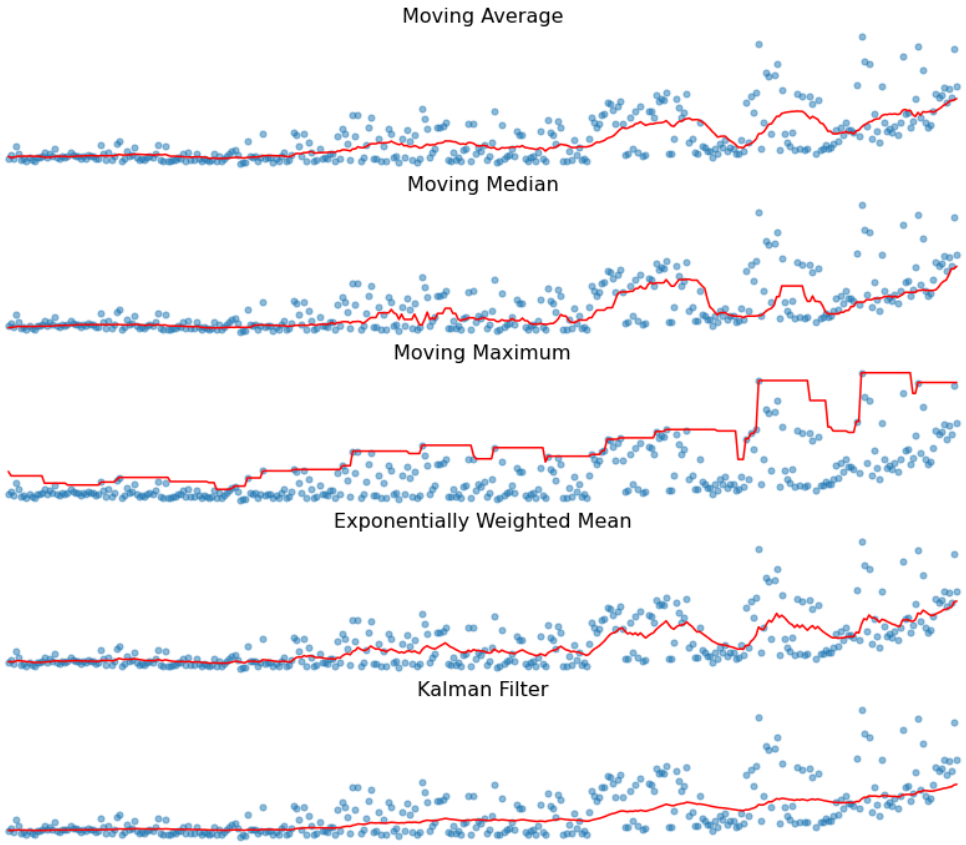


Figure 8: Effect of each filter for the same set of raw data

The moving mean and median are still noisy, given the variability of the physical process.

Moving maximum appears to have captured the trend better. Exponential weighted average presents some noise. Kalman filter results in a nicer increasing line, but not as pronounced as the raw data itself. All filters have parameters whose tuning might affect the results, however they were chosen to be compatible. For the moving window filters, the same size was chosen for all.

A single feature was chosen for an exploratory analysis on the filtered data, Feature 5, because it presents a greater difference in each target class. Bivariate boxplots of the filtered data are presented in figure 9. Each boxplot is represents the same feature, with the only difference being the application of the chosen filters.

