ISSN 2683-345X
journal homepage: http://ijiemjournal.uns.ac.rs/

International Journal of Industrial Engineering and Management

Volume 15 / No 4 / December 2024 / 320 - 337

Original research article

Case Studies of Survival Analysis for Predictive Maintenance in Manufacturing

S. Softic[a,*] 0000-0003-2949-8239, B. Hrnjica[b] 0000-0002-3142-1284

[a] *IT & Business Informatics, CAMPUS 02 University of Applied Sciences, Graz, Austria;*
[b] *University of Bihac, Faculty of Technical Engineering, Bihac, Bosnia and Herzegovina*

ABSTRACT

The Predictive Maintenance (PdM) as a tool for detecting future failures in manufacturing was recognized as an innovative and effective method. Different approaches for PdM have been developed to compromise the availability of data and the demanding needs for probability estimation. The Survival Analysis (SA) method was used in this paper for the probability estimation of machine failure. The paper presents the use of the two most popular SA models: Kaplan-Meier non-parametric and Cox proportional hazard models on two different datasets to present the methodology and the possibilities for applications in manufacturing. By using the first SA model, the results show the probability of a machine or component part to survive a certain amount of time. The Cox proportional model was used to find out the most significant covariates in the observed dataset which have an influence on survival time. The analysis showed that the use of SA in the PdM is a challenging task and can be used as an additional tool for failure analysis and maintenance planning.

ARTICLE INFO

Article history:

Received July 18, 2024
Revised August 11, 2024
Accepted September 14, 2024
Published online November 6, 2024

Keywords:
Survival analysis;
Predictive maintenance;
Machine learning

*Corresponding author:
Selver Softic
selver.softic@campus02.at

1. Introduction

Maintenance plays a vital role in the manufacturing process and can be defined as the set of activities that preserve the system in the functional state [1]. The maintenance can be classified in different ways depending on how it is performed on the manufacturing system [2]. Only a small number of maintenance types can be performed without interrupting the manufacturing process [3]. However, in most cases the maintenance must be performed only if the production process is shut down. When the maintenance requires production stopping it may lead to tension between the production and the maintenance departments in a way that the production needs quality and reliable maintenance without production interruption or at least with minimum stopping time [4]. Integrated production maintenance and quality management ensures maximum production efficiency and profitability [5]. The worst-case scenario may happen when the production stops due to the component failure [6]. Such a case should never happen in production with well-planned and organized maintenance. Being able to create a perfect maintenance plan can cost more than planned by budget [2]. It is obvious that good maintenance planning can offer nearly perfect production with minimal production interruption. The maintenance plan depends on adopted strategies and methods for maintenance

Published by the University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia.

https://doi.org/10.24867/IJIEM-2024-4-366

time calculation and prediction. Instead of relying on planned maintenance entirely the ability to predict possible failures in the manufacturing process has become largely popular in the last decades. Knowing the possible failure of the specific component before it occurs leads to multiple benefits in the production such as reduced maintenance costs improved production quality and increased the productivity [1].

Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) improved the maintenance process primarily in the ability to detect and predict errors before failure occurs [7]. This kind of maintenance is called Predictive Maintenance (PdM) and it requires the equipment to provide data from sensors monitoring the equipment as well as other operational data [2]. In other words, it is a technique used to determine (predict) the failure of the machine component in the near future so that the component can be replaced based on the maintenance plan before it fails and stops the production process. Different types of maintenance with the ability to improve the production process are presented in Figure 1.

The PdM uses data collected from various sensors installed on the machines. The sensors are built into the Internet of Things (IoT) devices that send data in the cloud. Once the data are in the Cloud different cloud solutions can use the data for processing and analysis. PdM can be implemented in the cloud solution as a part of Industry 4.0. Many different Cloud based PdM solutions have been implemented and published through the literature [11]-[14]. Annamalai, Udendhran, and Vimal [11] presented a detailed examination of cloud-based PdM and machine monitoring solutions specifically tailored for the auto-mobile industry. The relevance of this work is its specific focus on an industrial application, showcasing the practical benefits and challenges of implementing PdM solutions in a cloud environment. It provides a concrete example of how cloud technologies can be utilized to enhance maintenance operations and improve the reliability of manufacturing processes. Hrnjica and Mehr [12] discussed the application of deep learning techniques for energy demand forecasting, which is closely related to PdM. Their work highlights the importance of Explainable Artificial Intelligence (XAI) and Cloud Computing in making predictive models more transparent and accessible. The relevance here is the emphasis on the cloud's role in managing and processing large datasets necessary for accurate and reliable PdM models. Paolanti et al. [13] focused on the application of machine learning approaches for PdM within the context of Industry 4.0. The relevance of this work lies in its demonstration of how cloud-based systems can leverage machine learning algorithms to analyze vast amounts of sensor data for predicting machine failures. The integration of these systems into the cloud allows for scalable and flexible maintenance solutions, essential for modern manufacturing environments. Previous study explored the integration of PdM systems with Cloud Computing, emphasizing how cloud-enhanced systems can improve maintenance scheduling and fault detection [14]. The findings are particularly relevant because it outlines the practical implementation of cloud-based PdM systems, demonstrating the benefits of real-time data processing and analytics provided by cloud infrastructure.

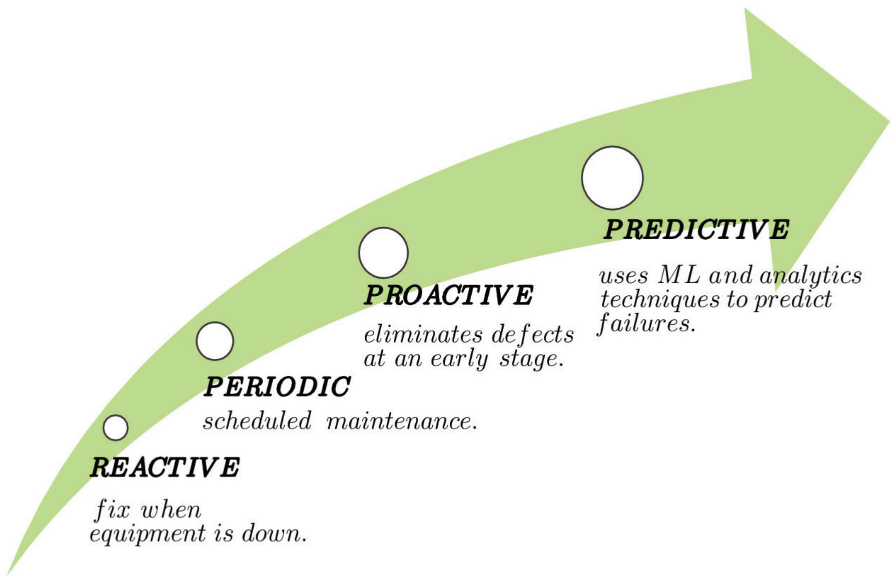


Figure 1. Different types of maintenance [2]

In the realm of PdM, leveraging advanced statistical and machine learning techniques has been a focal point of research, as evidenced by numerous studies emphasizing data-driven approaches for fault detection and maintenance optimization. Amidst this landscape, Survival Analysis (SA) presents a compelling methodology for several reasons. SA is a robust statistical method traditionally used in medicine to estimate the time of an event, such as the occurrence of a disease. SA is highly compelling for PdM due to its proven robustness in handling time-to-event data and censored observations, as evidenced in biomedical applications. Unlike traditional regression models, SA explicitly models the time until a machine component fails, enhancing maintenance scheduling and resource optimization. The Kaplan-Meier estimator and Cox proportional-hazards model adeptly manage censored data, which is common in industrial settings where preemptive maintenance occurs. Moreover, the Cox model identifies significant covariates affecting failure rates, providing valuable insights for targeted interventions. Integrating SA with Cloud Computing and AI, as highlighted in studies by Schmidt and Wang [14] and Hrnjica and Mehr [12], further refines predictive capabilities, making SA a versatile and powerful tool across various industries, from manufacturing to automotive. This integration aligns with the growing trend of data-driven and AI-enhanced industrial practices, demonstrating SA's potential to significantly advance PdM strategies. This paper explores the application of SA in the manufacturing sector for PdM, focusing on two of its most popular models: the Kaplan-Meier estimator and the Cox proportional-hazards model. These models are applied to two datasets: the NASA Turbofan Engine Degradation Simulation Data Set (NTED) and the Predictive Mainte-nance Modelling Guide Data Set (PRMM).

