FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

PROFIT-DRIVEN ANALYTICS IN A MAINTENANCE SETTING: THE PRMC

Aantal woorden / Word count: 11,667

Stamnummer / student number: 02109952

Promotor / supervisor: Prof. Dr. Matthias Bogaert
Co-promotor / co-supervisor: Drs. Shimanto Rahman

Masterproef voorgedragen tot het bekomen van de graad van:
Master's Dissertation submitted to obtain the degree of:

Master in Business Engineering: Data Analytics

Academiejaar / Academic year: 2022-2023

UNIVERSITEIT
GENT

Confidentiality clause and scientific integrity

The author declares that the research was conducted in accordance with the rules governing scientific and academic integrity. He has read, and acted in accordance with, the Code of Ethics of the Faculty.

Preface

Writing this dissertation was a challenging and interesting process. I am grateful for the opportunity that presented itself as this specific topic, as this topic allowed for a creative contribution that can be applied to many industries.

I thank my co-supervisor Drs. Shimanto Rahman for his flexibility and reflections on all aspects of the matter and my supervisor Prof. Dr. Matthias Bogaert for his additions on cost-sensitive learning which were implemented in code. This dissertation is the final work which concludes my pursuit of a Master's degree. Therefore, I would like to express my gratitude for the support of my friends and family. Not only during this final task, but rather all of my studies. However, this is not where my academic journey ends, as from next year onwards, I will be taking on a research position in the domain of Operations Research under the guidance of Prof. Dr. Maenhout and Prof. Dr. Eeckloo. I am looking forward to this opportunity.

Finally, I am convinced that this dissertation contributed to my general knowledge on many different matters. I personally believe that only when this work impacts the knowledge of others, its contribution can be relevant. Therefore, I hope that this work reaches many interested minds in the fields of profit-driven analytics or predictive maintenance.

Author: Samuel Bakker

Abstract

This dissertation proposes a new cost-based metric that is created for the field of maintenance prediction specifically. Whereas other approaches in this field only account for the direct observed costs of conducting maintenance, the metric in this dissertation (the PRMC) also accounts for the amount of lifetime that is lost because of non-reactive maintenance. When maintenance is conducted too early, too much of a machine's remaining useful life is lost. This is an indirect cost, that is accounted for in the PRMC. Two extension on the PRMC are also provided in this work. One is the so-called 'expected PRMC', which adds a stochastic meaning to the PRMC. The other is the 'iPRMC', which is to be used during inference.

A case study is provided in which models were trained on the CMAPSS-dataset. Each model was tuned with the MAE, MSE and the proposed PRMC. This case study shows that using the PRMC to tune a model leads to significantly lower costs as opposed to using the MSE. However, when comparing the PRMC and the MAE for the same task, no significant difference was found.

List of Tables

List of Figures

Glossary

C_p Cost of Predictive Maintenance.

C_r Cost of Reactive Maintenance.

CMAPSS Commercial Modular Aero-Propulsion System Simulation.

EMPC Expected Maximum Profit Criterion.

iPRMC Pure RUL Maintenance Cost during inference.

MAE Mean Absolute Error.

MCR Maintenance Cost Rate.

MP Maximum Profit.

MSE Mean Squared Error.

PDA Profit-driven analytics.

PdM Predictive Maintenance.

PM Preventive Maintenance.

PRMC Pure RUL Maintenance Cost.

R-EMP Robust Expected Maximum Profit.

RM Reactive Maintenance.

RUL Remaining useful life.

UL Useful Life.

Contents

1 Introduction

Production organizations regularly need to deal with maintenance of their equipment. Unfortunately, maintenance costs can be substantial. There are significant costs associated with maintenance (Mobley, 2002), which drive businesses to make intelligent decisions concerning maintenance policies.

Industry 4.0 is characterized by the growing amount of data, and the capabilities to transform this data into something useful. Leveraging these large volumes of data picked up by sensors of machinery in order to reduce costs is a critical part of becoming cost-effective. There are many different types of costs influenced by (not) performing maintenance. Examples are; the cost of maintenance itself, the cost of equipment downtime during maintenance, and even the degradation of product quality due to a lack of maintenance.

In the literature, there are many different ways in which maintenance is classified. For this dissertation, the author chose to make a clear distinction between three types; reactive, preventive, and predictive maintenance. This dissertation will focus itself on predictive maintenance. Specifically, the objective of this dissertation is to come up with a cost-based metric that can be used to build better predictive maintenance models.

2 Orientation

2.1 Types of maintenance

In this section the different types of maintenance will be introduced along with some of their advantages and disadvantages.

2.1.1 Reactive Maintenance (RM)

Reactive maintenance is definitely the least complicated type of maintenance. Often it is called 'Run-to-Failure'-maintenance. When equipment fails, it is repaired or replaced (Swanson, 2001). It is the most passive manner in which maintenance can occur. The cost of repair only occurs when a failure has happened.

A couple of disadvantages may be:

- Before a defect can be repaired, it needs to be found.
- Ripple-effects.
- Availability of spare parts.
- Higher levels of scrap products.
- Decrease in production capacity.

However, one advantage may be:

- Maximize the useful life of a component. (Zonta et al., 2020)

Driving a car until it breaks down is an example of reactive maintenance. Just as with the aforementioned car, the costs associated with repair can be higher (Swanson, 2001). When a component starts to fail, this might *ripple* to other components and damage these as well (Ran et al., 2019). Also, in order to repair the equipment, there need to be spare parts available. Because, if a machine were to fail, and the spare parts are not available (e.g. due to it being discontinued or supply chain issues), then the machine will be out of operation for a longer duration of time. For organizations that produce goods, it can be observed that, as wear increases, so does the amount of rejects (i.e. products that do not meet the quality standards) (Mobley, 2002).

The costs associated to reactive maintenance can average to be up to three times higher than those associated with preventive maintenance (Mobley, 2002).

2.1.2 Preventive Maintenance (PM)

Preventive maintenance is characterized by the regular intervals in which maintenance is being performed (Mobley, 2002; Ran et al., 2019). Maintenance occurs while the equipment is still operational, with as goal to reduce sudden downtime. The repairs that occur range from check-ups or lubrication, all the way to component replacement. The only constant in preventive maintenance is that maintenance is based on specific intervals of time (Mobley, 2002).

Advantages of Preventive Maintenance

- Less sudden downtime

- Fewer ripple-effects
- Lower Costs
- Less spare parts needed

Disadvantages of Preventive Maintenance

- No certainty whether full useful life of component was used.
- Does not keep in mind heterogeneous conditions of a component.

Since preventive maintenance can be scheduled upfront, there are less 'sudden downtimes' due to component failure. Also there are less ripple-effects throughout the equipment due to malfunctioning components (unwanted vibrations e.g.). However, there is no way to ensure that all remaining useful life was used. Furthermore, since assets may have different operating conditions on which the time of maintenance depend (humidity, temperature...), they may need maintenance at different moments in time.

2.1.3 Predictive Maintenance (PdM)

Setting up good schedules for Preventive Maintenance is rather difficult. The problem is twofold, an organization is minimizing the amount of 'sudden' failures, while also maximizing the amount of an equipment's useful lifetime used. If equipment is repaired or replaced before it has completely been used up, there is a cost associated to not using the complete lifetime of the equipment. This cost can be identified by the RUL (Remaining Useful Lifetime) and is difficult to quantify (Zonta et al., 2020).

Typically, management will try to answer the following question when setting up maintenance intervals: "When will a component *need* maintenance?". Often, the best strategy is to use existing data of the equipment's lifetime in order to estimate good time-intervals. A good example is the advertised lifetime by the manufacturer. However, treating failures as deterministic has its downsides. Due to the large amount of factors contributing to equipment wear (temperature, vibration, humidity...), failures may still occur before the planned maintenance happens. Such cases are no different from RM, as maintenance is planned after the failure of a machine.

Failures are stochastic processes and should be treated as such. Therefore, a more adequate way of dealing with maintenance is to predict it. This is what predictive maintenance (PdM) does. PdM is a data-driven approach to maintenance, and is sometimes classified as non-timebased Preventive Maintenance or Condition-Based Maintenance (CBM). However, in the current day and age of data, this classification is unjustified and more recent literature sees it as its own category of maintenance. PdM pushes the narrative of preventive maintenance even further and introduces more cost savings due to; fewer rejects, less downtime and an increase in production (Bahrin et al., 2016). There are different types of PdM, Ran et al., 2019 considers the following:

- Classification models to identify (early) failure
- Anomalous behavior detection
- RUL-prediction

Classification models to identify early failure answer the following question: "How likely is equipment to fail in the next **x** weeks?".

Anomalous behavior detection aims to find changes in vibrations, tribology or thermography of a system (Mobley, 2002; Ran et al., 2019). If an anomaly is detected, then maintenance is scheduled.

Remaining Useful Life prediction tries to estimate the RUL, the goal is to find how much longer equipment can be used until failure. If the predicted RUL ($\widehat{\text{RUL}}$) > the actual RUL, the machine will fail. If $\widehat{\text{RUL}}$ < the actual RUL, then the cost of replacing is equal to the amount of RUL that is lost. Lost RUL can be interpreted as 'lost productivity'.

2.2 Costs of maintenance

Some of the costs that are relevant for all types of maintenance are; the cost of spares, the wages of the people involved and the production that is lost because the machine is out of work.

Three important costs that are only relevant for reactive maintenance are: the cost of analysis, the costs (and especially availability) of spares, and the availability of repairmen. The cost of analysis are all costs that can be appointed to diagnosing the problem (e.g. time used for diagnosis). The spares cost depends on whether or not the business has spare parts readily available. When a business does not have spare parts in stock, they depend on the supply chain availability of these parts. Also, when a machine has a sudden failure, chances are that no skilled workers to repair the machine are available. If this is the case, a moment of repair will be scheduled in the future and the machine will be out of operation for a longer duration. These costs are averted because maintenance can be scheduled beforehand for PdM and PM.

This dissertation recognizes the importance of the RUL in the cost function. Therefore, RUL-prediction will also be the approach taken in this dissertation. When conducting preventive maintenance, it is vital to also account for the amount of useful life that is neglected by conducting the maintenance. That is, if a machine would fail in 5 months from now, replacing it right now would cost 5 months of useful lifetime. Of course, this is a cost that is not present in reactive maintenance.

3 Profit-driven analytics

3.1 Statistical measures

In data analytics, a model is often evaluated based on metrics such as the AUC, ROC, F-measure and so on. These measures are purely statistical and are based on model (statistical) (Verbeke et al., 2012) performance. For the ROC-curve, a model is evaluated on the True Positive Rate (TPR) and False Positive Rate (FPR) at a given threshold. The threshold can then be changed to create a new model that has a different number of misclassifications. For a given model, a threshold needs to be chosen that fits the use-case the best. Another statistical metric that is often used is top-decile lift (Verbeke et al., 2012), this metric only looks at the 10% of predictions of which the model is the most certain. To calculate the top-decile lift, a model's predictions are sorted in descending order. Of these sorted predictions, the 10% of which the model is the most certain are chosen. Then, for this top-decile, the lift is calculated as follows:

$$Lift = \frac{\text{TPR in top-decile}}{\text{Total rate of positives in the dataset}}$$

This metric is larger than 1 if for this decile the model does a good job of distinguishing. It is equal to 1 if the model's predictions are not better than random. If the total rate of positives in the dataset is equal to 5%, then a random model is expected to achieve a TPR of 5% for every decile possible in the dataset. If the constructed model is able to achieve a TPR of 20% for its top-decile, then the lift for this decile would be: $\frac{20\%}{5\%} = 4$. Essentially, this means that, for the top-decile, the model is able to distinguish 4 times more positive cases than a random model.

The AUC is an objective measure, however, it is rather ambiguous. For example, when two ROC-curves intersect, they cannot dominate each other anymore. In this case, the 'best' ROC will depend on the threshold that is best for the use case. However, the AUC does not take this into account, it is a global measure (Hand & Till, 2001). The issues that underly the AUC may lead to choosing the suboptimal model. There have been extensive studies describing the problems with the AUC, most notably that of Hand, 2009 which proposes an alternative measure, the so-called H-measure. This measure can be used in situations where the costs of misclassification are not the same for false positives and false negatives. This measure was proposed, because the AUC implicitly assumes different cost distributions for different classifiers (Buja et al., 2005; Hand, 2009). Essentially this means that, when evaluating and comparing models with the use of the AUC, two completely different underlying metrics are used. The costs should be defined as a part of the problem, and not vary per classifier (Hand, 2009).