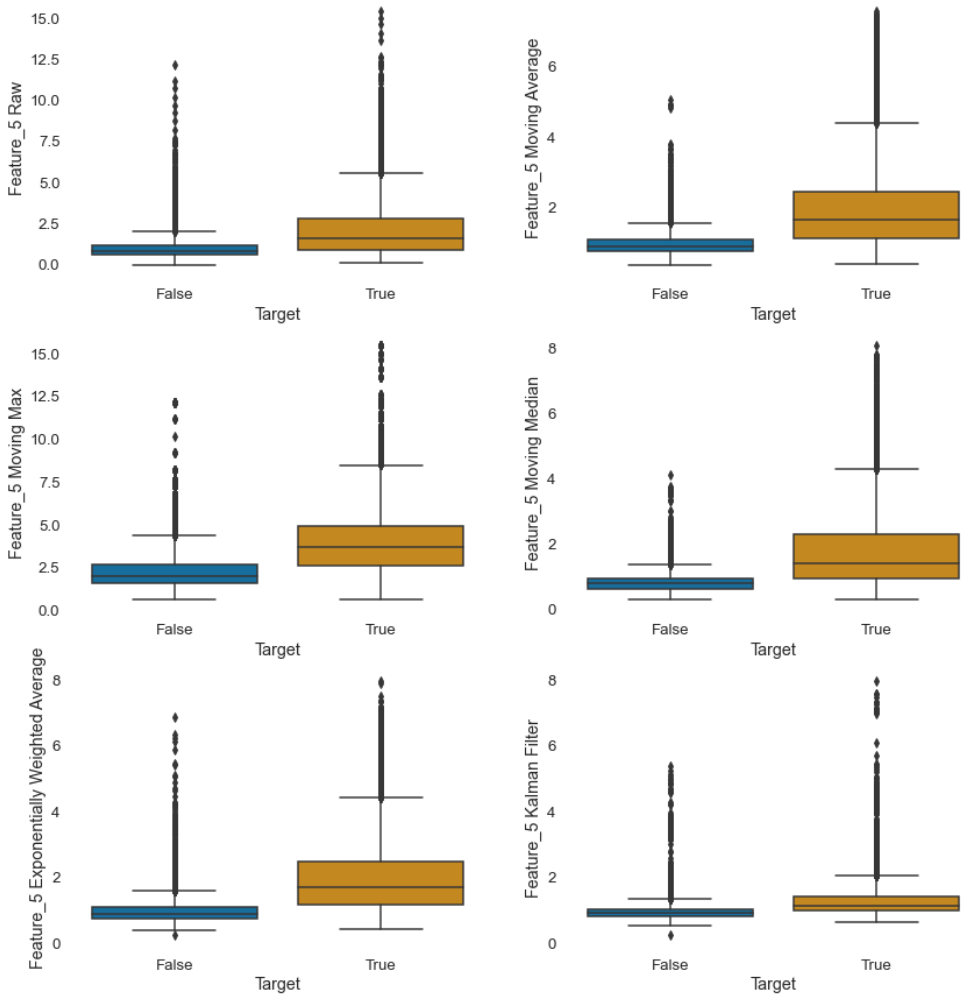


Figure 9: Boxplot of the filtered data

For the moving average, the first quartile for the *True* class begins after the third quartile ends for the *False* class, indicating that the overlap for this particular feature is decreasing. This means that the separation for the classes increases with the application of the moving average, indicating a possible improvement over the use of the raw data. Moving maximum presents the same trend, but the data for the *True* class is less skewed, as the central line of the boxplot indicates. The moving median and exponentially weighted average also are similar to the moving mean, which may indicate that the simple filter may be enough for the case. Kalman filter results in a stricter interquartile range, which may result in some misclassifications.

## 2.2 Machine Learning

The raw and filtered data, consisting of features that represent the flight, was then separated into training and testing with a target ratio of $75\%$ and $25\%$ respectively. The division must be at aircraft level, not on flight level, since the flights are sequential.

Since the focus is to evaluate the filters, a classification tree was configured with the Gini impurity criteria for the classification, which is the classic method for choosing a class when splitting nodes in a classification tree, no limits for the depth and no weights, which means that the process will not stop early and there are no preferences for any class. The only consideration was the minimal cost-complexity pruning in order to avoid an overfitted tree. Such concepts are discussed more in-depth in Breiman et al. (1984).

An analysis was conducted that returns the effective complexity parameter and the impurities at each step of the pruning process. An ideal value for the complexity must be chosen so that the tree is a good generalizer but also presents an acceptable error rate when deployed for use in test data. This process is illustrated by fig. 10, which shows that in order to avoid overfitting and increase the performance for the test set, a point that both minimizes the training accuracy and maximizes the testing accuracy must be chosen. The test set was then applied to the tree and the results are analyzed in the following section.

