Prognostic and risk of failure events using machine learning: An analysis based on onboard aircraft messages.

João Francisco Dos Reis Martins Rodrigues

Thesis to obtain the Master of Science Degree in

Aerospace Engineering

Supervisors: Prof. Elsa Maria Pires Henriques
Dr. Márcia Lourenço Baptista

Examination Committee

Chairperson: Prof. Fernando José Parracho Lau
Supervisor: Prof. Elsa Maria Pires Henriques
Member of the Committee: Prof. Francisco António Chaves Saraiva de Melo

October 2019

Acknowledgments

The learning process is a journey almost impossible to be accomplished alone. I was fortunate enough to be surrounded by those who always supported and believed in me.

First, I would like to thank my supervisors, Doctor Elsa Henriques and Doctor Márcia Batista, for all the support and guidance provided along this hard but enriching journey, which allowed me to broaden my knowledge in areas that I would hardly ever imagine myself studying. A special thanks to Portugália Airlines for allowing me the fulfilling opportunity to be integrated in the airline company for six months. I also wish to thank my supervisors in Portugália Airlines, namely to Engineer Pedro Figueira and Engineer Nuno Lima, for all the invaluable contributions, helpful advice, and shared knowledge. Also, a word to all the people I met at Portugália Airlines, for making me feel, from day one, an integrating part of the team.

Another special thanks to my friends, for all the pleasant moments we experienced together, and to Engineer Caio Cordeiro and Alexandre Batista, from AZUL airlines, for helping me to extend this study.

Last, but by no means the least, I would like to express my deepest thanks to my parents, for the unconditional support and profound belief in my abilities throughout this journey. I will always be grateful, and I genuinely believe that their encouragement made me halfway there.

# Resumo

As companhias aéreas têm vindo a exercer um grande esforço para encontrar formas de otimizar processos de manutenção para manter elevados padrões de segurança. A aplicação de métodos *data-driven* à manutenção tem sido impactante na aeronáutica, devido aos seus benefícios na redução de custos e aumento da segurança. Ao longo dos anos, os métodos de prognóstico baseados em dados têm-se tornado numa área de estudo relevante, complementar à ainda dominante na aeronáutica manutenção preventiva. Estes novos métodos permitem a aquisição de uma metodologia proativa em vez de reativa perante falhas, permitindo que sejam antecipadas e eliminadas antes que ocorram. Assim, há uma crescente preocupação no setor de manutenção em encontrar indicadores ou precursores de falhas futuras usando *machine learning* e inteligência artificial. Esta tese relata o trabalho realizado na avaliação das capacidades de prognóstico das mensagens do *Central Maintenance Computer* (CMC). Isto é alcançado através da comparação de diferentes tipos de modelos, variando propriedades dos conjuntos de dados analisados, permitindo associar as diferenças nos resultados com as diferentes características dos modelos. Usando dois conjuntos de dados reais de duas companhias aéreas (AZUL e Portugália), esta tese não somente visa a previsão do *Remaining Useful Life* (RUL) do equipamento, mas também a previsão do nível de urgência de uma intervenção num determinando momento. Os resultados demonstram que os dados analisados associados com as técnicas de *machine learning* aplicadas possuem capacidades de previsão de falhas, auxiliando no despoletar de ações de manutenção não planeadas.

Palavras-chave: Prognóstico, Manutenção preditiva, Machine learning, Aeronáutica, Mensagens do Central Maintenance Computer.

Abstract

Airline companies have been making a great effort to find ways to optimise maintenance processes in order to maintain high safety standards. The application of data-driven methods to maintenance has been introduced as a breakthrough in aeronautics, due to benefits in cost reduction and safety increase. Over the years, data-driven prognostics has become an important area of study, complementary to the still-dominant strategy in aeronautics, preventive maintenance. These new methods allow maintenance personnel and process engineers to take a proactive instead of a reactive approach to failures, where failures are anticipated and eliminated before they occur. In view of this, there has been a growing concern in the maintenance sector to find indicators or precursors of failures using machine learning and artificial intelligence. This thesis reports on the work carried out on the evaluation of the Prognostics and Health Management (PHM) capabilities of the Central Maintenance Computer (CMC) messages. This is achieved by comparing different types of models, varying several properties of the data sets allowing to relate differences in the results to differences in the characteristics of the models. Using two different real data sets from two airline companies (AZUL and Portugália), this thesis focuses not only on the prediction of the remaining useful life (RUL) of the equipment but also on the prediction of the urgency of an intervention at a given time. The results show that message data associated with the applied machine learning techniques have predictive failure capabilities, aiding the trigger of unplanned maintenance actions.

Keywords: Prognostics, Machine Learning, Predictive Maintenance, Aeronautics, Central Maintenance Computer Messages

Contents

List of Tables

List of Figures

Acronyms

ATA	Air Transport Association
AHM	Airplane Health Management
AOG	Aircraft On Ground
APU	Auxiliary Power Unit
ARINC	Aeronautical Radio Incorporated
ASCB	Avionics Standard Communication Bus
CAS	Crew Alerting System
CMC	Central Maintenance Computer
CMCF	Central Maintenance Computer Function
CMCS	Central Maintenance Computer System
EASA	European Union Safety Agency
FAA	Federal Aviation Administration
FHDB	Flight History Database
ICAO	International Civil Aviation Organisation
IT	Information Technology
KNN	K-Nearest Neighbours
MAE	Mean Absolute Error
MEL	Minimum Equipment List
MRB	Maintenance Review Board
MRBR	Maintenance Review Board Report
MRO	Maintenance Repair and Overhaul
MWF	Monitor Warning Function
PC	Principal Component
PCA	Principal Component Analysis
PGA	Portugália Airlines
RCM	Reliability Centred Maintenance
RMSE	Root Mean Absolute Error
RUL	Remaining Useful Life
SVM	Support Vector Machine

Chapter 1

Introduction

1.1 Motivation

The world is becoming largely more dependant on machines and systems that are crucial in individuals' daily routine. From the most common household devices, such as a coffee machine, a hair dryer, a dishwasher to the more complex and advanced forms of transportation such as cars, trains, and aircraft, maintenance plays an important role in a great part of the most usual to the more complex machines' useful life. The level of maintenance required to preserve the functionality and life of an equipment depends largely on how complex the running system is. An aircraft is a result of the continuous interaction of several highly complex systems, that allow the machine to provide the flying capabilities that human could only dream of a century ago. Capable of flying upwards of 40 000 feet, at a speed close to the speed of sound, over its long lifespan of, in some cases, more than 100 000 flight cycles, the commercial aircraft is one of the more complex equipment created and developed by human beings. Flying is no more a privilege like it was some decades ago. Nowadays, due to the diversified offers ans less expensive fares, flying is becoming a common mean of transportation. Despite the high risks involved in an aircraft operation, due to the numerous factors that may ultimately cause fatalities, the low number of accidents per million flights (13, in 2017 [1]) and the downward tendency of the number of fatalities per year in the commercial aviation over the last decades, showcased in figure 1.1, highlights the effect and influence of technology and advanced and rigorous maintenance programs on the increase in the aircraft transportation sector's overall safety. This decrease in fatalities is emphasised by the upwards tendency in both the number of flights and passengers per year [2]. Hence, the aircraft is changing the worldwide paradigm in transport safety and is considered one of the safest means of transportation, an important achievement considering the numerous risks associated with the operation.

As already mentioned, maintenance is one of the factors that influence the safety in the aircraft industry. There are two major types of maintenance: Unplanned maintenance and planned maintenance. The first is the oldest type of maintenance, and it may be denominated as "run to failure", as the machines under this type of maintenance are only subjected to corrective actions after the occurrence of a failure [3]. It is implemented when the equipment failure priority is low or the cost of performing a planned

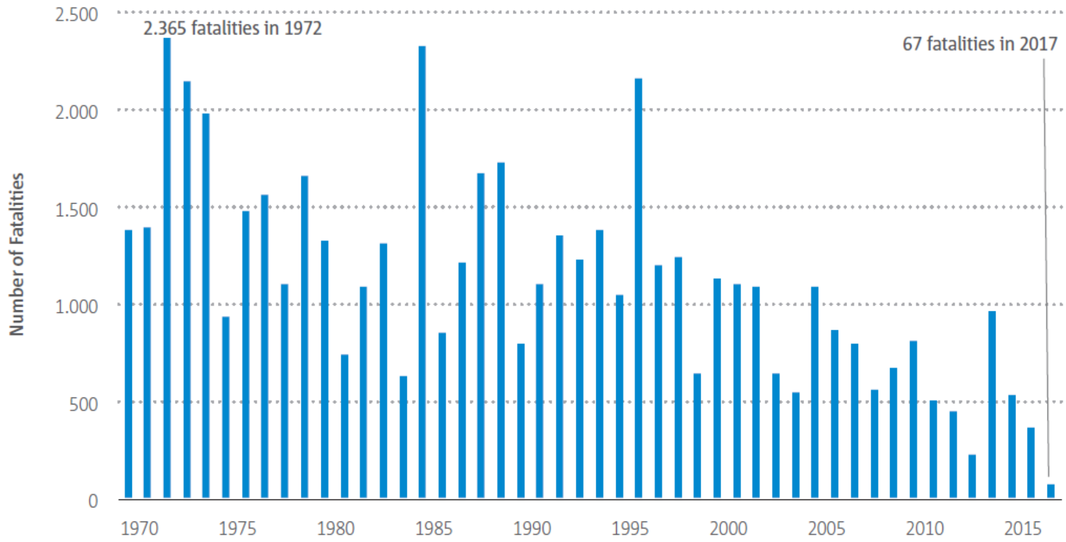


Figure 1.1: Number of Fatalities Involving Large Aeroplane Passenger and Cargo Operations Worldwide, 1970-2017. [1]

maintenance routine overcomes the cost of simply repairing the system whenever it fails. Also, some system components present a high level of failure unpredictability which turns the planned maintenance unjustifiable. On the other hand, planned maintenance refers to the type of maintenance where actions are performed in a planned manner. This type of maintenance is subdivided into four sub-types: Corrective maintenance; improvement maintenance; preventive maintenance; and predictive maintenance. The first refers to the type of maintenance carried out after fault recognition where the goal is to turn the item into a state in which the required function can be performed. It differs from the run to failure maintenance as in the corrective maintenance the repair is scheduled to whenever the item fails. For example, the maintenance team may decide to replace the light bulb whenever it burns out. The second intends to reduce or eliminate the need for maintenance by improving or upgrading the machine. Whether by performing activities that intend to eliminate the cause of failure, simplify or facilitate maintenance tasks, the end goal is always to eliminate or decrease the need for maintenance by the improvement of the machine under this type of maintenance.

Prevention is better than cure, that is the principle of the preventive maintenance. The maintenance actions of this type are carried out at pre-decided intervals or according to stipulated criteria, with the end goal of reducing the probability of failure, limiting the machine degradation or even the failure effects. Based on the previous location and identification of each machinery or system weak spots, periodic/scheduled inspections and actions performed reduce the risk/danger of unexpected or unanticipated breakdowns.

Finally, predictive maintenance is defined as the set of activities that intend to detect alterations in the condition of the equipment, or signs of failure, in order to perform the required maintenance actions to maximise the service life of the equipment, maintaining the same risk of failure. Predictive maintenance differs from the preventive maintenance as the former predicts the failure based on the condition of the equipment, whereas the later is not based on the machine's monitored state, as the failure is expected based on the equipment's age and previous failure data. Predictive maintenance relies on the condition monitoring of the equipment to make decisions on the occurrence of the next maintenance action. Through the most simple routine inspection of the ware state of a component, that enables to estimate its remaining useful life, or through the thorough analysis of the lubricant viscosity, predictive maintenance focuses on the current state of the equipment and decides when the next maintenance action would be scheduled. The evolution of technology is enabling the adaption of new tools, discussed in the following sections, that transform and increase the reliability and promptness of the task of deciding future interventions.

The aircraft is a highly complex machine and an integration of numerous systems, where some failures can lead to fatalities; therefore maintenance plays an essential role in its operation. All the above mentioned maintenance types are implemented in an aircraft useful life, however, airline companies follow a strict and rigorous hard time based preventive maintenance plan supplied by the aircraft manufacturer.

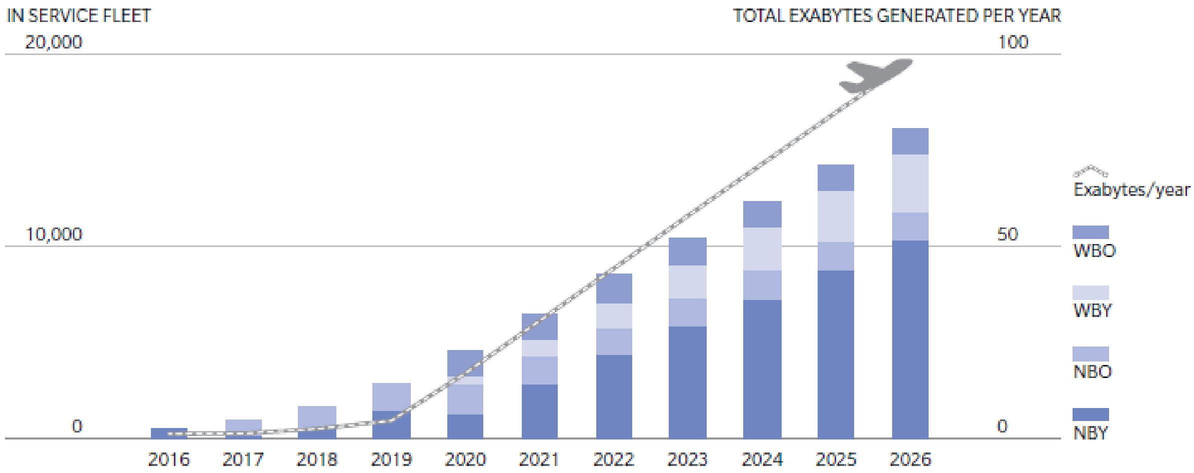


Figure 1.2: Data generated from projected global fleet [5]

Health monitoring systems

Nowadays, the commercial aircraft is a highly instrumented machine, capable of generating large amounts of data. Figure 1.2 showcases a forecast of the amount of data generated by the global fleet. Given the increasing tendency, the major players in the aircraft industry, namely the Original Equipment Manufacturers (OEM), operators and Maintenance, Repair and Overhaul (MRO) companies are trying to take advantage of the stream of data that the aircraft is capable of generating, investing mainly in Aircraft Health Monitoring and Predictive maintenance systems. The first consist of tools capable of constantly monitor and showcase the aircraft airworthiness through the analysis of data derived from the aircraft instrumentation. There are now several health monitoring softwares available, some directly from the aircraft manufacturer. Portugália Airlines, the Portuguese airline company currently operating a fleet of 13 Embraer aircraft, where the development of this thesis took place during a six month internship, is currently using a health monitoring software called AHEAD-PRO. It is an integrated tool that consolidates aircraft data from onboard systems and web-based databases. It provides the customer with aircraft usage information, real time preprocessed messages from the aircraft fault report system via satellite communication, and with the respective probable root cause associated with the most critical messages. It also provides the customer with up-to-date information from Embraer's maintenance team and a direct communication channel between the operator's and manufacturer's maintenance teams [4]. Other similar software solutions include the AIRcraft Maintenance ANalysis (AIRMAN), designed by Airbus, and Boeing's solution, the Airplane Health Management, or AHM.

Predictive maintenance systems

Although health monitoring systems represent an important complement to the maintenance and troubleshooting actions of an aircraft operation, its main goal is to identify changes that may indicate damage or imminent failure, showcasing important parameters to maintenance crews in an easy and simple to formats. These systems have some data interpretation capabilities, mainly based on the occurrence rate of certain predefined important failures, and have particular functionalities that, as an example, are able to alert operators when faced with certain situations. However, a significant part of the data interpretation rests on the knowledge of engineers and technicians since the software itself is incapable of predicting a failure occurrence. That's where predictive maintenance role lies, as with the significant amounts of data available, the main goal of this technique is to use past data to develop data-based or physical degradation based models that may interpret aircraft health monitoring parameters and predict when a failure might occur. Among the numerous advantages associated with this technique, a possible failure prediction would allow aircraft operators to improve the maintenance plans and reduce the unexpected failure events that, in some cases, hurt the brand's image and reputation. Furthermore, maintenance actions prior to failure would reduce out of base failures, that represent an important struggle, both in monetary terms as well as in a coordination point of view. Repairing an aircraft several thousand kilometres from the main maintenance base requires significant and unforeseen efforts. Better maintenance planning, resulting from predictive maintenance, would also eventually reduce the costs related to spare parts management, and production time spent on unforeseen maintenance actions. The enormous cost of delays or cancelled flights due to failure events would ultimately be lowered once a failure prediction model is capable of alerting maintenance crews that, in a certain number of flights, a certain system is likely to fail. Whenever an aircraft is not flying, it's costing money. However, safety being crucial in the aircraft sector, the ability to predict a future failure may ultimately avoid possible fatal accidents.

PROGNOS, developed by Air France Industries and KLM Engineering and Maintenance, is one of the many available predictive maintenance commercial software. Based on sensor data and maintenance logs, the ultimate goal is to predict failure occurrence of major aircraft systems, such as the engines and auxiliary power units [36].

1.2 Problem Statement and objectives

This thesis is the result of a collaboration with the airline company Portugália Airlines. It aims to study the failure predictive capabilities of a type of data produced and extracted from aircraft: Central Maintenance Computer (CMC) Messages. These warning messages represent decoded sensor data from several aircraft systems. They are stored by the aircraft internal database, downloaded periodically by the aircraft maintenance team, or streamed directly to the already mentioned AHEAD-PRO health monitoring software, giving the troubleshooters the ability to constantly oversee the fleet's state.

The, to some extent, inexplicable high rate of emission of CMC messages throughout Portugália Airlines' fleet and the constant search for data variables that could enable the development and implementation of a predictive maintenance model made Portugália Airlines recognise the need for a further in depth study of the aircraft's messages from the fault reporting system. Considering only the messages emitted between the offblock and onblock timestamps of each flight, i.e, disregarding messages from possible pre-flight checks and maintenance periods, over the 3 year database provided by Portugália Airlines, an average of 15 CMC messages per flight were emitted, reaching a maximum of 7751 CMC messages emitted in one specific flight. During the entire period of exactly 3 years and 27 days, 1.6 million CMC messages were issued, resulting in an average of 1435 records per day, over the entire 13 aircraft fleet. The importance of this high emission is augmented by the messages' reporting nature of the aircraft system's state. Therefore, this data type was considered to be a worthy candidate for a potential development of a failure prognostic tool.

The methodology employed in this dissertation to study the importance and the possible predictive nature of the CMC messages, is based on data mining techniques and machine learning methods. This will be measured by the magnitude of the resulting error associated with the development of machine learning models that, based on the messages' occurrence, aims to predict the time left until the succeeding failure event. This capability is also evaluated comparing the results from the machine learning models with a failure based approach, commonly used to define preventive maintenance plans - the Weibull analysis. The difference between the two approaches allows one to gather information about the messages' ability to predict future failure events. An improvement associated with the results from the machine learning models would also move the maintenance process one step forward from the usual preventive maintenance approaches.

1.3 Thesis outline

The first chapter aims to describe the motivation behind this thesis, the main objectives and the problem statement. It ends with an overview of the dissertation's structure. Chapter 2 starts with an introduction to aircraft maintenance, discussing the main maintenance types and how maintenance plans are elaborated. This is followed by a more thorough description of one of the maintenances performed in the aircraft industry - the condition based maintenance. The chapter ends with a brief analysis of three projects whose objectives are related with the main goal of this thesis. Chapter 3 begins by introducing Portugália Airlines, the Portuguese airline company, where this thesis was developed, during a six month internship. Due to the relevance of the message data to this dissertation, the chapter continues with the description of how the messages are generated by the aircraft. This is followed by an overview of the main subsystem studied throughout this thesis - the pneumatic system. The final sections of the chapter discuss the industrial challenges faced during the internship, and continues with the methodology applied. Chapter 4 describes the data processing cycle, from the acquisition and storage of data in the raw format until the final data preparation for machine learning. This chapter also includes a data's introductory exploratory analysis, which measures the quality of the data with more usual visual and statistical based methods. Chapter 5 presents the baseline model (Weibull analysis) and the machine learning models developed and used in this thesis.It also introduces the used machine learning framework. Chapter 6 presents and explores the results of all the different models discussed in chapter 5. Finally, the conclusions withdrawn from the study are presented. The dissertation ends with recommendations for future work to extend the research.

Chapter 2

Background

In section 2.1, a brief description of the general types of maintenance is presented. Some aspects of aircraft maintenance and current practices are discussed. The second section concerns the related work, or state of the art, reviewing some of the work done previously on the topic of prognostics.

2.1 Aircraft Maintenance

According to ICAO (International Civil Aviation Organization), the definition of maintenance is ”The performance of tasks required to ensure the continuing airworthiness of an aircraft, including any one or combination of overhaul, inspection, replacement, defect rectification, and the embodiment of a modification or repair” [11]. EASA (European Union Aviation Safety Agency) specifies maintenance as ”any one or combination of overhaul, repair, inspection, replacement, modification or defect rectification of an aircraft or component, with the exception of pre-flight inspection”; on the other hand, maintenance by FAA (Federal Aviation Administration) ”includes inspection, overhaul, repair, preservation, and the replacement of parts, but excludes preventive maintenance” [12]. Aircraft maintenance may be defined in several different ways, despite the main goal being common to all of the mentioned definitions - restore and preserve the aircraft's airworthiness. Aircraft maintenance services are generally provided by internal (subsidiary) or external (independent) Maintenance, Repair and Overhaul (MRO) companies, that adapt and apply to the aircraft actions, policies and concepts used in other traditional industries. With the aircraft being a highly complex machine, and safety being a crucial factor to any of the aircraft operators, maintenance is seen as fundamental to ensure the airworthiness of the fleet. In accordance with ICAO depending on the fleet age, size and usage, maintenance represents an average of 11% of operational costs of air operators up to 25% [13].

According to the European Standard [14], maintenance regards the ”combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function”. It is imbedded in the very definition of maintenance the separation of actions that (a) retain and (b) restore the required function. The actions that retain are denominated as Preventive maintenance, the ones that seek restoration are Corrective maintenance actions.

Aircraft maintenance uses different nomenclature to define the several types of maintenance. Preventive maintenance is known in the MRO sector as scheduled maintenance; whereas corrective maintenance is denominated as unscheduled maintenance. The first is defined by the European Standards [14] as ”maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or the degradation of the functioning of the item”, or by the FAA [12] by ” all the individual maintenance tasks performed according to the maintenance time limitations”, being the time limitations defined in the aircraft sector not only by calendar time itself but by other usage parameters such as flight cycles or flight hours [15]. The aircraft scheduled maintenance includes maintenance checks, which are a set of maintenance tasks performed regularly in equal intervals. The maintenance checks that are performed throughout the service life of the aircraft compose the aircraft maintenance program [16].

Corrective or unscheduled maintenance is also defined by the European Standards [14] as ”maintenance carried out after fault recognition and intended to put an item into a state in which it can perform”. It is not planned, but required whenever an item fails. Bird strikes, tail strikes, hard landings, or even faults identified on scheduled maintenance actions are among others unforeseen situations that lead to the necessity of unscheduled maintenance tasks.

The scheduled maintenance intervals and tasks are defined by the aircraft’s maintenance program. Its conduction must enable continuous safety and airworthiness throughout the aircraft’s operation. Nowadays, as the result of a mutual coordination between aviation authorities, manufacturers and operators, the establishment of the maintenance programs is defined by the Maintenance Review Board (MRB), which is an organisation composed of representatives of airline operators, manufacturers and members of regulatory authorities, responsible for the final approval of maintenance programs for specific aircraft [24]. According to EASA [17], the MRB is ”an acceptable means of compliance for developing a maintenance program”.

The methodology used nowadays, that aim the development of maintenance plans, is the so called MSG-3 (Maintenance Steering Group). It is a task oriented analytical method, or a decision logic process, recognised by the FAA and EASA, that aims to showcase a methodology that shall be used to develop the maintenance program of the aircraft. Under this methodology, ”activities are assessed at the system level” [24] and are separated due to security or economic reasons. It is currently the only methodology accepted by the aviation authorities for the development of maintenance programs.

The MSG-3 is based on the Reliability Centred Maintenance (RCM). The RCM is a method for planning maintenance, developed first for the aircraft industry and then spread out to the other industries [18]. It is defined as a systematic procedure that aims the effective and efficient identification of preventive maintenance tasks for items and their respective maintenance intervals, in a structured and traceable approach [18]. ”It is achieved through a detailed analysis of failure modes and failure causes” [18]. The RCM analysis considers the following questions: [20] ”What does the system or equipment do; what is its function?; What functional failures are likely to occur?; What are the likely consequences of these functional failures?; What can be done to reduce the probability of the failure, identify the onset of failure, or reduce the consequences of the failure?". The items subjected to the RCM end up following one or a combination of the four maintenance methods: Perform no Maintenance, or run to fail; perform preventive maintenance; perform condition based maintenance, which consists in predictive maintenance and/or real time monitoring; or redesign, whenever a system failure is an unacceptable risk and no other maintenance method is useful to mitigate the associated risk. Figure 2.1 present a block based diagram with the possible outcomes of the RCM analysis.

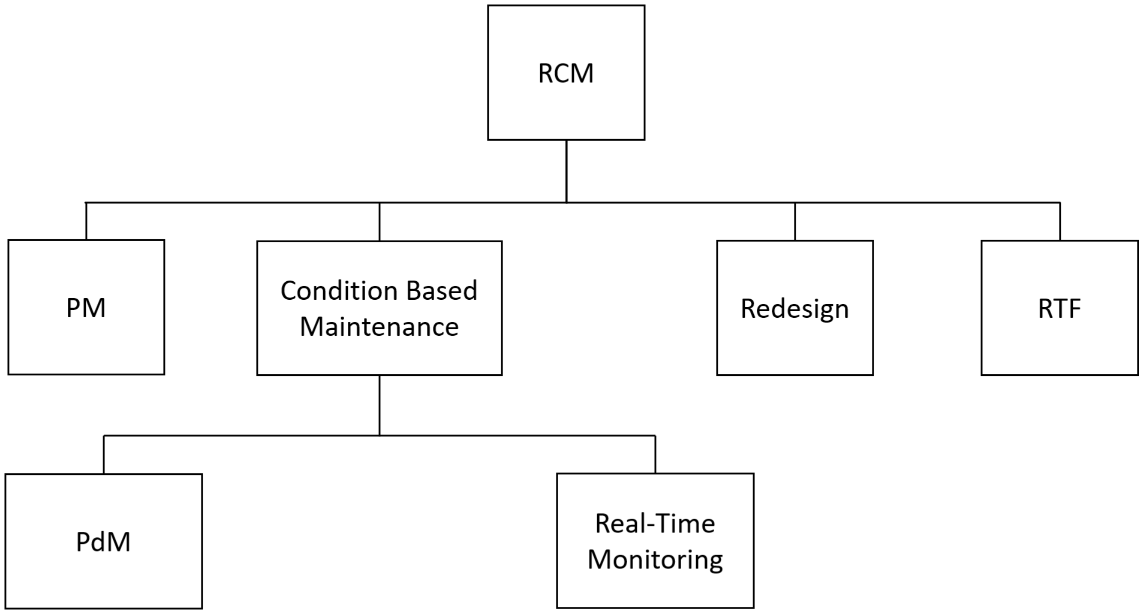


Figure 2.1: Elements of RCM. PM stands for Preventive Maintenance, RTF for Run To Failure and PdM for Predictive Maintenance. Adapted from [20].