For example, assume the case in which a classification-model is used to determine whether a given machine will need maintenance within the next three weeks. The costs of a False Negative (i.e. predicting a machine will not need maintenance, while in reality it should) is larger than the costs of a False Positive (i.e. predicting a machine will need maintenance, while it does not need maintenance). In the case of the former; a machine will not receive maintenance and will therefore probably break down. In the latter case, a machine will receive maintenance while it is not necessary, but this will not hinder production significantly. Clearly, the optimal threshold for a model depends on the trade-off between False Positives and False Negatives. Thus, the optimal threshold is situation-dependent. It depends on the costs associated with misclassifications. The optimal threshold for a model can be determined with the aid of the cost ratio and the ROC (Provost & Fawcett, 2001). With the aid of so-called 'iso-performance'-lines, the optimal threshold can be found on the ROC-curve. To construct the iso-performance line, the ratio of costs and the class-imbalance is used. All points on an iso-performance line

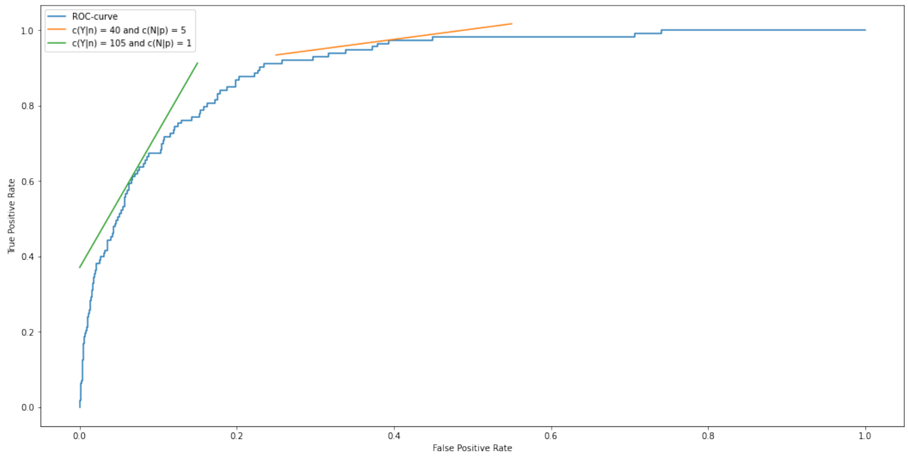

Figure 3.1: ROC-curve with two iso-performance lines at two different cost ratios.

share equal costs. The more north-west such a line is in ROC-space, the lower the cost of the iso-performance curve. So, when choosing the optimal threshold, one would have to find the most northwest point that is both on the iso-performance line AND on the ROC-curve. This point is the point at which the iso-performance line is tangent with the ROC-curve. The slope of such a iso-performance line is characterized by: $\frac{c(Y|n)p(n)}{c(N|p)p(p)}$, where $c(Y|n)$ and $c(N|p)$ are the costs of a false positive and false negative, respectively. $p(p)$ and $p(n)$ are the classes' prior probabilities (i.e. p(x)=$\frac{\text{total number of points belonging to class x}}{\text{total number of points in dataset}}$). Figure 3.1 shows an ROC-curve with two iso-performance lines. At different levels of costs, completely different thresholds are preferred. Note that, even though the AUC is related to the ROC-curve, the AUC does not encapsulate all of the characteristics of the ROC-curve, and may therefore lead to a suboptimal choice of model.

3.2 Profit-driven measures

In profit-driven analytics, instead of focusing on the statistical interpretation of a model's performance metrics, a monetary interpretation is given to it. In Verbeke et al., 2012, a profit measure was constructed to optimize a model's profit in function of its lift. The procedure was as follows: the predictions were sorted in descending order. Afterwards, of these ordered predictions the fraction which targeted the number of customers which maximizes profits was chosen. This metric was called the Maximum Profit criterion (MP). An important aspect which was accounted for in the MP, was the possibility of a correctly classified churner who, after receiving an offer from the retention campaign, decided to stay. This was modeled in a deterministic manner, by adding a variable 'γ' which was a number between 1 and 0 (i.e. the succes rate (Neslin et al., 2006; Verbeke et al., 2012)). The complement $(1-\gamma)$ can be interpreted as the fraction of would-be churners that decide to churn, even after receiving the offer. The MP criterion was heavily inspired by Neslin et al., 2006, who constructed a profit measure for a given churn management campaign (Π):

$$\Pi = N\alpha[\beta\gamma(CLV - c - \delta) + \beta(1-\gamma)(-c) + (1-\beta)(-c-\delta)] - A$$

Where (Neslin et al., 2006):

$$\begin{aligned}
N &= \text{Total number of customers,} \\
\alpha &= \text{The fraction of targeted customers who are targeted for the churn} \\
&\quad \text{management program,} \\
\beta &= \text{The fraction of targeted customers who are would-be churners,} \\
\delta &= \text{The cost of the customer incentive to the firm} \\
\gamma &= \text{The fraction of targeted would-be churners who decide to remain} \\
&\quad \text{because of the incentive (success rate)} \\
c &= \text{The cost of contacting a customer to offer him or her the incentive,} \\
CLV &= \text{The customer lifetime value} \\
A &= \text{The fixed administrative costs of running a campaign}
\end{aligned}$$

The MP criterion is defined as follows:

$$MP = \max_{\alpha} \Pi$$

So, in Verbeke et al., 2012, the MP criterion optimizes Π by targeting the 'top α%'-fraction of datapoints of which the model is the most certain. Optimizing α, is somewhat analogous to finding the optimal threshold for a model. This is an improvement on Neslin et al., 2006, since it accounts for the difference in lift-curves for two different models, as can be seen in Figure 3.2. In Neslin et al., 2006, the α was fixated at 10% (i.e. top-decile lift). However, fixating the α can potentially lead to suboptimal model selection. It is better to compare models based on their respective optimal values for the MP measure, and thus for the respective optimal value of α (Verbeke et al., 2012). In Figure 3.2, the two models have different fractions of included customers which maximize the profits (i.e. 5% and 10%).

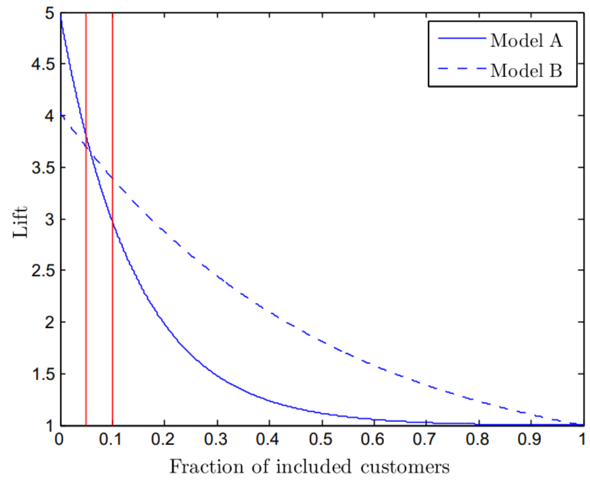

Figure 3.2: Lift curves for two different models. Directly taken from Verbeke et al., 2012.

Verbraken et al., 2013 took another approach, their research added a stochastic meaning to the aforementioned MP-criterion, the so-called 'Expected Maximum Profit Criterion' (EMPC). Instead of treating γ in a deterministic way, they treated it stochastically, as in the real world often the probability of a churner accepting an offer is not precisely known. The γ parameter is characterized by a beta distribution '$h(\gamma)$'. The EMPC can then be described by (Verbraken et al., 2013):

$$EMPC = \int_{\gamma} P_C(T(\gamma);\gamma, CLV, \delta, \phi) \cdot h(\gamma) d\gamma$$

Where $P_C(T(\gamma);\gamma, CLV, \delta, \phi)$ is the average classification profit of a classifier for customer churn and $T(\gamma)$ is the optimal cutoff for a given γ (Verbraken et al., 2013). Note that the expected maximum profit is simply the interpretation of the expected value (i.e. the statistical moment). So, the EMPC is the expected average classification profit of a classifier for customer churn, over the $h(\gamma)$ distribution.

The profits of course vary based on the threshold of a model and the fraction of customers that is being targeted. This removes a lot of ambiguity for the upper-management. Up until now, most of the work surrounding profit-driven analytics is in the field of Churn prediction or Default prediction (Óskarsdóttir et al., 2018; Verbeke et al., 2012; Verbraken et al., 2014; Verbraken et al., 2013). To calculate profits, both costs and benefits are needed. For churn and default cases, the benefits can be identified as the incoming cashflows that happen. For a churn case the incoming cashflows are the payments by a customer (Verbeke et al., 2012; Verbraken et al., 2013), whereas for a default case, the incoming cashflows are the interest payments for a loan (López & Maldonado, 2019).

In Garrido et al., 2018 the dependability of profits on internal or external events were accounted for with the use of random shocks. These random shocks affected the benefit and cost parameters in the EMP-measure, to create a more robust profit measure. For this, a new profit measure was created, the so-called R-EMP.

Profit-driven analytics has yet to be applied to maintenance prediction. Note that, most literature in maintenance prediction already incorporate some kind of monetary metric. However, there are no incoming cashflows associated with maintenance programs and therefore it is difficult to assess what benefits are in a maintenance context. So in this dissertation, instead of maximizing the profits, the focus will lie on minimizing the costs of maintenance. The goal is to create a measure that can be interpreted by the upper-management in a manner that shows how a model performs based on the costs associated with it.

4 Prior work

The costs that are associated with maintenance rely heavily on the type of maintenance strategy that an organization has implemented. Precise estimates depend on the context of an organization, however, there is evidence that on average the costs of RM are many times larger. In fact, multiple sources estimate the costs of PdM average to be atleast 3 times lower than those of RM (Mobley, 2002; Ran et al., 2019). In Yang et al., 2018 an extensive search for the optimal maintenance strategy was conducted on the simulated case of oil pipelines. To come up with the minimum cost of maintenance, they used a genetic algorithm which minimizes a cost function by setting the optimal number of maintenance check-ups and the intervals between these check-ups. The cost function in Yang et al., 2018 depends on the simulated probabilities of an event happening. This creates a very complicated equation, which was estimated via a discretization-algorithm.

Another approach is that of Florian et al., 2021, who treated the prediction of machine failure as a binary case. In their work, every hour the predictive model would be used to predict whether a machine was going to fail or not. The time interval in which the machine was going to fail, was assumed to be sufficiently large enough to allow planning of maintenance actions (Florian et al., 2021). After constructing a predictive model, the decision threshold was optimized with the use of iso-cost lines (cfr. Subsec. 3.1).

4.1 Random shocks

The authors of Yang et al., 2018, differentiated between age-based and degradation-based failures. The argument is that the condition of equipment is not only time-based, but also based on the presence of random sudden shocks that affect the current RUL of the machine. In the paper a case study on oil pipelines was presented, for which a model was created and simulated. This is inline with Garrido et al., 2018 who added random shocks to the EMP-measure, which affected the benefit & cost parameters. However, in a data-driven approach, adding random shocks to the RUL-prediction model, as Garrido et al., 2018 did, is rather difficult. It would alter the data directly, as the model is trying to predict the RUL. If the random shocks are not represented by the data, and no information of what the shocks are and when the shocks happened is available, then it would simply be impossible to add an extra variable representing the shocks.

As sudden shocks can have an effect on the estimated RUL, the RUL should be monitored actively. A sudden change in operating conditions can have detrimental effects on the estimated RUL. That is why in a maintenance setting the RUL should be estimated on regular basis. For example, the management of a production plant could decide to let a model predict the RUL of their machines every day. Then, the decision to conduct maintenance could be made if the predicted RUL is smaller than 7 days for example. It is important to also look at the change in RUL between two days. If it decreased significantly, then the reason as of why should be found. If the change in RUL can be attributed to specific behaviour, then the management can understand what behaviour to avoid. This would mean a shift from predictive maintenance to prescriptive maintenance, by communicating what actions to take (i.e. maintenance) and what actions not to take (i.e. undesired behaviour that shortens the lifespan).

4.2 The Maintenance Cost Rate

Another approach is that of Chen et al., 2022, here the researchers chose to only schedule one moment of maintenance. To find the optimal maintenance moment, the so-called *Mainte-*

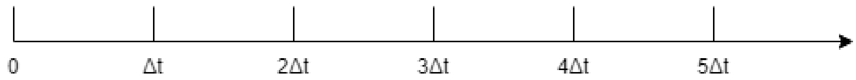

Figure 4.1: Timeline with 5 possible moments of maintenance.

nance Cost Rate (MCR) was constructed and optimized. The MCR is stochastic by definition, as the authors tried to account for the uncertainty in maintenance prediction. Chen et al., 2022 created a bi-LSTM that predicted the RUL prediction intervals of turbofan-engines from the CMAPSS-dataset simulated by NASA (Saxena et al., 2008). These prediction intervals were used to construct probability distributions that were used in the construction of the MCR. In a final phase, the MCR was optimized by varying the time of maintenance. For each turbofan engine the maintenance time that minimizes the MCR was chosen.

The idea behind the MCR is at the base of the cost function that will later be defined in Section 5. The MCR is constructed for each turbofan engine in the dataset by varying the t_m. For every possible time of maintenance an MCR is calculated. Afterwards the time of maintenance that corresponds to the lowest MCR is chosen for an engine.

First, a time interval Δt is chosen. Maintenance can only be scheduled at the end of such a time interval (i.e. Δt, $2\Delta t$, $3\Delta t$...). In Chen et al., 2022 this interval was set to 5 cycles. So, in their work, maintenance could only be scheduled at future moments that are multiples of 5.