This manuscript presents the following contributions to the field of PdM in manufacturing by (1) demonstrating the application of Kaplan-Meier and Cox proportional-hazards models to PdM showcasing their potential in estimating the survival probabilities and identifying significant factors influencing machinery failures, (2) providing a detailed methodology for transforming raw datasets into formats suitable for SA – this involves preprocessing steps to identify failure times and censoring events crucial for accurate survival modeling, (3) applying the SA methods to two distinct datasets the NTED and PRMM highlighting the differences in data collection approaches and their implications for PdM – this comparative analysis offers insights into how different data structures and information availability impact the effectiveness of SA models, (4) utilizing the Cox proportional-hazards model to identify significant sensor readings and other covariates that influence the survival times of machinery components – this information is valuable for maintenance planning and decision-making enabling targeted interventions to extend the lifespan of critical components, (5) discussing the practical implications and challenges of implementing SA models in a real-world manufacturing setting, and (6) highlighting the potential impact of Kaplan-Meier and Cox models on maintenance planning including optimizing the timing of maintenance activities and reducing the risk of unexpected failures.

The structure of the following sections is organized as follows. Initially, we provide a review of the existing literature on PdM, SA, and associated technologies. Subsequently, we detail the methodology applied in the case study. The manuscript concludes with two sections: a comprehensive description of the case studies, and the discussions and the conclusions that can be drawn, including the limitations and outlook.

2. Literature Review

2.1 Importance of Maintenance

According to recent investigations and depending on the industry, maintenance in general represents between 15% and 70% of overall costs in industry [15]. As stated by Mobley [16], the surveys conducted on maintenance show that about one third of spending on maintenance in the US is wasted because of unnecessary activities. Overall, the literature distinguishes two main categories of maintenance policies: corrective and preventive maintenance. Corrective maintenance (also referred to as reactive maintenance) follows the run-to-failure philosophy, which often leads to costly repairs and manufacturing reduction. The preventive maintenance policies are also called proactive maintenance, i.e. an attempt to prevent fatal failure occurrence [16].

2.2 Predictive Maintenance and Prognostic Technologies

PdM as such can be seen as a preventive maintenance policy which allows estimation and prediction of the remaining useful life of the machinery in production. The idea of PdM goes back to the early 1990's and expands regularly scheduled preventive maintenance. PdM as an approach is based on continuous monitoring of sensors of relevant machine parameters such as vibration and temperature. However, monitoring alone is not sufficient. Which maintenance tasks should be performed first and where to focus improvement efforts plays a very important role. Integration into a computerized system allows real-time task prioritization and performance monitoring, enhancing productivity and maintenance efficiency.

Prognostic techniques are a very important part of PdM. According to results of the recent survey on PdM in Industry 4.0, which analyzed over 150 papers on this topic, the number and diversity of prognostic technologies are enormous [17]. Clear classification of the different techniques is not easy to make. One common thing for all techniques is that they rely on monitored data. According to Krupitzer et al. [17], most of the approaches (around 74%) are data driven. Other approaches distinguished during the survey have been model-based approach and non-prognostic approach. Most applied techniques are regression based (multiple) linear regression, logistic regression, regression trees, random forest, and others). Likewise, widely used are the Bayesian probabilities. A special subcategory of data-driven approaches is artificial intelligence, which makes up 25% of investigated examples [17]. In this relation, the most frequent term occurs to be neural networks. In addition, around 37% of research papers used real data to evaluate the results and simulation-based evaluation.

In general, PdM extracts insights from the data and acts on them. The PdM can improve the production process and increase productivity. By successfully handling PdM, we can achieve the following:

- Reduce the operational risk of mission-critical equipment
- Control cost of maintenance by enabling just-in-time maintenance operations
- Discover patterns connected to various maintenance problems
- Provide Key Performance Indicators

Depending on the application scenario and variety of data as well as chosen techniques, different amounts of data have been applied to achieve the desired results. By using continuous production monitoring, the IoT, Big Data, and Cloud-based services can bring the latest states of production in the form of data sets. Once the data sets contain the latest information about the production, the developed models can extract information which is relevant to the current production state.

2.3 Predictive Maintenance and Industry 4.0

Without a doubt, the PdM plays a major role in Industry 4.0. The Industry 4.0 key components Big Data, AI, ML, Cloud Computing, and IoT offer possibilities to integrate PdM and connect it to other systems of the production process. Those technologies along with the produced data are also increasingly used by PdM to predict machine failure and related maintenance parameters. By using IoT and Cloud Computing, PdM reaches its full potential. Combination of Big Data ecosystem orchestrated through the IoT for various PdM approaches in industrial IoT-based smart manufacturing can be found in literature. In most cases, the IoT provides real-time data while the big data ecosystem provides predictive analytics algorithms in order to dynamically manage preventive maintenance and failures. Beside predictive algorithms, the ecosystem contains numerous technologies including big data ingestion, integration, transformation, storage, analytics, and visualization in a real-time environment using various technologies such as the Data Lake, NoSQL database, Apache Spark, Apache Drill, Apache Hive, OPC Collector, and other techniques [11], [18]-[20].

The data collection with noise (e.g. temperature and vibrations sensors) are nowadays easily applicable and cheap. In the case of PdM, Kommenda and Strumpf [21] provided a successful example of various regression-based predictions (e.g. Coke Quality Prediction, Design of friction systems as well as PdM). The IoT devices generate lots of data which are in the most cases identical, or the values are in the expected value range. This is because most of the time the production machines work in the same healthy conditions. From the dataset (software) point of view, such conditions generate identical copies of the production systems. However, in hardware sense (production machines) identical does not mean completely the same. This means that even if the systems are identical, it does not mean that they will behave in the same manner. It is more likely that depending on operating conditions and different humans as operators, most of them will, even by same parameters, behave differently. From the experience of real production systems, one can find out that the quality of manufactured parts and machine life also strongly depends on operators [22].

2.4 Using Survival Analysis for Predictive Maintenance

PdM based on SA can be implemented on different levels. In most cases, PdM based SA is used for the calculation of various survival curves and then incorporated into ML models. Survival curves are used to calculate the probability of defects derived from the Cox proportional-hazards model. Although in most cases SA is used in medical research of various diseases such as Alzheimer's, cancer, leukemia, and others, it is possible to find examples of successful application in production and related areas. There are few examples where PdM based SA is used in Cloud based Solutions using AI, ML, Big Data, IoT, and other Industry 4.0 based technologies. This paper presents an approach of using PdM based SA in the cloud-based solution [23]-[25].

There are few examples where PdM based SA is used in Cloud based Solutions using AI, ML, Big Data, IoT and other Industry 4.0 based technologies [20], [24], [26], [27]. This paper presents an approach of using PdM based SA in the cloud-based solution.

2.5 Survival Analysis as Method

SA is a popular data analysis method that first appeared in bio science and medicine science, and later expanded to other scientific fields. SA tries to estimate the time to event data. The time to event T can be anything related to maintenance such as: duration of proper operation of the machine, frequency of the machine failure, duration of the last maintenance etc. It is always a positive value. In the context of science, survival means probability [28]. There are four mayor probability functions frequently used in the SA. In SA, the survival function S is defined as function of time S(t). Generally, the survival function S(t) is defined as the probability for survival after time t of the random variable:

$$S(t) = \Pr(T > t), 0 < t < \infty \tag{1}$$

Obviously, S(0)=1, which indicates that the survival function is related to the lifetime distribution function. In SA, the cumulative distribution function F represents the probability that the event variable occurs earlier than t.

$$\mathrm{F(t)} = 1 - \mathrm{S(t)} \tag{2}$$

The first derivative of cumulative density function defines the death density function (DDF) which is expressed as:

$$f(t) = \frac{\mathrm{d}}{\mathrm{dt}}\,\mathrm{F(x)} = -\frac{d}{dt}S(t) \tag{3}$$

Hazard density function h(t) represents the probability for the event to occur in the next instant, given survival time t:

$$\mathrm{h(t)} = \frac{\mathrm{f(t)}}{\mathrm{S(t)}} = \frac{1}{\mathrm{s(t)}}\left[-\frac{\mathrm{d}}{\mathrm{dt}}\,\mathrm{s(t)}\right] = -\frac{\mathrm{d}}{\mathrm{dt}}\ln(\mathrm{s(t)}) \tag{4}$$

3. Methodology

This section provides a detailed explanation of the methodology adopted for applying SA to PdM in manufacturing. The research involved several key steps including data preprocessing, model application, and result interpretation as well as the conceptual design of the case studies conducted.

3.1 Conceptual Design of Case Studies

The research conducted in this manuscript revolves around two identical primary case studies using distinct datasets: the NASA Turbofan Engine Degradation Simulation Data Set (NTED) and the Predictive Maintenance Modelling Guide Data Set (PRMM). Detailed Datasets description is included in the case studies section. The conceptual design of these case studies includes the following phases:

(1) Selection of Datasets – Two datasets were selected based on their relevance and availability for PdM. NTED provides detailed sensor readings from aircraft engines while PRMM offers comprehensive telemetry and maintenance records for CNC machines.