The PM branch of figure 2.1 may be subdivided into four subtypes of scheduled tasks [18]:

- Scheduled on-condition task: regular scheduled inspection with the goal of localising any potential failures.
- Scheduled overhaul: known as hard time maintenance, consists on overhauling regularly an item in a scheduled manner before some specified age limit.
- Scheduled replacement: it is as the name suggests, a scheduled discard of part or the entire equipment or item, at or before some specific age limit.
- Scheduled function test: Consists on scheduled tests that aims the verification of the equipment's function.

The intervals and tasks that result from the MSG-3/RCM methodology are considered as "initial minimums" for any aircraft maintenance program. These requirements, as shown in figure 2.2, are published on the Maintenance Review Board Report (MRBR); whereas every aircraft has its own associated Maintenance Planning Document (MPD), a document that contains all the MRBR requirements "plus mandatory scheduled maintenance requirements that may only be changed with the permission of the applicable airworthiness authority" [24]. The MRBR is also periodically updated, with the approval of the aviation authorities, as a result of constant operational feedback between the operators and the manufacturers, which showcases possible adjustments that have to be performed to the maintenance processes.

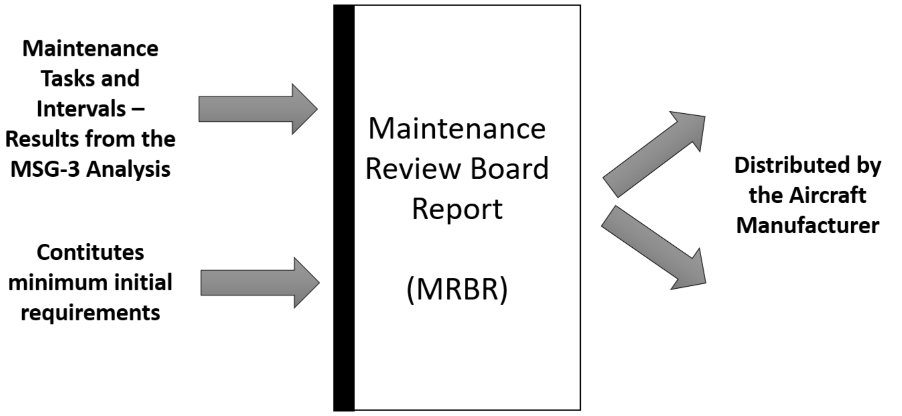


Figure 2.2: MRB Reports. Contains the results from the MSG-3 Analysis, with the corresponding minimum requirements concerning maintenance tasks and intervals. These reports are distributed by the aircraft manufacturer. Adapted from [24].

Scheduled aircraft maintenance is nowadays based on maintenance checks performed periodically, defined by the already mentioned aircraft maintenance programs. Checks are categorised into transit, ”A”, ”B”, ”C” and ”D” checks, subdivided in line and base categories [21]. Line maintenance, includes the transit, A and B checks; while the Base maintenance, includes the C and D checks. The main characteristics of the different maintenance checks are: [21] [22]

- Transitory check - Performed when the aircraft is on ground for more that 4 hours or after each stop. It includes visual inspections that search for evident damage or deterioration.
- ”A” check - Through the opening of main panels the service of certain items is accomplished. Crew oxygen system pressure check is an example of a task performed.
- ”B” check - Similar to the ”A” check, slightly more detailed and different tasks performed.
- ”C” check - It is an extensive check that inspects the serviceability of specific systems. Unlike the above checks, 3 to 5 days is the estimated duration, and it is required an interruption to the aircraft operation. The ”C” checks may also include ”A”, ”B” and transitory checks.
- ”D” check - Thorough and detailed inspection to the aircraft’s structure. It is known as the structural inspection, it includes non destructive testing and other visual and instrumented examination that intend to find structural damages such as corrosion, cracks and other deterioration indicators. This checks are time consuming, with a duration of 20 or more days. Like the ”C” check, this ”D” check may also include any of the other checks.

Figure 2.3 showcases the typical maintenance intervals of several aircraft from the world’s fleet. Some disparity is noticeable between the maintenance intervals for different aircraft. Time measures also vary according to the aircraft in question and its main operation type. For a short-haul aircraft, flight cycles are the determinant factor that define the maintenance intervals; as for long-haul aircraft, the schedule defining indicator is the aircraft's flight hours. For the aircraft that are, for long periods, not operating, calendar months is the variable that defines the maintenance checks. [22]

Aircraft	A check	B check	C check	D check
B737-300	275 FH	825 FH	18 MO	48 MO
B737-400	275 FH	825 FH	18 MO	48 MO
B737-500	275 FH	825 FH	18 MO	48 MO
B737-800	500 FH	n/a	4000–6000 FH	96–144 MO
B757-200	500–600 FH	n/a	18 MO/6000 FH/3000 FC	72 MO
B767-300ER	600 FH	n/a	18 MO/6000 FH	72 MO
B747-400	600 FH	n/a	18 MO/7500 FH	72 MO
A319	600 FH	n/a	18–20 MO/6000 FH/3000 FC	72 MO
A320	600 FH	n/a	18–20 MO/6000 FH/3000 FC	72 MO
A321	600 FH	n/a	18–20 MO/6000 FH/3000 FC	72 MO
ATR42-300	300–500 FH	n/a	3000–4000 FH	96 MO
ATR72-200	300–500 FH	n/a	3000–4000 FH	96 MO

Figure 2.3: Typical maintenance intervals for a number of aircraft present in the world's fleet. FH stands for Flight Hours, FC for Flight Cycles, whereas MO is the abbreviation used for Months. [22]

Unscheduled or corrective maintenance actions are performed, depending on the severity of the fault, during the aforementioned scheduled checks or, on higher severity situations, whenever the issue is reported. The significance of the situation dictates the airworthiness of the aircraft. There are several issues that, although present, do not jeopardize the operation. Each aircraft is equipped with the so-called Minimum Equipment List (MEL), which is a document that is developed and adapted by each aircraft operator and based on the Master Minimum Equipment List (MMEL), a document "developed by the applicant and holders of Type certificate" and approved by the aviation authorities, with the benchmark guidelines for the development of the MEL. The MEL is a document that enables to continue the aircraft operation for a certain period if certain listed critical systems or items did not fail. It defines the critical systems and equipment that, once failed, would lead to aircraft on ground situations, that put the aircraft's airworthiness in question. The MEL also defines, for the non-critical faults, the time window in which the failed system/item needs to go through a maintenance intervention [23].

The above-explained focused mainly on the preventive and corrective types of maintenance. Again, one of the possible outcomes of the RCM analysis is the condition based maintenance. It consists of performing maintenance based on the monitoring of an item to find possible evidences of imminent failures. This may be accomplished through real time monitoring using, for example, the already mentioned in chapter 1 health monitoring software, or by the application of predictive maintenance techniques, such as data based prognostics. This thesis focuses on the latter; hence, a description is presented in the following section.

2.2 Predictive Maintenance and Prognostics

"Most machine maintenance today is either purely reactive (fixing or replacing equipment after it fails) or blindly proactive (assuming a certain level of performance degradation, with no input from the machinery itself, and servicing equipment on a routine scheduled whether service is actually needed or not" [18]. These scenarios are extremely wasteful considering the level of instrumentation present in the modern days machinery. That is the reason why the maintenance world is adopting and moving forwards using new technologies and adopting the "predict and prevent" maintenance [18].

Predictive maintenance is based on the policy of only applying maintenance actions to the equipment once the magnitude of certain reliability indicators reach a predetermined level and lead to the possible imminent future failure. Using a combination of the available performance and diagnostic data, operation logs, or other available physical or digital data, the predictive maintenance uses a combination of human and technical skills to make decisions about maintenance procedures of certain equipment or systems [18]. Predictive maintenance is essentially "fitting a network of sensors to the aircraft or other equipment to measure condition signals" [25]. Those signals may be then used as condition monitoring variables that may be useful to decide the maintenance intervention to a specific item before it fails. This ability to schedule the intervention before a failure event is the main purpose of predictive maintenance [18].

Predictive maintenance is often associated with the most advanced forms of data acquisition and mining techniques. An example of a primary non-automated predictive maintenance action is the analysis of the aircraft's engine oil. According to the results, the next maintenance oil change is scheduled. Hence, although the new innovative technologies contribute to the development and the increasing interest in this technique, predictive maintenance can be performed in other ways.

Prognostics is the term used for the science of making predictions about engineering systems [30]. All the processes mentioned above that aim to predict the future behaviour of systems are considered a form of prognostics. One of the main goals of prognostics is the estimation of the time at which the system or the components is no longer capable of performing its task. Making use of various indicators of vibration, temperature, lubricant condition, among others, that may be extracted from the highly complex sensor network present in today's most complex machines, such as the aircraft, prognostics aims to correlate the indicators to a possible future failure event, hence reducing the system's unpredictability. It may be stated that the application of prognostics in maintenance results on the predictive maintenance.

A predictive maintenance program benefits several stakeholders present in the aircraft sector. Maintenance profits from an improved maintenance scheduling, an increase in the system's up-time and the minimisation of unnecessary maintenance actions. From the engineering point of view, the increased availability and reliability can assist on the validation of the design and may also be useful in evaluating systems' robustness. For the customers, less unexpected cancellations resulting from the scheduling of imminent failures can improve customer satisfaction and the brand's image. Logistically, better prediction of unforeseen maintenance actions may also reduce the inventory of spare parts and the respective environmental footprint [25]. Apart from previously mentioned benefits, cost reduction plays a major role in the emerging interest of airline operators in prognostics. Every time the aircraft is grounded, it represents a major cost. A better scheduling of maintenance actions, and a reduction of downtime will result in operation cost reduction. Beyond the above-mentioned points, safety is the most important factor assisted by these methods, as the capability of predicting failures may avoid fatalities.

There are three main classifications regarding the current prognostic approaches [18]:

- Model-based approach: It is a quantitative model that requires detailed knowledge of the physical relationships and characteristics between all the integrating components of a system. Physical variables are used to compare and to identify differences between the current and the predefined expected working state derived from the developed physical model. Maintenance actions are performed to the system if important differences are reported, and the reliability of the system falls below certain defined levels. Model-based approaches are often prohibited due to the high complexity levels of the systems in question, which turns the task of developing models based on the relationships and characteristics of all the related items within a system too complicated, leading to model accuracy that falls short of what is desirable. In addition, modeling the evolution of a failure from a phase that is still considered as operational to the phase that prevents the component from performing its task is not always reliable and feasible.

- Data-driven approach: This approach takes advantage of large amounts of data available from historical records, both from normal and faulty operations. No priori knowledge of the process is required as it only develops models from measured data from the process itself. However, general knowledge of the system or process may be useful to interpret results. Artificial intelligence, mainly machine learning or other pattern recognition methods, are tools used to treat and develop data-driven models that assist the decision-making process of maintenance teams. The low cost inherent to algorithm development, and the high adaptability and flexibility of data derived models are two additional advantages over the model-based approach.

- Hybrid approach: As the name suggests, it is the result of a fusion between the two above mentioned methods. Model-based and sensor-based models are combined with the end goal of obtaining models with better accuracy and reliability.

This thesis focuses on data-driven techniques. Prognostics rely on methods that can follow and analyse the trend in data, and forecast the next failure occurrence. Hence, machine learning is a very useful tool for prognostics. It is defined as a sub-field of computer science and artificial intelligence that explores the development of algorithms closely related to linear algebra, probability theory, statistics, and mathematical optimisation that can learn from data and make subsequent predictions. Machine learning allows data analysis otherwise not feasible with more conventional methods. It enables machines to learn by themselves based on provided data with the goal of making predictions [26]. The general workflow of machine learning consists in the reception, by the computer, of either textual, visual, audiovisual, or numeric data, such as historical data from health monitoring parameters and the corresponding system working state. The machine learning algorithm integrated into the computer is then able to learn from that data and develops a model that is capable of making predictions when faced with different input data, such as current health monitoring parameters. Basically, the goal is to turn a machine into a thinking being that can learn from data, develop a model, and proceed with the correct estimations or predictions when faced with new data.

Machine learning, depending on the nature of the learning data, may be classified as: [26]

- Unsupervised learning: It uses unlabelled learning data, i.e., learning data that has no description associated, only containing indicative signals. Anomaly detection, such as defective equipment, or to group customers with similar online behaviours are, among others, two of the applications associated with this category of machine learning.

- Supervised Learning: Unlike the above-mentioned category, supervised learning uses learning data denominated as labelled data, with description, targets, and desirable outputs. The models under this type of machine learning have the goal of finding the relation between the input and output of the training data, and then apply the found relation to make predictions using newly inputted data. Face and speech recognition, product recommendations, sales forecasting, and failure prognostics are, among others, the typical application of this type of learning method. Supervised learning may be subdivided into regression and classification. The first trains and predicts continued float point numbers, such as car prices, or the remaining useful life of a component, whereas the latter intends to predict class outputs, such as if the car price is above or below a certain value, or if the failure risk of a certain equipment is high or low.

- Reinforcement Learning: This type of learning method uses learning data that provide feedback ”so that the system adapts to dynamic conditions to achieve a certain goal in the end”. The system uses that feedback information to perform adjustments to the predicted output. Robotics and maintenance planning are two fields that use this type of learning.

Figure 2.4 represents graphically the fundamentals of what was referred above.

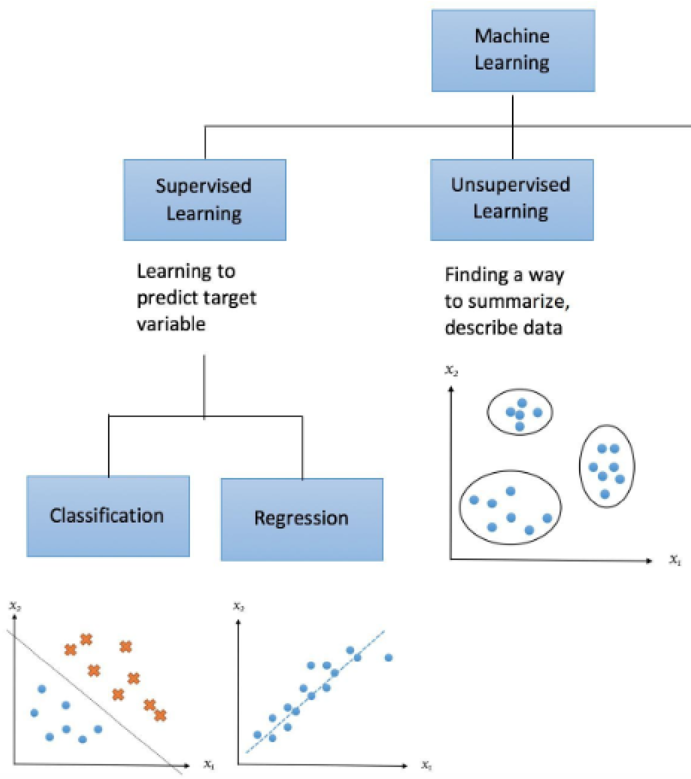


Figure 2.4: Machine leaning categories [26]

Apart from the three already mentioned subtypes of machine learning, if the training data is composed both by labelled and unlabelled data, then the method falls into the fourth type of machine learning, semi-supervised machine learning. It is used when the acquisition of fully labelled data is too expensive, or when labelling the data is a highly complex process that requires skilled experts.

There are two major components inherent to supervised machine learning analysis: features and labels. The features are the problem's descriptive attributes or variables, and the labels are the variables that the machine learning algorithm aims to predict. Again, depending on the label's nature, supervised machine learning may be regressive or classifying.

Additionally, machine learning englobes another subfield denominated as deep learning. Basically, it represents other techniques and algorithms of performing machine learning [29] that is inspired by the way the human brain processes information. The performance of deep learning is similar to the more classical machine learning algorithms whenever the available data is not excessive. The main difference arises for problems that require the analysis of much more significant amounts of data, while the machine learning performance stagnates, deep learning results continue to improve, as shown in figure 2.5

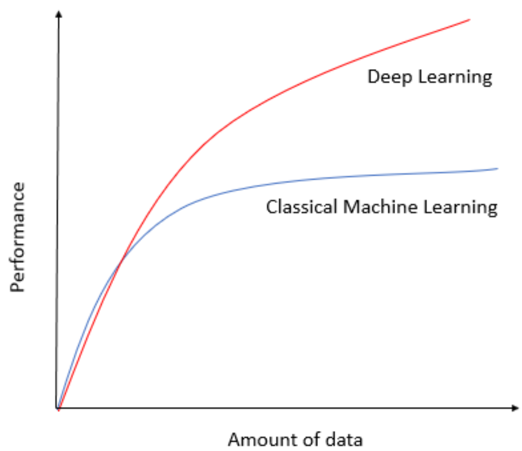


Figure 2.5: Evolution of performance with the amount of available data for classical machine learning and deep learning. Adapted from [30]

This thesis is focused on prognostics using supervised machine learning, as all the data available is labelled. As mentioned in chapter 1, the goal of this thesis is to study the messages from the aircraft fault report system using data-driven machine learning methods, in order to prognose the remaining useful life of an aircraft system. CMC messages are continually being emitted from the aircraft avionics, and it is believed that those messages may be used as degradation indicative variables. Therefore, they may be able to model/predict certain critical failures and therefore allow to schedule a priori maintenance interventions to the system in question. To study the predictive power of the CMC messages, both classification and regression supervised machine learning is used to model the influence between the CMC messages and certain predefined failures. The error associated with the models act as a measure of the predictive quality of the messages in question. The next section showcases some previous work done in this field and highlights both the similarities and differences between the work carried out on the selected articles and this thesis.

2.3 Related work on maintenance prognostics

This section aims to describe three selected projects ([27], [28], [25]), with objectives that are related to this thesis. The comparison is focused mainly on the used data types, the general methodologies, the algorithms implemented, and the concluding remarks. As a common characteristic, the three projects developed, in different ways, prognostic tools that aimed the prediction of the health state of a system. There are similarities in the three approaches, yet, the methodologies differ, and in the following paragraphs, a brief review is performed.

Starting with the data types, the authors of [27] used ”all the available parameters that could be related to a failure of the system/component under analysis”. These included data both from the crash protected and non-protected flight recorders. Basically, the data consisted in raw sensor data from the data recorders placed in the aircraft. Furthermore, the authors considered as failure indicative actions maintenance logs, such as replacements, cleanings and adjustment actions. Instead of using raw sensor data as the previous publication, the authors of [28] used, like in this thesis, central maintenance system messages as variables. Finally, the last project here presented [25] uses a combination of the last two, analysing the evolution of the results using each data type separately or both in the same analysis.

Focusing first on the project [27], the authors studied the left-hand bleed valve unit, present in a twin-jet aircraft model with a passenger capacity of approximately 100, due to the component’s low mean time between failures (MTBF) and criticality. The parameters considered, post the data pre-processing, included the ” bleed manifold pressure, temperature, the high-pressure compressor speed of the engine where the bleed is taken from (left engine).” These parameters were transformed into 24 features/ variables associated with each flight: ”mean, standard deviation, skewness, kurtosis, median” in the time domain; ”rms power, peak, and power over a 0.002Hz” in the frequency domain. ”To reduce the contribution of the flight profile on the bleed unit behaviour” only the stable cruise flight data from the flights with a minimum of 20 minutes of stable cruise flight phase were considered [27].

The study rested on the classification of the component’s health. The authors defined the two classifications on whether the component is within a certain period of a failure event. The authors defined this period as 30 days. Hence, the left-hand side bleed valve was considered as unhealthy if the nearest failure was within 30 days of a certain date reference, and considered as healthy if the flight was within 30 days after a replacement. Situations that didn’t fit in either classification are not considered. The definition of failure resulted from the filtering of maintenance logs. Due to the lack of importance of some, the only ones considered and defined as failures were those that were resultant from a bleed valve replacement [27].

Post the introduction of the data into a Support Vector Machine (SVM) classifier (algorithm explained in chapter 5), the authors decided to define a so-called degradation index, based on the binary predictions, that aimed to ”smooth the effect” of the misclassifications resultant from the predictions. It was computed as ”the rate of unhealthy classifications from the total amount of classifications” within 30 days. Figure 2.6 showcases the evolution of such degradation index from all the replacements of an aircraft [27].

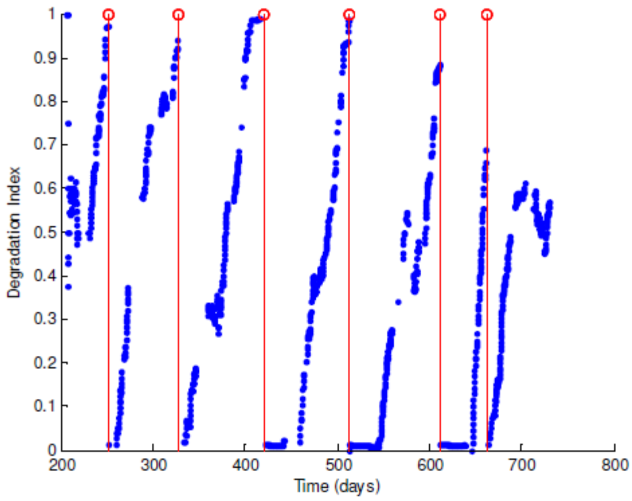


Figure 2.6: Degradation Index evolution. Vertical lines mark the replacement timestamps [27]

Also using a classification machine learning approach based on the SVM algorithm, the authors of the second article [28] stated that the criteria used to identify whether the objectives were or weren't met relied on a Notice Period (NP), i.e., time in advance of a fault occurrence. A minimal NP of 2 flights was defined and, to avoid too early alerts that might decrease the useful life of the equipment and therefore imply increased costs, 12 flights were defined as the maximum "Wasted Life (WL), identifying the largest amount of time a failure can be anticipated in a cost-effective manner". So, the overall goal of this project consisted in alerting the user if the failure occurrence is predicted to be between 2 and 12 flight legs prior to a particular reference time or flight [28].

The prognostic architecture used is different from the previous article. The authors computed an R-value which was considered as "the memory of the CMS fault recording process", defined by 2.1 [28]:

$$R(K) = \sum_{m} C\vec{M}S(m).C\vec{M}S(m-k) \tag{2.1}$$

$C\vec{M}S$ is defined as the number of messages emitted in a certain flight m. Therefore, the cardinality of the vector is equal to the number of different message types. The authors observed that R(20) was null, and consequently decided not to analyse groups of more than 20 flights. The general methodology consisted in the analysis of flights 2 days before every failure event, considered by the authors as replacement maintenance logs [28].

In the end, the authors presented a further analysis that aimed at the identification of the failed component, as the analysis above only predicted an unspecified replacement. The analysis included the usage of techniques such as the principal component analysis, eigenfaces, and others found for further detailed inspection in [28].

These first two projects met the desired objectives. In [27], the authors stated that the results were good, especially considering the limited amount of data. No quantitative results were demonstrated in the article; however, the good behaviour of the degradation index confirms the good results [27]. For the case of the second publication [28], the authors concluded that the results were overall good, as the main goal of developing a model with precision above 50% was achieved [28].

Finally, the author of the project [25] used, unlike the previous, a mix of classical and deep learning algorithms, and compared the results obtained with and without the application of one technique called Kalman filtering. Linear regression, K-nearest neighbours, Neural Networks, Support Vector Regression, and Random Forest were the classical machine learning algorithms in use. The deep learning algorithms consisted in six different recurrent neural network algorithms, such as the long short term memory and gated recurrent unit were applied to the data. Furthermore, the machine learning algorithms mentioned were used to predict the remaining useful life of the equipment and not to predict if it is between a specified time range. Hence, machine learning in question was regressive, unlike in [27] and [28].

The data used in the case study of interest consisted in 588 bleed valve removals, the majority being unscheduled, which defined the failure of the bleed valves. The author was also provided with around 100Gb of time-series of health monitoring signals of the engine bleed system (raw sensor data) and with maintenance messages from the central computer of the aircraft (in this thesis defined by the CMC messages). Based on these, four model categories were created based on the different combinations of data used.

According to the author, the goal was to develop models that would predict the remaining useful life of the bleed valves with a maximum mean absolute error (MAE) of 10 days, defined by the manufacturer Embraer. The major improvement of the results was obtained with the introduction of sensor data to the models. The errors were reduced to approximately half of what they were with the first two model types (containing only the message data), from a MAE of around 70 days for the first two category models to approximately 30 days with the introduction of sensor data to the models. The author analysed that ”the results confirm the intuition that is easier to extract information from sensory signals” than from message information. The algorithm with the overall better results was the Random Forest, with a minimum MAE of 22.7 days using all the available data combined [25].

The introduction of the Kalman filter, whose definition is explored in detail in [25], produced an overall improvement on the results, specially due to the smoothing of the data noise and allowing a better interpretability of the results. 15 days of mean absolute error obtained from the random forest algorithm was the best overall result of the case study in question, but still, it failed to reach the target of a MAE of 10 days. Also, the Kalman filter produces more accurate results near the failure improving the results for the most critical predictions [25].

Finally, the introduction of more advanced deep learning algorithms also contributed to the general decrease in prediction errors, despite the best result not reaching either the target of a MAE of 10 days or the 15 days achieved with the classical plus the Kalman filter. The author stated that ”deep learning models present a promising alternative to traditional machine learning models, especially for precision near the potential failure”. [25]

2.4 In Summary

The above three mentioned projects are useful to showcase the variety of prognostic approaches already applied in maintenance. It is, in fact, a complex field, with results that fall short of desirable total accuracy. Also, the variety of data used to develop the prediction models demonstrates both the high level of instrumentation present in today's aircraft and the overall uncertainty in the choice of the better data type to predict those failures. It is also important to notice that the two first classifying models from the two first articles explored fulfilled the authors' expectations, as the latter case study, possibly due to the regressive nature of the machine learning problem, missed the target proposed by the author. This latter described case study also demonstrated the evolution in the results due to the usage of different type of maintenance data, the most significant improvement came with the introduction of sensor signals to the models. The application of more complex recursive algorithms also led to an increase in model performance, especially for the prediction of the most critical near the failure remaining useful life.

Nothing is perfect, and these models are not an exception. Even if the models would be able to predict with almost perfect precision, the definition of failures used in the three projects, i.e., replacement maintenance logs, do not necessarily represent the end of the life of the equipment. Most replacements are performed, due to safety measures, according to human perception of health monitoring parameters or even due to complaints. As a result, not all the replacement actions are close to the true end of life of the equipment, and therefore, as the models learn according to those maintenance actions, predictions may fail to identify the real end of life indicators, meaning that the ground truth may fail to be captured.