Take Figure 4.1, for example. For every machine, maintenance can be conducted at each possible Δt. If the maintenance is scheduled at Δt, then the MCR depends on two possible scenarios. In the first scenario, a failure occurs before the scheduled maintenance (in this case in the $[0;\Delta t]$ interval). If a failure occurs before the planned maintenance, the cost that occurs is that of reactive maintenance (C_r). So, the expected cost for this scenario is defined as the product of the probability that failure occurs in $[0;\Delta t]$ and $\frac{C_r}{\Delta t}$ (cost of reactive maintenance per unit of operational time). The second scenario is that the failure occurs after the scheduled maintenance. The cost associated with this is defined as the product of the probability of the failure happening after the scheduled maintenance and the cost of predicted maintenance (C_p) per unit of operational time. The MCR is defined as the sum of these two, so algebraically;

$$MCR_{t_m=\Delta t} = P(0 < t_f \leq \Delta t) \cdot \frac{C_r}{\Delta t} + P(t_f > \Delta t) \cdot \frac{C_p}{\Delta t}$$

In the more general situation with $t_m = (h+k)\Delta t$, Chen et al., 2022 defined the MCR as follows:

$$MCR_{t_m=(h+k)\Delta t} = P((h-1)\Delta t < t_f \leq h\Delta t) \cdot \frac{C_r}{h\Delta t} + P(h\Delta t < t_f \leq (h+1)\Delta t) \cdot \frac{C_r}{(h+1)\Delta t} +$$

$$... + P((h+k-1)\Delta t < t_f \leq (h+k)\Delta t) \cdot \frac{C_r}{(h+k)\Delta t} + P(t_f > (h+k)\Delta t) \cdot \frac{C_p}{(h+k)\Delta t}$$

The intuition is that for every interval before the scheduled maintenance, the cost of failure should be calculated acknowledging the risk associated with it. For the intervals after the scheduled maintenance, the cost of failure is accounted for by C_p (and also the probability of a failure happening after the possible moment of maintenance). In Chen et al., 2022 the probability distribution that was constructed was a Gaussian distribution with parameters that were estimated with the use of a bi-LSTM. The bi-LSTM estimated 95% prediction intervals for the $\widehat{\text{RUL}}$, these estimates were then used to construct the RUL-probability distribution.

The key advantage of Chen et al., 2022 is that the authors account for the uncertainty in RUL prediction. However, the MCR does not carry information on the actual perceived costs. For a management audience, it might be more important to have information about the actual costs than on the forecasted costs. The goal of this dissertation is to construct a cost function that is comprehensible for a management audience.

5 The pure RUL maintenance cost

The proposed cost function, is based on the idea of 'misclassification costs' in 'Profit Driven Analytics'-literature and the 'RUL' in PdM-literature. It assumes that the management of an organization chooses a decision threshold which is used to decide whether to conduct maintenance. For PdM, the constructed model is a regression-model whose output is the predicted RUL ($\widehat{\text{RUL}}$). If $\widehat{\text{RUL}}$ is lower than or equal to the decision threshold, then maintenance should be conducted. To make the analogy with classification, this situation can be seen as a 'positive' prediction by the model. Mutatis mutandis for when $\widehat{\text{RUL}}$ is greater than the decision threshold. In the example presented in 4.1, the decision threshold would be 7 days. So everyday, the management compares the predicted RUL to the threshold in order to decide whether maintenance should be conducted.

There are two scenarios for maintenance, one is predictive (i.e. the machine undergoes maintenance before it fails) and the other is reactive (i.e. the machine fails and needs to be repaired). The costs associated with these maintenance types are as follows; C_p for predictive maintenance and C_r for reactive maintenance. The decision threshold T helps decide whether a prediction matches a positive or a negative case. If the model predicts $\widehat{\text{RUL}} > T$ (i.e. a negative case, thus no maintenance is scheduled). Then there are two possibilities for the true RUL. Either $RUL > T$ (i.e. a true negative, so the machine will not fail within the next T time units) or $RUL \leq T$ (i.e. a false negative, so the machine will fail within the next T time units). For the positive case, $\widehat{\text{RUL}} \leq T$, maintenance should be conducted. If $RUL \leq T$ (i.e. a true positive), machine will fail within the next T time units, if $RUL > T$ (i.e. a false positive) then the machine will not fail within the next T time units.

5.1 Possible cost scenarios

In the true negative case, no maintenance is planned and there is no sudden failure. So, no maintenance needs to be conducted. The cost in the true negative case is therefore equal to zero. However, for false negatives, no maintenance is planned, but the machine fails. In this case, reactive maintenance needs to be conducted. The cost of a false negative is therefore equal to C_r. In the positive case, maintenance will be conducted either way. Both for true positives and for false positives. Therefore the cost of true positives is equal to the cost of false positives. The cost of predictive maintenance is composed out of two parts: the fixed predictive cost (C_p) and the cost of the RUL that is not completely utilized (i.e. 'lost RUL').

The cost of the lost RUL depends on the amount of RUL that still remains and the time at which the maintenance is scheduled. This is best explained with the aid of an example; assume a case where a model has predicted RUL to be 13 days long. The threshold was chosen to be 14 days, so this is a positive case ($\widehat{\text{RUL}} \leq T$). The $\widehat{\text{RUL}}$ is a point-prediction and does not need to equal the true RUL, yet of course, ideally they are equal. If the true RUL is 9 days (even though the $\widehat{\text{RUL}}$ is 13 days), then if maintenance would be conducted right away, the lost RUL would be 9 days. However, if maintenance cannot be conducted right away, but needs to be scheduled in the near future (due to parts that need to be delivered or external repair men with busy schedules) then the machine will remain operational up until the moment of maintenance. So, if for example the maintenance would happen 5 days later, then the lost RUL would be 4 days

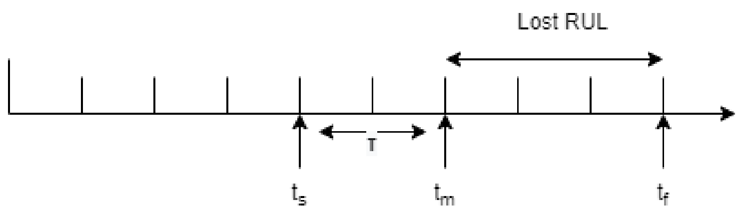

Figure 5.1: Process of determining the lost RUL

(9 days - 5 days). The time until maintenance can be scheduled 'τ' is defined as the number of days that pass between the moment at which a positive case is identified and the moment at which the maintenance takes place. If $\tau > RUL$, then the scheduled maintenance will be too late and reactive maintenance will need to take place.

So, the cost in positive cases is equal to:

$$I(\tau \leq RUL) \cdot ((RUL - \tau) \cdot \delta + C_p) + I(\tau > RUL) \cdot C_r$$

In this equation, $I(\tau \leq RUL)$ is an indicator function that is equal to one when the condition $(\tau \leq RUL)$ is met and is zero otherwise. $(RUL - \tau) \cdot \delta$ corresponds to the cost of the lost RUL, and will be further discussed in Section 5.3.

Figure 5.1 gives a simple overview of how the 'lost RUL' is identified. In the figure, t_s is the time at which maintenance is scheduled, t_m is the time at which maintenance was conducted (after waiting τ units). The lost RUL is determined by subtracting the true time of failure (t_f) by t_m.

5.2 The Pure RUL Maintenance Cost measure (PRMC)

The total cost of maintenance will be the sum of the costs of all True Positives, False Positives, True Negatives, and False Negatives. Therefore, the PRMC will be a piecewise function, defined as follows:

$$\begin{aligned} PRMC_i(T; \tau, \delta, C_r, C_p) = I(RM) \cdot C_r + I(PdM)[I(\tau \leq RUL_i)((RUL_i - \tau) \cdot \delta + C_p \\ + I(\tau > RUL_i) \cdot C_r] \end{aligned} \tag{1}$$

Where $I(RM)$ and $I(PdM)$ are indicator functions that indicate whether RM is conducted or PdM is scheduled. These indicator functions can either both be equal to zero, or at most one will be equal to one. This is simply because a machine cannot receive both PdM and RM. It is possible however, that no maintenance is conducted. Then, for this machine, the PRMC is equal to zero.

5.2.1 Adding a threshold

In the right-hand side of Eq. 1, the threshold T is still missing. The threshold is important in deciding whether maintenance should be conducted. The threshold can be added to Eq. 1 by substituting $I(PdM)$ with $I(\widehat{\mathrm{RUL}}_i < T)$. The management of an organization will likely be interested in how a model's PRMC changes as the threshold changes. More specifically, the management is interested in the threshold that minimizes the PRMC. Thus Eq. 1 becomes:

$$\begin{aligned} PRMC_i(T;\tau,\delta,C_r,C_p) = I(RM)\cdot C_r + I(\widehat{\mathrm{RUL}}_i \leq T)[I(\tau \leq RUL_i)(RUL_i - \tau)\cdot\delta + C_p \\ + I(\tau > RUL_i)\cdot C_r] \end{aligned} \tag{2}$$

Note that RM does not depend on a threshold, it simply occurs without prior notice.

5.2.2 Graphical representation of the PRMC

Figure 5.2 shows an example of the PRMC for one machine if maintenance is scheduled. For this representation, abstraction was taken from both τ and T, this was done by setting them equal to zero. The red line in the figure represents the true RUL, when $\widehat{\mathrm{RUL}} > RUL$ (right of the red line) the cost will be equal to C_r. In the other case, the cost will be equal to C_p. It becomes evident from Figure 5.2 that when a model underestimates the RUL by too much, the cost of PdM can well exceed that of reactive maintenance.

5.2.3 Total PRMC

The PRMC is calculated for one piece of equipment only. To get a complete overview of the costs of maintenance (i.e. the total PRMC), all PRMC's need to be summed up:

$$PRMC = \sum_{i}^{N} PRMC_i(T;\tau,\delta,C_r,C_p) \tag{3}$$

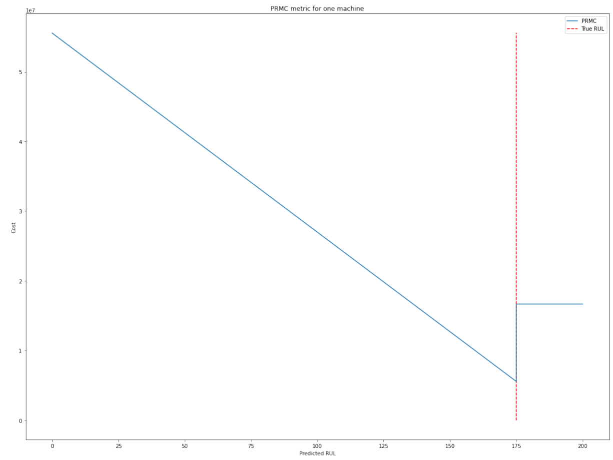

Figure 5.2: The PRMC metric, where $\tau = 0$ & $T = 0$

5.2.4 Minimum cost

For a given model, only the threshold which minimizes the PRMC will be relevant. An organization will therefore only be interested in the T which minimizes Eq. 3. Thus, comparing models should be done by using their respective minimum costs (MC):

$$MC(T_{opt}) = \min_{T} PRMC(T) \tag{4}$$

Where:

$$T_{opt} = \arg\min_{T} PRMC(T)$$

5.3 The cost of RUL

Defining the cost of RUL is rather difficult. It is different for different types of equipments, but it is detrimental for the use of a RUL-based costfunction. In a production setting, the cost of RUL for a machine could for example be the contribution to the revenue per time unit for that machine. The benefit of this approach is that cost of RUL will be measured by the actual value a machine generates for a business. If the machine has PdM, then the cost of the lost useful life can be expressed in the cost of the lost value for the business. However, this requires a thorough knowledge about the production processes. Another issue with this approach is that, in complex production environments, it is difficult to attribute the exact contribution of a single machine to the total revenues of a product or service.

This dissertation presents a more universal way of estimating the cost of RUL. In accounting, a common way of dealing with assets that require a large investment and have a certain lifespan, is spreading the costs of these investments over multiple years by the use of depreciation. Instead of adding the cost of the purchase of an asset to this year's costs, the cost is spread over its useful life (Tuovila, 2023). So, it is definitely related to the RUL. The cost of RUL will therefore be determined by an assets acquisition price and its useful life.

In accounting, the book value of a machine is not always equal to the value which that machine generates for a business. Take a car that is depreciated over a lifespan of five years, if the car still works after these 5 years and the firm still uses it, then it clearly still creates value for the business. However, this value is not depicted in the balance of the firm anymore, so the book value does not correspond with its true useful life. Luckily, as the PRMC is meant to be used during training, the true RUL (and thus also the true total useful life) will always be available when calculating the PRMC. So, instead of using a fixed lifespan for all machines (e.g. the aforementioned five years for a car), the true total useful life can be used on a per machine basis. This has two implications; first, the estimate of the cost of RUL can be derived from the true useful life. Second, every machine has its own value for the cost-of-RUL-parameter (δ):

$$\delta = \frac{\text{Acquisition Price}}{\text{Total Useful Life}} \tag{5}$$

5.4 The expected PRMC

There remain a couple of issues with the previous section. During training, the labels (the true RUL) are known. Since the PRMC is used for model selection, the true RUL can therefore be used in the calculation of the PRMC. However, during inference, calculating costs will be different, yet the management would probably still want to be able to monitor the PRMC. This will be discussed in Subsection 5.5.