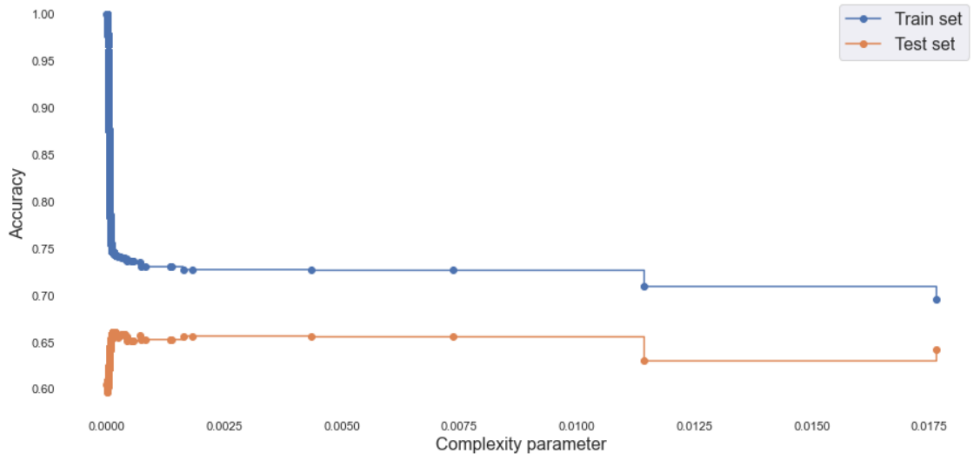


Figure 10: Minimal Cost-Complexity Pruning

## 3. Results and Discussion

Some important definitions are necessary for the evaluation of the performance of the filtering techniques considered. A confusion matrix, described in Stehman (1997), is a table that displays the results of a classification. The matrix represents how the output of the machine learning algorithm fares against the true class.

The classes *True* and *False* correlate to *High* and *Low* degradation states. The confusion matrix of table 2 reveals the basic performance of the classification tree for the raw data.

The figures in the confusion matrix represent the number of actual observations versus the predicted ones. The main diagonal represents the true positives and negatives. The opposite diagonal represents the false positives (type I error) and false negatives (type II error) Dekking (2005).

Table 2: Confusion Matrix

| TRAINING SET | | | | TEST SET | | | |
|---|---|---|---|---|---|---|---|
| True Label | FALSE | 16324 | 3531 | True Label | FALSE | 4837 | 1188 |
| | TRUE | 8103 | 14577 | | TRUE | 3142 | 3400 |
| | | FALSE | TRUE | | | FALSE | TRUE |
| | | Predicted Label | | | | Predicted Label | |

The final results will be indicators derived from the matrix that determine of the performance of the predictor as explained by Powers (2007).

- True Positive Rate: conditional probability of positive result if the data is truly positive.
- False Positive Rate: conditional probability of positive result if the data is truly negative.
- Precision: proportion of positive results that are truly positive.
- Accuracy: fraction of predictions that are correct.

In the case of failure prediction, a small amount of false positives is desired, since if the predictor labels a truly high degradation state as low, an unexpected failure might occur. False negatives mean that the operation may have been wrongfully halted. Table 3 presents the results for all filters.

Table 3: Results for the Classification Tree

| Index | Raw | Mov. Avg. | Mov. Max | Mov. Med. | EWM | Kalman |
|---|---|---|---|---|---|---|
| True Positive Rate | 0.744 | 0.894 | 0.898 | 0.887 | 0.905 | 0.876 |
| False Positive Rate | 0.174 | 0.247 | 0.341 | 0.216 | 0.252 | 0.242 |
| Precision | 0.721 | 0.687 | 0.615 | 0.713 | 0.686 | 0.687 |
| Accuracy | 0.795 | 0.806 | 0.749 | 0.823 | 0.807 | 0.802 |

For the case studied, the best scenario is a filter that boasts high precision and high true positive rate. However, this does not mean the others metrics should be ignored. The filter with the highest precision is the raw data itself, and the highest true positive rate is the moving average. It is desirable to capture the state of high degradation reliably, so a high true positive rate is better. Thus, the moving average was chosen as the most adequate filter for this case.