(2) Data Preprocessing – Raw data from both datasets were transformed to fit the requirements of SA models. This included handling missing values, normalizing sensor readings, encoding failure events and censoring information.

(3) Model Implementation – The Kaplan-Meier and Cox proportional-hazards models were applied to the preprocessed data to estimate survival probabilities and identify significant predictors of machinery failure.

(4) Evaluation and Analysis – The models were evaluated using statistical tests and performance metrics. The results were analyzed to draw meaningful conclusions about the applicability and effectiveness of SA in PdM.

3.2 Data Preprocessing

Preprocessing is a critical step to ensure the datasets are suitable for SA. For each dataset the following preprocessing steps were performed: (1) NTED Dataset and (2) PRMM Dataset.

3.2.1 In NTED Dataset

Event Definition: The time to event (failure) was recorded as the number of cycles until failure for each engine. Censoring was identified based on whether an engine survived beyond the observation period.

Data Transformation: The dataset was restructured to include columns for engine ID, cycles, failure event (1 for failure, 0 for censored), and sensor readings. Constant and empty columns were removed, and the analysis was constrained to the first 220 cycles.

3.2.2 In PRMM Dataset

Merging Datasets: Telemetry, maintenance, and failure records were merged using machine ID and date as key columns. This integration ensured that each record included comprehensive information about machine conditions and events.

Feature Engineering: Rolling means and standard deviations for sensor readings (voltage, vibration, pressure, and rotation) were calculated over various time intervals (3, 6, 12, 18, and 24 hours). These derived features were used as covariates in the SA models.

Event Definition: Failure events were identified based on maintenance records with censoring information derived from scheduled maintenance activities.

3.3 Model Implementation

The model implementation consists basically of applications of Models: Kaplan-Meier Estimator and Cox Proportional-Hazards Model on NTED and PRMM Datasets. In NTED Dataset, the survival probabilities for each engine were calculated showing the likelihood of engines remaining operational over the observed cycles. Confidence intervals were computed to assess the uncertainty of the survival estimates. As for the PRMM Dataset, Kaplan-Meier curves were plotted for different machine components allowing for comparison of their survival times. The model provided insights into the reliability of each component, informing maintenance schedules and spare parts management.

3.4 Evaluation and Analysis

The performance and reliability of the models were evaluated using various metrics: (1) Kaplan-Meier Estimator – Survival curves and confidence intervals were analyzed to understand the reliability and failure patterns of machinery. The area under the survival curve was assessed to compare the overall survival probabilities of different components. (2) Cox Proportional-Hazards Model – The significance of covariates was determined through p-values from statistical tests. The concordance index was calculated to measure the predictive accuracy of the model, with values closer to 1 indicating better performance.

3.5 Practical Implementation

The practical implementation of these models involved using relevant libraries for data processing, SA, and visualization.

4. Case Study

This section presents the application of SA in PdM of the manufacturing processes by using two common types of datasets, as referred to in above section 3.2. Usually, datasets for the SA analysis consist of the time to event and event status as well as the censoring status. The event status is usually a Boolean variable (true/false) or (1/0) indicating the occurrence of the event. By value 1 or "true" the event has occurred, but it is not censored whereas 0 or "false" indicates non-occurrence of the event which is also censored. To analyze any process by SA the two columns must be presented in the datasets. However, such information is naturally presented when dealing with drug analysis, patient treatment, or related subjects where the SA is used most. However, in the manufacturing processes, this is not the case because the recorded information is usually the consequence of the event e.g., information about the number of cycles before fail [29] or replacements of the component due to planned maintenance or failure [30].

To use SA in PdM the first step is to transform the data into SA compatible dataset which clearly presents the censored and non-censored events as well its duration. Besides the mentioned variables, the datasets should contain other related columns which can be used to build survival regression. By building survival regression models, SA can answer the question of what caused in a certain amount of the probability the duration of the event.

4.1 Data Set Description

The first dataset used in the paper is NASA Turbofan Engine Degradation Simulation Data Set (NTED) which is publicly available at NASA repository. The dataset contains several aircraft engines monitored throughout usage history. Each engine was employed under different flight conditions while the 21 sensors recorded various states of the engine. Depending on the recorded sensor values the amount and rate of damage accumulation can be obtained for each engine. Data generation and specific meaning of the sensors are described in the literature [29].

4.1.1 NTED Dataset

The dataset is grouped by operating conditions and fault modes which are stored in several text-based files. The paper presents the application of SA by using the first group of data set FD001 consisting of 100 different engines, one operating condition, and one fault mode. The first group of FD001 dataset consists of three files which are related to the training and testing sets. The complete list of all files for NTED dataset and other Meta information are shown in Table 1 [31].

The data set describes the usage history of the 100 engines in the form of run-to-failure event records. This means each record contains the time (cycles) until the engine failed. In addition to the engine identification number and cycle number, the dataset contains 21 sensor readings presented in Table 2 [31].

4.1.2 PRMM Dataset

The second dataset is Predictive Maintenance Modelling Guide Data Set (PRMM) which is also publicly available at Microsoft® Azure Gallery. Dataset contains information of telemetry, maintenance, failures, and machine properties about 100 CNC machines during the production process from January to December of 2015 [29]. Unlike previous, this dataset contains information about 4 components in each CNC machine which are subject of failure and replacement. The information about maintenance, failure and telemetry are stored in different csv files with proper machine ID and date. This means that every record contains the machine ID and date, which can be key columns for merging and joining data. The telemetry data contains sensor readings from 100 CNC machines about voltage, vibration, pressure and rotation. The replacement of the components in each machine was recorded in the maintenance table. The replacement is performed due to scheduled maintenance or due to failure. The failure information of each machine is recorded in the failure table. Each machine failure record has date, machine ID and failed component. Machine dataset contains information about model and age of 100 CNC machines. The error table contains the logged errors of each machine. The error recorded in the machine is usually a response of the machine because of some uncommon state, and in most cases is not caused by a machine failure. The details about PRMM datasets are presented in Figure 2.

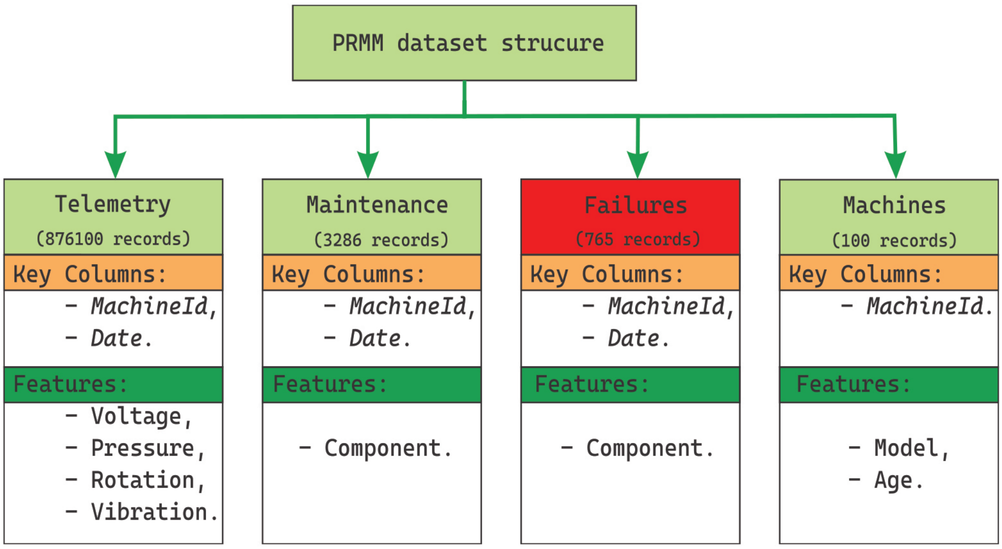
The datasets represent a different approach to collecting data about machine history. None of the datasets are ready to be included in the SA process and they should be carefully analyzed before the SA starts. In the first case, the dataset does not contain censored information. In the second case, the dataset does not contain the duration of the event. To successfully apply the SA on the presented datasets, the data preparation step should be performed to identify the event duration of both censored and non-censored events as well as set of covariates for building survival regressions models.