Prognostics using data science is an emerging field and is attracting the attention of many worldwide companies that seek new and innovative ways to increase safety, efficiency improving maintenance scheduling, and avoiding all the costs and struggles of unforeseen failure events. The aeronautic sector is one of them, and there is already several commercial software available such as the PROGNOS developed by Air France and KLM.

This thesis uses classical machine learning algorithms to inspect the predictive capabilities of the CMC messages. Notice that deep learning was not explored because the size of the available data did not justify the implementation of such algorithms. Based on the results shown in the previous section 2.3, there are indicative signs of the usability of CMC messages to formulate prognostic models. On the other hand, the explored case study of [25] showed that the significant improvement of results was obtained with the introduction of sensor signals to the model and that the results obtained solely with the messages were not satisfactory. This thesis aims to clarify, using real-world data sets, if the maintenance messages may be used as predictors to develop prognostic models. Worth mentioning that the development of this thesis took place during a six-month internship in the technology department of Portugália Airlines. The following chapters resume the methodology followed.

Chapter 3

Industrial Case Study

This thesis was developed in collaboration with Portugália Airlines, during a six-month internship. In the following section, a brief description of the company is presented. The messages already mentioned in the previous chapters represent decoded sensor data, pre-processed by two computers embedded in the aircraft's avionics, described in section 3.2. Also, the main initial focus of this thesis is the treatment of data regarding the aircraft's pneumatic system, and therefore, section 3.3 presents a brief overview of it. The rest of the chapter presents the methodologies and the initial remarks captured during the six-month internship in the airline company Portugália Airlines.

3.1 Portugália Airlines

In 1988, with the initiative of Coopav - Cooperativa de Pilotos (Pilots Cooperative), transconsult - Gabinete de Estudos e Projetos (Studies and Project Bureau for the Transport Sector) and Grupo Espírito Santo, a new airline company was founded. Two years passed until the inaugural flight was possible, more precisely, was the city of Porto, which received the first-ever flight from Portugália Airlines, operated then by a Fokker 100 that departed from Lisbon, on the day of 7 July of 1990. On the same day, the route Lisbon/Faro began [6].

In June 1992, Portugália Airline's internationalisation began, with regular flights from Lisbon to Cologne and Strasbourg. The acquisition of two new Fokker 100 meant that the company was able to explore market niches in Europe. The expansion continued in January 1993, with the full liberalisation of the European Airspace, expanding the route map to Brussels, Madrid, Hannover, Mulhouse/Basel and Stuttgard. This year also added two Fokker 100 to the fleet [6].

From 1997 to 2000, the route map grew, including 5 new destinations in France, 7 in the Iberian Peninsula and the United Kingdom, and the first non-European destination, Casablanca. The fleet also increased in 1997 with the purchase of six new flexible Embraer 145 and stabilised in the year 2000 with 6 Fokker 100 and 8 Embraer 145 [6].

In 2007, Portugália Airlines integrated the TAP Group, and left the independent exploration of the regional market. This acquisition strengthened the group, reinforced the Porto hub, and improved the group's offer. Worth mentioning that the contractual agreement between the companies is known as Wet Lease, as Portugália Airlines provides the aircraft, the crew, maintenance, and the insurance, and TAP is charged for each flight hour, the fuel, and the airport fees. That being said, both Portugália Airlines and TAP are still independent companies [6].

The year 2016 brought a fleet renewal. Both the Fokker 100 and the Embraer ERJ145 where replaced by 9 Embraer E190 (E190-100) and 4 Embraer E195 (E190-200). This increase in passenger capacity was followed by a brand image renovation [6].

Nowadays being operated under the TAP express brand, the current fleet stands since the last renewal in 2016, with 9 Embraer E190 (E190-100) and 4 Embraer E195 (E190-200), seating, in the current seat layout, 106 and 118 passengers respectively. Both these aircraft allowed Portugália Airlines to explore new markets, due to their high performance in terms of economy and efficiency. The main difference between the two variants is the number of seats, being the two variants similar in any other aspects. A brief overview of the aircraft specs is presented in the next section.

Embraer E190

The Embraer E190 is a narrow-body medium-range twin-engine jet airliner, capable of carrying a maximum of 118 passengers on a 2-class configuration (E190-200). Two General Electric turbofan CF34-10E power this aircraft, capable of providing 200,000 pounds each. Built by Embraer, the Brazilian manufacturer delivered more than 600 E190 aircraft all around the world since 2004, the date of the first flight. Table 3.1 lists the general specification of the E190-100 aircraft.

General Specifications (E190-100)	
Engine	2 x General Electric CF34-10E
Power	2 x 20,000 lbf
Avionics	Honeywell Primus Epic EFIS
Max Cruise speed	Mach 0.82
Service Ceiling	41,000 feet
Range	4,537 Km
Seating capacity	106 (E190-100) 2 class layout
Fuselage Length	36.24 meters
Fuselage Diameter	3.01 meters
Wingspan	28.72 meters
Maximum Take Off Weight (MTOW)	50,300 Kg

Table 3.1: Embraer E190-100 general specifications [9]

Apart from the number of seats (E190-200 holds 12 more seats) and the therefore increase in fuselage length (38.65 meters), the two variants are similar in any other relevant aspects.

In this thesis' context, it is important to understand how the CAS (Crew Alerting System) and the CMC (Central Maintenance Computer) messages are generated. The following subsection 3.2 explains just that.

3.2 Message generating system

Each aircraft is capable of generating several megabytes of data per flight. Mining this data is extremely important to understand behaviour and tendencies in the aircraft's health. Being the sensor data extremely difficult to analyse without any pre-processing, the aircraft has two computers whose role is to decode the sensor data and present it in a more concise and user-friendly way: MWF (Monitor Warning Function) and CMC (Central Maintenance Computer).

The MWF, present in the aircraft's Module Avionics Unit (cabinet that holds multiple avionic systems) continuously monitors the avionic systems status and alerts the crew in a message form, known in this thesis by CAS messages. These messages represent decoded raw data from several aircraft systems. Due to their informative importance, they are automatically emitted on the CAS display, which is a fixed size window on the right-hand top corner of the EICAS display, present in the center region of the cockpit. Worth mentioning that these messages have a severity spectrum that can range from simple indicative messages to fatal failures. Pilots are trained to respond to these situations.

On the other hand, the CMC is a module integrated into a major system in the aircraft - Central Maintenance System (CMS). This system is a combination of a fault recording system and a maintenance access system. Each aircraft subsystem supplies real-time and captured data to the CMS. A layout overview of the CMS in presented in figure 3.1.

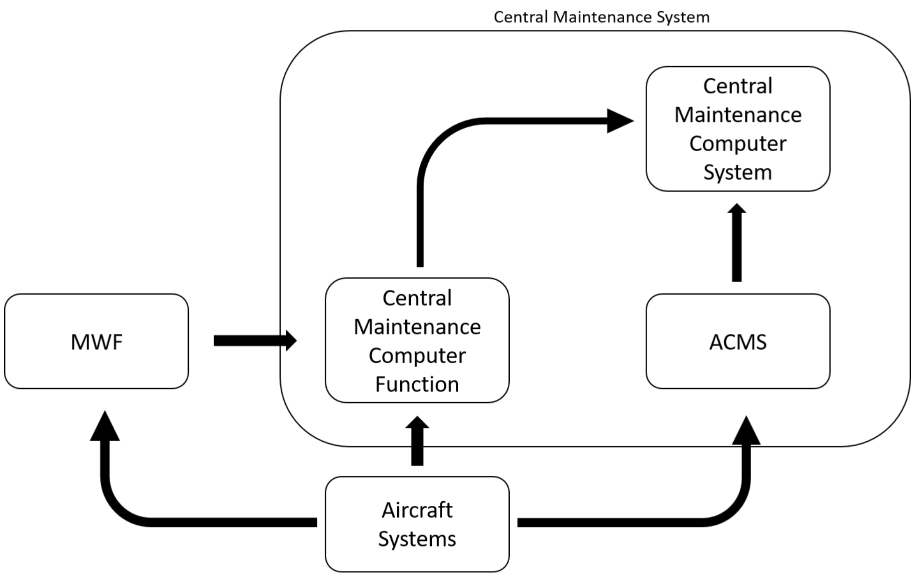


Figure 3.1: Central Maintenance System - Overview. Adapted from [8]

The CMCS (central message computer system) allows access to the member systems (any system installed on the aircraft that both comply with the CMC requirements and implements the features in the CMC) from a single user interface. The computer is powered and functional in all flight phases and is always communicating and receiving data from the member systems, through two different standard communication data formats: ASCB and ARINC-429. Data is transferred to the ACMS (Aircraft Condition Monitoring System) and CMFC (Central Maintenance Computer Function). These modules are responsible for the data analysis and output reports in a message format. The two data variants are then joined by the Central Maintenance Computer system, where all the decoded sensor data may accessed as maintenance or CMC messages.

The CAS messages from the MWF are also integrated into the CMS, as one of the CMCF tasks is to compare and analyse the interaction between CAS and CMC messages. Worth mentioning that despite being accessible from the MFD (Multi Function Display, present in the aircraft's cockpit), this system also integrates a database where both the maintenance and CAS messages are stored, denominated as Fault History Database, or FHDB. In the CMCS interface, one of the options available is the ”Export Fault History”, and regularly, the recorded messages are downloaded through the aircraft's LAN port (Embraer E190-200) or the Data Management Unit (Embraer E190-100). Once this regular task is completed, these files are stored in the company's AMOS software, the Maintenance, Repair and Overhaul (MRO) software solution used by Portugália Airlines.

Every single message has an ATA chapter and section associated, which eases the process of separating the messages through the aircraft's several systems. According to the company's engineers, one of the most recurrent ATA chapters is the ATA 36, referring to the aircraft's pneumatic system. Although this thesis' overall goal may be adapted to any aircraft system, as a first approach, the main focus is the analysis of the pneumatic system's data. Hence, a brief overview of this system is presented in section 3.3

3.3 Pneumatic system

The pneumatic system supplies controlled bleed air to either the anti-ice, air conditioning and engine starting systems. The AMS (Air Management System) is a embedded controller in the aircraft avionics which responsibility is to manage the bleed air flow through the mentioned systems. The pneumatic system includes three subsystems: Air Bleed Distribution; Indicating; and Ozone Converters [8].

3.3.1 Air Bleed Distribution

The air bleed distribution comprises the bleed air manifold components. Its role is to receive, control, and distribute the hot and compressed air to other aircraft systems. This system receives compressed and hot air from: the engines' low stage supply port (5th stage of the engines' compressor); engines' high stage supply port (9th stage of the engine's compressor); the APU; and from a ground source. This hot and compressed air is directed to several systems, such as engine start, air conditioning, ice protection, and water tank pressurisation.

Figure 3.2 showcases a block diagram of the air bleed distribution. It includes, as subsystems, the engine pneumatic bleed system, which provides bleed flow selection between the high pressure and low-pressure engine bleed ports, and acts as a regulator and controller to the bleed pressure and temperature prior to delivery to the pneumatic system bleed air manifold. Other subsystems include the APU (Auxiliary Power Unit) pneumatic bleed system, the ground air supply, control, and water tank pressurisation. The first uses and controls pneumatic power from the APU. Whenever the aircraft is not

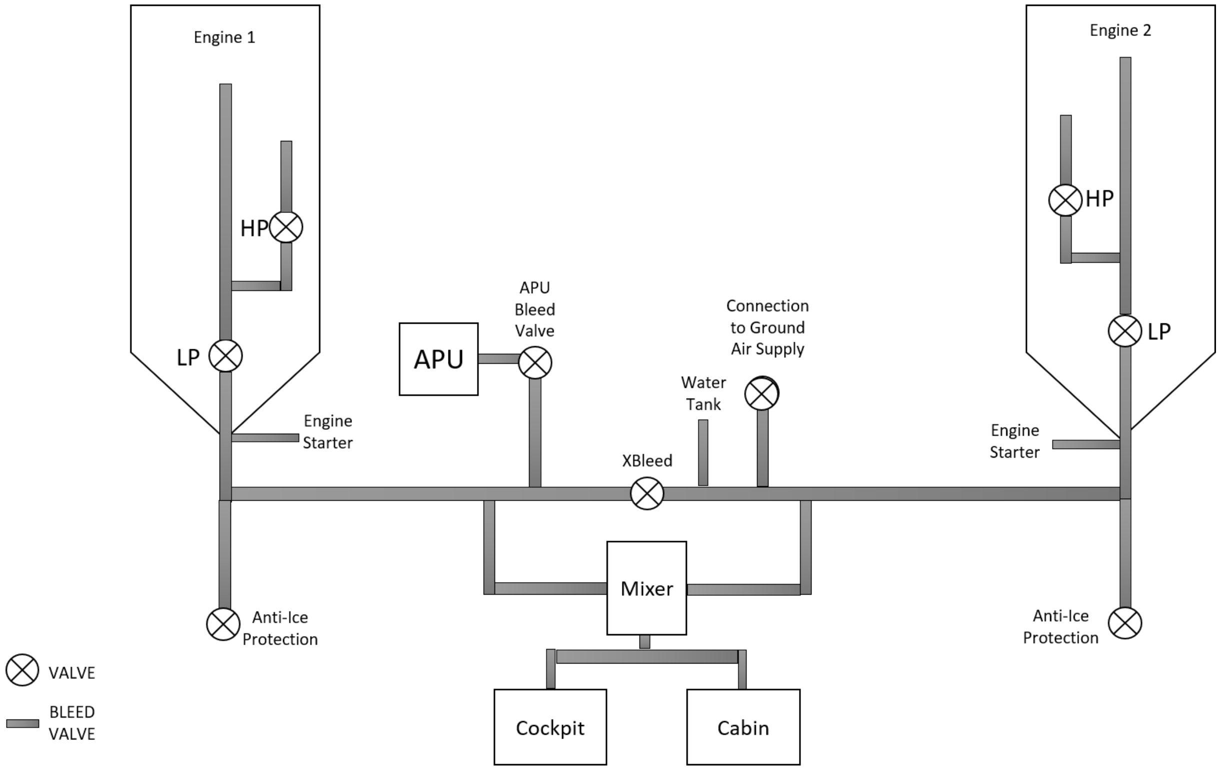


Figure 3.2: Air Bleed Distribution - Simplified Block Diagram. HP stands for High Pressure, and LP for Low Pressure. Adapted from [8]

operating the main engines, for instance, at the gate or even in emergency situations, the APU is capable of supplying air to the aircraft's pneumatic system. The ground air supply allows the connection of an external high-pressure source to the aircraft, the control provides all elements and components to control the air bleed system operation, manifold pressure and temperature, having a built-in logic software that continuously monitors the system components' performance. Finally, as the aircraft is equipped with a portable water tank that supplies water to water heaters, lavatory faucet, water spigots, and coffee makers in the galleys, the pressurisation of this water reservoir is accomplished by the bleed system.

Worth noticing the symmetry of both bleed channels, as air from either the left and right-hand side engine, are sources to any of the receiving systems. Several different valves complete the functioning of the air bleed distribution system, as well as temperature and pressure sensors located by the air channels. Furthermore, the air bleed distribution is also controllable using the air conditioning/pneumatic control panel in the cockpit's overhead panel, and the EICAS display acts as an external interface of the system's state. The switches that control the pneumatic systems are identified in the same way that the CAS messages are, being divided into bleed 1, bleed 2, bleed APU, and xbleed. The bleed 1 refers to the left-hand side of the air bleed distribution system; likewise, the bleed 2 corresponds to the right-hand side. Bleed APU, as the name indicates, regards the APU subsection of the system, and the Xbleed the valve that acts as a connector to the left and right-hand side of the bleed distribution system, named cross bleed valve.

3.3.2 Indicating

The indicating subsystem is composed by all elements and components that are able to provide the system status indication and overheat or leak detection throughout the left hand (Bleed 1), right hand (Bleed 2), and APU bleed lines. It includes two subsystems: Engine pneumatic Indicating; and APU pneumatic indicating. The first is responsible for the engine bleed status indication, and the second for the status of the APU bleed subsystem. This system acts as a complement to the air bleed distribution as it monitors and continuously emits reports of its working state. It is interlinked to the CMS, and MWF mentioned in the previous subsection that decodes all the sensor data provided by the sensors and report the status via either maintenance or crew alerting system messages.

3.3.3 Ozone Converters

Not as important in this thesis' context, the ozone converters are used to convert the ozone present in the air during high altitude flight into oxygen. There are two converters present in the aircraft bleed system: one in the left-wing, for the left side of the bleed system; and one in the right-wing, for the right-hand side bleed system.

3.4 Internship and Industrial Challenge

As before mentioned, this thesis was developed during a six-month internship in the Portugália Airlines' Engineering/IT strategy department. Its main task is to apply innovative, technology-based, up to date solutions to the airline company. Using various open-source and proprietary software, this team automatised several manual processes and allowed the visualisation and analysis of previously unprocessed data. Fuel data treatment, crew scheduling assistance, sensor data treatment and visualisation, quantification of the total pilot's cosmic radiation exposure are some examples of the department's implemented tasks, made much easier and sometimes only possible due to the integration of technology-based solutions to the company.

The development of the thesis during an internship embedded in a company department allowed the identification and experience in person some of the complaints and worries transmitted on the first thesis meetings, such as the high rate of message emission and the maybe excessive dependency on the knowledge and interpretation of the messages by the troubleshooting staff, without having implemented data analysis techniques as assistance.

The messages are emitted by the aircraft's internal computers explained in 3.2. These messages are mainly directed to the troubleshooting department. Between other responsibilities, this department continuously monitors, analyses, and takes the necessary responsive actions when faced with certain alerting situations, such as emitting corrective work-orders. For that, one very useful tool used by the airline company is called AHEAD-PRO. It is a health monitoring system that is continually showing the fleet status to any of the departments with access to the platform. The status is mainly focused on if some message that represents aircraft on ground risk was emitted. Figure 3.3 showcases the AHEAD-

PRO's interface. Each block represents the current status of each aircraft in specific according to the respective aircraft box colour. Along with the status, the system displays each aircraft's flight number and whether the aircraft is airborne or at one of the destination's airports.

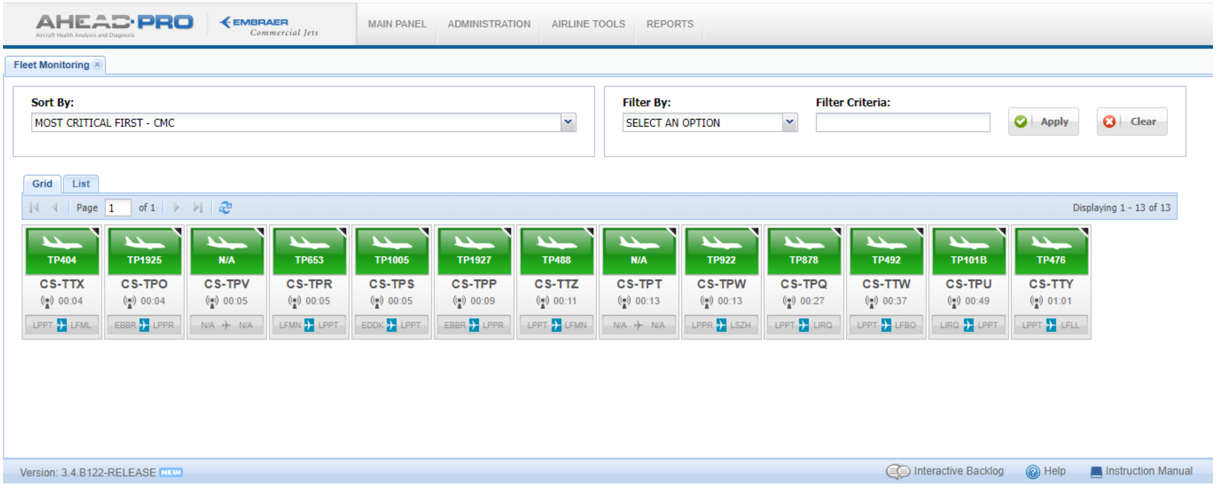


Figure 3.3: AHEAD-PRO main interface [32]

Clicking on any of the boxes guides the user to the interface shown in figure 3.4.

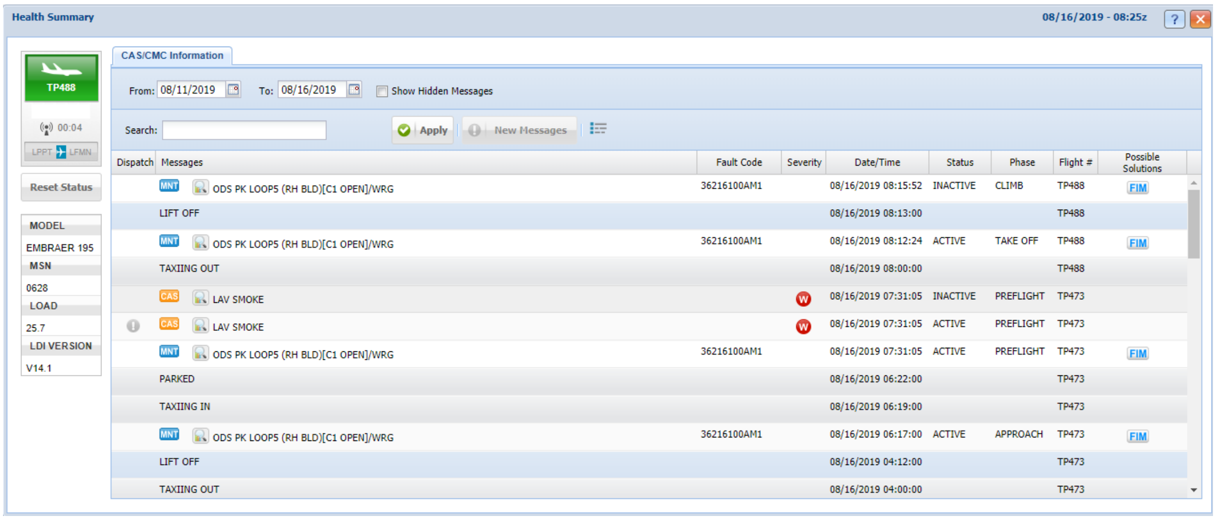


Figure 3.4: AHEAD-PRO interface - message reports from a specific aircraft [32]

This interface is the one troubleshooters use to inspect the messages emitted by the fleet. Either CMC or CAS, they are categorised according to severity, which is defined by Embraer and classifies the messages according to the risk of jeopardising the aircraft operation. The severity is then used to define the current aircraft state, indicated by the box's color in the figure 3.3. As shown in figure 3.4, the interface also includes other information. Associated with the CMC messages, there is always a link to the fault isolation manual, which indicates possible resolutions to the likely problems associated with the message. The other columns represent the fault code of the maintenance messages, the status (active or inactive), and the flight phase of the aircraft at the time of the messages' emission. Another option available is the "show hidden messages" select box. The AHEAD-PRO software applies a filter that only keeps messages considered as relevant. This filter's software insights are not known by the airline operators, that have to trust the algorithm to use the filter. The feedback given by the troubleshooters indicates that the filter is used by Portugália Airlines as a first approach and, whenever there is the need for a further investigation, all the messages are considered. Apart from the AHEAD-PRO, the company has at their disposal other similar software, such as the Embraer Decoder. Data visualisation regarding the messages relies on software built by a former employee, which plots the occurrence of a specific message and compares it with the average of the rest of the fleet. The troubleshooters also use another inherent functionality of the AHEAD-PRO, which enables the download of an excel file with all the message history of the last 3 months. The excel may then be used to filter particular messages and allows, to some extent, data visualisation. Notice that, according to the feedback provided by the troubleshooting team, this latter message analysis method is only applied punctually to investigate some specific situations. Worth recalling that the messages are regularly downloaded directly from the aircraft and stored as compressed folders in the airline company's MRO software - AMOS.

3.5 Case study/Methodology

CMC and CAS messages represent several megabytes of aircraft health reports. The aircraft airworthiness may be dependent on whether a fatal CAS message may or may not be emitted by the avionic system. Some messages represent an AOG (Aircraft On Ground) risk, and are capable of interrupting the otherwise smooth operation on a daily basis. Although the failure CAS messages may be able to alert the crew and indicate what the next steps to troubleshoot the issue are, they are not capable of predicting the future failure event, as their appearance only occurs after the system failed. Due to the high incidence of maintenance messages which represent, in greater detail, the system's behaviour, this thesis intends to answer the following industrial hypothesis: Do the maintenance messages have any predictive power over the crew alerting system's failure messages?; Is it possible to predict a future failure event based on the appearance of maintenance messages?.

The high number of emitted messages per flight turns the task of visually finding patterns in the emitted data almost impossible, not allowing maintenance personnel and process engineers to take a proactive instead of a reactive approach to failures. Considering only the messages emitted between the off-block and on-block timestamps of each flight, i.e, disregarding messages from possible pre-flight checks and maintenance periods, over the 3 year database provided by Portugália Airlines, an average of 15 CMC messages per flight were emitted, reaching a maximum of 7751 CMC messages emitted during one specific flight. It's easy to understand the struggles of finding justification for a short to medium-haul flight with 7751 emitted messages. Over the entire period of exactly 3 years and 27 days, 1.6 million CMC messages were emitted, resulting in an average of 1435 records per day, over the entire 13 aircraft fleet. The significance of this high emission rate is augmented by the messages' reporting nature of the aircraft system's state.

Despite the important assistance of software such as the already explained AHEAD-PRO, there is still a major component of the interpretation of the messages that rely on the knowledge and experience of the troubleshooting staff. As an example, there are several messages with the highest level of severity associated, such as smoke indications on numerous aircraft compartments, that are ignored based on the troubleshooter's experience because, most likely, they represent routine pre-flight checks. Therefore, taking advantage of the groundbreaking data-based methodologies available, the development of a failure prognostic tool would be an important or even crucial complement to the decision-making process that precedes the triggering of specific maintenance actions. Hence the importance of studying the predictive capabilities of one data type emitted by the aircraft - CMC messages.

This thesis' approach to clarify the predictive capabilities of the message data follows six main steps:

1. Define how the messages' predictive power will be quantified
2. Choose the evaluation measures and the desired results
3. Combine different pre-processing techniques and apply to the data
4. Train and optimise the selected machine learning algorithms
5. Evaluate the models' performance
6. Compare the different results from the different data sets and models built

The methodology used in this thesis rests on the application of machine learning algorithms on the available data, with the outcome being the estimation of the remaining useful life of the system, i.e., the time left until the next failure event. Using data pre-processing, feature engineering, and applying the methodology first to the aircraft's pneumatic system, the algorithms' end goal is to estimate the future appearance of the CAS messages "Bleed 1 Fail" and "Bleed 2 Fail" based on previous maintenance messages. The quality of results is adopted as an indicator of the predictive capabilities of maintenance messages. This ability is also measured by comparing the results from a Weibull distribution based analysis with the most advanced machine learning analysis. As the former only considers the failure events, disregarding the influence of the CMC messages, the comparison of the results aims to conclude about their specific influence on the failure prognostics. In addition, to measure the importance of different failure definitions, the replacement maintenance logs of two valves embedded into the pneumatic system are also considered as failure events.