The lowest false positive rate was for the raw data, even if the application of any filter increases it, the false negatives are reduced, a desirable trait for predictions of this problem class, where unnoticed states of high degradation might be more impactful than the labeling of low degradation as high. Finally, the Kalman filter did not present a result significantly better than the moving average, which could point to the fact that a simpler approach is enough for this type of analysis.

## 4. CONCLUSIONS

This paper has presented an analysis of the impact of data filtering on machine learning classification for aircraft equipment failure, using a case study of the Pressure Regulating Shut-Off Valve of the bleed air system. An exploratory data analysis revealed trends in the observed component degradation over time and identified important variables that indicative of the state of health of the valve.

Oscillations present in the data were reduced by the successful application of various data filtering techniques. These actions improved the machine learning performance. The conclusion drawn from analyzing the performance of the filters is that the simple moving average was able to out-perform the Kalman filter, a complex technique very common among those looking to filter noisy data. This means that simple techniques should not be discarded, and may be better suited from a given class of analysis.

The research takeaway brings valuable insights for the industry, in particular where PHM would be helpful. The conclusions contribute to the development of better machine learning predictions, allowing for the improvement of maintenance strategies. Furthermore, this work shows the importance of using field data, even with all the issues that arise from it.

Future work suggests the exploration of additional filters or a deep dive in their parametrization. The study of others machine learning techniques more adequate for each filtering technique may prove fruitful.

***Acknowledgements***

We thank Embraer for kindly providing the data and supporting this study and INPE for the opportunity to develop this research as part of the doctoral program.

## REFERENCES

M. L. Baptista. *Machine Learning and Deep Learning for Prognostics and Predictive Maintenance of Aeronautical Equipment*. PhD thesis, Universidade de Lisboa - Instituto Superior Técnico, 2018.

L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen. *Classification and Regression Trees*. Taylor and Francis, 1984. ISBN 9780412048418.

F.M. Dekking. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer, 2005.

Thayna França, Arthur Martins Barbosa Braga, and Helon Vicente Hultmann Ayala. Feature engineering to cope with noisy data in sparse identification. *Expert Systems with Applications*, 188:115995, 2022.

Julio Cesar Graves, Wallace Hessler Leal Turcio, and Takashi Yoneyama. Degradation analysis of an aeronautical pneumatic actuator using hysteresis-based signatures. *Journal of Control, Automation and Electrical Systems*, 29(4):451–459, 2018.

Longcheen Huwang, Yi-Hua Tina Wang, Arthur B. Yeh, and Ze-Shiang Jason Chen. On the exponentially weighted moving variance. *Naval Research Logistics*, 56(7):659–668, 2009.

R. P. Moreira. Prognóstico de sistemas aeronáuticos utilizando o algoritmo svm treinado com dados de voo e registros de manutenção. Master's thesis, Instituto Técnológico de Aeronáutica, São José dos Campos, 2012.

Renata Monteiro Pascoal. Estimação de parâmetros de falha de apu empregando regressão linear e redes neurais artificiais. Master's thesis, Instituto Tecnológico de Aeronáutica, 2013.

David M. W. Powers. Evaluation: From precision, recall and f-factor to roc, informedness, markedness and correlation. Technical report, Flinders University of South Australia, 2007.

Devarshi Shah, Jin Wang, and Peter He. Feature engineering in big data analytics for iot-enabled smart manufacturing - comparison between deep learning and statistical learning. *Computers and chemical engineering*, 141:106970, 2020.

Stephen V. Stehman. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62(1):77–89, 1997.

J.W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.

G. Vachtsevanos, F.L. Lewis, M. Roemer, A. Hess, and B. Wu. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Wiley, 2006.