The real issue lies in the fact that the time until maintenance (i.e. lead-time) τ is typically not known and is subject to many different factors. For example, it could be the case that the parts and expertise are not available right away. τ can be treated as a deterministic variable, however, this may lead to suboptimal model-selection as τ can differ and a model should be robust to these differences in τ. Figures 5.3 and 5.4 illustrate how different approaches to τ can result in different costs. In Figure 5.3, the deterministic approach is taken. In this situation τ is fixed and smaller than the remaining useful life of the asset. This will mean that the moment at which maintenance will take place, will be before the moment of failure. The cost will therefore be the cost of maintenance plus the cost of the lost RUL. In Figure 5.4, τ is treated as a stochastic variable which can lie anywhere in the red box. In this situation, there are two possible outcomes; either τ can be smaller than or equal to the RUL, or τ can be larger than the RUL. In the former, the cost will be calculated in the same way as in Figure 5.3. In the latter, however, the machine will fail before the maintenance can be conducted and therefore the cost will be equal to the cost of reactive maintenance. Ideally, a model accounts for the uncertainty in maintenance planning, which is why treating τ as a stochastic variable might be more suitable.

Note that, even though one could come up with a good estimation of τ, if it is consistently incorrect, the PRMC will not be a good indicator of model performance. Also, it is rather difficult to come up with a good estimation for the τ as it depends on many different external factors. Therefore, it is easier to come up with a range or distribution of possible τ's. For every τ in this range, there will be a different optimal threshold T. This is rather logical, as when the lead times increase, the threshold will need to increase as well to have less cases in which reactive maintenance occurs. Therefore, the optimal threshold can be written as a function of τ: $T(\tau)$.

For the lead-times, a lognormal distribution is proposed to model the distribution of τ (Das & Abdel-Malek, 2003). Assume that $h(\tau)$ is the probability distribution of τ. Since τ is a parameter in the $PRMC$, the expected value of the minimum cost (see Eq. 4) can be calculated with respect to $h(\tau)$.

The expected $MC(T(\tau); \tau, \delta, C_r, C_p)$ can then be defined as;

$$\begin{aligned} EMC(T) &= \mathbb{E}[MC(T(\tau))] \\ &= \mathbb{E}[\sum_i^N PRMC_i(T(\tau))] \\ &= \sum_i^N \mathbb{E}[PRMC_i(T(\tau))] \\ &= \sum_i^N \int_\tau PRMC_i(T(\tau)) \cdot h(\tau) d\tau \end{aligned} \tag{6}$$

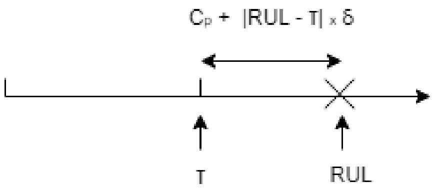

Figure 5.3: Deterministic τ. Note that τ can be greater than the RUL as well, in which case reactive maintenance will occur.

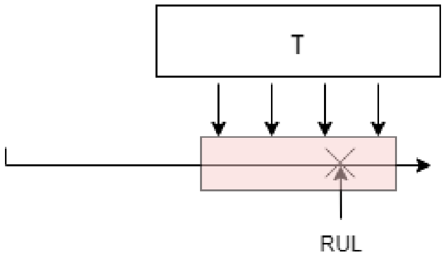

Figure 5.4: Stochastic τ.

5.5 The PRMC during inference (iPRMC)

As mentioned in Subsection 5.4, an organization's management will want to monitor the costs of maintenance actively during inference. The true RUL is not known during inference, due to this, the cost of the lost RUL cannot be calculated. Therefore the only costs that can be directly calculated are the actual perceived costs of maintenance (C_p and C_r). The problem is twofold, on the one hand the true RUL is needed to calculate the lost RUL and on the other hand it is needed to calculate the cost of the lost RUL ($\delta = \frac{\text{Acquisition Price}}{\text{Useful life}}$). A solution could be to let the person responsible for the maintenance come up with an estimate of how much longer the machine could have lived. These estimates can then be used to estimate the PRMC during inference. Another advantage of this is that these RUL-estimates could be used later on in the training of a different model. This requires a lot of expertise from the workers who conduct maintenance, which may not be a realistic expectation. Also, whether these RUL-estimates are useful should be examined.

During inference, τ is not relevant anymore, as all costs are à posteriori. Either a machine has suddenly broken down with cost C_r or a machine has undergone predictive maintenance at cost C_p. If predictive maintenance has been conducted, then logically, the time until maintenance τ is not relevant anymore as maintenance has already happened. However, during inference, a stakeholder will probably also be interested in the cost of the lost RUL. Omitting τ from the equation, the cost of the lost RUL can be denoted as:

$$RUL \cdot \delta$$

But, because the true RUL is still not known, the cost of the lost RUL cannot be calculated.

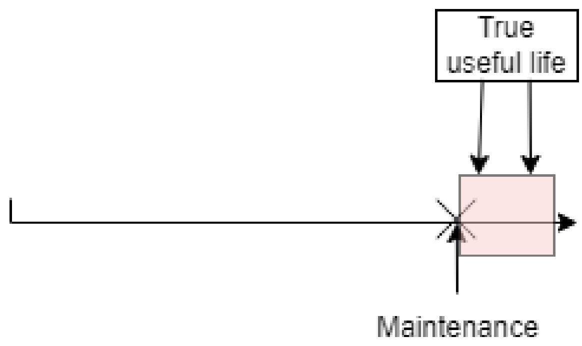

Figure 5.5: The true useful life will often be greater than the moment at which maintenance was conducted.

To combat this, one could for example come up with a good estimation of the lifetime of a machine, such as: the advertised lifetime of an asset. The issue is that this estimation will surely always be smaller or greater than the actual lifetime of an equipment. This is definitely the case for a large plant, with many machines. These machines will not all have the same lifetime. So, treating the expected lifetime as a deterministic variable will lead to subpar conclusions. Also, often, instead of coming up with a single anticipated lifetime, it is easier to come up with a range of possible lifetimes. The expected lifetime can be derived from a distribution of lifetimes. So, perhaps a more correct way of handling the cost of lost RUL is to construct a probability distribution for the RUL of a machine and to use its expected value in the calculation of the PRMC.

The expected RUL can be calculated by subtracting the expected useful lifetime with the moment at which maintenance is conducted, this will further be discussed in Section 5.5.2.

5.5.1 The probability distribution

The expected useful life needs to be calculated from a probability distribution. For machines that have undergone predictive maintenance, the true RUL will always be greater than or equal to zero. This is rather logical, if a machine has not yet failed, then the only moment at which it can fail will be in the future (hence RUL ≥ 0). This means that the expected lifetime of a machine will always be at least as large as the moment at which maintenance is conducted. This is depicted in Figure 5.5, the red area on the timeline is the possible range for the expected lifetime of a machine. This range (or distribution) will be different per machine, as the moments of maintenance will also be different. A good choice for a distribution of lifetimes of assets is the Weibull-distribution (Erumban, 2008). The distribution can be constructed using expert knowledge or with the use of an existing set of datapoints (i.e. training data). Figure 5.6 shows a Weibull distribution fitted on the total useful lives of the Turbofan-dataset (Saxena et al., 2008). It is important to emphasise that the distribution in Figure 5.6 is fitted on the total Useful Lives (UL) in the dataset. It is the lifetime distribution function for the machines in the dataset. At any given moment in time t_0, the probability that the machine will fail at a later time than t_0 is denoted by $P(UL > t_0)$ and can be derived from the distribution in Figure 5.6. The useful life of a machine will always be greater than or equal than 0, this is also a characteristic of the Weibull-distribution. For a Weibull-distributed random variable ($X \sim Weibull$), the following will always hold:

$$P(X < 0) = 0$$

The RUL-distribution can directly be derived from the 'useful life'-distribution (UL). If a machine has underwent PdM at time step t, then the total useful life will be larger than t. So, at time step t, the probability that the RUL is larger than x (and also the cumulative RUL-distribution) will be equal to the probability that the total useful life is larger than $(x + t)$, given that the total useful life is larger than t:

$$P(RUL > x) = P(UL > x + t \mid UL > t) \tag{7}$$

Using Bayes' rule; this becomes:

$$= \frac{P(UL > x + t) \cap P(UL > t)}{P(UL > t)} \tag{8}$$

If UL > x+t, then automatically; UL > t, hence:

$$= \frac{P(UL > x + t)}{P(UL > t)} \tag{9}$$

$P(UL > t)$ will always be a constant, namely the grey area in Figure 5.7.

The cumulative distributions are as follows:

$$\int_x^\infty f_{RUL}(x)dx = \frac{\int_x^\infty f_{UL}(x + t)dx}{P(UL > t)} \tag{10}$$

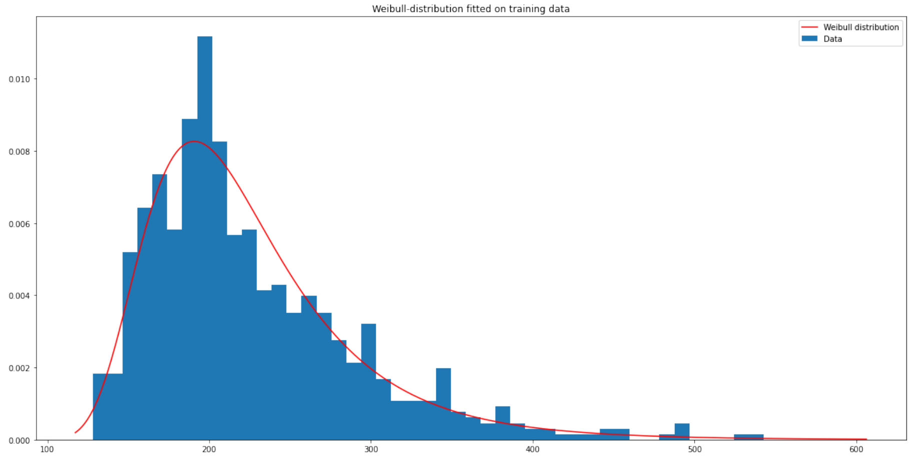

Figure 5.6: Weibull distribution.

Parameters: a: 2385.53, c: 0.56, loc: 0 and scale: 5.01

and thus, the probability distributions:

$$f_{RUL}(x) = \frac{f_{UL}(x+t)}{P(UL > t)} \tag{11}$$

This is a rather important derivation, as this means that during inference, only one distribution needs to be fitted to the data in order to come up with a RUL-distribution for each machine. The intuition is as follows; the UL-distribution is shifted to the left by t (the time at which maintenance was conducted), of course the RUL-distribution does not sum up to 1. Therefore, it is divided by $P(UL > t)$, this 'normalizes' the RUL-distribution. If this had not been the case, then a new distribution would have needed to be fitted for every machine that underwent PdM. Also, $P(UL > t)$ can be calculated from the UL-distribution and will always be a constant:

$$P(UL > t) = \int_t^{\infty} f_{UL}(x)dx \tag{12}$$

Note: When constructing a 'Useful Life'-distribution, an important assumption is made. It is assumed that all machines come from the same distribution.

5.5.2 The expected value

For a machine that has undergone predictive maintenance at timestep t, the range of true possible useful lives must be larger than or equal to t. Looking at the constructed probability distribution, only the part of the distribution after t is relevant. The probability of the true useful life being larger than t is denoted as: $P(UL \mid UL \geq t)$ where UL is the useful life.

Figure 5.7 shows an example of a machine that has underwent PdM at time step 200. Since the machine was repaired before failure occured, the true RUL will be greater than or equal to zero (and thus its useful life will be greater than or equal to 200). In Figure 5.7, the time of maintenance is depicted by the black line. The grey area under the curve, represents $P(UL > x \mid UL \geq t)$ (i.e. the probability that the useful life will be larger than or equal to 200 timesteps) and is equal to 59%. The true RUL, to be used in the calculation of the iPRMC, will lie somewhere in the grey area.

In order to calculate the cost of the lost RUL, a good approximation of an asset's RUL is needed. Section 5.5.1 showed how for every machine a different RUL-distribution can be derived from the initial UL-distribution. Sampling a useful life from the grey area might seem like a good option, however, knowing that for every machine the grey area (and thus the RUL-distribution) will be different, this adds too much variability. So, instead of sampling a RUL from the distribution, the expected value is used.