The remaining useful life prognostics is, as explained in chapter 2, a regression problem in machine learning. As an alternative, this thesis also concerns the classification machine learning approach, as instead of trying to predict the number of days left until the system fails, the model may also classify the risk of failure as High or Low, whether the system is within 20 days of a failure or not.

Two structurally similar data-sets are analysed in this thesis. Portugália Airlines data set includes the full fault history database from the fleet's 13 aircraft over the last 3 years. This thesis' analysis extends to another data-set kindly provided by the Brazilian airline company AZUL airlines. Although the data also includes stored CMC and CAS messages, AZUL airlines only stores the relevant messages filtered by the AHEAD-PRO software. Therefore, one of the thesis' possible analysis is to compare the outcome results of the data set filtered versus the full Portugália Airlines' data set. Another aspect to consider is the increased number of failures and CMC message records that constitute the AZUL airlines data set, being substantially more extensive, with around 3.3 million recorded CMC messages.

As mentioned before, the prediction power of the maintenance's messages is quantified based on the predictions' errors. It is important to clarify by this stage what are the evaluation measures considered either on the regression and the classification problem. For the regression methods, the considered error quantifiers are the mean absolute error (MAE), and root mean squared error (RMSE), defined in the table 3.2 as:

Table 3.2: Performance metrics. T_{pred} stands for the predicted remaining useful life, T_{actual} for the observed value, n the number of observations, and i is the observation identifier.

Performance measure regression

Root Mean Squared Error	RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_{pred_i} - T_{actual_i})^2}$
Mean Absolute Error	MAE = $\frac{1}{n}\sum_{i=1}^{n}\lvert T_{pred_i} - T_{actual_i}\rvert$

For the classification approach, the performance metrics are different. The evaluation is based on the number of true positives, false negatives, and false positives. These concepts are relevant and need to be defined before the evaluating parameters. Notice that the meaning of positive in this context is the cases when there is a "HIGH" risk of failure. The definitions are the following [10]:

- True Positive (TP) : The predicted and the actual result are both positive, or indicate that there is a high risk of failure.
- True Negative (TN) : The predicted and the actual result are both negative, or indicate that there is a low risk of failure.
- False Positive (FP): The predicted result suggests a positive result (high risk) but the actual result is in fact negative (low risk). This is also known as a type I error.
- False Negative (FN) : The predicted result suggests a negative result (low risk) but the actual result is in fact positive (high risk). This is also known as a type II error.

The evaluating metrics for the classification problem are showcased in the table 3.3

Table 3.3: Performance metrics for classification.

Performance measure classification

Precision	Precision = $\frac{TruePositives}{TruePositive + FalsePositive}$
Recall	Recall = $\frac{TruePositives}{TruePositive + FalseNegatives}$
Accuracy	Accuracy = $\frac{TruePositives + TrueNegatives}{TotalObservations}$
F1 Score	F1 score= $2 \times \frac{Precision \times Recall}{Precision + Recall}$

In both the classification and regression problems, the main goal is to optimise all the evaluation measures, and therefore obtain the best performance out of the developed models. Regular machine learning algorithms explained in chapter 5 are used to quantify the predictive power of the maintenance messages. This ability is considered in this thesis as proportional to the quality of the obtained models.

The following chapters resume the work carried out on the messages' analysis. Chapter 4 starts by describing the performed steps that allowed to identify and decode the data downloaded from the aircraft. It shows the development of automatic procedures to store and decode the data in the Portugália Airlines' IT department database. After this, section 4.2 presents an initial exploratory analysis that aims to extract, from the data, some initial conclusions, more specifically from the Portugália Airlines's data-set. The final section of the chapter explains some pre-processing techniques applied that seek the best possible outcome from the machine learning algorithms that are subsequently applied to the data. Chapter 5 explains the different models developed that allow the comparison and maximisation of the performance obtained from the available data sets. Finally, chapter 6 showcases the results from the followed analysis, and the subsequent chapter 7 explores the final conclusions.

Chapter 4

Data

This chapter is divided into three sections, the first explaining the data decoding and integration, the second showcasing a preliminary data overview, and a third illustrating the pre-processing techniques applied to prepare the data for a further machine learning-based analysis explained in the chapter 5.

The aircraft fault history database is periodically, typically every 4 days, downloaded and stored in the airline company's AMOS software. Currently, these data files are only stored for possible future consulting. However, the goal of Portugália Airlines is to use that perhaps wasted information to improve the fleet's monitoring and safety. Therefore, two initial goals were defined and included in the chapter:

1. Development of an automatic tool that periodically uploads and integrates the data downloaded from the AMOS software into the internal IT (Information and Technology) database.
2. Development of a graphical interface that automatically interprets the data from the internal IT database and periodically publishes throughout the company's departments.

The first is explained in section 4.1, along with all the steps followed to decode the initial data format. The second complements the exploratory analysis section, where a preliminary overview of the data is performed.

4.1 Data Integration

Routine to upload data to the IT database

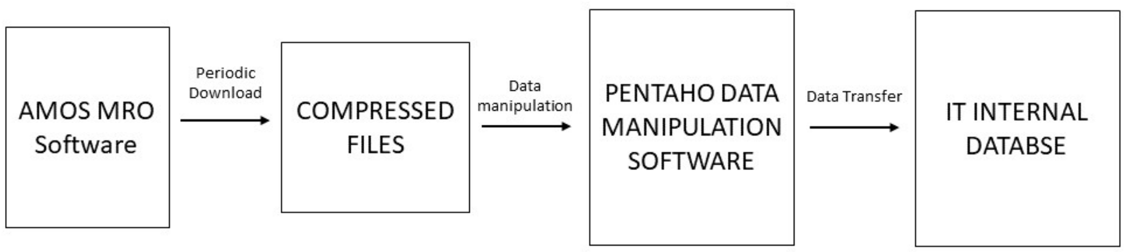


Figure 4.1: Routine workflow

Figure 4.1 showcases the implemented data flow. First, the AMOS software is programmed to periodically download the previously uploaded FHDB (Fault History Database) data to the company's specific network folders. The data is formatted as compressed zipped files, named after a code that corresponds to the source aircraft registration and the date of the download. Each folder contains four different text files: One storing the CMC (Central Maintenance Computer) messages; other the CAS (Crew Alerting System) messages; a third with the list of the flight legs that the aircraft performed during the period of interest; and a fourth blank file, named after the already mentioned code defining the aircraft registration and download's date.

Figure 4.2 showcases the data format of the text file concerning the CMC messages. Worth noticing the raw and coded nature of the data. With the knowledge from the Portugália Airline's employees and further research and analysis of the aircraft's documents, the identification of the columns was possible. The fourth is identifiable as the timestamp of the message emission.

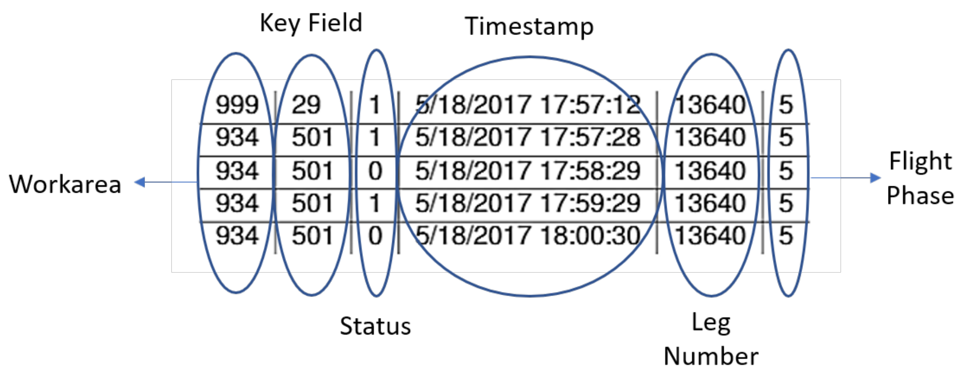


Figure 4.2: Subset of the input data from the text file concerning the stored CMC message history

The third column distinguishes whether the messages' status is active (1) or inactive (0). The fifth column corresponds to the flight leg number on which the message was emitted. Worth mentioning that this number is internally defined by the aircraft, being independent from the flight leg number used by the airline company. The sixth column corresponds to the flight phase the aircraft was performing when the particular message was emitted. All flight phases are described in a code number that ranges from 0 to 8 corresponding respectively to the uncertain, maintenance, pre-flight, taxi, take-off, climb, cruise, approach, and roll out flight phases. Finally, the first and the second column are denominated as "Workarea" and "Key Field" respectively. Instead of storing the messages as strings in the FHDB that eventually would increase the size of each file and decrease the storing capacity of the internal aircraft memory, this codification method is adopted. It allows the latter decoding of the messages with the assistance of a so-called master table that relates the combination of the two numbers to the corresponding messages.

Apart from the messages' decoding, the master table also enables the linkage of the messages to the corresponding ATA (Air Transport Association) chapter and subsection. Consequently, each message may be associated with a major aircraft subsystem, and also even more specifically with the ATA subsection. Furthermore, this linkage also enables, for CMC messages, the association of every message to the respective message code, that is relevant whenever there is a need to consult the fault isolation manual.

The succeeding step consists on using an open-source software denominated as Pentaho. Developed by Hitachi, it is a business intelligence software that enables data integration and data mining through a user-friendly environment. It is based on box diagrams, requiring little to no coding [33]. This step is crucial, as it is responsible for all the data manipulation that enables the periodic pre-processing and transfer to the IT internal database.

Figure 4.3 details the adopted functionalities of the Pentaho software. Firstly, the zipped files are unzipped, and the already mentioned four files are stored in a temporary folder. The file regarding the recorded CMC messages is imported. To allow the cross-reference of every single imported message to the source aircraft, a sub-string of the previously introduced coded file is extracted and linked with the corresponding aircraft registration. Hence, a new column is added to the existing six columns.

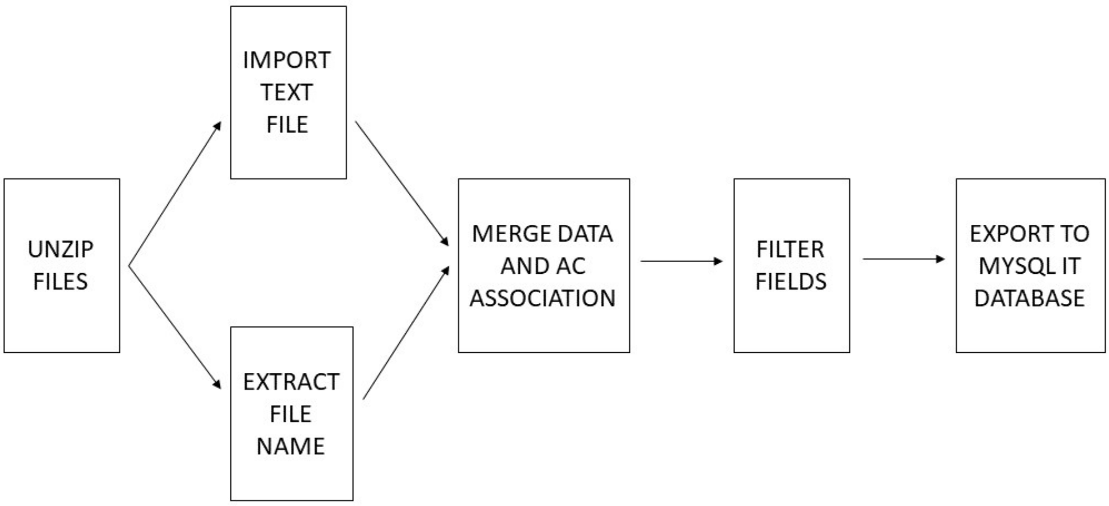


Figure 4.3: Pentaho data manipulation flow

The next step consists in the filtering of some useless fields created during previous steps. Finally, the exportation of the data to the IT database is possible. The temporary folder is deleted at the end of the process, and all the data is stored in the database. Notice that one of the main advantages of using this software is the ease to program the scheduling of the actions, requiring no user intervention other than the download of the data from the aircraft to the MRO (Maintenance Repair and Overhaul) software AMOS.

The data being stored in the IT department's database also allows cross-referencing other already existing tables, with other important data such as the flight information of every flight leg number recorded. That way, it is also possible to associate every single message to the leg number, and hence the flight the aircraft was performing at the time that message was emitted. Notice that this leg number is different from the above mentioned internally generated leg number. The former is defined by the airline company, as the latter is a sequence internally generated by the aircraft.

This described process is in all aspects similar to the CAS message data manipulation. The final outcome is the important storage of two tables, one regarding the reporting history of the CMC messages and other the CAS messages. To complement these two tables, two master tables were also uploaded to decode and link the data. For the sake of memory saving, it was decided to upload the master tables and the message data separately to the IT database, maintaining its coded format.

Once uploaded to the database, the data is easily accessible. The two main purposes of this stored data are demonstrated in figure 4.4. One is the prognostic analysis, the main goal of this thesis. The other is to act as a source to, the also developed on this thesis' behalf, graphical data visualisation interface, explained bellow.

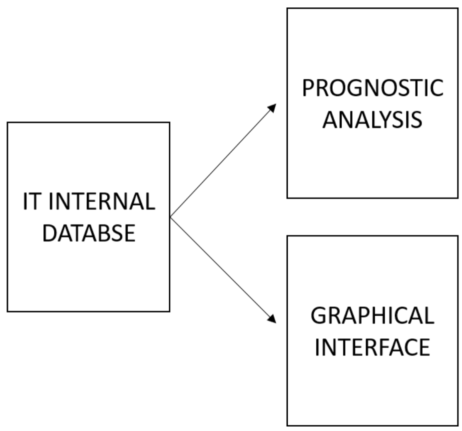


Figure 4.4: Continuation of the data flow of figure 4.1

4.2 Exploratory Analysis

This section presents a preliminary analysis of the data, using more conventional visual and statistical methods. As mentioned before, the initial approach of this thesis focuses on the study of the pneumatic system, more specifically, the air bleed distribution system. Worth recalling that the main failure events considered in this thesis regard the emission of failure CAS messages, that for the pneumatic system are "Bleed 1 Fail" and "Bleed 2 Fail". This exploratory analysis considers the message based failure definition, and it is performed to the Portugália Airline's data set.

Also worth mentioning that, between every single flight, the aircraft performs some pre-flight checks. Those routines and internal verifications are often the cause of misleading messages. Also, during maintenance actions, the aircraft is often not able to distinguish the flight phases of maintenance and pre-flight, often misclassifying the messages. Therefore, this thesis disregarded messages emitted during the on-ground flight phases, and only considered the airborne messages.

As a starting point, it is in the next subsection described the development of the second initial objective mentioned at the beginning of the current chapter - the graphical interface integrating messages' evolution.

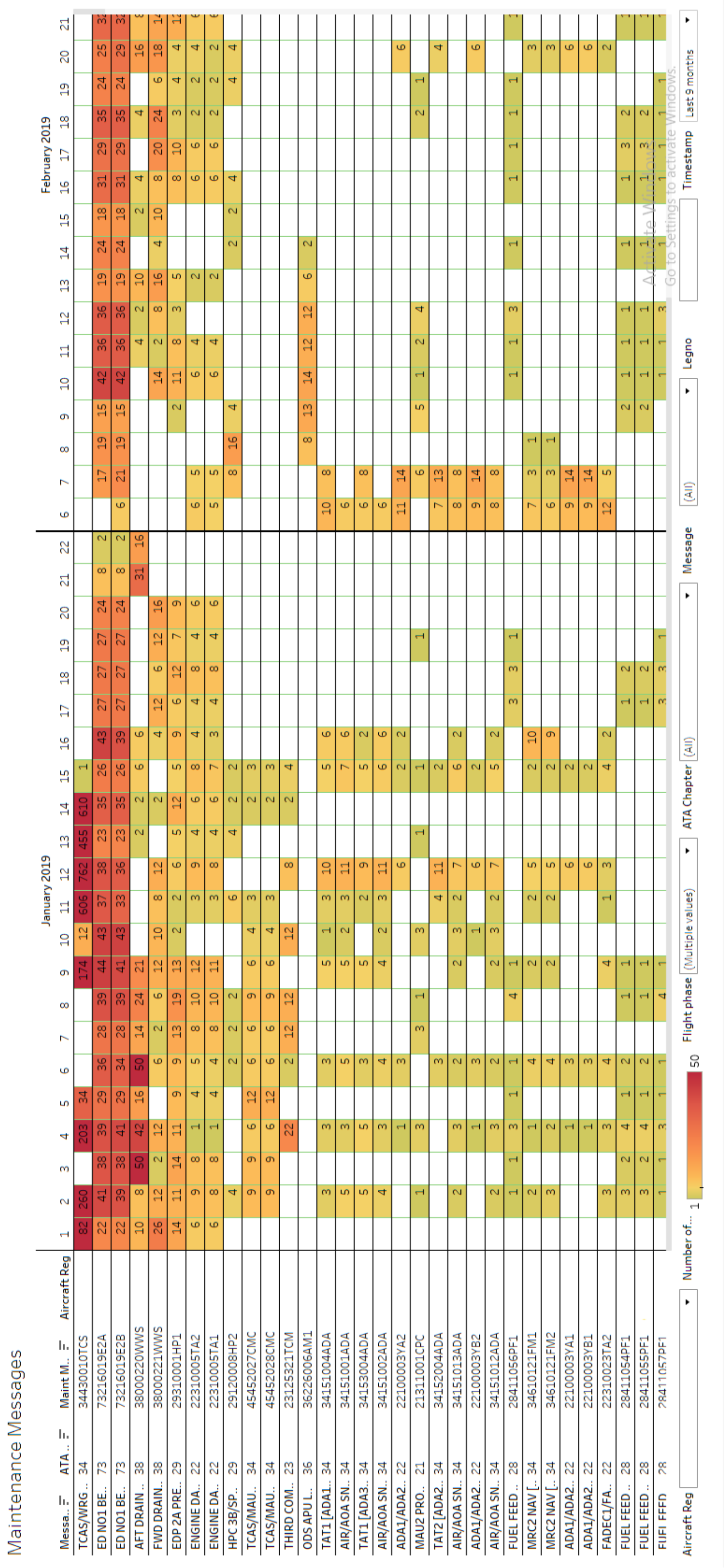


Figure 4.5: Interface of the developed graphical visualisation tool

4.2.1 Graphical Interface for Visual Data Analysis

This goal was achieved using one of the proprietary software available at Portugália Airlines - Tableau. It is a business intelligence and data visualisation software that aims to improve and enlarge the scope of data analysis, enabling the development of several types of graphical dashboards [35]. The results are also easily shareable throughout the company's departments using the internal company's network. Data refresh is also passive of intervention, and so are the publications that may or may not be programmed to be performed periodically.

The software extracts the data periodically directly from the database, manipulates, and graphically transforms the original data format into an engaging and easy to understand dashboard. Figure 4.5 showcases an example of the interface of the developed tool.

The evolution of the number of emitted messages was the focus during the construction process of this graphical dashboard. Each coloured and numbered cell details the number of messages of a particular type that were issued on a specific day, designated by the top header row. Furthermore, apart from the message name, the columns on the left-hand side of the dashboard detail the respective message code, ATA chapter, and the source aircraft registration. The cells are subjected to colour scale applied according to the number of issued messages to highlight sudden variations in the rate of message occurrence.

One of the main benefits of using this particular software is the interactive capabilities of the produced dashboards. As the system is connected directly to the database, there is a high level of customisation and diversification. This is provided by the filters located at the bottom of the dashboard. There, the user may filter the shown data according to either the message name, code, ATA chapter, aircraft registration, flight phase, timestamp range or a combination of any of the latter. The shown data adapts, and only the relevant data according to the filters applied, is exhibited.

The informative columns at the left-hand side are also sorted according to the total number of messages issued during the chosen period. Hence, the first upper message that appears in the dashboard is, according to the filters applied, the most frequent message on the period of interest. Therefore, it is more natural for the user to identify the most recurrent messages without having to perform any search. Furthermore, there is also the possibility of alternating the message type, allowing the graphical analysis of both the CAS and the CMC messages.

This health monitoring tool is implemented in the Portugália Airline's network and acts as a complement to the already other software such as the AHEAD-PRO. Despite not having the capability of real-time monitoring like the AHEAD-PRO, this tool showcases the recorded history in a more engaging and graphical manner, allowing the verification of tendencies based on the emission frequency of the messages.

The figures 4.6 and 4.7 show the functionality of this solution to inspect the stored message records. It also serves as a first analysis of the data regarding the system in study. Starting by figure 4.6, it showcases the message evolution several days before the emission by the aircraft AA of the BLEED 1 FAIL message, on the 10th of January of 2019. The interest messages are highlighted by the horizontal red rectangles, shown in the figure. The other rows regard the emitted messages by other aircraft.

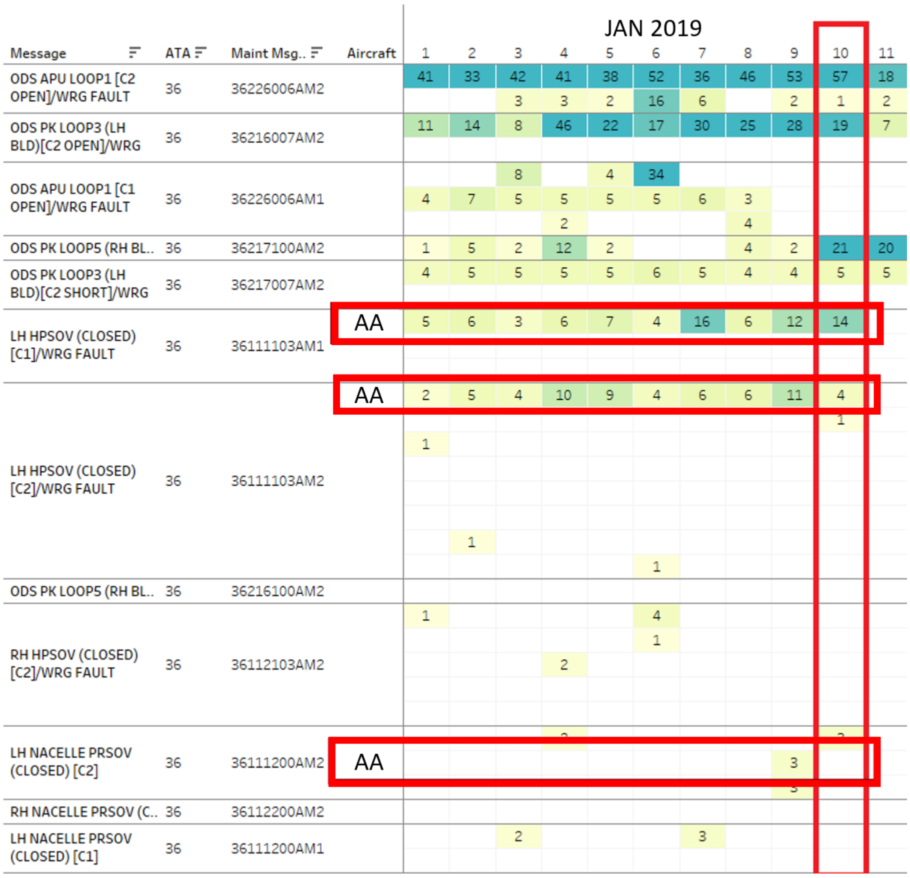


Figure 4.6: Message evolution several days before the emission of the failure CAS message BLEED 1 FAIL on the 10th of January 2019. The aircraft registration is hidden due to confidentiality.

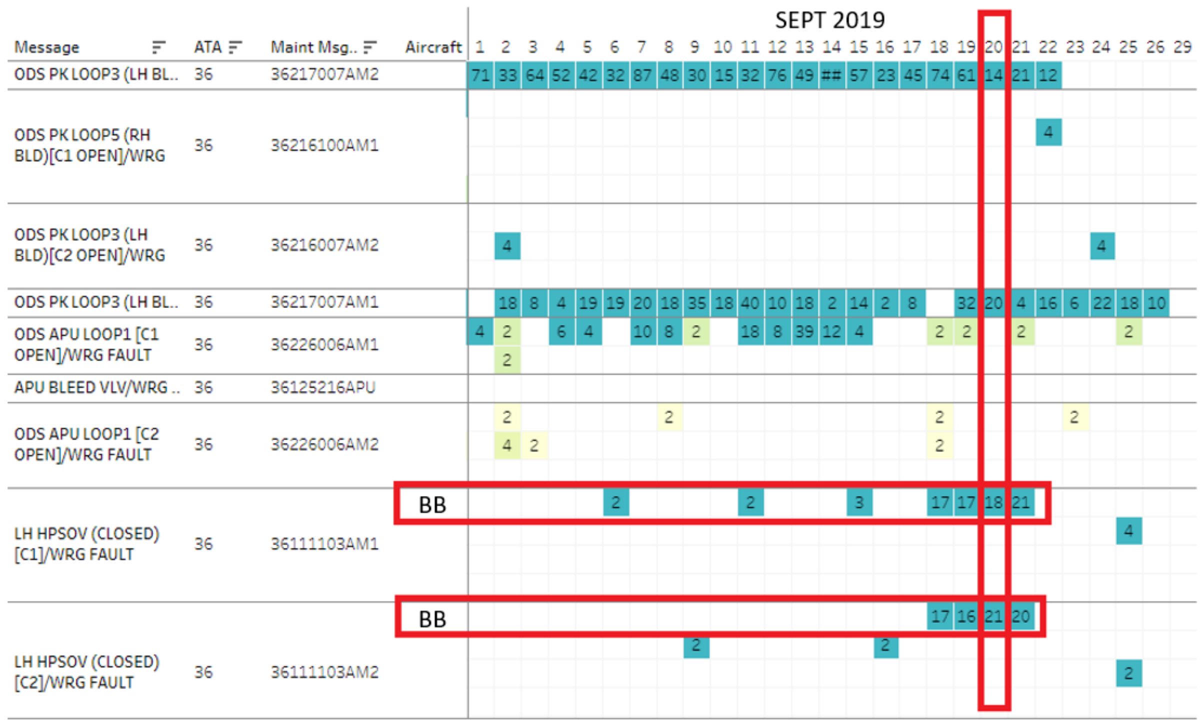


Figure 4.7: Message evolution several days before the emission of the failure CAS message BLEED 1 FAIL on the 20th of September 2019. The aircraft registration is hidden due to confidentiality.

It is not obvious that there is an upwards evolution of the three highlighted messages during the days that anticipated the failure. There is, however, a constant emission, in low numbers, of the two top messages during at least 10 days before the failure. This behaviour ceases after the day of the failure, which indicates that there was, in fact, some intervention that fixed the problem. Also worth noticing that other aircraft presented an emission rate higher than the aircraft AA, without emitting the failure CAS messages.

The second case also presents the evolution of messages of an aircraft (here coded as BB) that emitted a failure message BLEED 1 FAIL on the 20th of September of 2019 (figure 4.7). This case highlights a reaction to the messages' emission that is almost coincident with the failure event. There is a peak on the message rate, starting only two days prior to the failure message emission.