Table 1. NTED dataset description

Dataset parameter	FD001	FD002	FD003	FD004
# of engines	100	260	100	249
training size	20631	53579	24270	61249
test size	100	259	100	248
# of columns	26	26	26	26
avg. lifespan	206	206	206	206
# operations cond.	1	6	1	6
# faults cond.	1	1	2	2

Table 2. Description of columns for the NTED dataset

Sensor ID	Symbol	Description	Unit
1	T2	Total temperature at fan inlet	°R
2	T24	Total temperature at LPC outlet	°R
3	T30	Total temperature at HPC outlet	°R
4	T50	Total temperature at LPT outlet	°R
5	P2	Pressure at fan inlet	Psia
6	P15	Total pressure in bypass-duct	Psia
7	P30	Total pressure at HPC outlet	Psia
8	Nf	Physical fan speed	Rpm
9	Nc	Physical core speed	Rpm
10	Epr	Engine pressure ratio	-
11	Ps30	Static pressure at HPC outlet	Psia
12	Phi	Ratio of fuel flow to Ps30	Pps/psi
13	NRf	Corrected fan speed	Rpm
14	NRc	Corrected core speed	Rpm
15	BPR	Bypass ratio	-
16	farB	Burner fuel-air ratio	-
17	htBleed	Bypass enthalpy	-
18	Nf_dmd	Demanded fan speed	Rpm
19	PCNfR	Demanded corrected fan speed	Rpm
20	W31	HPT coolant bleed	lbm/s
21	W32	LPT coolant bleed	lbm/s



Note. The structure of the data set depicted as one at the top and four at bottom frames connected by arrows. The top frame holds the name of the dataset, while lower frames hold the information of the table name, key columns and feature columns.

Figure 2. PRMM dataset structure

4.2 Data Preparation for SA

To perform SA on the presented datasets, the time to event and the indicator for censoring occurrence variables have to be defined. From the NTED data set description file, the train dataset collects information which monitored each engine till the failure time. This means that for each engine the training dataset contains failure information, presenting the maximum cycle. In order to prepare the dataset to contain either non-censored or censored information, the SA should be defined as the time interval [0, tn]. Details about data preparation can be found in the literature [32]. Table 3 shows NTED dataset after preparation where constant and empty columns were re-moved and the analysis constrained to 220 cycles.

The data preparation for the PRMM dataset requires more data transformation than the previous case. The datasets must be joined, and the information should be transformed. The rolling mean and standard deviation were calculated on the telemetry data set for each three, six, twelve, eighteen and twenty-four hours. Subsequently, the telemetry data was merged with the failure, maintenance and machine data. The most important tables of the PRMM dataset are maintenance and failure. The two tables are merged and created a new column which represents the calculated time till component replacement. By joining the failure table, each replacement record is identified if caused by the failure or it was scheduled maintenance event. The first event is identified as failure event without censoring, while former event was identified as non-occurring event with censoring status. Unlike the previous one, PRMM dataset contains information about censoring events, so the SA was performed on whole time interval.

4.4 Kaplan-Meier Non-Parametric Model

The Kaplan-Meier survival model [33] represents a non-parametric model that estimates the survival function from the lifetime data set. The model defines the survival function S(t) which represents the probability that life is longer than time t:

$$\hat{S}(t) = \prod_{i,t_i \leq t} \left(1 - \frac{d_i}{n_i}\right), \quad (5)$$

where,
$\hat{S}(t)$ - Kaplan-Meier survival probability function,
t_i - is a time that at least one event happened,
d_i - the number of events occurred at t_i,
n_i - the number of machines at risk at time t_i,

Table 3. The first several rows of the NTED after transformation and analysis

Id	Cycles	event	T24	T30	T50	P30	Nf	Nc	Ps30	Phi	NRf	NRc	BPR
1	192	1	643.54	1601.41	1427.2	551.25	2388.32	9033.22	48.25	520.08	2388.32	8110.93	8.5113
2	220	0	642.87	1600.76	1406.68	552.92	2388.09	9068.44	47.54	520.77	2388.11	8148.15	8.4123
3	179	1	643.51	1604.8	1428.23	551.91	2388.14	9197.52	48.09	519.53	2388.2	8255.34	8.5056
4	189	1	644.53	1612.11	1432.55	551.93	2388.13	9198.32	48.15	519.84	2388.16	8259.42	8.5246
5	220	0	642.63	1591.54	1412.16	553.48	2388.13	9102.11	47.59	521.68	2388.11	8173.86	8.4664
6	188	1	643.47	1600.63	1434.92	550.7	2388.34	9027.96	48.21	518.98	2388.34	8102.82	8.5358
7	220	0	642.89	1596.6	1416.42	553.2	2388.14	9073.62	47.75	520.9	2388.16	8145.22	8.4495
8	150	1	643.87	1602.84	1432.31	551.08	2388.27	9046.16	48.24	519.57	2388.26	8121.27	8.5509
9	201	1	644.04	1595.36	1428.43	552.3	2388.5	9239.76	48.11	520.28	2388.56	8289.63	8.5156
10	220	0	643.46	1604.91	1422.76	551.73	2388.25	9115.15	47.95	520.23	2388.25	8181.11	8.5182
..													
99	185	1	643.93	1598.42	1421.56	550.64	2388.29	9050.61	48.29	519.99	2388.24	8127.53	8.5425
100	200	1	643.85	1600.38	1432.14	550.79	2388.26	9061.48	48.2	519.3	2388.26	8137.33	8.5036

Table 4. The first several rows of the PRMM after merging machine info, telemetry, maintenance and failure data

machineID	datetime	comp	failure	time	model	age	voltmean_3hrs	rotatemean_3hrs	...
1	5/Jan/2015 6:00	comp4	1	4152	model3	18	186.444687	452.190186	
1	5/Jan/2015 6:00	comp1	0	552	model3	18	186.444687	452.190186	
1	20/Jan/2015 6:00	comp3	0	4152	model3	18	169.748245	436.644806	
1	20/Jan/2015 6:00	comp1	0	360	model3	18	169.748245	436.644806	
1	4/Feb/2015 6:00	comp4	0	720	model3	18	174.022934	463.676208	
...									
100	9/Dec/2015 6:00	comp2	0	3240	model4	5	162.885956	396.215607	
100	24/Dec/2015 6:00	comp2	0	360	model4	5	158.8302	471.187683	

In non-parametric models there are no any specific mathematical relations for the probability density, and they are known to be distribution-free. One of the widely used such models is Kaplan–Meier estimator for modelling survival distributions. During the time t_i, the parameter n_i has cumulative values of the total number of machines, so that it can be simply calculated as:

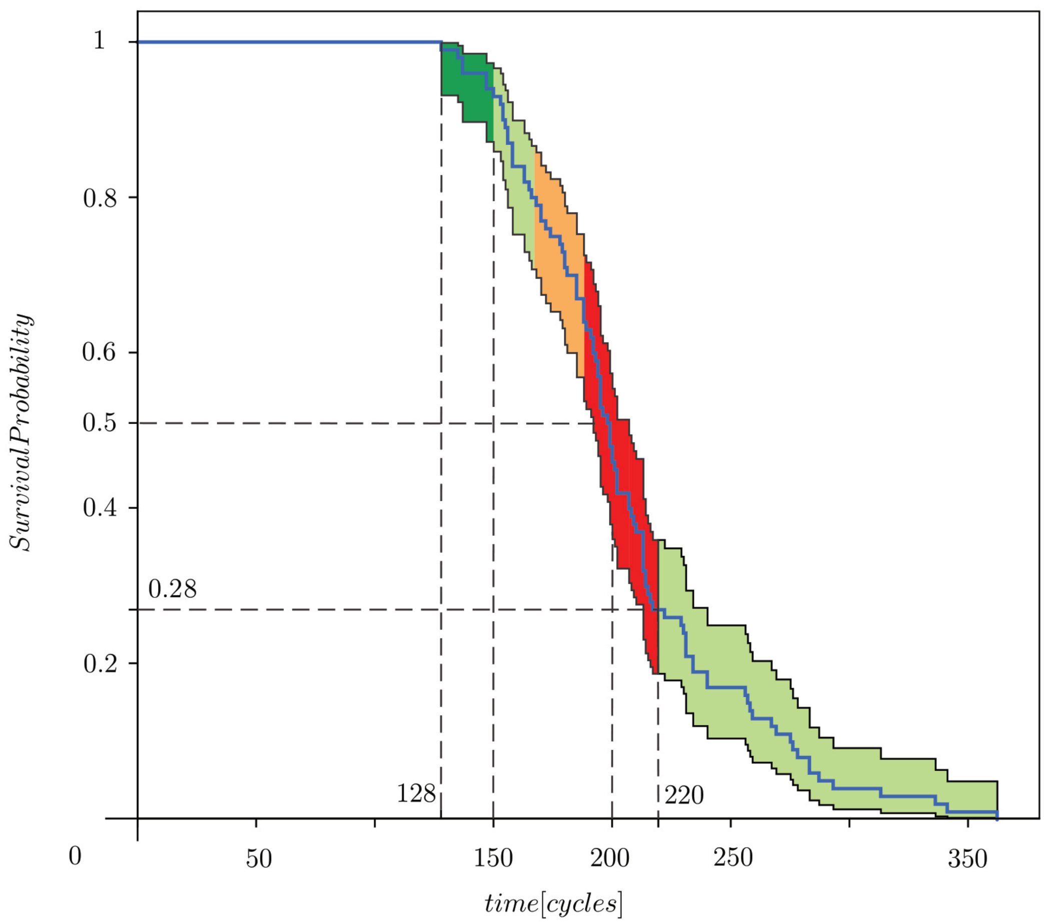
$$n_{i+1} = n_i - d_i. \quad (6)$$

Figure 3 shows the Kaplan-Meier probability function for NTED dataset [32].