The expected value for a continuous random variable X (such as the Weibull distribution) is defined as:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

So, the expected value for the RUL-distribution would become:

$$\mathbb{E}[RUL] = \int_{-\infty}^{\infty} x f_{RUL}(x) dx \tag{13}$$

Plugging Eq. 11 into Eq. 13 results in:

$$\mathbb{E}[RUL] = \int_{-\infty}^{\infty} \frac{x f_{UL}(x+t) dx}{P(UL > t)} \tag{14}$$

$$= \frac{1}{P(UL > t)} \int_{-\infty}^{\infty} x f_{UL}(x+t) dx \tag{15}$$

Because the Weibull-distribution is equal to zero for x < 0, the following holds for Eq. 15:

$$= \frac{1}{P(UL > t)} \int_{0}^{\infty} x f_{UL}(x+t) dx \tag{16}$$

Essentially, Eq. 16 means that the expected RUL can be calcultated by shifting the initial UL-distribution to the left by t time units and dividing by the surface under the UL-distribution

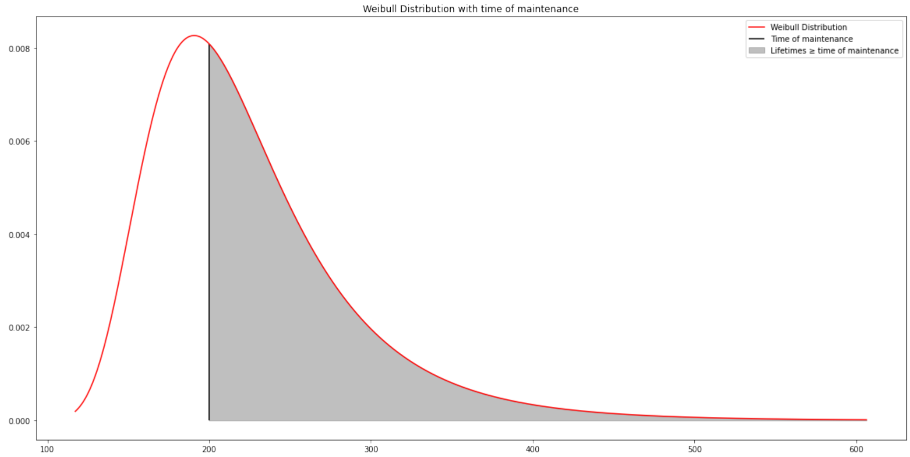

Figure 5.7: Fitted Weibull, with time of maintenance.

from point t onwards (Eq. 12). So, this shows that the expected value of the RUL of any given machine can directly be calculated from the initial UL-distribution. The initial UL-distribution can directly be fitted on a dataset containing failure-times and will be the same for every machine.

So the iPRMC can be identified as:

$$iPRMC_i = I(RM) \cdot C_r + I(PdM) \cdot (C_p + \mathbb{E}[RUL_i] \cdot \frac{\text{Acquisition Price}}{\mathbb{E}[RUL_i] + t}) \tag{17}$$

and the total iPRMC:

$$iPRMC = \sum_{i}^{N} iPRMC_i \tag{18}$$

Essentially, when PdM is conducted, the estimated cost of maintenance (during inference) for a machine will depend on the expected RUL of that specific machine. So in Eq. 17, the denominator is not equal to the actual UL of the machine, since it is not known. Instead, the opportunity cost is spread over the amount of time it has functioned up until the moment of maintenance t, plus the remaining useful life it is expected to have had.

6 Case study

In this section, the PRMC metric will be applied using a case study on the synthetic C-MAPSS dataset (Saxena et al., 2008). Five types of models were trained, each of which was tuned using three different metrics; the MAE, the MSE and the PRMC. The objective of this study is to analyze whether using the PRMC leads to selecting models with lower maintenance costs.

Variable	Value
Acquisition Price	50 000 000.00
C_r	16 666 666.67
C_p	5 555 555.56
τ	12
α	0.05

Table 6.1: Parameters used in the case study

All variables used in this case study are shown in Table 6.1.

6.1 Dataset

The dataset that was utilized was created with the use of the C-MAPSS (Commercial Modular Aero-Propulsion System Simulation) tool by NASA (Saxena et al., 2008). The machines that were simulated with this tool are turbofan engines from the 90,000 lbs thrust class. Examples of engines in this thrust class are the Rolls-Royce Trent 800, General Electric GE90 or Pratt & Whitney PW4000. These engines power large commercial aircraft carriers such as the Boeing 777 (Praveen, 2020). For every engine in the dataset, 21 sensor readings are reported starting from a point in time (with a non-zero value for the initial wear) up until engine failure. The initial wear is arbitrarily chosen for every engine, this makes the dataset more in line with real-world data. This is because, in the real world, the initial condition of a machine often depends on factors during the production processes or transport. An explicit assumption that is made for the data, is that the degradation behaviour of an engine can not be solely attributed to one flight. Instead, it is attributed to the usage over all flights. As a result, Saxena et al., 2008 only report one measurement per flight. So, in the dataset, every data point represents one moment during one flight of an engine. As flights can have different durations, there is no data on the true lifetime of an engine (in time units). So, this dissertation will need to take abstraction from 'time units' and treat the useful life in 'flight units'. If, for example, an engine has a RUL of 74, then this means that the engine still has 74 flights left. It does not mean that the engine has 74 days (or weeks) of lifetime left in this data set. In the following sections, the words 'flight units' and 'time units' will be used interchangeably, yet it will always refer to the number of flights that are remaining for a given engine.

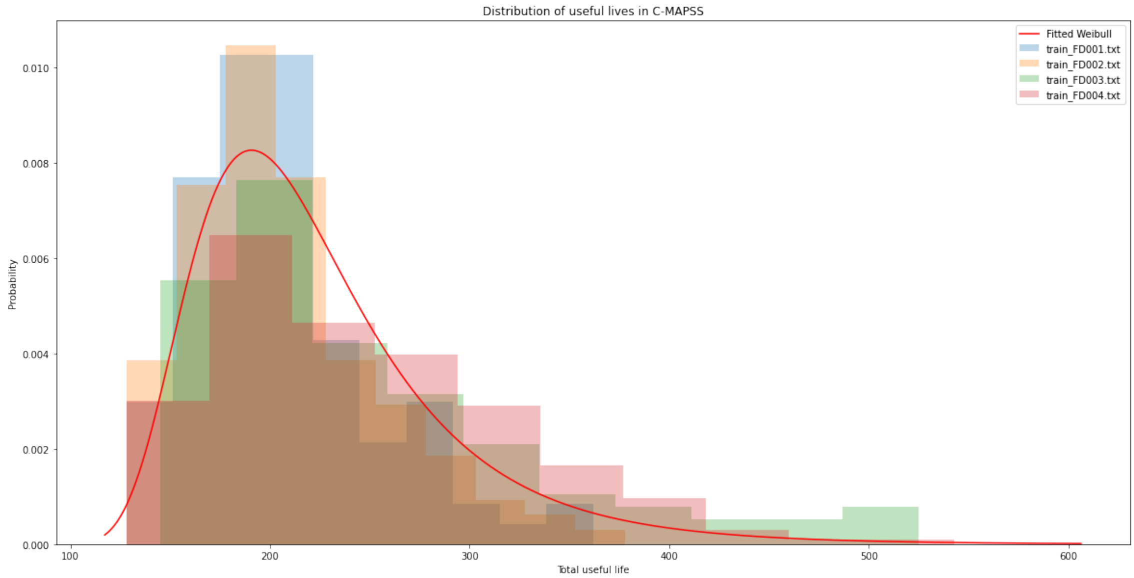

Figure 6.1: Distribution of useful lives for all files.

The dataset consists out of four train files which, combined, account for 709 engines. Figure 6.1 shows the distributions of total useful lives for the four separate files, along with a Weibull-distribution fitted on the distribution of the four files combined. The dataset also contains eight files for the test set, four for the independent and dependent variables each. However, these are not used in this dissertation, as these are unsuitable for the maintenance strategy used, this will be discussed in Section 6.2.1. A table containing all variables is attached in the Appendix (Table A.6).

6.1.1 Choices of cost values

The cost of RUL depends on the acquisition cost and the the total useful life of an engine. Since the total useful life is known for every machine, only an estimate for the acquisition cost needs to be provided. As there are no catalogs available on the price of large commercial turbofan engines, an estimate needs to be made based on information on expert opinions (Ezik, 2003). Estimates for the GE90 engine vary from the high 30 million EUR (General Electric, 2019; Memon, 2022) to the low 65 million EUR range (Price, 2023). For this reason, an acquisition cost of 50 million EUR was chosen for every engine.

The reactive maintenance cost is chosen to be a third of the acquisition cost. This is based on Steel, 2020, however, note that sudden engine failures might have implications on the entire aviation industry. For example, consider the Boeing 737 Max groundings (Isidore, 2020). Even

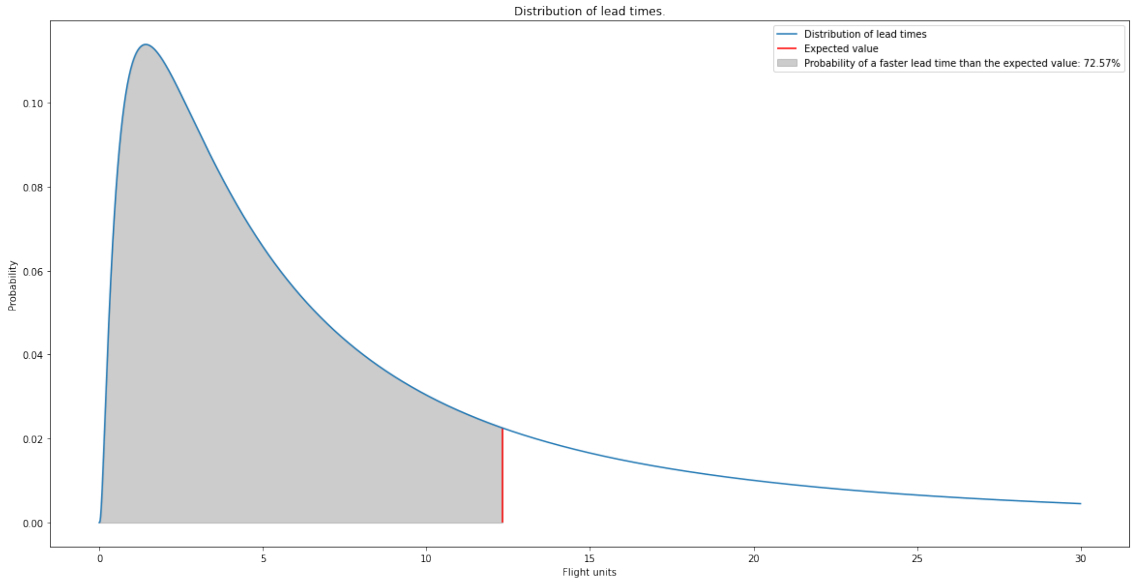

Figure 6.2: Lead time distribution

though these unforeseen costs are very substantial, they need to be abstracted from. The cost of predictive maintenance is set to a third of the reactive cost, this is supported by literature in the predictive maintenance field (Mobley, 2002).

6.1.2 Choice of lead time

The lead time τ is chosen to be equal to 12 flight units for the calculation of the PRMC (deterministic). This assumption is based on the availability of parts and skills, and on the time it takes to diagnose the issue within an engine. For the calculation of the EPRMC (stochastic) a log-normal distribution (cfr. Subsec. 5.4) is constructed, with an expected value of 12 (shape $= \sigma = 1.2$ and scale $= e^{\mu} = 6$). This distribution is depicted in Figure 6.2, where the probability of having a lead time smaller than the expected value is equal to 72,57% in this specific case. Unfortunately, calculating the EPRMC was not possible due to computational limitations, as will be explained in Section 8.

Note that, in a real setting the lead time distribution should be constructed based on expert knowledge, data, or a combination of the two.

6.2 Methodology

6.2.1 Maintenance strategy

Instead of focusing on, for example, optimizing the maintenance intervals and frequency directly (Bris et al., 2003) or on predicting the future optimal moment of maintenance (i.e. predicting the RUL) (Chen et al., 2022), this dissertation focuses on continuous monitoring of a system. That is, it is assumed that during the usage of machinery, their $\widehat{\text{RUL}}$ is actively monitored. As soon as the $\widehat{\text{RUL}}$ is lower than a threshold T, maintenance is scheduled. As mentioned in Section 5.2.4, this allows an organization to optimize a model's performance by choosing the threshold for a model which achieves the Minimum Cost (MC, as defined in Section 5.2.4).

Therefore, such an environment was simulated using the data that is available in the train files. The test files themselves were not suitable for this approach, as the engines in these files did not contain sensor measurements up until the moment of failure. Instead, the data in these files end before an actual engine failure was perceived. This, of course, does not allow for the simulation of the complete engine degradation process. In order to set a threshold, the moment of failure needs to be known. The reason for this is evident; if the moment of failure is not known for a machine, then it could be possible that the PRMC for this machine is equal to zero. As long as no failure is perceived (C_r) or maintenance is scheduled ($C_p + \delta * (RUL - \tau)$), there will be no cost associated with the machine. Hence, it would influence the optimal threshold in a very negative way. For an engine i without failure data and a final prediction $\widehat{\text{RUL}}_t$, the $PRMC_i$ will always be optimal for $T < \widehat{\text{RUL}}_t$. This essentially means that if such non-failure data is added to the data set, the optimal threshold will be pushed downwards. This is unwanted, as the model will generalize worse to unseen data.