This short alerting period might indicate some lack of prognostic capabilities of the CMC messages. Prognostic data aims to predict or alert to a future failure with some anticipation, based on the evolution of a specific parameter over time. The failure analysed in figure 4.7 does not show any message predictors at any point in time prior to the data of analysis, questioning the predictive nature of these messages. Also, this shows that the developed graphical tool might enable data diagnosis and not data prognostics, allowing the analysis of the roots of the failure but not the failure prediction. Besides, considering the messages' turn around period of 4 days, since the emission of the messages until they are uploaded to the database, it would fail, in this particular case, shown in the figure 4.7, to notify the user to some imminent failure event.

To summarise, the developed graphical interface may serve as a diagnostic and hardly as a prognostic tool for the airline. It may allow one to find the causes that led to a specific failure, therefore being essential to study the failure modes of the several aircraft systems. The rest of this thesis focuses on the study of the predictive capabilities of CMC messages using more advanced approaches, based not only purely visual analysis. Notice that the graphical interface may be integrated into the outcome of this thesis. If it is verified that CMC messages have, to some extent, predictive power over the system's failure, it may be possible to integrate the results into this tool, highlighting the most predictive messages and alerting the user once some pre-defined threshold is surpassed.

4.2.2 Statistical Exploratory Analysis

To continue the seek of any evident relation between the evolution of the message emission rate at the void of a failure event, figure 4.8 may be scrutinised. Each line represents the sum of the number of messages from 0 to 30 days before failure. Notice that this analysis only regards the left-hand side of the bleed system. As it is noticeable, there is no evident increase in the message emission rate when approaching the failure, nor a good behaviour in the evolution. Instead, there is an almost random behaviour of most of the variables. Worth highlighting, however, the general upwards tendency of some variables imminent to the failure event, the overall peak obtained by the message ”ODS PK LOOP 3 (LH BLD)[C1 SHORT]/WRG” at the day of the failure, and its sudden increase the 19 days from the failure. On the other hand, the variable ”ODS PK LOOP 3 (LH BLD)[C2 OPEN]/WRG” presents a peak 20 and

23 days before the failure even, and a downwards tendency from this point on.

This plot also points out an important issue - a possible lack of data. Notice that the magnitude of the majority of the lines falls under the 50 emitted messages in total, from the entire data set available.

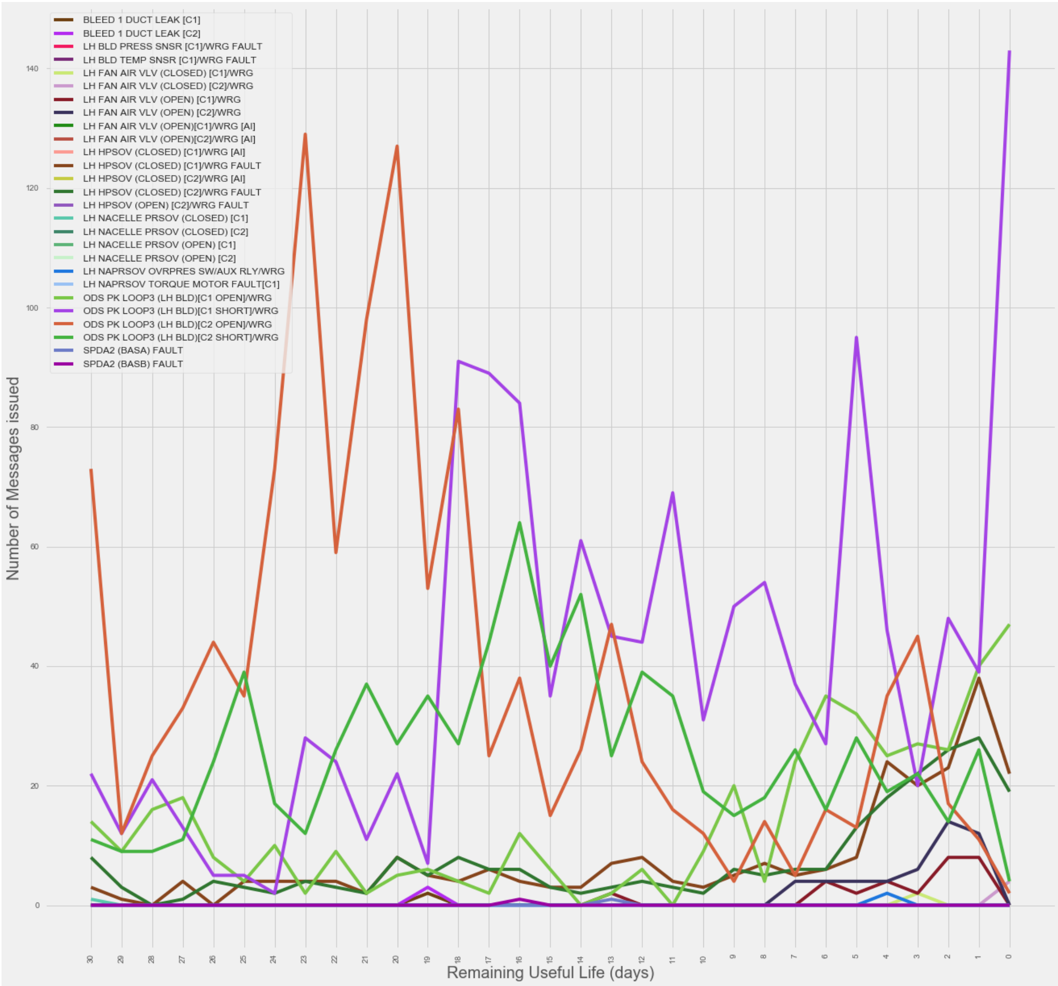


Figure 4.8: Accumulative sum of the number of messages emitted by the aircraft from 1 to 30 days prior to the Bleed 1 failure

Figure 4.9 presents the Pearson correlation heatmap between all the messages regarding the pneumatic system. More specifically, this graph presents the correlation between each number of messages issued by the aircraft in a certain period (1 day). Furthermore, both the ”Bleed 1 Fail” and ”Bleed 2 fail” messages were introduced as variables, aiming the verification of possible correlative relations between these and the CMC messages.

The Pearson correlation aims to measure how linearly correlated two variables may be. The method determines the Pearson coefficient, which is the statistical measure of the linearity strength of two variables. It ranges from -1 to 1, allowing the distinction of negative and positive linearly correlated variables.

Due to space and rendering limitation, not every variable is shown in figure 4.9. However, with this subset, it is possible to extract some valuable information. Notice the last two rows of the heatmap. Worth highlighting the high correlation coefficient between specific messages and the indicative failure ones. It is also confirmed a high correspondence between the messages regarding the left-hand side,

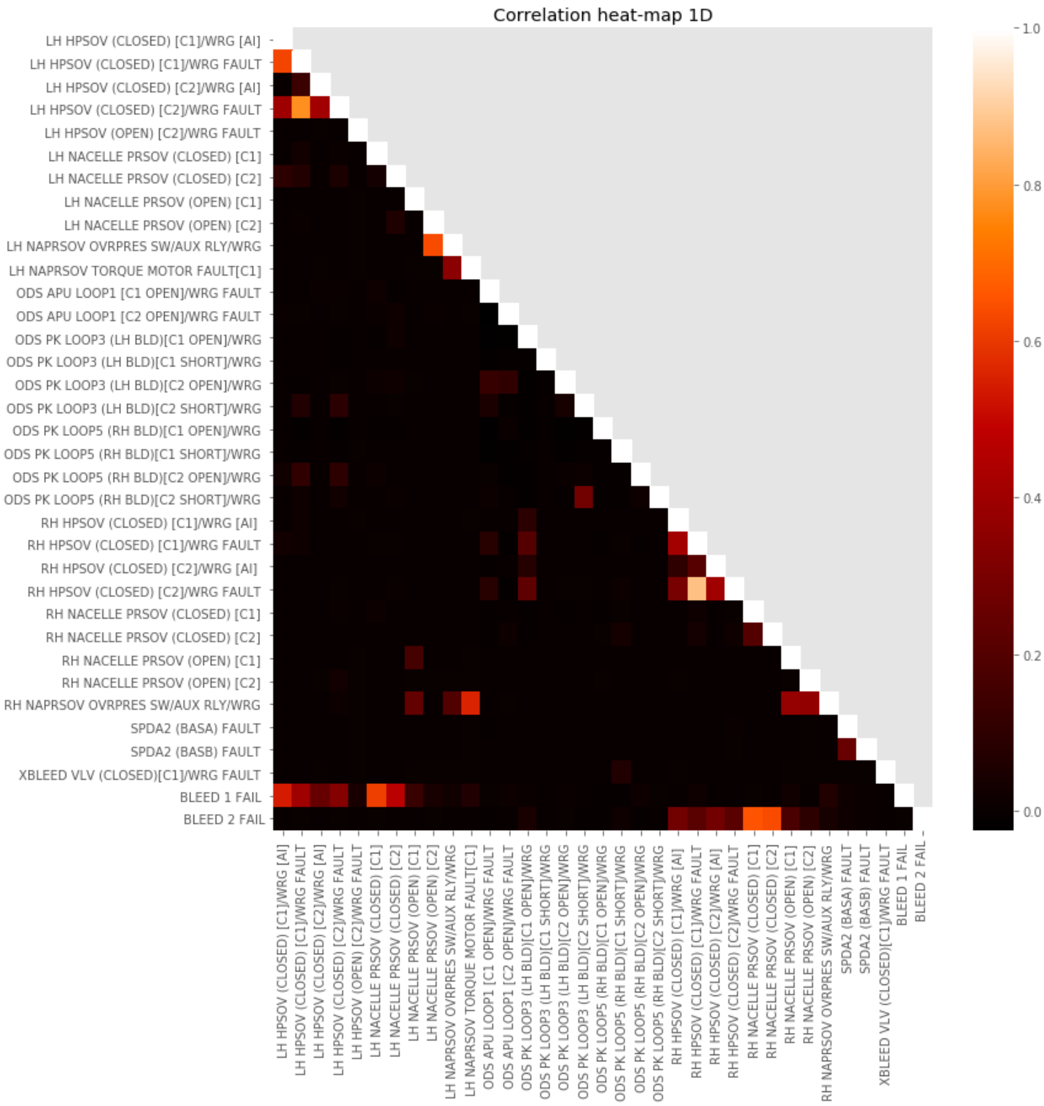


Figure 4.9: Linear correlation between CMC messages and failure indicative "Bleed 1 Fail" and "Bleed 2 Fail" messages

marked as "LH", and the failure of the Bleed 1 system, also referring to the left-hand side bleed system. The same happens to the right-hand side messages (RH) and the Bleed 2 Fail. The two variables substantially correlated to the failure are denominated as "LH NACELLE PRSOV CLOSED [C1]/[C2]" and "RH NACELLE PRSOV CLOSED [C1]/[C2]", for the left and right-hand side bleed system respectively. That indicates that often, the emission of bleed fail messages is highly linked to these. Furthermore, high correlation coefficients may be observed between specific pairs of CMC messages, which are addressed in future sections.

The high correlation coefficients between CMC messages and failure indicative messages, in this case, indicates that the CMC messages in question are often emitted within 1 day of the failure messages. Hence, the predictive power is not significant, as any prognostic tool aims to predict the occurrence of failure with an antecedence greater than 1 day.

Worth recalling that, in this thesis, the remaining useful life of the system regards the time difference between a reference timestamp and the corresponding nearest failure timestamp. Regarding the message based failure definition, from this point on, this thesis adopts just a single failure message, the "Bleed Fail", i.e., combines the failure of the left-hand side and right-hand side of the bleed system and defines a general failure state. To support this assumption, as shown in section 3.3, the system is symmetric, and the CMC messages are categorised according to the bleed system side. Besides, the already mentioned lack of correlation between messages referring to different sided systems corroborates this assumption. Notice that the influence of the CMC messages is still dependant on the bleed system's side. The messages regarding the left-hand side of the bleed system are the ones considered as variables and possible predictors to the failure of the left-hand side of the bleed system, and likewise for the right-hand side system. The denomination of the CMC messages is, from this point on, undifferentiated regarding the side of the bleed system. This assumption is also useful to increase data availability.

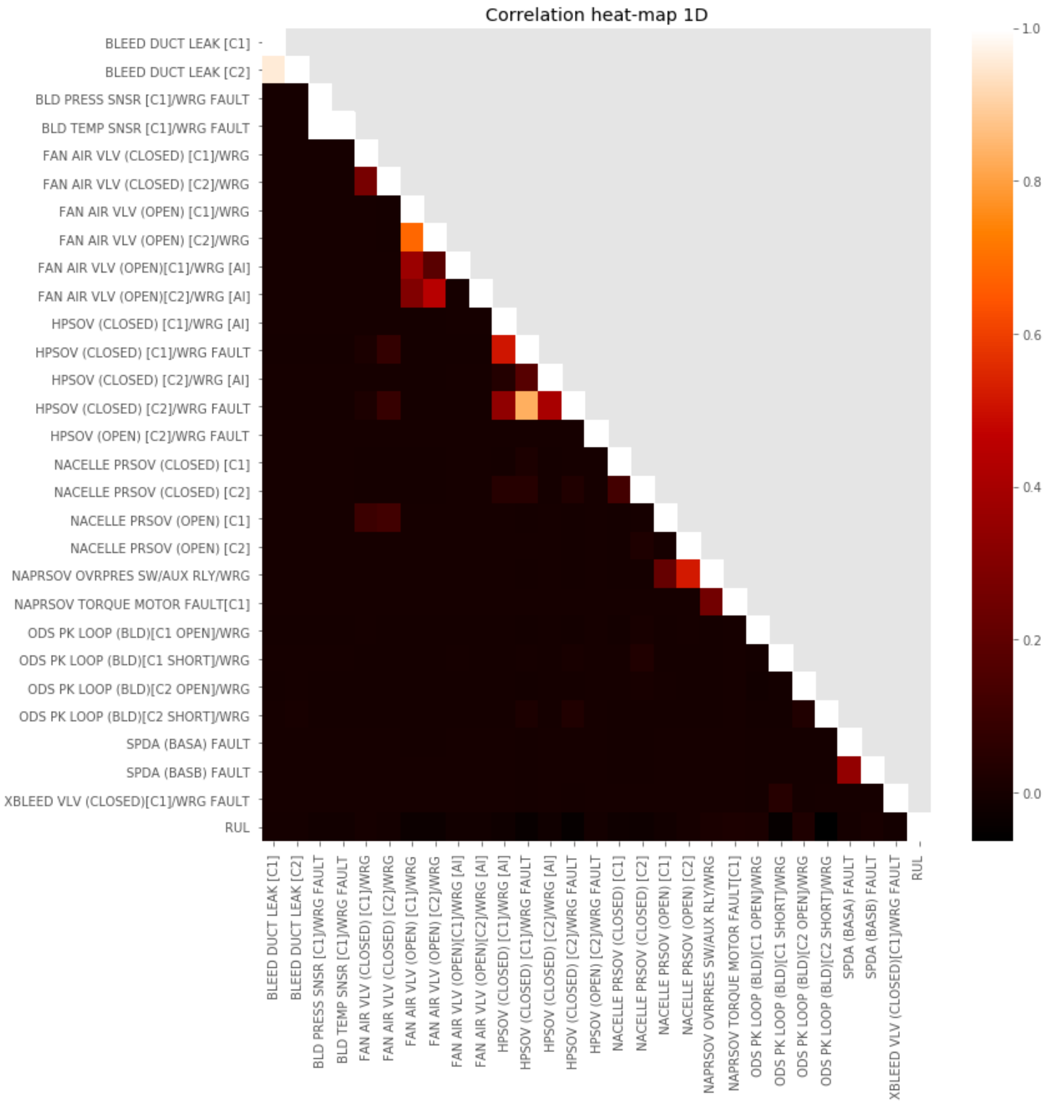


Figure 4.10: Correlation heatmap between the number of CMC messages emitted during a reference day. The last row also correlates the Remaining Useful Life (RUL) of the system and all the CMC messages.

Figure 4.10 explores the correlation between the number of messages emitted during a reference day and the remaining useful life of the system. Again, due to rendering visualisation issues, not every variable is present. Focusing mainly on the last row of the plot, some fluctuations of the correlation coefficient may be hardly observable. The ideal situation would be to have correlation coefficients close to -1, which would indicate that the increase in the number of emitted messages would be correlated with the decrease in the RUL.

The computation of non-linear correlation coefficients fails to provide substantial result improvements. The lack of correlation strength between the RUL and the CMC messages does not allow one to identify or conclude that the messages may be modelled as influential predecessors of failures. Therefore, this first data analysis fails on the identification of bright candidates for a robust prognostic tool. There is, however, like in the previous heatmap, some strong correlations between CMC messages. Again, this is discussed later in the document.

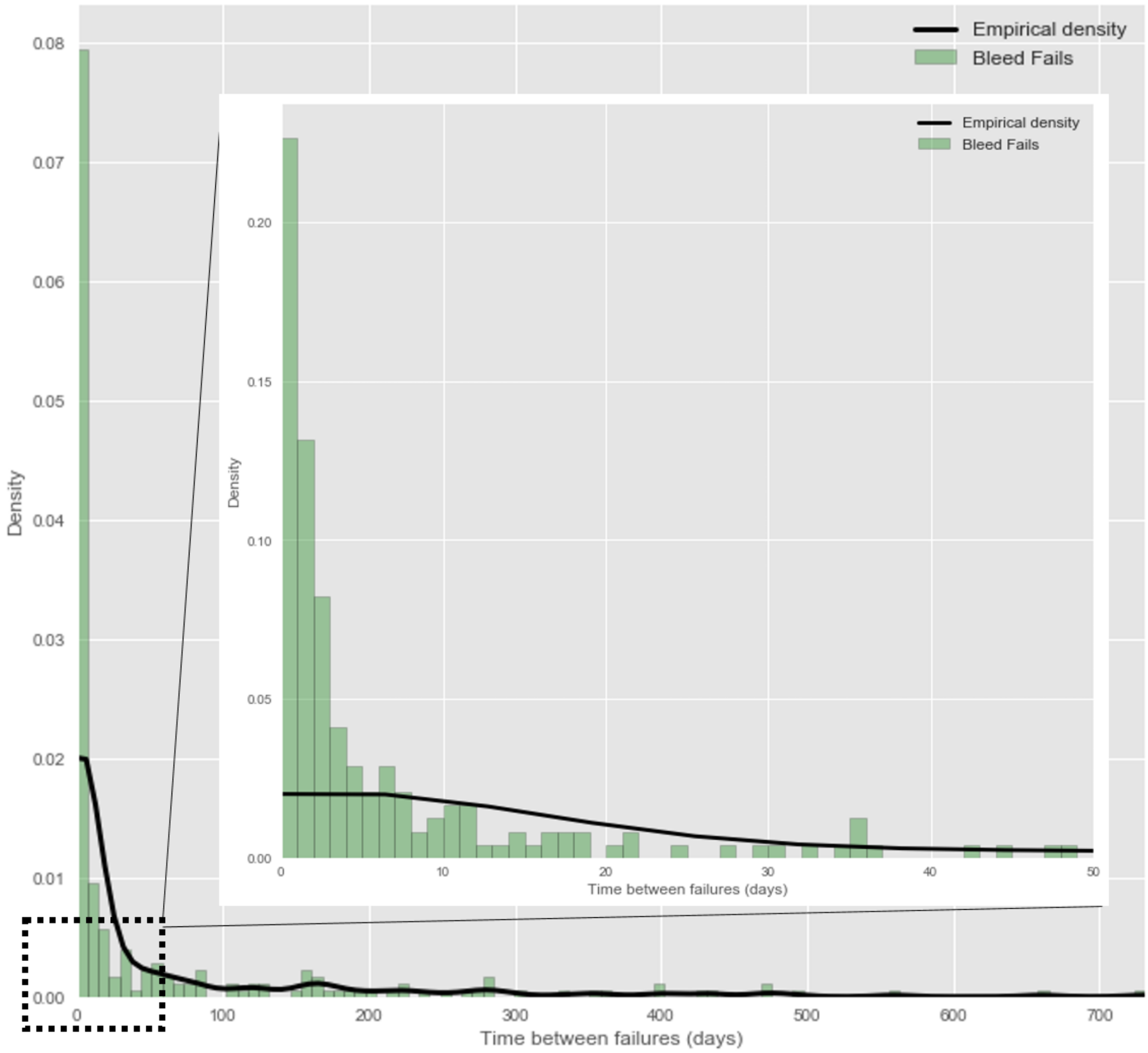


Figure 4.11: Failure histogram and the respective density function plot It includes a sub-figure that shows in detail the distributions of failures close to the origin. Results from the data-set provided by Portugália Airlines

Figure 4.11 displays the histogram and the respective density function plot of the time between failures. It highlights the disparity, ranging from 0 to 731 days, of the time between failures during the analysed operation period. It is also noticeable the dominance of the small failures, as there is an upwards tendency of the density function when the x-axis approaches the origin. The continuous detection and messages' emission by the aircraft may explain this high density of small failures that persists until a maintenance action is performed. Considering that there is a policy of performing maintenance actions within 3 days of the detection of a failure message, and due to the high density of failures separated by less than 3 days, this thesis assumes that these failures are a result of the constant failure state detection by the aircraft, and therefore are only accounted once. This assumption aims to minimise the failures' misclassification.

To summarise, the initial data exploratory analysis does not allow to state any evident conclusions regarding the relationship between the CMC and the failure of CAS messages. The rest of the thesis aims the usage of more advanced machine learning techniques to determine the predictive capabilities of CMC messages. To extract the best performance out of the machine learning algorithms, some pre-processing techniques are required and explained in the following section.

4.3 Data Pre-processing Techniques

The application of machine learning techniques requires a pre-processing phase. Especially for real-world data sets, the original format is often not compatible with most machine learning algorithms. Also, surplus and unreliable data only contribute to the contamination and the weakening of the models. Therefore, the main steps in data pre-processing allow extracting the best possible performance with the available data set. These include data filtering, cleaning, normalisation, transformation, feature construction, extraction, and feature selection [36]. These techniques enable the transformation of the raw data to be studied into the final training set, ready to be introduced to the machine learning algorithms. This section illustrates the application of these techniques to the real-world data-set provided by Portugália Airlines. Worth noticing that the definition and application of pre-processing techniques is an interactive process, as the constant seek for better results extend the search and the application of the numerous techniques available.

4.3.1 Data Filtering and Cleaning

Section 4.1 already exposed how the raw data is manipulated and uploaded into the IT department's database. Two main tables, one corresponding to the report history of the CMC messages and the other the CAS messages, are continually being updated. Two further tables are used to cross-reference the two coded fields that allow to identify, among others, the name of the messages.

Figure 4.4 demonstrated the two primary outputs for the stored data: The already explained health monitoring tool, and the data analysis using more advanced machine learning algorithms. Hence, the next step consisted in the exportation of the data from the database, and the usage of the two master tables to cross-reference. R was the programming language chosen to perform this step, due to its capabilities in data manipulation and easy to understand programming mechanics.

With the four tables merged into two, the next step consisted in data filtering. The final data structure was composed, for both the CMC and CAS message table, by six columns regarding the Message name, the timestamp and the flight phase of the emission, the aircraft registration, the respective ATA chapter, and subsection. Notice that these tables were defined as *dataframes*, a two-dimensional array-like structure, where data can be accessed by both the positional index pair or by the respective column and row names.

Not all the available information present in the *dataframes* are meaningful for the first analysis concerning the pneumatic system. The field ATA chapter is handy for the needed filtering process, as only the data linked to the ATA 36 was maintained. Also essential to recall that the data considered for the analysis only contemplated messages emitted in the airborne flight phases. As discussed in section 4.2, this aims to reduce the impurity of the data, as, in certain situations, the aircraft is unable to distinguish maintenance phases from pre-flight phases. Also, some routine pre-flight checks may cause a misleading appearance of messages. For this purpose, the flight phase field was used in this stage to disregard the messages with the flight phase coded as 0 or 1. All the remaining data was considered meaningful and trustworthy, as unlike other data types, such as raw sensor data, this data results from a manipulative and processing phase performed by the aircraft's internal computers.

4.3.2 Feature Engineering

Until this point, the data has been described by two *dataframes* containing the report history of the messages. Machine learning techniques require a previous data framework that allows the method to distinguish what is to be predicted and what are variables used as predictors. The latter are the features and the first the labels.

The definition of a feature is not always easy to understand. A feature is a characteristic, a measurable property defined by the analyst that is used to power the machine learning algorithm. It may also be defined as constructed variables based on the data to be analysed. Feature engineering is often compared to art. The endless possibilities of constructing different kinds of features, and its direct impact on the success of the machine learning project, turns this process of finding and building feeding variables crucial and sometimes dependent on the experience, sensibility, and skill of the analyst.

Feature Construction

The intent of this thesis is the analysis of the interaction between the CMC messages and failure events, and to measure the feasibility of using the CMC messages as failure predictors. It was decided to base the analysis on the message's emission frequency. Therefore, the main features are set as the sum of emitted messages in a specific period. This is further detailed in the section 5.3. The time reference is defined as a timestamp that ranges from the first and the last recorded message timestamp. The time-step of the time reference is variable, and in this case, only definable measuring the accuracy of the models.

It was also decided to add the aircraft registration as categorical variables. This aimed to measure and compare the influence of the messages' source to the frequency of the messages' emission. A categorical variable differs from the numerical variables as the first contains labelled values. For this purpose, it was used the technique denominated as *one-hot encoding*. This consists in a binary approach, assigning the value of 1 to the column that refers to the actual categorical variable. Imagine that one row in the *dataframe* to be analysed regards the number of emitted messages during a certain period of the aircraft "AA". Having already n columns defined as the n aircraft registrations in the fleet, all but one column is assigned with 0. The one that isn't refers to the aircraft "AA". This is the general categorical variable integration of this technique.

Another way to create categorical variables consist in assigning a number to each variable, from a sequence that ranges from 1 to the n categorical variables needed. This feature is consolidated into one column, reducing the feature space of the machine learning problem, which may or may not improve the results. In this thesis, both methods are considered and used the one that improves the results.

Apart from the above mentioned, the system's lifetime was also included in the features' set to add temporal sense to the model. The new feature measured, in days, the time difference between the previously recorded failure event and the time reference.