The Kaplan-Meier model shows that all 100 engines will survive the first 128 cycles. After 150 cycles, machines have 90% chance to be in functional state, and almost 50% chance to survive after 200 cycles. The survival probability after 200 cycles decreases rapidly, to 220 cycles which shows the probability of 28%. The model was calculated with 95% confidence interval, as depicted in Figure 3. The confident level is colored by different color at different time cycles indicating if the probability confidence is good. The confidence level at around 200 cycles is larger than at other time cycles indicating more uncertainty.

The NTED dataset-based Kaplan-Meier model can improve maintenance in the context of planning and reduce the number of engines at risk based on the probability distribution and confidence level. The impact of the model can be in reducing the maintenance cost due to decreased number of the engine at risk.

The Kaplan-Meier model for PRMM dataset was created using different component failures. In this way, the model can differentiate between component failures as well as detect which component has



At risk	100	93	46	17	1
Censored	0	0	0	0	0
Events	0	7	54	83	99

Note. The figure contains a survival curve plot with confident intervals around it colored in four different colors at different time cycles indicating if the probability confidence is good or bad. The below part represents information about the number of machines a risk, the number of censored and number of events each certain number of cycles.

Figure 3. Kaplan-Meier probability function diagram for NTED dataset

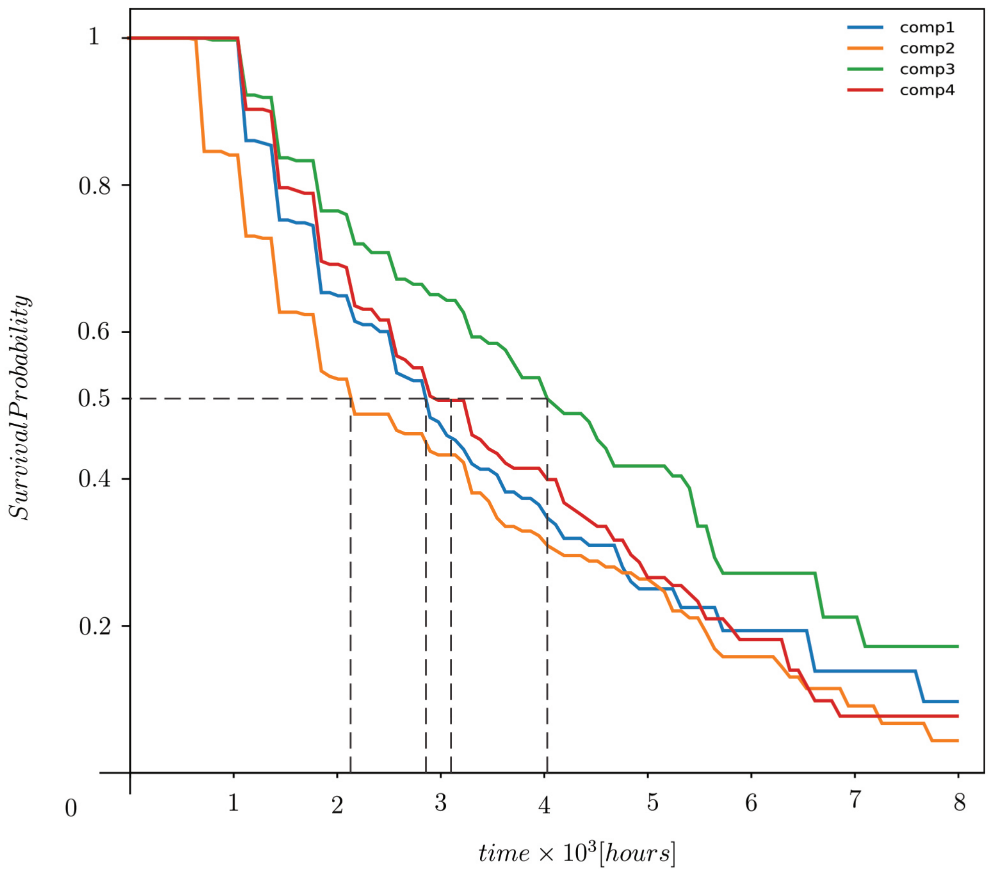
less chance to survive certain time duration. Figure 4 shows Kaplan-Meier models for four components of the PRMM dataset. Component “comp2” has lower survival time than the other. All components have 100% probability to survive 1000 hours, except “comp2”. Furthermore, component “comp2” has a 50% probability to survive 2200 hours, while component “comp3” has the same probability of surviving 4200 hours, which is double than the former.

As mentioned above, component “comp3” has the longest survival time, as identified in the survival plot. The analysis of the model also shows that except component “comp2”, other components have the same initial time of approximately 1200 hours with no failure. The superiority of the component “comp3” can be identified after 2000 hours, whereas component “comp2” has 20% higher probability to survive certain numbers of working hours.

To differentiate working hours between machine components, we can estimate working hours for 50% changes that the component will survive. Figure 4 depicts that component “comp2” can last 2500 hours before fail, whereas component “comp3” can last 4200 hours, which means it has approximately double work life. This information is important for the maintenance department, for orders of specific machine components.

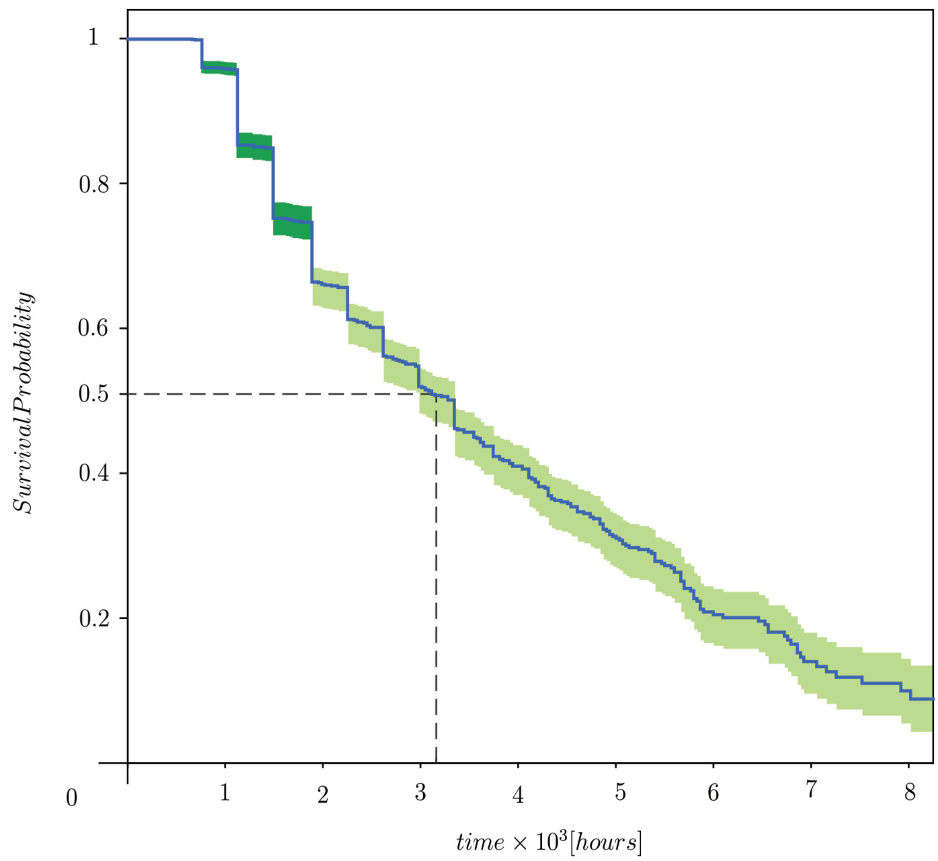
Figure 5 shows Kaplan-Meier model of PRMM dataset in total. The confidence interval is calculated at every point along the probability line. Below the plot in Figure 5, there is information about the number of machines at risk, the number of censored and number of events at every 1000 hours. The confidence interval is much tighter than in the previous dataset, since the censored information is present, and the dataset has more information than the previous one. The censored information was created due to scheduled maintenance. In fact, when maintenance is activated by scheduled time, the component is replaced regardless of the failure, thus the failure time is not known. With this kind of replacement, the failure time is greater than event and represents typical right censored event.