6.2.2 Data leakage

During the tuning of the models, the data was split in to a train, a validation and a test set. The splits need to be done with care in order to prevent data leakage. A machine can not be present in multiple subsets at once. In this dissertation, every engine ID was assigned to only one subset. Afterwards, only sensor measurements belonging to these engines were added to their respective subset. 70% of all data was in the train set, 15% in the validation set and 15% in the test set. After tuning the hyper parameters, the models were fitted on the train and validation sets combined and evaluated on the test set.

6.2.3 Simulation of the online-phase

The performance of a model needs to be evaluated based on the context in which it will operate. Since this dissertation follows the assumption of 'active monitoring', a model needs to be evaluated in such context. Therefore, an online-phase was simulated by letting a model make predictions for the entire operational lifetime of all engines in the test set. Then, using these predictions, the impact of changing the threshold on the total *PRMC* was analyzed. This was done by calculating the *PRMC* for every possible threshold, and then finding the threshold which minimized it. Figure 6.3 shows the total cost of maintenance as a function of the threshold. For this specific model, the optimal threshold should be set to 27 'flight units'. Figure 6.4 shows the simulation process for one engine, where the grey dotted line represents the optimal threshold as found in Figure 6.3. At time step 83, $\widehat{\text{RUL}}$ was smaller than the threshold for the first time. Hence, this was the time step at which maintenance was scheduled. According to Eq. 2 the cost for this specific engine is equal to € 9,882,478.63. In the case of reactive maintenance, the cost would be equal to € 16,666,666.67. This means that for this engine, the model managed to save € 6,784,188.04.

Note; it may be possible that $\widehat{RUL}$ deviates too much from RUL. As a result, maintenance can be scheduled at such an early moment in time, that the cost of maintenance still exceeds the cost of reactive maintenance.

6.2.4 Preparation

Standard-scaling was used for all models. For neural networks, scaling is necessary to improve convergence. For tree-based ensembles scaling is not necessary, because decision trees are not sensitive to transformations that preserve the order of values within a variable (Filho, 2023). To calculate the PRMC, the cost of RUL needs to be known. Therefore, the acquisition cost was divided by its total useful life for every engine (as depicted in Eq. 5). This resulted in a vector containing a cost of RUL for every engine in the data set.

The data was windowed with different window sizes. The window size was treated as a hyper parameter, for different values up to 40. Every model favoured a window size of 40 time steps (i.e. the maximum possible window size). This effectively means that for every prediction, all

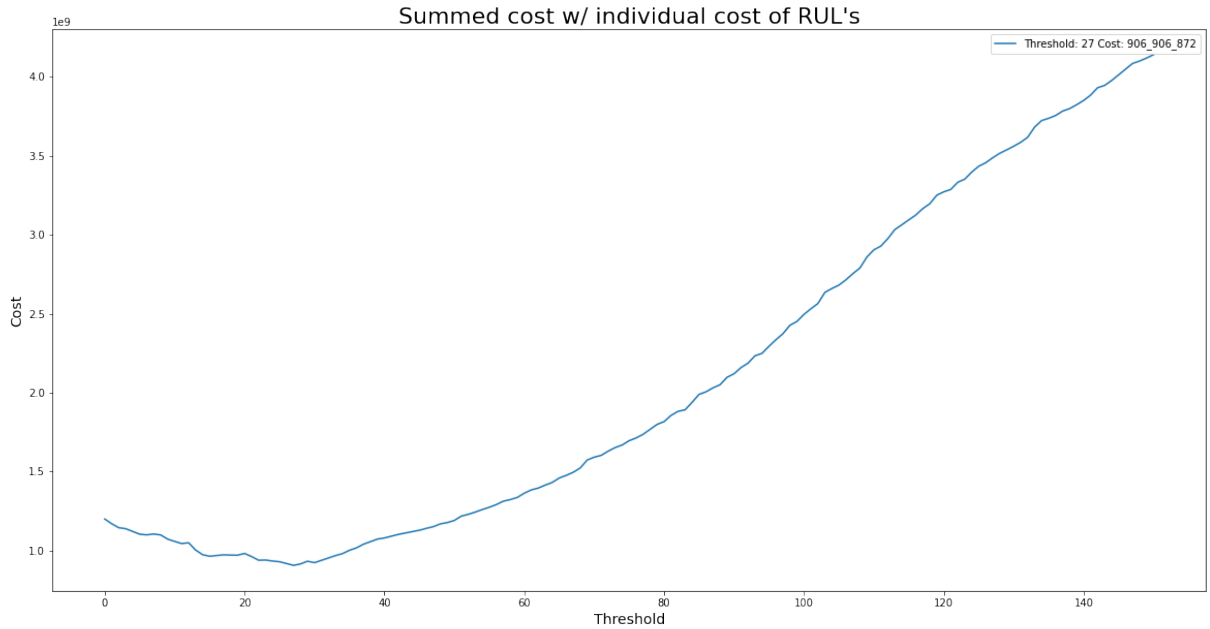

Figure 6.3: PRMC plotted as a function of the threshold.

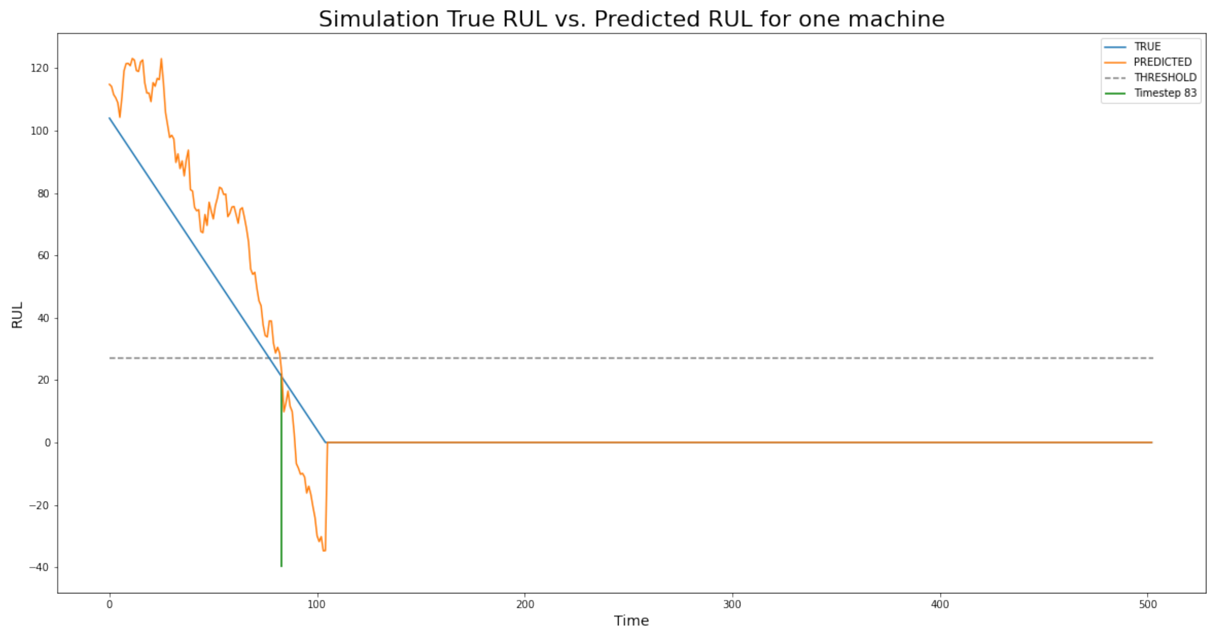

Figure 6.4: Overview of the simulation run for one engine in the test set.

models took in to account 24 variables for the last 40 flights of an engine (i.e. 960 variables for one RUL-prediction). The reason why the maximum possible window size was set to 40, is because a larger window size might not be realistic in real situations. A window size of 40 means that before 1 datapoint can be created for an engine, the engine has to have been used in 40 flights.

6.3 Models

Five types of models were trained for this dissertation, an LSTM-model, a Ridge Regression, a Support Vector Regressor, a Random Forest and an XGBoost model. Every model was tuned three times, using three different evaluation metrics. The evaluation metrics that were used, were the MAE, MSE and the PRMC. So, for every model, three variants with different hyper parameters were created. The hyper parameters were chosen based on the evaluation on the validation set. Evidently, optimizing models on three different metrics, resulted in three different combinations of hyper parameters per model (i.e. 15 models in total).

All hyper parameters were tuned using the grid search procedure on the validation data set. As the Random Forest, Boosting and LSTM models allow for choosing loss functions, these were also tuned. For the MAE and MSE metrics the MAE and MSE losses were chosen, respectively. For the PRMC-metric, the loss function was treated as a hyper parameter. This way, the loss function which optimizes the PRMC could be identified. The final selections of hyper parameters are attached in the appendix of this dissertation (Appx. A.1).

6.4 Results & analysis

To verify whether the models tuned with the PRMC achieve lower costs than the models tuned with the MAE and MSE, a Wilcoxon signed-ranks test will be performed.

6.4.1 Interpretation of the chosen metrics

Residual-based metrics

Residual-based metrics are metrics that are based on the difference between the target variable and its prediction by the model. Both the MSE and MAE are residual-based metrics and can take values anywhere in the interval $[0; \inf]$, where the closer they are to zero, the better the performance of the model.

PRMC-based metrics

The chosen PRMC-based metric is the total PRMC (i.e. the summed cost for the optimal threshold). One could, for example, also calculate the mean PRMC and the PRMC-ratio. However, these metrics are linear transformations from the total PRMC and therefore will have the same conclusion as the total PRMC. The total PRMC is equal to the sum of the costs of all machines in the test set. The mean PRMC is equal to the mean cost for a machine in the test set (i.e. $\frac{\text{Total PRMC}}{\text{\# of machines in the test set}}$). The PRMC-ratio is equal to the total PRMC divided by the total cost if all machines were to fail (i.e. $\frac{\text{Total PRMC}}{\text{\# of machines in test set}\cdot C_r}$). If this ratio is equal to 1, the model does not perform better than a pure-failure strategy (i.e. letting all engines run until failure (RM)).

	MSE	**MAE**	**PRMC**
Ridge Regression	906,906,872	897,858,196	900,880,997
Support Vector Regressor	892,764,602	878,793,443	871,155,565
XGBoost	1,058,218,285	1,050,319,016	951,047,867
Random Forest	1,107,902,264	1,783,333,333	1,073,288,321
LSTM	974,013,838	901,702,882	892,080,081

Table 6.2: Final models and their corresponding minimal costs.

6.4.2 Model performance

The goal of this section is to compare the differences in model choice for every metric. Especially, whether using the PRMC-based metric proves to be more suitable by selecting better hyper parameters for a model. Table 6.2 shows the costs of maintenance for each model.

Testing significance

The significance is tested with the aid of a Wilcoxon signed-ranks test at the 5% level. When comparing the PRMC metric to any other metric M, the hypotheses become:

$$\begin{cases} H_0 & : PRMC = M \\ H_a & : PRMC < M \end{cases}$$

Which can be rewritten as:

$$\begin{cases} H_0 & : PRMC - M = 0 \\ H_a & : PRMC - M < 0 \end{cases}$$

If the null hypothesis is rejected, then there is a significant difference between the models tuned with aid of the PRMC as opposed to the models tuned with another metric. As the goal is to minimize the costs, the PRMC-tuned model is significantly better when its total cost of maintenance is *lower* than its total cost when tuned for another metric. This is tested by comparing the pairwise differences in PRMC for the observations. An example of such a pairwise difference would be to subtract the cost for Ridge regression in the PRMC column by the cost in the MAE column (Table 6.2).

6.4.3 Wilcoxon signed-rank test

Tables 6.3 and 6.4 show the differences and ranks calculated for the test, along with the corresponding test statistics and p-values. The differences are depicted in the left column and were calculated as follows: $Cost_{PRMC} - Cost_{Other}$.

The T_+ test-statistic is used to determine the p-value under the null hypothesis. The test statistic for the comparison with the MAE (Table 6.3) is equal to 1, with a corresponding p-value of 0.062. This is not smaller than the chosen significance level, and therefore the null hypothesis cannot be rejected. As a consequence, we cannot conclude that using the PRMC metric led to a better tuning of hyper parameters than using the MAE metric. Table 6.4, however, shows the Wilcoxon signed-rank test between the PRMC and the MSE. For these two metrics, the test statistic has a p-value of 0.031. This leads to a rejection of the null hypothesis and the conclusion that tuning with the PRMC metric led to the creation of models that achieve lower costs.

	Difference MAE	**Absolute Difference**	**Rank**
Ridge Regression	3022800.822	3022800.822	1
SVR	-7637878.158	7637878.158	2
XGBoost	-99271149.41	99271149.41	4
Random Forest	-710045012.8	710045012.8	5
LSTM	-9622800.47	9622800.47	3

T_+	1
T_-	14
p-value	0.062

Table 6.3: Wilcoxon signed-rank test: PRMC vs MAE.