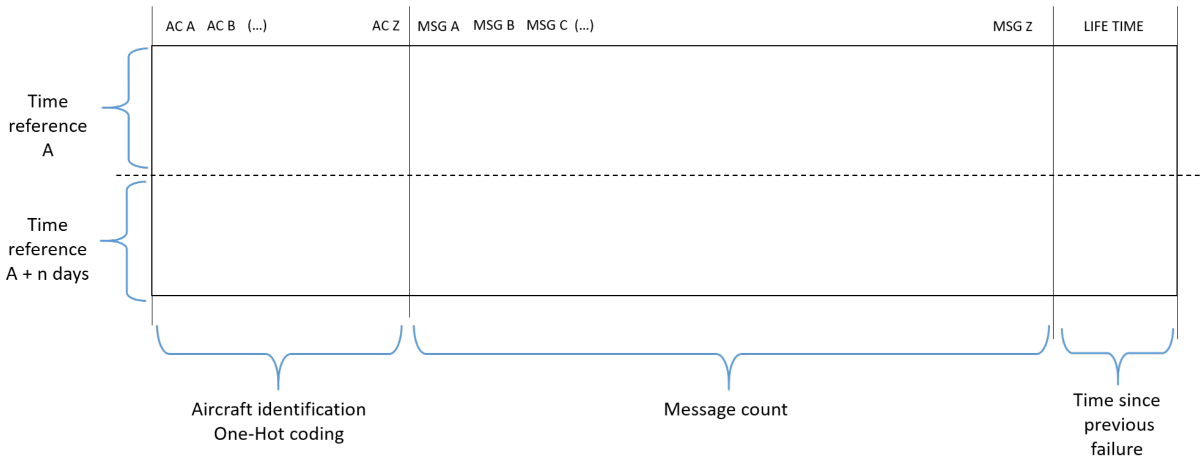


Figure 4.12: Final *Dataframe* Structure. The first column set represent the one-hot coding of the aircraft registration. The following column block refers to the message count within a specified time period. Finally, the last column introduces the life time of the system, i.e, the period, in days, between the previous failure event and the time reference in vigour. Each row block refers to a certain time reference, that ranges from the latest and the earliest ever recorded message timestamp. MSG stands for MeSsaGe, AC for AirCraft.

Figure 4.12 resumes the final data frame structure. There are three main different column blocks: the aircraft one-hot encoding; the message count between a certain period; and the lifetime of the system. Notice that a certain time reference defines each row block, that increasingly progresses through the period of analysis, with time steps of n days. Each row within each block is computed by the values of the features considering a certain aircraft and the specified time reference. This is only the data structure of the first analysis. As shown in section 5.3, several other structures were derived from this one. The other technique for encoding categorical variables replaces the block identified as one-hot encoding with a single column that differentiates, with a number, the different categorical variables.

Among the many built variables, some may fail to add new information to the model. Therefore, there are many techniques that prevent the inclusion of unnecessary features to the analysed feature set, or techniques that allow to perform a dimensional resuction to the feature set, retaining the most information possible. The following sections present such pre-processing techniques.

Feature statistics

Section 4.2 already presented the correlation concept. It measures the similarity of the variables' behaviour. Although section 4.2 focused mainly on the correlation between the predictors and the target, this section explains how to treat two predictors highly correlated between themselves. Whenever there are two highly correlated variables, they may be considered as redundant, as both explain the same variance. Therefore, the standard procedure in this situation is to define a correlation coefficient threshold and disregard one of the two highly correlated variables. The threshold considered in this thesis is 0.90. Maintaining highly correlated variable pairs would only increase the dimensionality of the problem, without improving the final model. One of the main objectives inherent to machine learning consists in conserving the highest possible variable variance with the least amount of variables possible. This also prevents *overfitting*, a common problem in machine learning. It occurs when there is excessive learning of the training data, turning the model biased and not performing well when faced with the test data. On the other hand, when the model fails to train with the training data, the model is *underfitted* [37]. All these pre-processing techniques contribute to avoid both cases.

Feature Extraction

Feature extraction consists of extracting new features based on the existing ones. These new features may be used to complement the feature set or to reduce the dimensionality of the problem. One of the most well-known methods is the Principal Component Analysis (PCA). It is a dimension-reduction tool that aims to keep the information existing in a specific feature set into a new smaller one. The new uncorrelated variables are denominated as Principal Components (PC), sorted by the amount of information each one store from the initial feature set. This thesis uses *PCA*() function available in the *python's* library *sklearn*.

The PCA concept may seem at first complex. Essentially, the method consists in finding the directions of the highest variance in the feature space, by decomposing the eigenvalues (values that define how spread the variables are) in a specific feature space direction, the eigenvector. In this thesis' context, the formulation of the method is not as important as the results it may provide. Therefore the mathematical formulation is omitted.

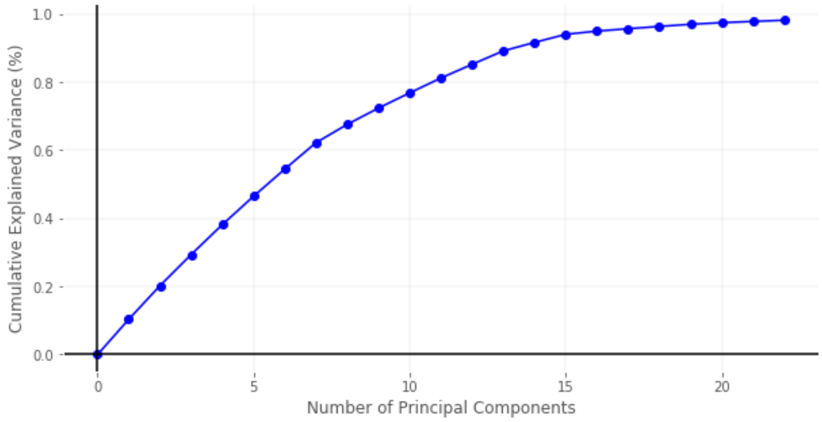


Figure 4.13: Cumulative explained variance evolution with the number of principal components

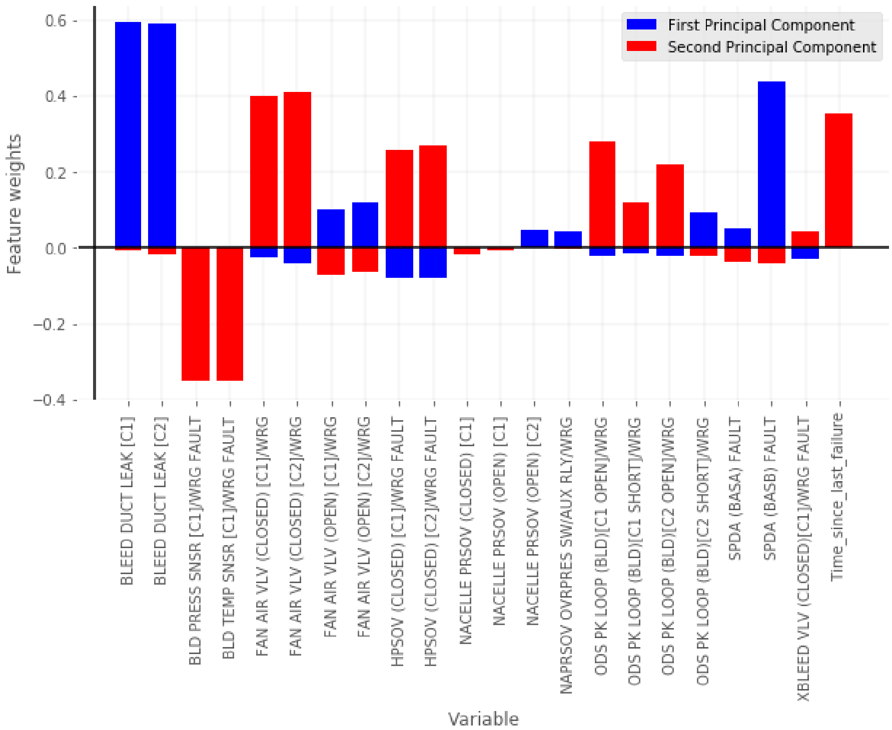


Figure 4.14: Variable contribution to the first two principal components

To reduce the feature space, preserving most of the information possible, the plot relating the percentage of variance explained with the number of principal components is shown in figure 4.13. As shown, the 15 principal components represent a cumulative explained variance of approximately 95 %, and only using the entire feature set is possible to achieve an explained variance of 100%. This information may be useful to define the number of principal components elected for re-running the algorithms. This choice is not always straight forward, as the balance between the benefits that come from the problem's dimensional reduction and the loss in information must be met.

The slope of the plotline tends to reduce as the number of principal components increases. This shows that the first principal components retain the most significant percentage of information compared with the rest. In fact, the explained variance of the two first principal components totals as 20.9%.

Figure 4.14 illustrates the features' contributions to the first two principal components. The blue bars explain how the original features explain the variance stored by the first principal component. It is noticeable that the variance of the first principal component is well explained by the variables with the highest absolute magnitude blue bars. The reasoning is similar for the second principal component, only this time, the bars measuring the contribution are red.

Although the principle of reducing the problem's dimensionality, keeping the majority of information possible, may, in theory, improve the overall machine learning results, in practice, this was not observed. Using the fifteen first principal components, the results deteriorated. This might be explained by the decrease in the explained variance, from 100% to 95%, without the dimensionality reduction of the feature space having benefits that would justify the adoption of this technique. Consequently, the inclusion of this technique in the adopted framework was disregarded. Another possibility to improve the machine learning problem's dimensionality is to classify the variables according to their importance, and focusing the analysis on those. This is what the next section explains.

Feature Selection

This section aims to select and transform the data so that the machine learning algorithms may learn and predict in the best way possible. The previous step aimed to reduce the dimensionality of the problem using PCA. In this case, the goal is to only retain the strictly necessary original features in the features' space. This may be done by making use of the general knowledge of the messages, eliminating some that are already excluded for not be related to the problem. In fact, the already performed filtering of the messages may be, to some extent, considered a process of feature selection. However, this section presents a further complement to the general goal of reducing the problem's dimensionality. It consists in measuring how useful the variables may be to the result. For that, the feature importance concept is in this section introduced. It may be measured in several different ways. However, this thesis employed three main approaches to measure importance: *random forest* decision tree Mean Decrease in Impurity (MDI) feature importance; column drop feature importance; permutation feature importance. The first measures how effective is the variable at reducing variance or uncertainty, for regressive and classification models, respectively. The second computes the importance of the variable comparing the model's performance after and before the feature's values are permuted. The third follows the same reasoning of the latter, but instead of randomly replacing the values of each variable, the column is disregarded. The application of three different methods allows the cross confirmation of the results.

Figures 4.15, 4.16 and 4.17 illustrate the features' importance by the already mentioned three different methods. Notice that the models derived for this analysis were all regressive and based on the *random forest* algorithm, from the python's library *sklearn*. The influence of the source of the messages is also taken into consideration, with the first 13 columns. The real aircraft registration is omitted due to confidentiality. The predicted label is the remaining useful life (RUL) of the pneumatic system, being the failure defined by the emission of the already mentioned CAS messages.

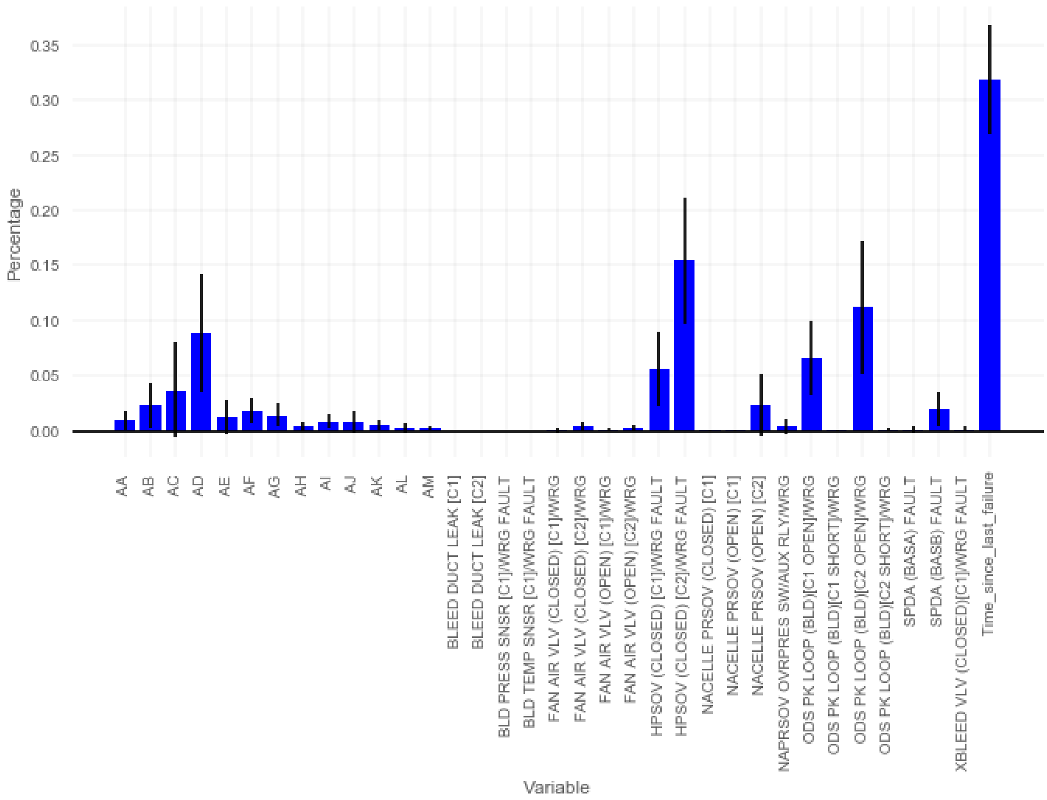


Figure 4.15: Feature importance via mean decrease in impurity

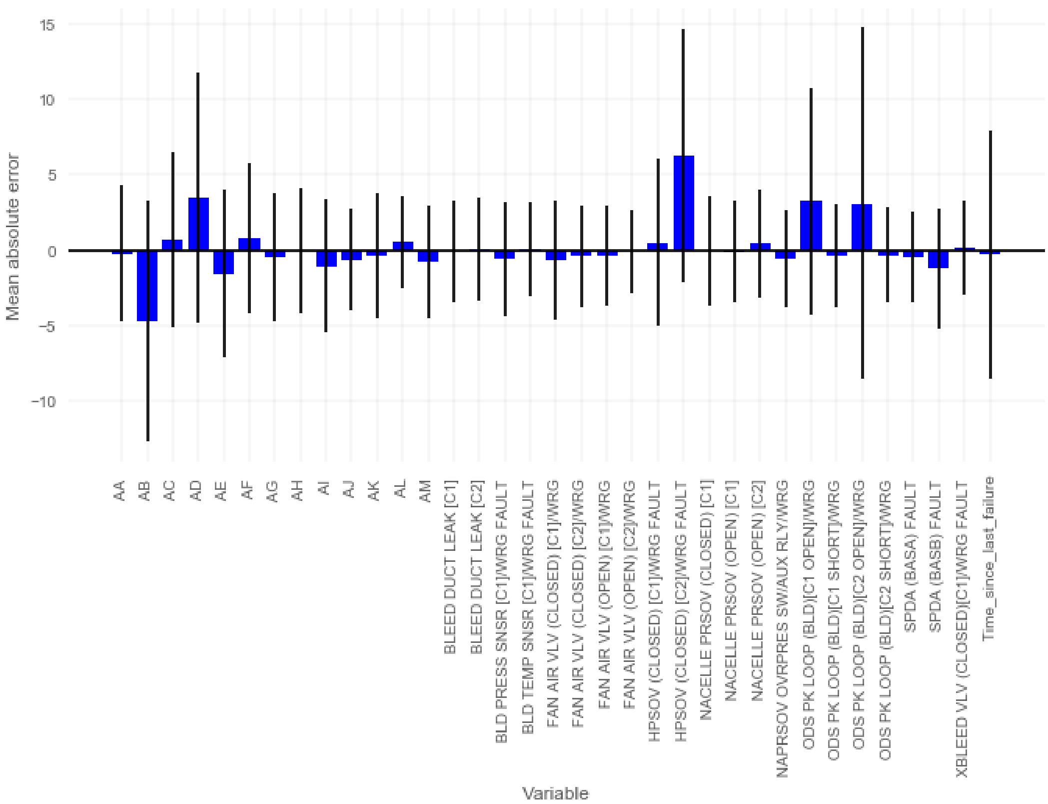


Figure 4.16: Feature importance via column permutation

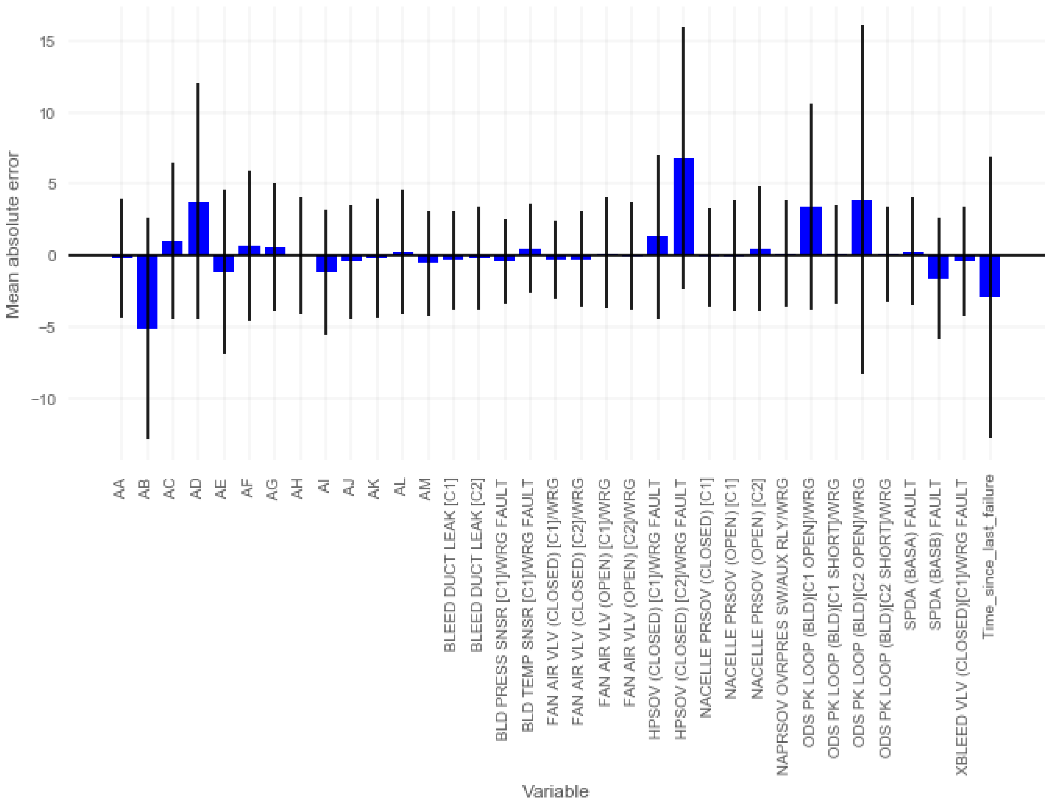


Figure 4.17: Feature importance via column drop

The first plot (figure 4.15) highlights a high relevance attributed by the model to the variable referring to the lifetime of the system. This may indicate some lack of predictive capability of the rest of the features, as the model may be predicting the RUL over-trusting the time passed since the previous failure event and not the number of emitted messages in itself. However, this importance is not shown by the other two methods measuring variable importance, therefore questioning what was obtained by the MDI method. There is however mutual peaks on the variables ”HPSOV (CLOSED) [C2]/WRG FAULT”, ”ODS PK LOOP (BLD) [C2 SHORT] /WRG” and ”ODS PK LOOP (BLD) [C1 OPEN] /WRG”. Worth mentioning that the shown importance attributed to each feature was the mean importance over 100 iterations, each one dividing, training, and testing the data in several random partitions. It was noticed significant variations in the importance's evolution during the iteration, highlighted by the standard deviation of each variable represented by the vertical black lines of the bar plots. This, especially noticeable by the column manipulative methods, shows that the variables' significance is highly dependant on the way the training and test set is derived. These incoherence turns the task of selecting features based on this method untrustworthy. Hence, it was decided to keep the not very extensive feature set for future analysis.

4.3.3 Data Scaling

This step aims to maximise the machine learning techniques' performance redefining and scaling the features' values. Usually, due to the different nature or sources, the variables in the variable set have very different scales. As an example, consider a feature that measures time, in days, between two events, and others that represent the price of a component. It is easy to understand that these illustrative features have different scales and variances. The same is observable with the different ranges of messages emitted in a certain period. So converting these variables into the same scale is propitious to the overall performance of the machine learning algorithms. Sometimes, more than advisable, it is required by some algorithms implemented in the library used in this thesis - *sklearn* - that the features all vary in comparable scales. Gradient-based and metric-based algorithms all assume that the data is standardised.

This thesis applies the *StandardScaler* function available from the *scikit-learn preprocessing* library. The *StandardScaler* re-scales the features so that they end up in a normal distribution with a standard deviation of one and a mean of zero, i.e, $\sigma = 1$ and $\mu = 0$. The values, or z-scores are then given by:

$$z_j = \frac{x - \mu_t}{\sigma_t} \tag{4.1}$$

This technique is applied to post the division of the data into training and test datasets. The function's fit is performed to the training set, and the transformation is performed subsequently based on the parameters learned in the fitting stage. These parameters are also utilized to perform the scaling of the test set. Worth pointing out the missing fit stage on the scaling process of the test set. This is done to keep the ”unseen” nature of this data set, thus being scaled based on the same parameters as the training set.

On the other hand, the target of the analysis is, in this thesis, not scaled, either for the classifying (would not be possible) or the regressive approaches. On the latter, the author considered that it would not make much sense limiting the range of the label through scaling. Also, explained in the following section, the one-hot encoding of the categorical variables is not passive of scaling methods, but on the other hand, the method that compresses the categorisation into one column is.

The following chapter aims to showcase the implementation phase, using the data pre-processing techniques explained.

Chapter 5

Models

This chapter aims to present the implementation steps carried out to develop the baseline and the prognostic models. These are used to evaluate the predictive capabilities of the message data over system failures. It begins by explaining the implementation of a more usual approach based on the Weibull distribution applied to reliability. This is used as the baseline approach and compared, in chapter 6, with the more advanced machine learning models.

The second section of this chapter explains the applied machine learning framework and introduces a brief description of the machine learning algorithms. Apart from the methods explained in the previous chapter, there are further procedures that need to be executed before the introduction of the data to the machine learning algorithms.

Finally, the different machine learning model variants are thoroughly explained, dividing the problem into 5 different approaches' types. As with every machine learning project, different iterations lead to different results. Therefore, this section presents the development 5 alternative analysis that aim to extract the best possible results from the available data.

5.1 Baseline

To compare the results from the more advanced approaches based on machine learning algorithms with a more conventional life data-based methodology, this thesis applies the Weibull distribution to reliability engineering. It has been used for many years for failure analysis, and it is still used to define maintenance intervals for several preventive maintenance plans. Instead of considering all the already discussed variables referring to the message evolution, the distribution aims to fit the life data of the system and extract some useful reliability results to compare with the more advanced methods presented in this thesis.

The main goal of the life data analysis is to predict the life of, in this case, a system by fitting a statistical distribution to a representative sample of data. The fitted distribution may then be used to estimate important reliability results such as the probability of failure in a certain future time, or the failure rate. The lifetime data is, in this case, the time between failures of the system. To be comparable with the machine learning models discussed in future sections, the failure definition is common to both approaches.

The Weibull distribution may be applied in a variety of forms. This thesis applies a two-parameter Weibull distribution (shape β and scale η). These parameters are computed by the fitting stage of the process, where, based on a selection of time between failure data, the parameters are estimated to best suit its evolution. Those parameters are then fundamental to calculate the already mentioned reliability results. This thesis adopted the fitting function "*fitdist*" from the package "*fitdistrplus*" implemented using the R programming language.

One of the several reliability measurements possible to analyse using the Weibull distribution is the BX life. The BX life refers to the life point in time, that being days or cycles, when less than X% of the population has failed. It may as well be defined as the time when the probability of failure reaches X%. For the sake of argument, consider the B10 life. That is the time, in days, or the number of cycles, when 10% of the population has failed, or the time when there is a 10% probability of failure. The same reasoning applies to whatever the X percentage may be defined.

Consider a highly important member of a system, which failure may jeopardise the safety of the operation. In these cases, the maintenance plan may consider the level of criticality of the equipment and therefore define the maintenance intervals with a higher level of safety possible. Therefore, the X percentage of the BX life may be reduced, to make sure that at the computed time, there is a low percentage of failure. Meanwhile, for other not as critical components, this time may be extended.

The BX life may be computed using the Cumulative Density Function (CDF):

$$F(t) = 1 - e^{-(\frac{t}{\eta})^{\beta}} \tag{5.1}$$

F(t) being the cumulative probability of failure, t the time, and η and β the scale and shape Weibull parameters. Being the BX life the time at which the cumulative probability of failure is X, it may be derived from the equation 5.1, and it is defined as:

$$BX = \eta[-\ln(1 - X)]^{-\frac{1}{\beta}} \tag{5.2}$$

Equation 5.2 is nothing more than equation 5.1 deriving t (BX) as a function of F(t) (X). Therefore, post the fitting stage, any BX may be calculated with the equation 5.2.

It was decided, in this thesis's behalf, to compare the results from the more advanced machine learning approaches with the BX life evolution considering different X, ranging from 10 to 90 percent, in steps of 10. The value considered for the consequent comparison with the data driven models will be the one that has the least error associated when compared with an unseen subset failure data, hence reducing the uncertainty of only considering a unique value for X.

The evaluation of the baseline approach follows a cross-validation method. The failure data is randomly divided into 5 folds, each one containing a subset of the entire failure data. Over five iterations, one of the folds is defined as the test set, the other 4 combined to form the training set. The latter are used, in each iteration, to fit the Weibull distribution, and therefore compute the shape and scale parameters. Based on these parameters, the BX, X ranging from 10 to 90, in steps of 10 percent, are calculated, resulting in the times when the probability of failure is X. In the same iteration, the BX is compared with the test set, containing time between failure data, unused to fit the distribution, and therefore independent of the previous steps. The comparison allows computing the error measures discussed in the next chapter that are common to those used in the machine learning approach.

Post the 5 iterations, the final results consist in the average of the error measures obtained over the iterations. This process aims to reduce the variance of the results obtained by the different ways to split the test and train sets, and therefore improve the reality and thrust-worthiness levels of the method. This process is, to some extent, similar to the one used in the machine learning framework, explained in the following section.

This baseline analysis also aims to measure the influence of the message data predictive power. The uncertainty associated with only using failure data showcases to what extent is the committed error if the failure event and subsequent maintenance action are expected to occur on a fixed periodic basis defined by the BX value. Therefore the failure based analysis disregards any added information concerning the evolution of messages. Hence, the comparison between the approaches' results helps to conclude about the predictive capabilities of the message data.

5.2 Machine Learning Framework

This section presents the framework responsible for the application of data into the machine learning algorithms. Although the previous chapter, in the section 4.3 presented and explained the pre-processing techniques necessary to extract the best possible performance out of the algorithms, there are some important steps yet not mentioned. Figure 5.1 presents the general workflow of one of the iterations of the process.