In the case of PRMM dataset, the Kaplan Meier model contributes to improving the management and maintenance planning, both in optimizing the number of machines at risk and in planning and optimizing the quantities of spare parts (components).



Note. Four colored survival curves are related to four different machine components.

Figure 4. Kaplan-Meier probability function diagram for PRMM dataset separated by components



At risk	2877	1464	637	346	202	124	52	20	9
Censored	0	1333	1779	1949	2025	2062	2104	2123	2130
Events	0	80	461	582	650	691	721	734	738

Note. The survival curve plot with confident interval colored in two different colors.

Figure 5. Kaplan-Meier probability function diagram for PRMM dataset

The impact of the preventive maintenance model is also reflected in the reduction of costs both through the reduction of the number of machines at risk and the stock of spare parts.

4.5 Cox Proportional-Hazards Model

The Cox proportional-hazards model [33] represents the regression model used for investigating the association between the survival time of the engine and predictor variables. The model can be expressed based on the hazard function (4):

$$h(t) = h_0(t) \cdot exp(b_1 x_1 + b_2 x_2 + \cdots + b_k x_k), \quad (7)$$

where:
$h(t)$ - hazard function estimated by the set of k covariates (x_1, x_2, ..., x_k),
b_1 ,b_2 ,..., b_k - regressors for measuring the influence of the covariates,
h_0 (t) - baseline hazard related to the value of the hazard if all the x_i are equal to zero.

In this paper, the Cox proportional model was used to provide influence of several covariates (predictors) among 21 sensor readings at rate of particular event. The Cox regression test results show several important indicators which define the influence between the covariate and the hazard rates. The first indicator is statistical significance which defines the influence between each covariate on the hazard rate.

The Cox regression was performed on the NTED dataset first. Figure 6 shows three covariates “Total temperature at LPT outlet”, “Total pressure at HPC outlet” and “Bypass ratio”, among 21 sensor readings with the highest influence in the model.

The statistical significance (lines 23-25) indicate that the model is significant since all three p-values of the tests (probability, Wald and score) are far lower than 0.05. The tests also prove that regressors are significant, as shown in lines 6-10. The Concordance index 0.77 shows that the model is much higher than random.

```
1 ***********************************************************************
2 ********* Summary of Cox-proportionate model for NTED dataset!**********
3 ***********************************************************************
4   n= 100, number of events= 73
5
6     |  coef      exp(coef)    se(coef)     z      Pr(>|z|)     Signif.
7 -----------------------------------------------------------------------
8 T50|  8.986e-02  1.094e+00    2.711e-02    3.314  0.000919     ***
9 P30| -8.306e-01  4.358e-01    2.631e-01   -3.157  0.001592     **
10 BPR|  2.673e+01  4.041e+11    6.812e+00    3.923  8.74e-05     ***
11 -----------------------------------------------------------------------
12 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
13
14
15    |  exp(coef)   exp(-coef)  lower .95   upper .95
16 -----------------------------------------------------------------------
17 T50| 1.094e+00    9.141e-01   1.037e+00   1.154e+00
18 P30| 4.358e-01    2.295e+00   2.602e-01   7.298e-01
19 BPR| 4.041e+11    2.474e-12   6.430e+05   2.540e+17
20 -----------------------------------------------------------------------
21
22 Concordance= 0.773  (se = 0.03 )
23 Likelihood ratio test= 91.27  on 3 df,   p=<2e-16
24 Wald test            = 44.61  on 3 df,   p=1e-09
25 Score (logrank) test = 56.62  on 3 df,   p=3e-12
```

Note. Two tables with coefficients of the NTED model and different statistical values related to them. Beside tables there are short description of the model, different test values as well as significance meaning of the p-value.

Figure 6. The Cox regression test results for NTED dataset

From lines 6-11, one can see the p-value for T50 is 0.000919, with hazard ratio HR=exp(coef)=1.094 (line 9) indicating the strong influence between the total temperature at LPT outlet and increased risk of failure. A similar statement can be defined for the BPR. The negative value of the total pressure at HPC outlet indicates that the hazard ratio has a strong influence between BPR and decreased risk of failure (line 10).

Cox regression analysis performed on PRMM dataset shows that, among more than 30 covariates made during the data transformation and feature engineering, only 5 can be identified as significant (Figure 7). The average values of voltage, rotation, pressure and vibration in the past 3, 6, 9, 12, 18 and 24 hours were calculated and defined as covariates. However, the test shows that average values of all covariates for the last 24 hours have significant influence on the hazard rate with p-value lower than 2e-16. Besides telemetry covariates, the machine age also has significant influence in the model, and it is selected in the Cox model. Lines 26-28 show that all three tests have passed with p-values less than 0.005 and define the influence between each covariate on the hazard rate. On the other hand, the Concordance index of 0.76 shows a much higher index value of the Cox model than random model.

In general, SA-based models predict events with probability value as well as with a safe interval. If the safe intervals are wider, the probability is less certain and vice versa. Since there is no model which can predict all failures, developing more specific and diverse models can improve the maintenance process in more segments. It is well known that the correlations between the values of sensors and the malfunctions of production machines are very common from long time ago [34] to quite recent time [25]. Developing PdM models would be to estimate the correlations between sensors and failures, but also to estimate the working life of the machines.

5 Discussion

The paper applied the SA method for the PdM to estimate the probability of survival of 100 engines of the NTED data set, and 100 CNC machines of the PRMM dataset. The datasets provide two different approaches to data collection and none of them are specially made for SA. This means that additional data transformation and feature engineering must be performed in order to create SA ready dataset.

The two most popular SA models, Kaplan-Meier and Cox proportional models were created and

```
**********************************************************************************
********* Summary of Cox-proportionate model for PRMM dataset!*******************
**********************************************************************************
 n= 2624, number of events= 682

                      |   coef     exp(coef)  se(coef)    z      Pr(>|z|)   Signif.
-----------------------------------------------------------------------------------
age                   | 0.0320040  1.0325216  0.0075165   4.258 2.06e-05      ***
pressure_mean_24hrs   | 0.0300729  1.0305296  0.0034614   8.688  < 2e-16      ***
volt_mean_24hrs       | 0.0478541  1.0490176  0.0039618  12.079  < 2e-16      ***
rotate_mean_24hrs     |-0.0143361  0.9857662  0.0009131 -15.700  < 2e-16      ***
vibration_mean_24hrs  | 0.0809867  1.0843565  0.0090829   8.916  < 2e-16      ***
-----------------------------------------------------------------------------------
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                      | exp(coef)    exp(-coef)      lower .95       upper .95
-----------------------------------------------------------------------------------
age                   |  1.0325      0.9685          1.017           1.0478
pressure_mean_24hrs   |  1.0305      0.9704          1.024           1.0375
volt_mean_24hrs       |  1.0490      0.9533          1.041           1.0572
rotate_mean_24hrs     |  0.9858      1.0144          0.984           0.9875
vibration_mean_24hrs  |  1.0844      0.9222          1.065           1.1038
-----------------------------------------------------------------------------------


Concordance= 0.762  (se = 0.01 )
Likelihood ratio test= 416.9   on 5 df,   p=<2e-16
Wald test            = 439.5   on 5 df,   p=<2e-16
Score (logrank) test = 466.6   on 5 df,   p=<2e-16
```

Note. Two tables with coefficients of the PRMM model and different statistical values related to them. Beside tables there are short description of the model, different test values as well as significance meaning of the p-value.

Figure 7. The Cox regression test results for PRMM dataset

tested [2]. The implementation of SA models, specifically the Kaplan-Meier estimator and the Cox proportional-hazards model, in the context of PdM for manufacturing processes, has yielded insightful results.

Analysis indicates that the Kaplan-Meier model for NTED dataset shows much wider confidence interval than that for the PRMM dataset. This is caused by the fact that PRMM dataset holds more information about failure event than NTED dataset, as well as censored information.

By using Cox proportional hazard model, the influence of different sensor readings on hazard rates were estimated. Both models reach similar performance and test indicators. The Cox proportional hazard model for NTED data set identified three main covariates among 21 sensors reading, whereas the Cox proportional hazard model for PRMM data set identified 5 covariates as significant. Both models pass probability, Wald and score test, and reached the concordance index around 0.77. Furthermore, all models can be included in the analysis during the creation of the maintenance plan and equipment ordering since they give valuable information about future failures.

When comparing Cox proportionate models with other machine learning methods that we or they used in previous works on this topic such as, e.g., LightGBM or Deep Learning, it can be stated that the former is inferior since they give less accurate prediction [32], [35]. The use of such models therefore can be treated as an additional tool for other classic PdM methods like Deep learning, Random Forest, et cetera.

5.1 Validity and Appropriateness of Models

In the context of the considered case studies, the non-parametric Kaplan-Meier estimator and the Cox proportional-hazards model have been selected due to their specific advantages and proven effectiveness in reliability estimation. Both models (Table 5) are extensively documented in the literature and have demonstrated robustness and accuracy in various fields, including medical research and industrial applications.