Analysis of the model with the lowest cost

The model which achieves the lowest total cost of maintenance is the PRMC-tuned SVR. It achieved a minimal cost of maintenance of 871,155,565 EUR at a threshold of 17. This means that when the model predicts the remaining flights to be less than or equal to 17, maintenance needs to be scheduled. This is also visible in Figure 6.5.

When maintenance is scheduled for a machine, it can happen that it is scheduled at such an early moment in time that the cost of predictive maintenance exceed those of reactive

	Difference MSE	Absolute Difference	Rank
Ridge Regression	-6025874.653	6025874.653	1
SVR	-21609037.29	21609037.29	2
XGBoost	-107170417.6	107170417.6	5
Random Forest	-34613943.32	34613943.32	3
LSTM	-81933756.49	81933756.49	4

T_+	0
T_-	15
p-value	0.031

Table 6.4: Wilcoxon signed-rank test: PRMC vs MSE.

maintenance. This is because in such case, too much RUL is lost. Hence the cost of predictive maintenance can be subdivided in to two classes. Either the cost of PdM is smaller than or equal to the cost of reactive maintenance ($PdM \leq C_r$), or it is not ($PdM > C_r$). This inner cost distribution is depicted in Figure 6.7. The yellow bar corresponds to the case where the cost of PdM exceeds that of RM, the green bar corresponds with cases for which the cost of PdM is lower than that of RM and the red bar depicts the amount of observed failures (RM).

Figure 6.8 shows the distribution of costs for the MSE-tuned Ridge Regression model. Even though the overall costs for this model are larger than those achieved by the PRMC-tuned SVR, it has less instances for which reactive maintenance was scheduled. The Ridge-model is more risk-averse. Although a higher cost is associated with it, in some cases it might be more attractive to select this model. In the commercial aviation industry, for example, it is probable that a more risk-averse model is better because of the safety of the people on board. Figure 6.6 shows the total cost of maintenance for varying thresholds for the MSE-tuned Ridge Regression model.

Conclusion

The Wilcoxon signed-rank test does not provide enough evidence to conclude that using the PRMC leads to models with a significantly lower cost as opposed to using the MAE. However, the difference between the PRMC and the MSE *is* significant at the 5% level. This means that using the PRMC would lead to selecting a model with a lower overall cost of maintenance as opposed to using the MSE.

Even though using the PRMC could lead to lower costs, that does not necessarily mean that there will be fewer failures. It can very well be the case that more machines fail, however, that less RUL is lost for the machines that receive PdM. This can be seen in Figures 6.7 and 6.8. Figure 6.8 corresponds to a model tuned with the use of the MSE, and has fewer cases of RM than the PRMC-tuned model in Figure 6.7, however, a higher cost is associated with it. The costs of these two models can be observed in Figures 6.5 and 6.6 and also in Table 6.2.

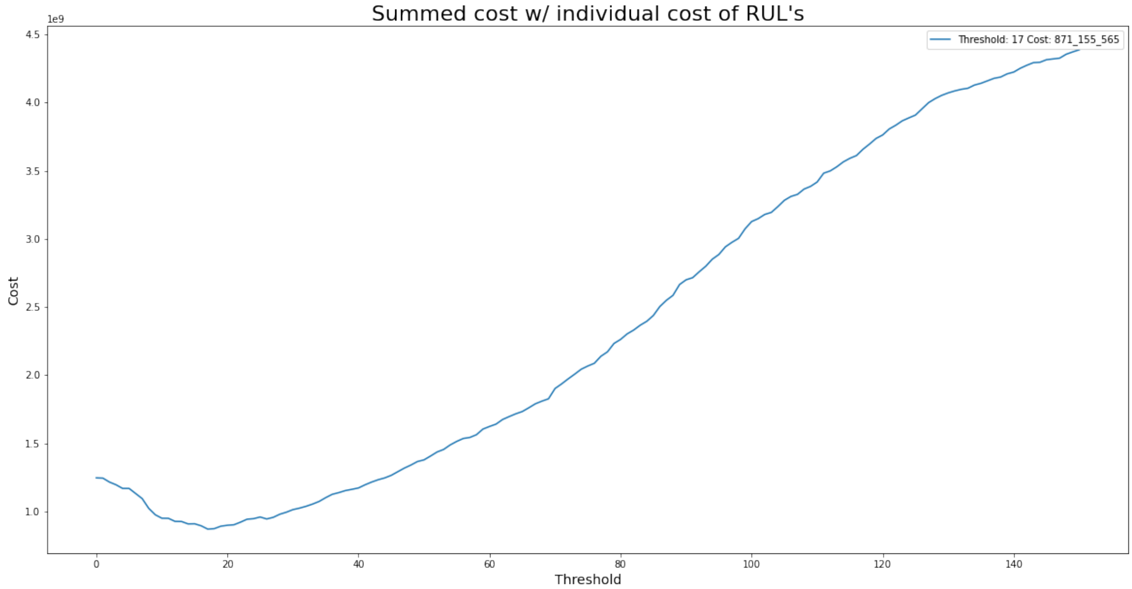

Figure 6.5: Total cost of maintenance per threshold for the PRMC-tuned SVR.

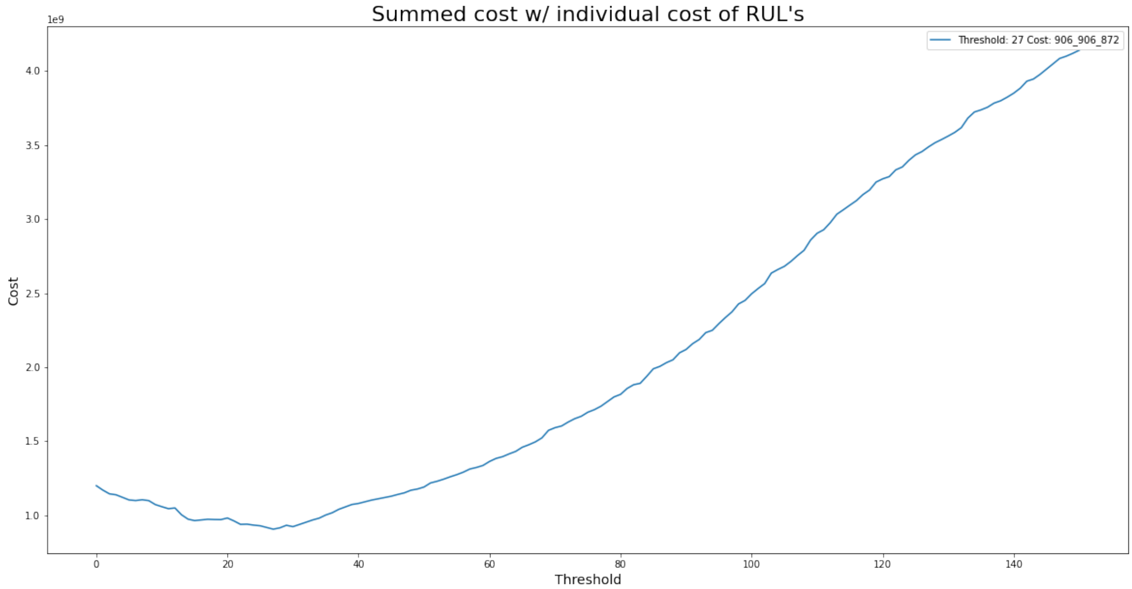

Figure 6.6: Total cost of maintenance per threshold for the MSE-tuned Ridge Regression.

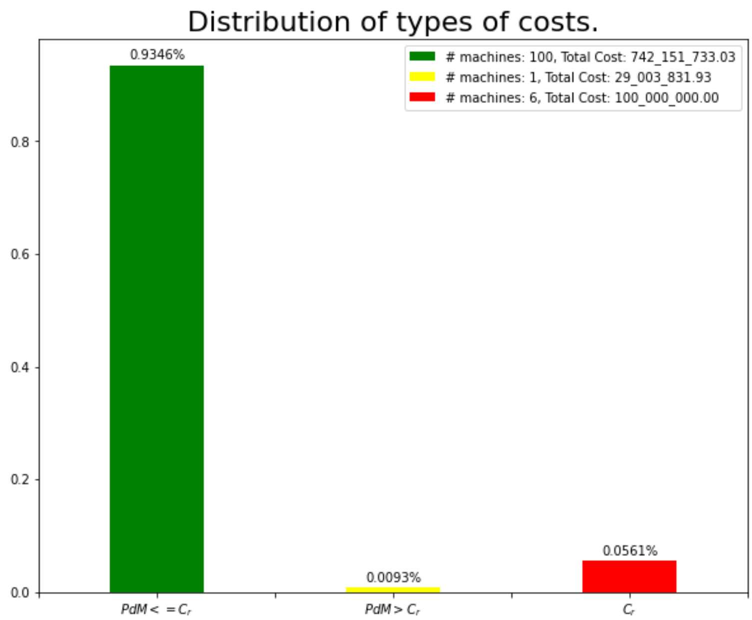

Figure 6.7: Distribution of costs for the PRMC-tuned SVR.

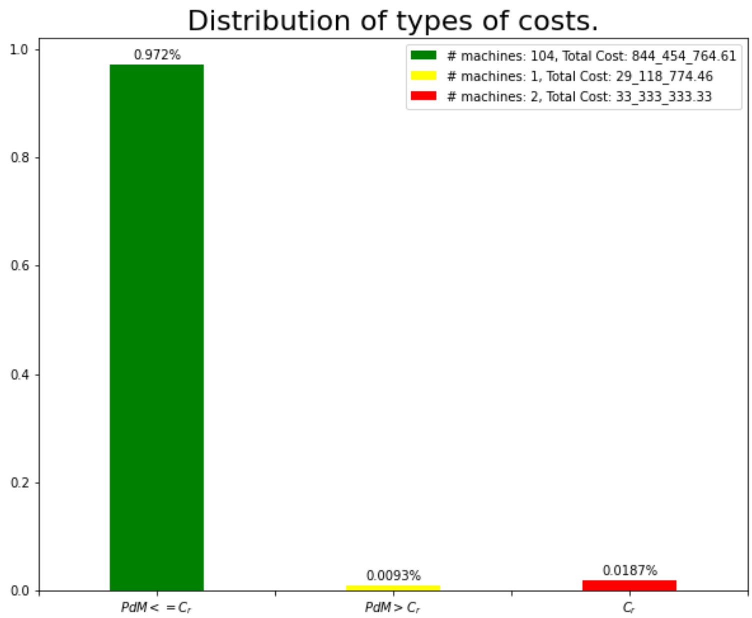

Figure 6.8: Distribution of costs for the MSE-tuned Ridge Regression.

7 Conclusion and further research

In this dissertation, a cost-based metric was proposed for maintenance prediction. As there are significant costs associated with maintenance, the choice of a predictive model needs to be justified by the minimization of these costs. The PRMC reflects the total cost a predictive model accumulates during inference, and hence can be used to determine which model achieves the lowest total cost. As maintenance often needs to be scheduled upfront, a point of scheduling needs to be determined. This is accounted for by means of a threshold, as the PRMC finds the threshold which minimizes the costs for a given model. A new addition to the field is that the PRMC accounts for the lost useful time of a machine, instead of only accounting for C_p and C_r. Another important benefit of the PRMC is that it is easily interpreted by a management audience.

A case study was provided in Section 6, where the PRMC showed to be significantly better than the MSE. There was not enough evidence to reject the null hypothesis when comparing the PRMC and the MAE. It is important to remark that the Wilcoxon signed-rank test has less power than other statistical tests (such as the t-test). This means that, the probability of falsely accepting the null hypothesis is larger for the Wilcoxon signed-rank test than it is for other statistical tests. This is especially a problem for small sample sizes, as is the case in this dissertation (Whitley & Ball, 2002). Therefore, further research could use a more powerful test to analyze the difference between the PRMC and the MAE metrics.

A final and very interesting addition of this dissertation, is the implicit shift towards prescriptive analytics. The management of an organization should not only be interested in what model lowers the costs of a maintenance strategy, but also what actions lower the costs of maintenance. That is, with the continuous monitoring strategy that this dissertation proposes, a manager can gather more information on what actions have a large influence on the $\widehat{\text{RUL}}$. Take Figure 6.4, by identifying the reason for the sudden decrease at time step 83, the manager could get a better understanding of what actions should be avoided.

7.1 Further research

The following points could be interesting for further research:

First, since the PRMC focuses on quantifying maintenance costs, using it as a cost metric will help an organization choose the most cost-effective predictive maintenance model. However, if other factors (such as safety) need to be accounted for, then using the PRMC might be unsuitable. Unless, of course, these safety costs can be quantified in monetary terms.

Second, the PRMC was only analyzed on a data set with homogeneous costs (because all engines had the same acquisition cost). In reality, an organization might buy some of their machines at different prices. This can directly be implemented in the PRMC, however, this was not done in this dissertation.

Third, this dissertation does not analyze the thresholds in a range around 'the optimal threshold' for a model, which could be an alternative to the computationally infeasible EPRMC. Perhaps choosing a model that has a wider range of 'good thresholds' is better in an empirical setting, as the costs deviate less for thresholds near the 'optimal threshold'.