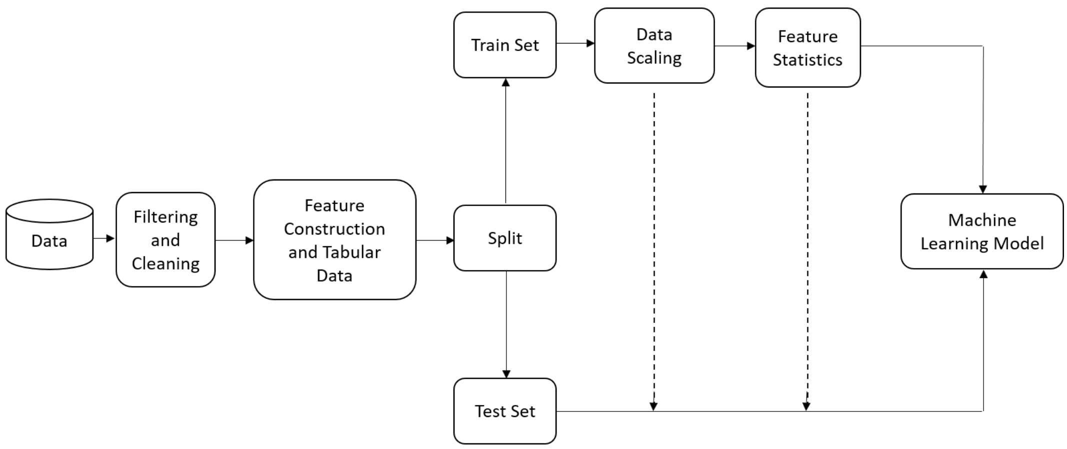


Figure 5.1: Machine learning process framework

The first two steps, regarding the data filtering, cleaning, and the feature construction and reorganisation of the data into a tabular form, are already explained in section 4.3. The only variants from what was explained in chapter 4 regard the several feature alternatives explained in section 5.3.

The following step in the process consists in dividing the data into training and test data sets. Before mentioning the cross-validation method used, it is important to highlight the specificity of this otherwise simple process to the problem in question. When dealing with models treating multiple independent time-series data, to keep the unseen nature of the test data, the temporal dependencies may be as well divided. This thesis deals with failures, to ease the explanation, consider a failure period as the time between two subsequent failure events. To train and test with the highest possible reality, the test and train data have to be divided chronologically [38], to prevent the training on data concerning a failure that is also included in the test set. Basically, the model cannot train and test data regarding the same failure period. Hence, to achieve this goal, every single row on the *dataframes* has a failure id associated that differentiates the failure that the RUL (Remaining Useful Life) is being measured on is. Therefore, the data in this thesis is divided based on the failure id, 80% allocated to the training data, and the rest to the test data set.

The following steps consist in scaling the data, and in the application of the feature statistics to eliminate one of each pair of highly correlated variables. The dashed arrows indicate that this process is firstly applied to the training data set, and, based on the results and parameters, the same application is performed to the test data set. Notice that the feature extraction and selection, as mentioned in the previous chapter, were not applied due to the reasons already explained.

Finally, the last step in the machine learning framework consists in creating and testing the models. Therefore, it is by this point important to mentioned and provide a brief introduction to the machine learning algorithms implemented. This thesis uses algorithms implemented in python's library *sklearn*. The selection of the best algorithm before the actual implementation of data is impossible. Therefore, the reasoning followed was based on the implementation of five different algorithms for the regression problem, and four for classification. The end goal of trying several different algorithms is to extract the best performance out of the models developed on the available data. The algorithms for the classification and the regression problems are the following:

- From *sklearn.linear_model* - *LinearRegression*
  - It fits the data with the end goal of minimising the residual sum of squares between the predicted responses by the linear approximation and the observed responses in the data. The result of this minimisation is a set of coefficients $z = (z_0, z_1, z_2, ..., z_p)$ that complete the linear relationship between the inputs and the outputs $y(z, x) = z_0 + z_1x_1 + z_2x_2 + ... + z_px_p$. It solves, mathematically, the following expression : $min_z||X_z - y||^2$. This algorithm is only used for the regression problems [39].
- From *sklearn.ensemble* - *RandomForestRegressor* and *RandomForestClassifier*
  - It is an averaging ensemble method that employs a combination of several base estimators to increase the generalizability and robustness of the method. It uses two different ways to decrease the variance of the forest estimator: each tree is built from a sample drawn with replacement from the training set; the other is to consider either all features or a random selection when splitting each node. Also, this tends to reduce the overfitting problem compared

with simple decision trees. The main difference between the regressor and classifier is that for the discrete labels, the algorithm predicts by averaging their probabilistic prediction, as for the continuous, the result is purely based on arithmetic average [40].

- From *sklearn.neighbors* - *KNeighborsRegressor* and *KNeighborsClassifier*
  - It is founded on the principle of the nearest neighbours methods, which aims to find a predefined number of training samples nearest to the new point and predict based on these. The number of labels is, in this case, a user-defined constant K, and the distance is the standard Euclidean distance. For the regressor, the label is computed based on the mean of the k nearest neighbours' labels. For the classifier, the labels are purely based on the clustering of the K nearest neighbours [41].
- From *sklearn.svm* - *SVR* and *SVC*
  - *SVC*, used for the classification problem, follow the reasoning of separating data point using a hyperplane with the most amount of data. It finds an optimal hyperplane iteratively, that presents the maximum segregation from the nearest data points. A kernel is what allows the implementation of the algorithm as it transforms an input data space into the required form. Using a technique called kernel trick, it transforms a low-dimensional input space to a higher-dimensional space. The *SVR* is the adaption of the algorithm for regression problems. The input is mapped as a multidimensional feature space using some fixed mapping, and consequently, a model is constructed that relates the inputs and the outputs [42].
- From *sklearn.ensemble* - *GradientBoostingRegressor* and *GradientBoostingClassifier*
  - Like the random forest algorithm, this is an ensemble method. By contrast, this algorithm is sub-classified as a boosting method, where the building of the estimators is performed sequentially, to try to reduce the final bias of the combined estimator. It supports both classifying and regressive predictors [40].

To optimise the validation and performance evaluation, a cross-validation methodology is employed. It consists in averaging the results from each analysis loop and hence getting a more accurate evaluation metrics. Due to the variance inherent to the models' results obtained from the numerous ways to split the data into training and test, the cross-validation intends to minimise this effect by training and testing different portions of data in each iteration, over n iterations, and the end results of the analysis are considered the average of the iteration results.

Noticing that all the data has a fault id associated. The reasoning for the cross-validation application was based on these to define, on each iteration, one of the K failure groups as the test set, and the other as the training set. Therefore, cross-validation consists in iteratively running the machine learning process over the K groups of data. The end results come from the average of the K iterations. Notice that, in the figure 5.1, the iterative approach encompasses the processes forward of the block regarding the "Feature construction and Tabular Data".

All the algorithms are trained and tested according to the cross-validation method, and the results are shown in the following chapter 6. Also worth pointing out that in this thesis, the default hyperparameters of each algorithm are used. Each algorithm contains parameters that need adjustment depending on the data to try to extract the best performance possible when applied to a specific data set. Unfortunately, due to computational limitations, this was not feasible. The conclusions are not influenced by the lack of model tuning, as the improvements do not reach a level that would alter the outcome of this thesis.

5.3 Implementation

5.3.1 Regressive Models

As mentioned before, the goal of this dissertation is to measure the predictive capabilities of message data over failure events of a certain system. It was decided to investigate this perspective by looking at the emission frequency of the CMC (Central Maintenance Computer) messages. The general reasoning behind this decision was that the more critical messages would increase its frequency when closer to the failure event. In fact, the feedback provided by the engineers at Portugália Airlines supported this assumption.

Realising that the methods would have to follow tendencies to extract the best performance out of the machine learning algorithms, it was decided that, due, to some extent, to data shortage, the most suitable approach for the type 1 models would be to measure the number of emitted messages since the previous recorded failure event. Instead of providing the model with the messages emitted for a certain period of time, as for example, each day, it was decided to sum the number of messages over the period since the previous recorded failure. Therefore, no matter what is the evolution of messages issued per day, the tendency is always positive. The irregularities inherent to the messages' emission rate, also shown in the figures 4.6 and 4.7, would otherwise promote the model's confusion, due to, in some cases, the decrease in the message count when approaching the failure event.

Apart from the mentioned features regarding the message count, the aircraft was also considered as a variable, to measure the influence of the source of the messages and compare it with the rest of the features. In addition, to add temporal sense to the model, a new feature was also added measuring the time since the previous failure event. These considerations were mentioned in section 4.3.2, with the description of the methods used to categorise the variables and the organisation of the resulting data frame. In addition, the decision to keep both the categorical variables and the time since last failure feature was supported by the results which showed considerable improvement when these features were included in the model.

The target variable was defined having in mind the final prediction goal. Noticing that the predictive power of message data would be used to define the worthiness of a future prognostic tool, the target of the analysis was considered to be the remaining useful life of the equipment. The remaining useful life is in this thesis defined as the period between a specific time reference and the nearest failure event of a certain system. In this type 1 iteration of the problem, the failure event is considered as the appearance of failure indicative CAS (Crew Alerting System) messages. For the pneumatic system, the two messages considered were the ”BLEED 1 FAIL” and ”BLEED 2 FAIL”. As mentioned in section 4.2, these two messages were merged into one - ”BLEED FAIL” -, keeping the side influence between the CMC messages and the failure side of the system.

The constant search of new data formulations that intend to maximise the machine learning methods' performance led to the redefinition of the data structure of the problem. Having in mind the two different data sources, there are several different comparisons possible.

Apart from the analysis mentioned in the previous paragraph, it is defined as the type 2 analysis, the one that instead of considering the sum of messages from the previous failure event until the time reference, it only counts the messages emitted within 24 hours prior of the reference timestamp. Again, the rest of the feature set remains the same. Although the type 1 solution was built to always present to the algorithm upwards tendencies of the data, it was considered a good comparison measure to keep this on the set of analysed approaches.

In the aeronautical field, the time may be measured in several different ways. The first approaches consisted in predicting the time, in days, left until the next failure event, and the so-called reference time also evolved with time steps of one day. Another approach to this problem consists in considering the evolution metrics as flight cycles or flight legs. One flight leg consists in a performed cycle, where an aircraft performs all the possible flight phases. A usual flight from Lisbon to Porto is considered a flight leg. Based on that, the from this point on denominated as type 3 approach consists in redefining the remaining useful life of the problem as the number of cycles left until the next failure event. In addition, the reference time redefined as a flight leg. Hence, the features regard the number of messages emitted during a certain flight leg by a certain aircraft. The additional feature denominated as ”time since last failure” now counts the number of cycles the aircraft has performed since the previous failure event. This analysis intends to minimise the effects of maintenance periods on the determination of the RUL. Although not very frequent, the day-based approaches were influenced by maintenance periods, as they were unable to distinguish if the aircraft was in fact flying.

Apart from the already mentioned analysis, type 4 and type 5 approaches are in all aspects similar to the type 1 and type 2 respectively, except the definition of failure. Until this point, the RUL was determined as the time difference between a certain time reference and the timestamp of the emission of failure indicative CAS messages, more specifically, the ”BLEED FAIL” message. To measure the influence of considering one of the many failure definitions, this fourth and fifth analysis considers the failure events as the replacements of the two valves present on each side of the air bleed distribution subsystem. One is the valve at the high-pressure stage of the engines' compressor, and the other is the low-pressure valve.

Table 5.1 presents the main characteristics of the several models developed. The type 1 and type 2 approaches may be comparable between the data sets provided by Portugália Airlines and AZUL airlines. Unfortunately, due to data unavailability, the type 3, type 4, and type 5 analysis may only be performed using the data from Portugália Airlines.

Also, worth mentioning that the feature set resultant from the AZUL airlines data is more extensive as

Table 5.1: Summary of all the regressive models described in the section 5.3.1. PGA stands for Portugália Airlines

Type	Data	Failure Definition	Feature Definition (Units)
1	PGA & AZUL	Failure CAS Messages	Since Previous Failure (Days)
2			Periodic (Days)
3	PGA		Since Previous Failure (Flight Cycles)
4		Unscheduled Valve Replacements	Since Previous Failure (Days)
5			Periodic (Days)

the one resulting from the Portugália Airlines' data. This is because, due to the substantially larger fleet of the Brazilian airline company, the recorded messages include some message types not yet emitted by the Portugália Airlines' fleet. This increase in the feature space may be prejudicial to the models. However, the increase in the sample, acknowledging that the report history data size is double compared with the Portugália Airline's data set, the overall model performance is improved. The following chapter presents the results and, subsequently, chapter 7 the respective conclusions.

5.3.2 Classification Scheme

Until this point, all the considerations were based on the development of regressive model, that intend to, in one way or the other, estimate the time or number of fights left until the next failure event. Therefore, what the models are aimed to predict are floating point numbers, that have an infinitive range.

To broaden the scope of the analysis, it was decided to try a classification approach. Two out of the three projects discussed in section 2.3 applied classifying approaches to develop prognostic models based on the available data, both with success. Therefore, to complement the five different types of regressive analysis already discussed, this thesis uses a similar structure to those mentioned case studies. The main difference rests on the definition of the label. Figure 5.2 showcases the reasoning followed for the classification problem definition.

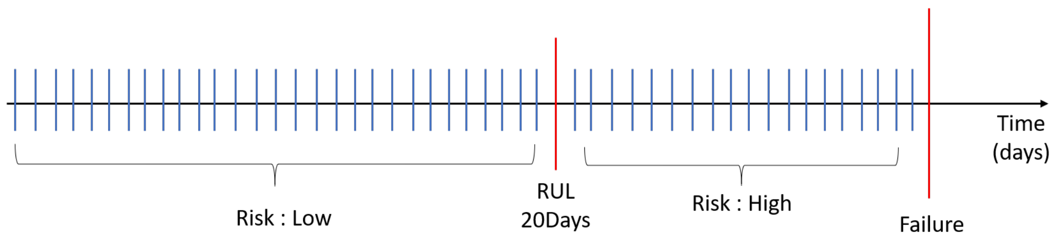


Figure 5.2: Classification of the system's risk. RUL stands for Remaining Useful Life.

Considering the real world operation requirements, it was decided to define the classification approach's target as the system's failure risk. Hence, the system's jeopardy is considered high if the failure is predicted to occur within 20 days of a certain time reference. Otherwise, the risk is considered as low. This attribution of the system's risk is what the models are design to predict. The main objective of the analysis performed in this thesis is to obtain the best possible results from the models. The quantification of what is considered as a good result is discussed in the next chapter.

All the processes in the framework are common to regression and classification, as the only aspect that varies is the definition of the labels and the application of the algorithms.

Chapter 6

Results

This chapter presents the main findings of this dissertation and investigates how they relate to the research questions. As a starting point, the first section presents the results from the baseline approach, which concerns the application of the Weibull distribution to failure data using a 5-fold cross-validation scheme. The second section concerns the regression approach, which tries to predict the remaining useful life of the pneumatic system. Several approaches to this problem are explored, starting by the already mentioned five different types of analysis, and the comparison between the results from the two available data sets: Portugália Airlines and AZUL airlines.

The third section aims to showcase the results from the classification approaches, in all aspects similar to the previous, but this time with a binary predictive goal. Worth mentioning, however, that the first results concerning the data's exploratory analysis were shown previously in section 4.2. No conclusions are retrievable from that preliminary analysis.

6.1 Baseline Results

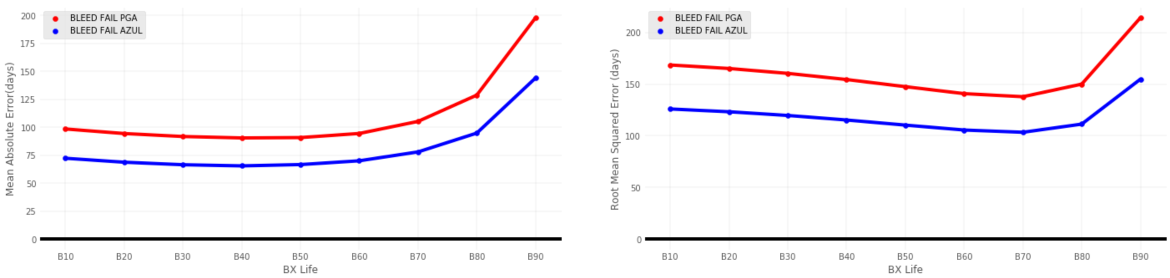
Table 6.1: Performance metrics. T_{pred} stands for the predicted remaining useful life, T_{actual} for the observed value, n the number of observations, and i is the observation identifier.

Performance measure regression

Root Mean Squared Error	RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_{pred_i} - T_{actual_i})^2}$
Mean Absolute Error	MAE = $\frac{1}{n}\sum_{i=1}^{n}\lvert T_{pred_i} - T_{actual_i}\rvert$

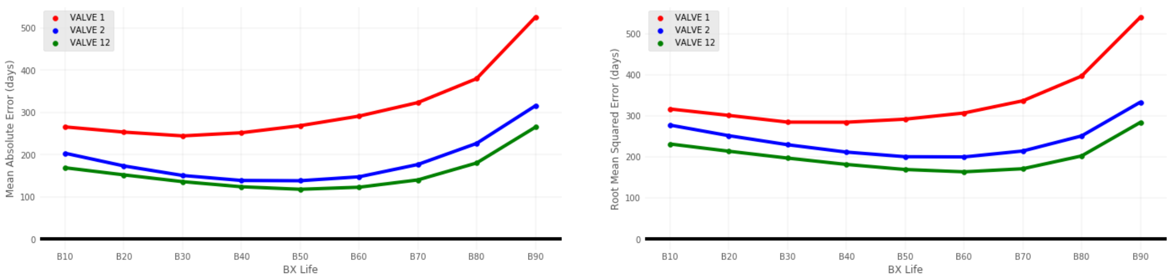
This section aims to showcase the results derived from the baseline approach explained in the previous chapter. It is important by this stage to recall the evaluating measures used in this thesis. For the baseline and the regressive approaches, table 6.1 presents the two different selected error measures considered. Both the MAE and the RMSE express the error in the units of the variable of interest, and both express the best behaviour of the models with the lower values. However, since the errors from the RMSE are squared, it penalises the larger errors, therefore it indicates how well the predictions are able to follow the test results. For the specific case studied in this thesis, the RMSE is relevant as the large errors associated with the prediction of the remaining useful life may have a great impact in an operational point of view.

Considering the 5 different approaches explained in the section 5.3, there are two different failure data and therefore two baseline results. The first is inherent to the type 1, 2 and 3 approaches, because all base the analysis on the appearance of failure indicative CAS (Crew Alerting System) messages. Depending on the source, two different baseline results from the first two approaches are possible, resulting respectively from the two data sets in study. Finally, the change in the failure definition of the last two model types (4 and 5) also lead to a different baseline results.



(a) Mean Absolute Error (b) Root Mean Squared Error

Figure 6.1: Evolution of the mean absolute error and root mean absolute error for the different BX parameters. Failure data from either the Portugália Airlines and AZUL airlines data sets.



(a) Mean Absolute Error (b) Root Mean Squared Error

Figure 6.2: Evolution of the mean absolute error and root mean squared error the different BX considered. Valve failure data from the Portugália Airlines data set.

Figure 6.1 and 6.2 showcase the evolution of the mean squared error, and the root mean squared error respectively for the different BX life. Recall that these plots show the committed error by assuming that the failure occurs in a constant periodical basis. The interval depends on the different probabilities of failure, defined by BX. For the sake of argument, consider the B10 life. According to the fitted Weibull distribution, it is assumed that the failure occurs every t days. The measures presented in the plots show the error committed by following this assumption. For the sake of comparison, this thesis considers the BX which provides the lowest MAE and RMSE. Analysing figure 6.1, the minimum MAE is derived from the B40 life, with a MAE = 90.4 days and MAE = 65.4 days for the Portugália airlines and AZUL airlines failure data sets respectively. On the other hand, the minimum RMSE results from the assumption of the

B70 life, with a value equal to 137.8 days and 103.3 days for the Portugália airlines and AZUL airlines data sets, respectively.

The final failure data set concerns the replacements of two main valves of each side of the engine pneumatic system. What is represented from this point on by valve 1 is the high-pressure valve represented by the figure 3.2, and valve 2 refers to the low pressure also shown in the same figure. Other possible analysis consists in not differentiating the valves, therefore considering a failure event as an unscheduled replacement of one out of the two valves (Valve 12). The results are shown in figures 6.2.

Analysing first the evolution of the MAE (figure 6.2 a)), it is noticeable that the predictions of the valve 1 have far more uncertainty that the others. The minimum of the evolution of MAE as a function of the chosen failure probability obtained is equal to 244.2, 137.8, and 117.6 days respectively, for valve 1, valve 2, and valve 12 replacements. The latter two cases present a minimum when considering a BX life of B50, whereas the valve 1 replacement error is minimised with a B30 life. When considering the RMSE, the values increase to 284.3, 199.7, and 163.4 days respectively, for valve 1, valve 2, and valve 12 failures. Likewise, the BX life that minimises this error differs for the valve replacements (B40 for the valve 1, and B60 for the valve 2 and valve 12).

To summarise all the results that are used as a comparison measure to what is obtained from the machine learning-based models, table 6.2 compiles what was mentioned in the previous paragraphs. There is also the connection with the five different approach types, discussed in section 5.3.

Table 6.2: Minimum values of the error's evolution with the BX life range considered. All the units are in days, PGA stands for Portugália Airlines, MAE for Mean Absolute Error and RMSE for Root Mean Squared Error

TYPE	SOURCE		ERROR/BX MIN			
			MAE	BX	RMSE	BX
1&2&3	PGA		90.4	B40	137.8	B70
	AZUL		65.4	B40	103.3	B70
4&5	PGA	Valve 1	244.2	B30	284.3	B40
		Valve 2	137.8	B50	199.7	B60
		Valve 12	117.6	B50	163.4	B60

6.2 Regression Results

The prognostic of the remaining useful life of the pneumatic system is the second analysis to be discussed. All the steps inherent to the pre-processing are already explained in the previous chapters. However, there are some aspects worth recalling.

Once again, the cross-validation allows the to reduce the variance of the results from different splits. Therefore, it retrieves the most real possible results from the models. The number of iterations is a variable that might influence the quality of the results, therefore leading to a mis-representation of the model's skill. A trade-off between high bias and high variance results is crucial in the choice of the final value. Usually, according to the literature, 5 or 10 are the most common choices [43]. Therefore, also due to the limited computational resources, the number of iterations for the cross-validation considered in this thesis is 5.

All aspects considered, table 6.3 showcases the results obtained from the five algorithms considered in this approach, comparing the results from the AZUL and Portugália Airlines data set, and the type 1 and 2 methods mentioned in the section 5.3.

Table 6.3: Results from the type one and two approaches, using the two different data sets in analysis. All the units are in days.

TYPE	ALGORITHM	DATA			
		PGA		AZUL	
		MAE	RMSE	MAE	RMSE
1	LinearRegression	151.5	186.7	95.5	125.1
	RandomForest	138.4	182.2	84.1	118.8
	KnearestNeighbors	139.6	184.8	91.9	127.5
	SupportVectorMachine	106.1	143.5	77.3	107.8
	GradientBoostingMachine	128.4	168.1	74.1	99.1
2	LinearRegression	132.7	182.6	4.1e+11	1.68e+12
	RandomForest	146.3	187.7	94.9	128.2
	KnearestNeighbors	149.1	189.1	93.4	126.2
	SupportVectorMachine	108.7	147.2	75.3	105.7
	GradientBoostingMachine	130.8	169.8	78.6	103.9

To recall the different types of approaches, the main difference between type 1 and 2 regards the counting period of the messages. Type one's features consider the sum of messages from the previously recorded failure event until the time reference in consideration, whereas type two counts the number of emitted messages per day.

It is noticeable a general improvement of the results from the AZUL's data set in comparison with the results from the Portugália Airlines' data set. The larger fleet, and therefore the higher number of failure events in the AZUL data set, may be a significant factor to explain the observable difference. Also, recall that the AZUL airlines data set is filtered by the already mentioned AHEAD-PRO software, being composed of only the messages that it considers as relevant. Considering the the type 2 results using the AZUL data set, there is a highly noticeable lack of ability of the linear regression algorithm to output meaningful results. This might confirm the lack of linearity between the features and the label, as in some iterations the model totally fails to predict meaningful results. In addition, the results also highlight a general improvement from the type 1 approach in comparison with type 2. This might be explained by the constant upwards evolution on the number of emitted messages leading into a failure, that may not occur on the type 2 analysis due to some irregularities of the data.

Having compared the results from the two available data sets, it is by this point important to showcase the comparison between all five types of approaches, applied to the Portugália Airlines' data set. Starting with the type 3 analysis, recall that this is in all aspects similar to the type 1 approach, but measures the evolution in flight cycles instead of days. Table 6.4 presents the results. In order to allow the comparison with the other approaches, and considering that, on average, each aircraft performs 5.7 cycles per day (January to September 2019), the second column block on the right-hand side of the table computes the errors in days.

It is noticeable a general improvement on the results comparing the type 3 with the type 1 and 2

Table 6.4: Results from the type three approach.

TYPE	ALGORITHM	PGA			
		Cycles		Days (5.7 cycles/day)	
		MAE	RMSE	MAE	RMSE
3	LinearRegression	760.3	960.6	133.4	168.5
	RandomForest	579.1	813.5	101.6	142.7
	KnearestNeighbors	606	859.6	106.3	150.8
	SupportVectorMachine	465.6	618.2	81.7	108.4
	GradientBoostingMachine	501.5	647.8	88	113.6

outcomes, showcased on table 6.3. This might be explained by the lack of some inaccurate remaining useful time definition caused by maintenance periods that the day based analysis is unable to filter, whereas the flight cycles are immune to those misconceptions of operative periods. From the first three analyses applied to the Portugália Airlines data set, the best model comes from type three approach, with a minimum MAE of 81.7 days and a minimum RMSE of 108.4 days. Once again, the results from the linear regression are the worse, however not as abnormal as in the cases presented in table 6.3.

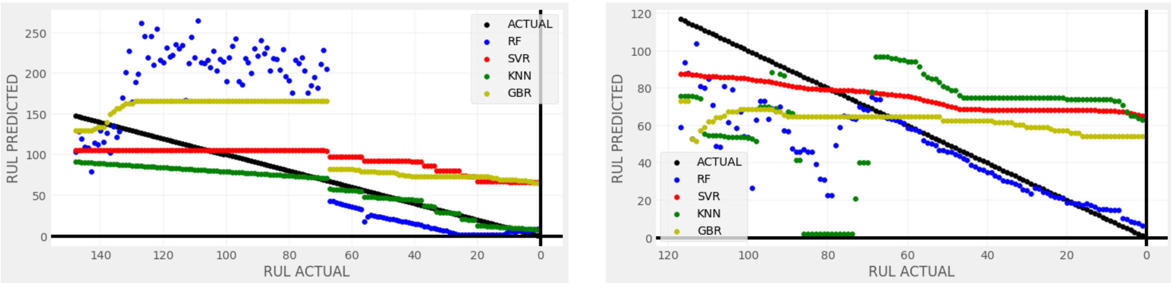
The types 4 and 5 analysis are in all aspects similar to the first two (type 1 and 2) analysis, respectively, except on the definition of failure. Until this point, the failures were defined by the emission of the already mentioned failure CAS messages. However, another possibility is to consider the unplanned replacements of two main valves of the pneumatic system as the failure events. These two valves are responsible for controlling the extraction of air from the high (valve 1) and low (valve 2) pressure stage of the engine's compressor.