5.2 Comparison with Other Models

While the Kaplan-Meier and Cox models are well-documented and widely used, it is essential to acknowledge other models that could potentially be used for reliability estimation. Parametric Survival Models (e.g., Weibull, Exponential, Log-normal

Table 5. Kaplan-Meier Estimator and Cox Proportional-Hazards model

	Kaplan-Meier Estimator	Cox Proportional-Hazards Model
Validity	The Kaplan-Meier estimator is a non-parametric model that does not assume any specific distribution for the time-to-event data, making it highly versatile and applicable to various types of data. This flexibility is particularly useful in our case studies, where the exact distribution of failure times is unknown.	The Cox proportional-hazards model is a semi-parametric model that assumes proportional hazards over time but does not require the baseline hazard function to be specified. This assumption allows the model to accommodate varying risk factors while maintaining flexibility in its application.
Appropriateness	In the NTED and PRMM datasets, the Kaplan-Meier estimator effectively estimates the survival probabilities of machinery components over time. This model provides clear, interpretable survival curves, which are crucial for understanding the reliability and expected lifespan of the components under study. The estimator's ability to handle censored data, where not all components have failed within the observation period, adds to its suitability for the datasets used.	The Cox model is particularly suitable for our case studies because it can incorporate multiple covariates to assess their impact on the survival time. In the NTED dataset, it identifies significant sensor readings that influence failure rates, while in the PRMM dataset, it evaluates the impact of telemetry data on the hazard rates. The model's ability to handle both time-dependent and time-independent covariates makes it a powerful tool for understanding the underlying factors affecting machinery reliability.

[36]) assume a specific distribution for the survival times, which can provide more precise estimates if the chosen distribution accurately represents the data. However, their applicability is limited when the true distribution is unknown or varies across different datasets. Accelerated Failure Time (AFT) [35], [36], models offer an alternative by directly modelling the relationship between covariates and the survival time. They assume that the effect of covariates accelerates or decelerates the lifetime by some factor. While useful, AFT models require distributional assumptions, making them less flexible than Cox models in diverse data scenarios. Machine Learning Approaches (e.g., Random Survival Forests, Deep Learning) can handle complex interactions and non-linear relationships within the data, potentially offering higher predictive accuracy [10]. However, they often require large amounts of data and computational resources, and their interpretability can be limited compared to traditional statistical models.

5.3 Effectiveness of Kaplan-Meier and Cox Models

The Kaplan-Meier and Cox models were chosen for their balance of flexibility, interpretability, and robustness. They provide clear, actionable insights into the reliability of machinery components without the need for extensive assumptions about the underlying data distribution. Their ability to handle censored data and incorporate multiple covariates ensures that they can effectively capture the complexities of the case studies considered.

5.4 Practical Implementation

The application of SA models to the NASA Turbofan Engine Degradation Simulation Data Set (NTED) and the Predictive Maintenance Modelling Guide Data Set (PRMM) has demonstrated several practical benefits and challenges. First, successfully transforming raw datasets into SA-compatible formats was critical. For NTED, identifying failure times and censoring events required careful handling, while for PRMM, merging multiple data sources and performing feature engineering were necessary steps. The process highlighted the importance of data quality and completeness. Missing values, inconsistent records, and noise in sensor readings posed significant challenges that had to be addressed through preprocessing techniques. Second, the Kaplan-Meier estimator provided clear survival probabilities, which are intuitive and useful for maintenance planning. However, its non-parametric nature limits its ability to account for multiple covariates. The Cox proportional-hazards model offered deeper insights by identifying significant predictors of failure [31]. This model's ability to handle multiple covariates made it particularly valuable for understanding the impact of different factors on machinery survival.

6. Conclusion

The current study builds on previous research in PdM and SA and contributes to the body of knowledge by demonstrating the practical application of SA models and addressing specific challenges related to data transformation and model implementation. By applying transparent and interpretable models like Kaplan-Meier and Cox, this study reinforces the need for explainability in PdM, enabling maintenance planners to make informed decisions based on model outputs.

Despite the successful implementation, several limitations aspects were identified: data limitations, model assumptions, generalizability. Regarding data limitations and datasets used we can conclude that the lack of censoring information in the NTED dataset restricted the analysis, leading to wider confidence intervals and potentially less reliable survival estimates. Additionally, the dataset's limited scope (only 100 engines) may not fully capture the variability in real-world scenarios. Although PRMM provided more comprehensive data, including censoring information, the complexity of merging different sources introduced potential errors. The derived features, such as rolling means and standard deviations, might not fully represent the underlying failure mechanisms. The Kaplan-Meier estimator assumes independence between events, which might not be held in all manufacturing environments. Similarly, the Cox model's proportional hazards assumption may not be valid if the hazard ratios change over time. Both models assume that the covariates are correctly specified and that there are no significant interactions or nonlinear effects that are not accounted for. According to the findings if the case studies regarding the generalizability we can say that the models were applied to specific datasets with characteristics. The generalizability of the results to other datasets or industries may be limited. Different types of machinery, operational conditions, and maintenance practices could significantly impact the applicability of the findings.

To address the limitations and build upon the current work, several directions for future research are proposed. Future studies should aim to collect more comprehensive and higher-quality data, including detailed censoring information and a broader range of sensor readings. Integrating data from different sources in a more automated and robust manner could improve the reliability of the models. Exploring advanced SA models that can handle non-proportional hazards, such as time-varying coefficients, could provide more accurate and flexible modeling. Machine learning techniques, such as survival random forests or deep learning models, could also be investigated to capture complex interactions and nonlinearities. Implementing and validating the models in real-world manufacturing environments would provide valuable insights into their practical applicability and effectiveness. Collaborating with industry partners to conduct pilot studies could help refine the models and demonstrate their value in operational settings. Integrating SA models with existing PdM systems and Industry 4.0 technologies, such as IoT and Cloud Computing, could enhance their utility. Developing user-friendly tools and dashboards for maintenance planners to interact with the models and interpret the results is crucial for practical adoption.

The implementation of SA for PdM in manufacturing has shown promising results, providing valuable insights into machinery reliability and failure patterns. Despite the identified limitations, the study offers a robust framework for applying SA models to real-world datasets. Future research should focus on enhancing data quality, exploring advanced modeling techniques, and validating the models in operational environments to maximize their practical impact.

Data Availability Statement

The data sets used in the paper are openly available. The NTED dataset is available in “Turbofan Engine Degradation Simulation Data Set” repository at https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository, as reference [29]. The PRMM dataset is available in the “Azure AI Gallery” portal at https://gallery.azure.ai/Experiment/Predictive-Maintenance-Modelling-Guide-Experiment-1, as reference [30].

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] N. Hafidi, A. El Barkany, and A. El Mhamedi, "Joint optimization of production, maintenance and quality: A review and research trends," Int. J. Ind. Eng. Manag., vol. 14, no. 4, pp. 282–296, 2023, doi: 10.24867/IJIEM-2023-4-339.

[2] B. Hrnjica and S. Softic, “Explainable AI in Manufacturing: A Predictive Maintenance Case Study,” in Advances in Production Management Systems. Towards Smart and Digital Manufacturing, B. Lalic, V. Majstorovic, U. Marjanovic, G. von Cieminski, and D. Romero, Eds., IFIP Advances in Information and Communication Technology, vol. 592, Springer Cham, 2020, pp. 580-588. doi: 10.1007/978-3-030-57997-5_8.

[3] I. Lopes, M. Figueiredo, and V. Sá, "Criticality evaluation to support maintenance management of manufacturing

systems," Int. J. Ind. Eng. Manag., vol. 11, no. 1, pp. 3-18, 2020. doi: 10.24867/IJIEM-2020-1-248.

[4] W. Yu, T. Dillon, F. Mostafa, W. Rahayu and Y. Liu, "A Global Manufacturing Big Data Ecosystem for Fault Detection in Predictive Maintenance," in IEEE Transactions on Industrial Informatics, vol. 16, no. 1, pp. 183-192, Jan. 2020, doi: 10.1109/TII.2019.2915846.

[5] J. Wang, C. Li, S. Han, S. Sarkar and X. Zhou, "Predictive maintenance based on event-log analysis: A case study," IBM Journal of Research and Development, vol. 61, no. 1, pp. 121-132, 2017, doi: 10.1147/JRD.2017.2648298.

[6] D. G. Kleinbaum and M. Klein, Survival Analysis: A Self-Learning Text. New York, NY, USA: Springer Science & Business Media, 2005.