Last, a very interesting addition to the field of predictive maintenance could be applied to situations where no run-to-failure data is readily available (i.e. the true moment of failure is unknown). These situations make it difficult to construct predictive maintenance models and are, unfortunately, commonplace. In such cases, the true moment of failure could perhaps be estimated by use of the methodology in Section 5.5. This would open the door to predictive maintenance for many organisations who do not have data on failures available.

8 Limitations

This dissertation also proposes the EPRMC as a more robust alternative to the PRMC (where τ is treated stochastically) and the iPRMC as a metric of evaluating model performance during inference. Even though both of these metrics were implemented in code, these could not be reported due to computational limits (approximating the integral in the EPRMC simply took too long) and because further research is needed. For the iPRMC, it needs to be empirically verified whether the residuals of a predictive model follow the same distribution as the UL-distribution, whereas for the EPRMC finding a closed form for the integral would be very beneficial.

To use the PRMC as a metric, data on the failures of machines is needed. If there is no data available for the moments of machine failure, then RM will never happen and C_r will not be used. As explained in Section 6.2.3, the PRMC simulates the entire lifetime of a machine up until failure. Therefore, the PRMC cannot be calculated in cases where the failure data is not available.

Lastly, a limitation regarding the case as explored in Section 6. The cost of RUL was derived from the acquisition price, however in the aviation sector, the cost of RUL can also be quantified in 'lost flights'. The cost of not utilizing 1 RUL-unit (i.e. one flight), should also account for the missed revenues of 1 flight. This was not done in this dissertation, but would probably have impacted the results.

9 Final remarks

For an earlier version of this dissertation, custom objective functions were constructed for Tensorflow and XGBoost. This allows for the direct optimization of the PRMC within a model. These objective functions were created with both the possibility for shared costs of RUL and instance-dependent costs of RUL in mind. Instance-dependent costs of RUL are interesting for the field of predictive maintenance, because machines can have different total lengths of life (and thus the cost of RUL would be different for these machines).

All code will be made available on GitHub at: https://github.com/HunkBenny/PRMC

References

Bahrin, M. R. K., Othman, M. H. D., Azli, N., & Talib, M. L. (2016). INDUSTRY 4.0: A REVIEW ON INDUSTRIAL AUTOMATION AND ROBOTIC. Jurnal teknologi, 78(6-13). https://doi.org/10.11113/jt.v78.9285

Bris, R., Châtelet, E., & Yalaoui, F. (2003). New method to minimize the preventive maintenance cost of series–parallel systems. Reliability Engineering & System Safety. https://doi.org/10.1016/s0951-8320(03)00166-2

Buja, A., Stuetzle, W., & Shen, Y. (2005). Loss Functions for Binary Class Probability Estimation and Classification: Structure and Applications. Technical Report, The Wharton School. http://www-stat.wharton.upenn.edu/~buja/PAPERS/paper-proper-scoring.pdf

Chen, C., Shi, J., Lu, N., Zhu, Z. H., & Jiang, B. (2022). Data-driven Predictive Maintenance Strategy Considering the Uncertainty in Remaining Useful Life Prediction. Neurocomputing, 494, 79–88. https://doi.org/10.1016/j.neucom.2022.04.055

Das, S. K., & Abdel-Malek, L. (2003). Modeling the flexibility of order quantities and lead-times in supply chains. International Journal of Production Economics, 85(2), 171–181. https://doi.org/10.1016/s0925-5273(03)00108-7

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. Advances in neural information processing systems, 9.

Erumban, A. A. (2008). LIFETIMES OF MACHINERY AND EQUIPMENT: EVIDENCE FROM DUTCH MANUFACTURING. Review of Income and Wealth, 54(2), 237–268. https://doi.org/10.1111/j.1475-4991.2008.00272.x

Ezik, O. (2003). Calculation of the Actual Cost of Engine Maintenance.

Filho, M. (2023). Do Decision Trees Need Feature Scaling Or Normalization? https://forecastegy.com/posts/do-decision-trees-need-feature-scaling-or-normalization/

Florian, E., Sgarbossa, F., & Zennaro, I. (2021). Machine learning-based predictive maintenance: A cost-oriented model for implementation. International Journal of Production Economics, 236, 108114. https://doi.org/10.1016/j.ijpe.2021.108114

Garrido, F., Verbeke, W., & Bravo, C. (2018). A Robust profit measure for binary classification model evaluation. Expert Systems With Applications, 92, 154–160. https://doi.org/10.1016/j.eswa.2017.09.045

General Electric. (2019). Qatar Airways Orders Additional GE90-powered Boeing 777 Freighters | GE Aerospace. https://www.geaerospace.com/press-release/ge90-engine-family/qatar-airways-orders-additional-ge90-powered-boeing-777-freighters

Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. Machine Learning, 77(1), 103–123. https://doi.org/10.1007/s10994-009-5119-5

Hand, D. J., & Till, R. C. (2001). A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning, 45(2), 171–186. https://doi.org/10.1023/a:1010920819831

Isidore, C. (2020). Boeing's 737 Max debacle could be the most expensive corporate blunder ever. https://edition.cnn.com/2020/11/17/business/boeing-737-max-grounding-cost/index.html

López, J., & Maldonado, S. (2019). Profit-based credit scoring based on robust optimization and feature selection. Information Sciences, 500, 190–202. https://doi.org/10.1016/j.ins.2019.05.093

Memon, O. (2022). How Much Do Jet Engines Cost? Simple Flying. https://simpleflying.com/how-much-do-jet-engines-cost/#:~:text=A%5C%20jet%5C%20engine%5C%20can%5C%20cost,the%5C%20size%5C%20and%5C%20power%5C%20capability.

Mobley, R. K. (2002). An Introduction to Predictive Maintenance.

Neslin, S. A., Gupta, S., Kamakura, W. A., Lu, J., & Mason, C. H. (2006). Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. Journal of Marketing Research, 43(2), 204–211. https://doi.org/10.1509/jmkr.43.2.204

Óskarsdóttir, M., Baesens, B., & Vanthienen, J. (2018). Profit-Based Model Selection for Customer Retention Using Individual Customer Lifetime Values. Big data. https://doi.org/10.1089/big.2018.0015

Praveen. (2020). The most powerful commercial jet engines. https://www.aerospace-technology.com/features/featurethe-most-powerful-commercial-jet-engines-4434003/

Price, M. (2023). Breaking Down the Cost of a Jet Engine. Executive Flyers. https://executiveflyers.com/how-much-does-a-jet-engine-cost/

Provost, F., & Fawcett, T. (2001). Robust Classification for Imprecise Environments. Machine Learning, 42(3), 203–231. https://doi.org/10.1023/a:1007601015854

Ran, Y., Zhou, X., Lin, P., Wen, Y., & Deng, R. (2019). A survey of predictive maintenance: Systems, purposes and approaches. arXiv preprint arXiv:1912.07383.

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. 2008 International Conference on Prognostics and Health Management, 1–9. https://doi.org/10.1109/PHM.2008.4711414

Steel, M. (2020). Your Guide to Jet Engine Maintenance. https://www.marlinwire.com/blog/your-guide-to-jet-engine-maintenance#:~:text=On%5C%20average%5C%2C%5C%20a%5C%20full%5C%20overhaul,provide%5C%20a%5C%20better%5C%20value%5C%20today.

Swanson, L. L. (2001). Linking maintenance strategies to performance. International Journal of Production Economics, 70(3), 237–244. https://doi.org/10.1016/s0925-5273(00)00067-0

Tuovila, A. (2023). What Is Depreciation, and How Is It Calculated? https://www.investopedia.com/terms/d/depreciation.asp

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. European Journal of Operational Research, 218(1), 211–229. https://doi.org/10.1016/j.ejor.2011.09.031

Verbraken, T., Bravo, C. G. B., Weber, R. W., & Baesens, B. (2014). Development and application of consumer credit scoring models using profit-based classification measures. European Journal of Operational Research, 238(2), 505–513. https://doi.org/10.1016/j.ejor.2014.04.001

Verbraken, T., Verbeke, W., & Baesens, B. (2013). A Novel Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models. IEEE Transactions on Knowledge and Data Engineering, 25(5), 961–973. https://doi.org/10.1109/tkde.2012.50

Whitley, E., & Ball, J. K. (2002). Statistics review 6: Nonparametric methods. Critical Care, 6(6), 509. https://doi.org/10.1186/cc1820

Yang, L., Zhao, Y., Peng, R., & Ma, X. (2018). Hybrid preventive maintenance of competing failures under random environment. Reliability Engineering & System Safety, 174, 130–140. https://doi.org/10.1016/j.ress.2018.02.017

Zonta, T., Da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the industry 4.0: A systematic literature review. Computers & Industrial Engineering, 150, 106889.

A Appendix

A.1 Hyper parameter tuning

A.1.1 XGBoost

The hyper parameters for the XGBoost models are presented in Table A.1.

’Columns sampled per tree’ is a number between 0 and 1 indicating what fraction of features should be passed to a weak learner to conduct its splits on. The ’Subsample’ parameter is also a number between 0 and 1, and relates to the number of training instances that are passed to each individual tree to learn on.

	MAE	**MSE**	**PRMC**
Columns sampled per tree	1	1	0.9
Learning rate	0.3	0.1	0.2
Maximum depth	11	4	15
Number of estimators	150	100	100
Subsample	0.9	1	0.6
Window size	40	40	40
Objective	MAE	MSE	MSE

Table A.1: XGBoost hyper parameters

A.1.2 Random Forest

Table A.2 shows the hyper parameters for the Random Forest model.

A.1.3 LSTM

The LSTM-model consists out of 4 layers; the first layer is an LSTM-layer, the second and third layers are dense-layers and the final layer is also a dense-layer with one neuron. For the first, second and third layer, the amount of neurons was treated as a hyper parameter and tuned according to its performance on the validation data set.

Apart from the number of neurons, the amount of dropout for each layer was also tuned. Dropout is a regularization method to avoid overfitting. It is a number between 0 and 1 and it

	MAE	**MSE**	**PRMC**
Maximum depth	15	18	6
Number of trees	100	300	75
Features considered per split	21	10	19
Maximum samples	0.9	1	1
Window size	40	40	40
Objective function	MAE	MSE	MSE

Table A.2: Random Forest hyper parameters

controls the number of neuron outputs that are 'ignored' during training. All hyper parameters are displayed in Table A.3.

	MAE	**MSE**	**PRMC**
Loss function	MAE	MSE	Huber
LSTM neurons	96	64	64
First dense neurons	96	64	64
Second dense neurons	64	128	64
Dropout LSTM	0	0	0
First dense dropout	0.1	0.1	0
Second dense dropout	0	0.1	0
Recurrent dropout	0	0	0
Window Size	40	40	40
Batch Size	256	256	256
Epochs	100	75	150

Table A.3: LSTM Hyper parameters

A.1.4 SVR

Support vector regression is a procedure developed by Drucker et al., 1996. SVR's excel in high-dimensional feature spaces (Drucker et al., 1996). An important hyper parameter is the selected 'kernel', a kernel maps the feature space in to a higher dimensional space. For this dissertation, the best results were achieved with a linear kernel. This kernel has two other important hyper parameters: ϵ and C.

The so-called 'error sensitivity parameter' ϵ defines a 'tube' that is used in the loss function. If a data point lies within the tube, its contribution to the loss function is equal to zero. If it lies outside of the tube, its contribution to the loss function is proportional to its distance from the tube (Drucker et al., 1996). C is a non-zero and strictly positive regularization parameter The larger this parameter is, the more emphasis will lie on the minimization of the squared error Drucker et al., 1996.

The final hyper parameters are depicted in Table A.4.

	MAE	**MSE**	**PRMC**
C	5	11	18
Epsilon	5	6.5	6.5
Window size	40	40	40
Max iter	800	1000	1000

Table A.4: SVR hyper parameters

A.1.5 Ridge Regression

The final hyper parameters can be seen in Table A.5. λ is a hyper parameter that controls the amount of regularization. If λ is equal to zero, Ridge regression will be the same as OLS-regression.

	MAE	**MSE**	**PRMC**
Lambda	1.5	0.05	9
Window size	40	40	40
Max iterations	100	500	100

Table A.5: Ridge Regression hyper parameters

A.2 Dataset

Variable name
Operating condition 1: altitude
Operating condition 2: Mach level
Operating condition 3: temperature
Total temperature at fan inlet
Total temperature at LPC outlet
Total temperature at HPC outlet
Total temperature at LPT outlet
Pressure at fan inlet
Total pressure in bypass-duct
Total pressure at HPC outlet
Physical fan speed
Physical core speed
Engine pressure ratio (P50/P2)
Static pressure at HPC outlet
Ratio of fuel flow to Ps30
Corrected fan speed
Corrected core speed
Bypass Ratio
Burner fuel-air ratio
Bleed Enthalpy
Demanded fan speed
Demanded corrected fan speed
HPT coolant bleed
LPT coolant bleed

Table A.6: All variables in the dataset