Table 6.5: Results from the fourth and fifth approaches. All units are in days

TYPE	ALGORITHM	PGA					
		Valve 1		Valve 2		Valve 1 and 2	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
4	LinearRegression	444.8	482.1	280.8	371.2	206.1	399.6
	RandomForest	330.6	384.3	159.8	217.9	151,2	198.5
	KnearestNeighbors	355.6	395.2	180.4	255.8	147.1	191.9
	SupportVectorMachine	212.4	244.2	125.9	159.8	113.4	151.6
	GradientBoostingMachine	352.4	400.4	141.9	191.7	128.6	166.5
5	LinearRegression	294.1	328.9	168.9	212.5	184.9	236.5
	RandomForest	301.4	351.7	144.8	201.4	130.9	173.1
	KnearestNeighbors	305.8	351.9	140.8	196.9	131.4	174.1
	SupportVectorMachine	204.9	232.1	138.4	197.1	108.9	148.3
	GradientBoostingMachine	293.9	337.2	131.2	182.9	119.4	159.9

Three different variants of the last two types of approaches are presented in table 6.5. The first tries to predict a future unplanned replacement of valve 1, other the valve 2, and finally the last aims to predict the need for a replacement, without distinguishing the valve. Comparing with the results from the first three approaches, there are no improvements result wise. This may lead to the conclusion that considering the replacement maintenance logs may not improve the results over the emission of failure indicative CAS messages. However, the differences may be caused by the lower number of valve replacement logs comparing with the emitted failure CAS messages, which may indicate that there are situations of indicated failures in the pneumatic system that are not necessarily correlated to the replacement of the valves. The human interpretation inherent to the decision making regarding the replacement maintenance actions may also explain the disparity. The results tend to improve with the increasing number of failures studied (valve 2 compared with valve 1) which may also be noticeable analysing the improvement of the results when combining the replacement logs of the two valves.

It is clear to conclude that the best results are generally from the Support Vector Machine algorithms, and the worse from the Linear Regression, which indicates the lack of linearity of the problems in question. The best overall mean decrease in accuracy of 74.1 days is provided when using the more extensive AZUL data set, and applying the Gradient Boosting Machine algorithm. There is a constant tendency on the RMSE to be substantially higher than the MAE, which indicates the presence of large errors on the model's predictions.



(a) Failure predictions PGA

(b) Failure predictions AZUL

Figure 6.3: Evolution of the Remaining Useful Life (RUL) predictions and comparison with the actual RUL. This evolution concerns a selected failure from the Portugália Airlines (left hand side) and AZUL airlines (right hand side) failure set. Extracted from a type 1 model. All units are in days.

Figure 6.3 presents the plot comparing the evolution of the RUL estimation of a selected failure. It's possible to compare the predictions with the actual values, and how the predicted times follow the real evolution. The predictions are extracted from a type 1 problem. Most of the algorithms are unable to follow the RUL's real evolution, sometimes predicting constant values over the failure period. However, in both figures, there is a noticeable well behaved evolution from the results from the random forest algorithm and also from the KNN (K Nearest Neighbours) in the figure 6.3(a). This ability to follow the true evolution of the remaining useful life may indicate that there is, to some extent and in certain cases, predictive capabilities in the developed models. Otherwise, nor the algorithms would be able to learn from the training data, nor such well behaved predictions would be possible.

Having explained all the comparisons between the regressive approaches, it is by this point relevant to compare them with the baseline results. Recall that the baseline results simulate the committed error by expecting a failure event every certain fixed period. This period depends on the failure probability decided to establish as critical, and is in this thesis defined as the BX life. Due to the uncertainty of the generally admitted BX life for the pneumatic system, it is considered for further comparisons the one that produces the least amount of error when faced with an independent failure test data set. These results are shown in the figures 6.1 and 6.2. Comparing the results from the AZUL airlines data set, the best MAE from the baseline approach overcomes the most favourable result shown in the table 6.3 (65.4 from the baseline approach, 74.1 from the type 1 approach). However, if the main comparison measure is the RMSE, then the machine learning results are better, with a minimum of 99.1 days when compared with the 103.3 days of the baseline approach. This means that, for this case, the source of the best results is dependant on the desired evaluation measure.

Moving to the comparison between results from the Portugália Airlines data set, the baseline results shall be faced with both the type 1, 2 and 3 approaches from the machine learning results, due to the common failure events. This case differs from the former as both the evaluating criteria are improved with the machine learning approach. The overall minimum is achieved by the type 3 approach, with 81.7 days of MAE and 108.4 days of RMSE.

Last but not the least, considering the valve replacement data, the results follow the same trend, with the baseline results not reaching the best results from the machine learning-based approach (type 4 and 5). This is the prevailing tendency of the results, except for the first motioned case regarding the AZUL data set. The conclusions withdrawn are exposed in the chapter 7.

6.3 Classification Results

This section presents the results from the classification models built based on the available data. For the sake of recalling, these approaches are in all aspects similar to the previous already presented except on the definition of the problem's label. Instead of requiring the algorithms to predict float point numbers, referred to the remaining useful life of the system, this second approach aims to classify the risk of failure inherent to the system. If a failure is expected to occur within 20 days of a certain timestamp, the system's risk is considered as high or low otherwise. Unlike the evaluation measures analysed for the regressive approaches, in this case, the goal is to maximise the values, defined by the following expressions shown in the table 6.6.

Table 6.6: Performance metrics for classification.

Performance measure classification	
Precision	Precision = $\frac{TruePositives}{TruePositive + FalsePositive}$
Recall	Recall = $\frac{TruePositives}{TruePositive + FalseNegatives}$
Accuracy	Accuracy = $\frac{TruePositives + TrueNegatives}{Totalobservations}$
F1 Score	F1 score= $2 \times \frac{Precision \times Recall}{Precision + Recall}$

The F1 score combines the influence of other two important performance measures: precision and recall. The precision measures what is the proportion of correct positive predictions compared with all the positive predictions from the models. On the other hand, the recall compares, once again, the correct positive predictions with all the samples that should have been predicted as positives. Finally, the accuracy computes the percentage of correct predictions between the total number of predictions performed. Notice that the positive results are in this case considered as the prediction of a high risk system.

The result demonstration in this section adopted follows the same structure of the previous section. Firstly, table 6.7 showcases the results obtained from the first two analysis: type 1 and type 2.

Table 6.7: Classification results for the type 1 and 2 problems

TYPE	ALGORITHM	DATA							
		PGA				AZUL			
		Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
1	RandomForest	0.533	0.153	0.39	0.205	0.616	0.289	0.572	0.380
	KnearestNeighbors	0.542	0.157	0.351	0.183	0.613	0.266	0.477	0.337
	SupportVectorMachine	0.764	0.205	0.158	0.151	0.782	0.581	0.160	0.229
	GradientBoostingMachine	0.699	0.265	0.241	0.188	0.778	0.480	0.279	0.344
2	RandomForest	0.638	0.179	0.276	0.185	0.602	0.271	0.487	0.338
	KnearestNeighbors	0.656	0.151	0.216	0.151	0.649	0.282	0.416	0.331
	SupportVectorMachine	0.771	0.391	0.09	0.090	0.810	0.665	0.173	0.273
	GradientBoostingMachine	0.733	0.247	0.159	0.149	0.760	0.398	0.258	0.309

Unlike the regressive approach, the best results are generally from the random forest algorithm. The maximum F1 score obtained from the analysis results from the application of the AZUL data set, with a value of 0.380. Also, it is possible to notice a slight improvement result wise comparing the results from Portugália Airlines and the AZUL airlines data set, and comparing the type 2 and type 1 models. This confirms what was already noticeable through the regressive approach showcased in the preceding section. The maximum recall from the first two analysis comes at 0.572. In fact, this might be the evaluating measure with the highest operational importance, because a low recall score shows that there are several high-risk situations that the models were not able to predict correctly, defining the system as healthy when in fact it was about to fail.

Table 6.8 shows the results from the analysis that redefines the evolution measure as flight cycles instead of days. Notice that, following the same reasoning as in the previous section, it was considered that each aircraft performs 5.7 cycles per day, consequently the threshold of the system's risk is in this approach redefined as 114 cycles. Therefore, the system's risk is high if it is within 114 cycles of a failure event.

Table 6.8: Classification results from the type 3 approach

TYPE	ALGORITHM	PGA			
		Accuracy	Precision	Recall	F1 Score
3	RandomForest	0.467	0.199	0.640	0.3
	KnearestNeighbors	0.501	0.203	0.588	0.297
	SupportVectorMachine	0.507	0.182	0.450	0.250
	GradientBoostingMachine	0.640	0.190	0.307	0.212

This approach provides the best results considering the recall evaluation measure, with a maximum of 0.640. It also allows one to confirm the general improvement of the results when considering cycles instead of days, once again following what was already perceptible with the regressive models. Notice that, although the accuracy suffers a slight decrease, the precision and the recall are improved. The accuracy is not always the best measure. A prediction that has the majority of the test labels as negative, or, in this case, low risk, the high number of true negatives results have a substantial contribution to the end accuracy result. For the sake of example, consider that the model predicts all the situations as low risk. Due to a high number of low-risk situations in the test set, the accuracy result is always enhanced by the close to one fraction between the true negatives and the number of total observations. Therefore, the accuracy is not considered in this thesis as meaningful as the other evaluating measures.

The last two methods' results are shown in table 6.9. The fault definition is here considered as unscheduled replacement logs instead of the emission of fault indicative CAS messages. Three different analysis are here once again performed, with the valve subject to the replacement being differentiated as valve 1, 2, and undifferentiated in the third result set. The best results in terms of the F1 score are obtained predicting the replacement of the valve 2, using a type 5 approach. Once again, the results generally fail to reach the what is shown on the table 6.8. The next chapter aims to discuss the conclusions extracted from the results showed in this chapter.

Table 6.9: Type 4 and 5 classification results.

TYPE	ALGORITHM	PGA							
		Valve 1/Valve 2				Valve 1 + Valve 2			
		Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
4	RandomForest	0.36/0.610	0.025/0.155	0.3/0.485	0.046/0.183	0.687	0.155	0.388	0.205
	KnearestNeighbors	0.411/0.633	0.008/0.104	0.1/0.426	0.015/0.166	0.695	0.176	0.488	0.235
	SupportVectorMachine	0.74/0.733	0.0/0.194	0.0/0.335	0.0/0.193	0.877	0.102	0.015	0.02
	GradientBoostingMachine	0.272/0.771	0.044/0.267	0.4/0.348	0.077/0.215	0.760	0.153	0.268	0.174
5	RandomForest	0.466/0.748	0.078/0.167	0.214/0.466	0.095/0.241	0.734	0.167	0.358	0.210
	KnearestNeighbors	0.468/0.788	0.012/0.180	0.119/0.369	0.022/0.236	0.769	0.193	0.337	0.225
	SupportVectorMachine	0.635/0.912	0.01/0.399	0.1/0.0277	0.0183/0.051	0.894	0	0	0
	GradientBoostingMachine	0.399/0.870	0.099/0.294	0.314/0.324	0.209/0.302	0.813	0.222	0.240	0.185

Chapter 7

Conclusions

This chapter aims to discuss the achievements and final remarks of this dissertation. There is also a section that addresses what the author considers important for future work.

7.1 Achievements

This thesis proposed to analyse the potential predictive capabilities of one type of data continuously being emitted by the aircraft - the CMC (Central Maintenance Computer) messages. This was performed by an initial data's exploratory analysis, in the hope of finding clear evidence of such potential. Failing to find that, the process of preparing the data for a more advanced approach was described. Several model alternatives were considered, including a non-machine learning-based baseline approach, which only considered failure data and the corresponding time between failures. The result comparison between all the different models allows to extract some conclusions.

Starting with the analysis of the regressive results, although not very pronounced, there is an improvement in the methods that adopt the counting period as time difference since the previously recorded failure (type 1) in comparison with the approach that accounted messages emitted within 24 hours of the certain time reference (type 2). This improvement in the results might indicate a better reaction by the models to an always positive evolution of the message count along each failure period. Although this behaviour is also noticeable for the first two classification approaches, the type 4 and 5 (similar to the type 1 and 2, only with a different failure definition) do not follow the same tendency. Therefore, the best feature construction strategy appears to be dependant on the failure definition.

Noticeable in both the regressive and classification approaches is the improvement results-wise when treating the evolution of the data in flight cycles instead of days (type 3). This, already mentioned in the previous chapter, may indicate that the machine learning models are sensitive to non-operative periods, for instance, in maintenance, and that cycles represent in a more accurate way the degradation of the aircraft. The type 1 and 2, although not considering messages emitted when the aircraft is on the ground, the day based feature construction may fail to ignore maintenance periods, considering them with 0 emitted messages. Type 3 data structure only considers performed cycles, therefore avoiding such misconceptions.

The comparison between the results gathered from the two analysed data sets is also of interest. The noticeable improvement from the AZUL airlines data set may indicate that with a higher number of data records, which ultimately results in more samples for the algorithms to learn from, results in a significant improvement of the models. This may be true, but there is also another factor that needs to be taken into account: the filtered nature of the messages from the AZUL airlines' data set. Hence, two main influences do not allow one to compare the results from the two data sets directly. It would be interesting to equal the size of the two data sets and therefore conclude on the influence of the filter applied to the data, which only keeps what the software considers as relevant.

The alteration of the failure definition from failure indicative CAS (Crew Alerting System) messages to replacement maintenance logs of two main constituting valves of the pneumatic system (type 4 and 5 approaches) did not improve the results. Like what was discussed in the previous paragraph, the lower number of failure events considered influences the ability of the models to learn from the data in the best way possible. Also, the distinct failure modes that may trigger the failure message may not all be associated with the necessity of a replacement action to the considered valves. In addition, the human influence inherent to the valve replacement decision may also affect the results.

Analysing purely the quantification of the errors, the best results from the regressive models reach an MAE (Mean Absolute Error) of 74.1 days and an RMSE (Root Mean Squared Error) of 99.1 days. This would mean if one takes into consideration any prediction in specific, one should expect an error of 74.1 days. In a real-world situation, such high error is prejudicial to the trustworthiness of an autonomous model that would eventually predict the remaining useful life of the equipment. Two different scenarios are worth considering. In the circumstances where the failure is imminent, a misprediction of, on average, 74.1 days may be highly more prejudicial than on situations far from the failure. Consequently, for the remaining useful life prognosis, errors in the order of magnitude of what was obtained in this thesis would jeopardise the trustworthiness of a fully independent prognostic tool based on the models obtained. Regarding the RMSE, due to its definition, the results are always equal or greater than the MAE, equality being the best possible outcome. The RMSE results from this thesis suggest the presence of high predictive errors, shown by the tendency of the RMSE being substantially higher than the MAE.

Comparing the regressive results with the ones obtained from the baseline analysis allows to extract some relevant conclusions. Worth recalling that the baseline approach only considered failure data combined with the Weibull distribution to define the best preventive maintenance intervals possible. The regressive models, along with the failure data, used the message data to prognose the remaining useful life of a system. Almost all the regressive approaches presented, considering the best models from each comparable case, presented the best results when applying the machine learning-based models. The only exception came from the MAE obtained by the AZUL data set. Nevertheless, the undeniable improvement means that a machine learning-based solution would estimate the failures in a more reliable manner than a more conventional preventive maintenance scheduling. Therefore, it proves that such techniques help maintenance move one step forward, allowing to reduce risks, increase safety, and helps to build a more accurate maintenance plan. Also, this improvement enables to conclude that, associated with more complex methods, the message data assists in providing more accurate predictions regarding the system's remaining useful life.

In view of this, from the figures 6.3 it is noticeable a well behaved evolution of the RUL's prediction, obtained from the random forest (6.3(a) and (b)) and KNN (K Nearest Neighbors) (figure 6.3(a)) algorithms, specifically within 60 days of the failure event. This indicates that, in fact, the models were able to extract some information from the learning data that allowed to rightfully predict about the remaining useful life of the system. If the models' data were fully uncorrelated to the failure event, those predictions would not be possible. Consequently, although these predictions are not frequent in the predictive set obtained in this thesis, it may indicate that the data has in fact predictive capabilities.

The classification results also support this idea. The best overall result in terms of the F1 score was 0.380, from the type 1 approach applied to the AZUL airlines data set. However, the type three approach results in a maximum overall recall score of 0.640. This might be considered the most important evaluating measure for a prognostic tool applied to the aeronautics, as it is defined as the ratio between the correct positive scores and the situations that were supposed to be classified as positives. It has an importance that maybe should be more pronounced than the precision score to the final F1 score, as a low recall indicates that there are a significant amount of situations where the model considers the system as healthy, but in fact, it is not. The precision, on the other hand, measures how often the predictions indicated a high failure risk when, in fact, the failure is not imminent. Therefore, a low precision might lead to unnecessary maintenance actions, implying monetary waste. Though, safety being, in this case, a priority, it might be a reasonable assumption to consider the recall as the number one score to maximise.

The results of this thesis indicate that the inherent predictive capabilities of the CMC messages are not enough to develop a fully independent prognostic tool. By fully independent, the author means a tool that would act as the only reference to decide whether a maintenance action is or is not necessary at a given moment. Despite this, the results show that there are capabilities associated with the CMC messages that enable the improvement of failure prognosis when compared with the more conventional baseline approach. Also, the classification results reach a recall of 64%, which would not be achievable if the models had no failure predictive capabilities. Therefore, such models may be suitable as a complement to the decision-making process of the unplanned maintenance interventions, and not as the unique decisive factor. From an operational point of view, it is believed that those models may act as alerts to upcoming failure events and that the experienced personnel may also be able to judge the predictions of the models with other indicators, therefore concluding on the truthiness of the prognosis. The results show that the predictions often misjudge the real state of the system, hence the need for the human intervention may not be dismissed. Therefore, a prognostic tool based on CMC messages, although not fully independent, may aid the health management of the fleet, improving the proactiveness to failure events.

7.2 Future Work

As mentioned in the previous paragraph, it would be interesting to compare the results from the two airline companies equalling the data set size and to compare the influence of the software filter fairly. Any improvement would suggest that the filter ignores some irrelevant messages the only contribute as noise to the models and therefore degrade their performance. It would also be interesting to study the messages' evolution from the AZUL airlines data set in terms of flight cycles. This implied a meaningful improvement result-wise to the treatment of the Portugália Airlines data set, hence it might also contribute to the improvement of the models using the AZUL data set.

Finally, to improve quantitatively the results, a future possibility would be to redefine the predictive data. Instead of considering message data that fails to be a promising source to an autonomous prognostic tool, the author suggests considering raw sensor data instead. The message data is, to some extent, unreliable, often misrepresenting the actual failure state of the system, as there are replacement events not associated with the "bleed fail" message nor associated with an evident increase on the CMC messages' emission rate. Therefore, this leads to conclude that CMC messages, in some circumstances, fail to be the most accurate indicator of the system's state. This is supported by the significant improvement result-wise with the introduction of sensor data to the models, discussed in the project [25]. Raw sensor data represents the behaviour evolution of the system continuously and in greater detail, therefore having greater possibilities in improving the performance of machine learning models, and hence increasing the chances of developing a fully automated failure prognostic tool.

References

[1] European Aviation Safety Agency (2018). Annual Safety Review 2018. pp.16-19.

[2] Icao.int. (2019). Facts & Figures. [online] Available at: https://www.icao.int/sustainability/Pages/FactsFigures.aspx [Accessed 19 Jul. 2019].

[3] Infante, V. (2018). Reliability, Maintenance and Availability.

[4] Flyembraer.com. (n.d.). Aircraft health Analysis and Diagnosis. [online] Available at: https://www.flyembraer.com/irj/go/km/docs/download_center/Anonymous/AHEAD/HTML/ahead.html [Accessed 13 Jul. 2019].

[5] T. Hoyland, C. Spafford, and A. Medland. MRO BIG DATA – A Lion or a Lamb? Oliver Wyman, 2016. Innovation and Adoption in Aviation MRO.

[6] Portugalia-airlines.pt. (2019). PGA > About Us. [online] Available at: https://www.portugalia-airlines.pt/en/Pages/AboutUs.aspx [Accessed 9 Jul. 2019].

[7] Skybrary.aero. (2019). Ground Operations - SKYbrary Aviation Safety. [online] Available at: https://www.skybrary.aero/index.php/GroundOperations [Accessed 10 Jul. 2019].

[8] Embraer. Aircraft Maintenance Manual - System Description Section. Embraer S.A., Sao José dos Campos, Sao Paulo, Brasil, August 2008. Revision 25 - 22 September 2017.

[9] Embraer. Maintenance Training Manual. Embraer S.A. Embraer ERJ-170/190 (GE CF34) General Familiarisation.

[10] Fortney, K. (2018). Machine Learning — An Error by Any Other Name - Towards Data Science. [online] Towards Data Science. Available at: https://towardsdatascience.com/machine-learning-an-error-by-any-other-name-a7760a702c4d [Accessed 18 Jul. 2019].

[11] Manual of procedures for operations inspection, certification and continued surveillance. (2010). 5th ed. Montreal: International Civil Aviation Organization, p.xvi.

[12] Faa-aircraft-certification.com. (2019). FAA Definitions. [online] Available at: http://www.faa-aircraft-certification.com/faa-definitions.html [Accessed 6 Aug. 2019].

[13] Papakostas, N., Papachatzakis, P., Xanthakis, V., Mourtzis, D. and Chryssolouris, G., 2010. An approach to operational aircraft maintenance planning. Decision Support Systems, 48(4), pp.604-612.

[14] BSI Standards Publication. (2010). Maintenance — Maintenance terminology (Standard No. BS EN 13306:2010). Retrieved from http://irma-award.ir/wp-content/uploads/2017/08/BS-EN-13306-2010.pdf

[15] A320 Maintenance Planning Document, Revision 31, Issue 0. (2010). 31707 Blagnac Cedex France: Maintenance Planning and Services, p.6.

[16] Dinis, D. (2018). Integrated Methodology for Aircraft Maintenance Forecasting and Capacity Planning Under Uncertainty. PhD. Instituto Superior Técnico, Universidade De Lisboa.

[17] Robelin, O. (2010). Maintenance Review Board Process (MRB) And Instructions for Continued Airworthiness.

[18] Kobbacy, K.A.H. and Murthy, D.P. eds., 2008. Complex system maintenance handbook. Springer Science & Business Media, p57,81.

[19] Mishra, M. (2018). Prognostics and Health Management of Engineering Systems for Operation and Maintenance Optimisation. Phd. Luleå University of Technology Luleå, Sweden.

[20] Rose, A., 2002. Reliability Centered Maintenance. Cadick Corporation TB014, August, 1, p.2002.

[21] Kinnison, H.A. and Siddiqui, T., 2012. Aviation maintenance management.

[22] Saltoğlu, R., Humaira, N. and İnalhan, G., 2016. Aircraft scheduled airframe maintenance and downtime integrated cost model. Advances in operations research, 2016.

[23] Easy Access Rules for Master Minimum Equipment List (CS-MMEL). (2018). [ebook] pp.11-12. Available at: https://www.easa.europa.eu/sites/default/files/dfu/Easy%20Access%20Rules%20for%20 CS-MMEL.pdf [Accessed 24 Jul. 2019].

[24] Ackert, S.P., 2010. Basics of aircraft maintenance programs for financiers. URL: http://aircraftmonitor.com/uploads/1/5/9/9/15993320/basics_of_aircraft_maintenance_programs_for_ financiers___v1.pdf [Accessed 25 Jul. 2019].

[25] Baptista, M. (2018). Machine Learning and Deep Learning for Prognostics and Predictive Maintenance of Aeronautical Equipment. Phd. Instituto Superior Técnico, Universidade De Lisboa.

[26] Liu, Y. (2019). Python machine learning by example. 2nd ed. Python Machine Learning By Example: The easiest way to get into machine learning, pp.5-20.

[27] de Pádua Moreira, R. and Nascimento, C.L., 2012, March. Prognostics of aircraft bleed valves using a SVM classification algorithm. In 2012 IEEE Aerospace Conference (pp. 1-8). IEEE.

[28] Nicchiotti, G., 2018, July. Data-Driven Prediction of Unscheduled Maintenance Replacements in a Fleet of Commercial Aircrafts. In PHM Society European Conference (Vol. 4, No. 1).

[29] J. Garbade, M. (2019). Clearing the Confusion: AI vs Machine Learning vs Deep Learning Differences. [online] Towards Data Science. Available at: https://towardsdatascience.com/clearing-the-confusion-ai-vs-machine-learning-vs-deep-learning-differences-fce69b21d5eb [Accessed 31 Jul. 2019].

[30] Mahapatra, S. (2019). Why Deep Learning over Traditional Machine Learning?. [online] Medium. Available at: https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063 [Accessed 8 Aug. 2019].

[31] Sharma, A. (2019). Machine Learning & Deep Learning. [online] DataCamp. Available at: https://www.datacamp.com/community/tutorials/machine-deep-learning [Accessed 3 Aug. 2019].

[32] Embraer.com. (2019). Portal Embraer. [online] Available at: https://embraer.com/global/en [Accessed 5 Aug. 2019].

[33] Hitachivantara.com. (2019). Pentaho Data Integration - Create Data Pipelines. [online] Available at: https://www.hitachivantara.com/en-us/products/big-data-integration-analytics/pentaho-data-integration.html [Accessed 9 Aug. 2019].

[34] Afiklmem.com. (2019). Air France Industries KLM Engineering & Maintenance - PROGNOS - Predictive Maintenance. [online] Available at: http://www.afiklmem.com/AFIKLMEM/en/g_page_standard/MRO_lab_Innovations/BigData.html [Accessed 13 Aug. 2019].

[35] Tableau Software. (2019). Tableau: Business Intelligence and Analytics Software. [online] Available at: https://www.tableau.com [Accessed 23 Aug. 2019].

[36] Kotsiantis, S.B., Kanellopoulos, D. and Pintelas, P.E., 2006. Data preprocessing for supervised leaning. International Journal of Computer Science, 1(2), pp.111-117.

[37] Al-Masri, A. (2019). What Are Overfitting and Underfitting in Machine Learning?. [online] Towards Data Science. Available at: https://towardsdatascience.com/what-are-overfitting-and-underfitting-in-machine-learning-a96b30864690 [Accessed 31 Aug. 2019].

[38] Cochrane, C. (2019). Time Series Nested Cross-Validation. [online] Towards Data Science. Available at: https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9 [Accessed 2 Sep. 2019].

[39] Scikit-learn.org. (2019). 1.1. Generalized Linear Models — scikit-learn 0.21.3 documentation. [online] Available at: https://scikit-learn.org/stable/modules/linear_model.html [Accessed 8 Sep. 2019].

[40] Scikit-learn.org. (2019). 1.11. Ensemble methods — scikit-learn 0.21.3 documentation. [online] Available at: https://scikit-learn.org/stable/modules/ensemble.html#forest [Accessed 10 Sep. 2019].

[41] Scikit-learn.org. (2019). 1.6. Nearest Neighbors — scikit-learn 0.21.3 documentation. [online] Available at: https://scikit-learn.org/stable/modules/neighbors.html#regression [Accessed 14 Sep. 2019].

[42] Kernelsvm.tripod.com. (2019). Support Vector Machine Regression. [online] Available at: http://kernelsvm.tripod.com/ [Accessed 18 Sep. 2019].

[43] James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. An introduction to statistical learning (Vol. 112, p. 184). New York: springer.