[7] W. Xu, H.Y. Sun, A.L. Awaga, Y. Yan, and Y.J. Cui, "Optimization approaches for solving production scheduling problem: A brief overview and a case study for hybrid flow shop using genetic algorithms," Advances in Production Engineering and Management, vol. 17, no. 1, pp. 45-56, 2022, doi: 10.14743/apem2022.1.420.

[8] P. Prerna and N. Singh, "Evolution and impact of artificial intelligence in chatbots," J. Tr., Chal. Art. Intell., vol. 1, no. 4, pp. 139-142, 2024, doi: 10.61552/JAI.2024.04.004.

[9] A. Nekoonam, R. F. Nasab, S. Jafari, T. Nikolaidis, N. Ale Ebrahim, and S. A. Miran Fashandi, "A Scientometric Methodology Based on Co-Word Analysis in Gas Turbine Maintenance", Tehnicki Vjesnik, vol. 30, no. 1, pp. 361 - 372, 2023, doi: 10.17559/TV-20220118165828.

[10] V. Kumar, R. P. Tewari, R. Pandey, and A. Rawat, "Triboinformatic Modeling of Wear in Total Knee Replacement Implants Using Machine Learning Algorithms," J. Mater. Eng., vol. 1, no. 3, pp. 97-105, 2023, doi: 10.61552/JME.2023.03.001.

[11] S. Annamalai, R. Udendhran, and S. Vimal, “Cloud-Based Predictive Maintenance and Machine Monitoring for Intelligent Manufacturing for Automobile Industry,” in Novel Practices and Trends in Grid and Cloud Computing, P. Raj and S. Koteeswaran, Eds. New Your, NY, USA: IGI Global, 2019, pp. 74-89. doi: 10.4018/978-1-5225-9023-1.ch006.

[12] B. Hrnjica and A. D. Mehr, “Energy Demand Forecasting Using Deep Learning,” in Smart Cities: Performability, Cognition & Security, F. Al-Turjman, Ed., Springer, 2020, pp. 83-101. doi: 10.1007/978-3-030-14718-1_4.

[13] O. Salunkhe and Åsa Fasth Berglund, “Industry 4.0 Enabling Technologies for Increasing Operational Flexibility in Final Assembly”, Int. J. Ind. Eng. Manag., vol. 13, no. 1, pp. 38–48, 2022, doi: 10.24867/IJIEM-2022-1-299.

[14] B. Schmidt and L. Wang, “Cloud-enhanced predictive maintenance,” Int. J. Adv. Manuf. Technol., vol. 99, pp. 5-13, 2018, doi: 10.1007/s00170-016-8983-8.

[15] MM. -Y. You, F. Liu, W. Wang and G. Meng, "Statistically Planned and Individually Improved Predictive Maintenance Management for Continuously Monitored Degrading Systems," in IEEE Transactions on Reliability, vol. 59, no. 4, pp. 744-753, Dec. 2010, doi: 10.1109/TR.2010.2085572.

[16] R. K. Mobley, An Introduction to Predictive Maintenance. Woburn, MA, USA: Elsevier, 2002.

[17] C. Krupitzer et al., “A Survey on Predictive Maintenance for Industry 4.0,” arXiv, 2020. [Online]. Available: https://arxiv.org/abs/2002.08224. [Accessed: 17-Jul-2024].

[18] Y. He, X. Han, C. Gu, and Z. Chen, “Cost-oriented predictive maintenance based on mission reliability state for cyber manufacturing systems,” Adv. Mech. Eng., vol. 10, no. 1, pp. 1-15, 2018, doi: 10.1177/1687814017751467.

[19] S. T. March and G. D. Scudder, “Predictive maintenance: strategic use of IT in manufacturing organizations,” Inf. Syst. Front., vol. 21, pp. 327-341, 2019, doi: 10.1007/s10796-017-9749-z.

[20] J. Wang, L. Zhang, L. Duan, and R. X. Gao, “A new paradigm of cloud-based predictive maintenance for intelligent manufacturing,” J. Intell. Manuf., vol. 28, pp. 1125–1137, 2017, doi: 10.1007/s10845-015-1066-0.

[21] M. Kommenda and S. Strumpf, "KI & Predictive Maintenance: Maschinelles Lernen & Datenanalyse," WKO.at [Online]. Available: www.wko.at/service/ooe/innovation-technologie-digitalisierung/Scheuch-FH.pdf. [Accessed: 10-Feb-2022].

[22] F. J. Lacueva-Pérez et al., "Comparing Approaches for Evaluating Digital Interventions on the Shop Floor," Technologies, vol. 6, no. 4, p. 116, 2018. doi: 10.3390/technologies6040116.

[23] M. de Almeida Costa, J. P. de Azevedo Peixoto Braga, and A. R. Andrade, "A data-driven maintenance policy for railway wheelset based on survival analysis and Markov decision process," Quality and Reliability Engineering International, vol. 37, pp. 1176-1198, 2021, doi: 10.1002/qre.2729.

[24] O. Aydin and S. Guldamlasioglu, "Using LSTM networks to predict engine condition on large scale data processing framework," 2017 4th International Conference on Electrical and Electronic Engineering (ICEEE), Ankara, Turkey, 2017, pp. 281-285, doi: 10.1109/ICEEE2.2017.7935834.

[25] K. Prażnowski, A. Bieniek, J. Mamala, and A. Deptuła, "The Use of Multicriteria Inference Method to Identify and Classify Selected Combustion Engine Malfunctions Based on Vehicle Structure Vibrations," Sensors, vol. 21, no. 7, p. 2470, 2021, doi: 10.3390/s21072470.

[26] P. Korvesis, S. Besseau and M. Vazirgiannis, "Predictive Maintenance in Aviation: Failure Prediction from Post-Flight Reports," 2018 IEEE 34th International Conference on Data Engineering (ICDE), Paris, France, 2018, pp. 1414-1422, doi: 10.1109/ICDE.2018.00160.

[27] P. G. Ramesh, S. J. Dutta, S. S. Neog, P. Baishya, and I. Bezbaruah, “Implementation of Predictive Maintenance Systems in Remotely Located Process Plants under Industry 4.0 Scenario,” in Advances in RAMS Engineering. Springer Series in Reliability Engineering, D. Karanki, G. Vinod, and S. Ajit, Eds., Springer, Cham, 2020, pp. 293-326, doi: 10.1007/978-3-030-36518-9_12.

[28] F. Emmert-Streib and M. Dehmer, "Introduction to survival analysis in practice," Mach. Learn. Knowl. Extr., vol. 1, no. 3, pp. 1013-1038, 2019, doi: 10.3390/make1030058.

[29] A. Saxena, K. Goebel, D. Simon and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," 2008 International Conference on Prognostics and Health Management, Denver, CO, USA, 2008, pp. 1-9, doi: 10.1109/PHM.2008.4711414.

[30] U. F. Boylu, "Predictive Maintenance Modelling Guide Data Sets," Azure Gallery Article, 2016. [Online]. Available: https://gallery.azure.ai/Experiment/Predictive-Maintenance-Implementation-Guide-Data-Sets-1. [Accessed: Dec. 25, 2021].

[31] W. Hong Chang, C. Lee, K. Lee, M.-S. Ko, D. E. Kim, and K. Hur, "Remaining Useful Life Prognosis for Turbofan Engine Using Explainable Deep Neural Networks with Dimensionality Reduction," Sensors, vol. 20, no. 22, p. 6626, 2020, doi: 10.3390/s20226626.

[32] B. Hrnjica and S. Softic, “The Survival Analysis for a Predictive Maintenance in Manufacturing,” in Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems, A. Dolgui, A. Bernard, D. Lemoine, G. von Cieminski, and D. Romero, Eds., IFIP Advances in Information and Communication Technology, vol. 632, Springer Cham, 2021, pp. 301-309. doi: 10.1007/978-3-030-85906-0_9.

[33] G. Budai, R. Dekker, and R. P. Nicolai, “Maintenance and Production: A Review of Planning Models,” in Complex

System Maintenance Handbook, K. A. H. Kobbacy and D. N. Prabhakar Murthy, Eds., London, UK: Springer, 2008, pp. 321-344, doi: 10.1007/978-1-84800-011-7_13.
[34] A. Muszynska, "Vibrational Diagnostics of Rotating Machinery Malfunctions," Int. J. Rotating Mach., vol. 1, no. 3-4, pp. 237-266, 1995, doi: 10.1155/S1023621X95000108.
[35] A. Spooner, E. Chen, A. Sowmya, P. Sachdev, N. A. Kochan, J. Trollor, and H. Brodaty," A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction," Sci. Rep., vol. 10, p. 20410, 2020, doi: 10.1038/s41598-020-77220-w.
[36] S. Lemeshow, S. May, and D. W. Hosmer Jr, Applied Survival Analysis: Regression Modeling of Time-to-Event Data. Hoboken, NJ, USA: Wiley-Interscience, 2008.