CHALMERS
UNIVERSITY OF TECHNOLOGY

A study on remaining useful life estimation for predictive maintenance of a robot cell

DOYEL JOSEPH

TILANI GALLEGE

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se

Master's thesis 2021

A study on remaining useful life estimation for predictive maintenance of a robot cell

DOYEL JOSEPH
TILANI GALLEGE

Department of Industrial and Materials Science
Chalmers University of Technology
Gothenburg, Sweden 2021

A study on remaining useful life estimation for predictive maintenance of a robot cell

DOYEL JOSEPH
TILANI GALLEGE

Supervisor: Catarina Dudas, Volvo Cars AB
Supervisor: Ebru Turanoglu Bekar, Chalmers University of Technology
Examiner: Anders Skoogh, Chalmers University of Technology

Master’s Thesis 2021
Department of Industrial and Materials Science
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Neural neworks which works like a human brain, and provides artificial intelligence

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2021

A study on Remaining useful life estimation for predictive maintenance of a robot cell

DOYEL JOSEPH & TILANI GALLEGE
Department of Industrial and Materials Science
Chalmers University of Technology

Abstract

With the introduction of digitization, a vast amount of data is available for industries, which can be used for their future sustainability and competitive advantage. The amount of data generated by the modern machines exceeds the capacity of manual analysis. At the same time, the improvement in computational power and advancements in the field of Artificial Intelligence (AI) provides insightful analysis, which can enable data driven decision making in numerous fields such as manufacturing, automobile, construction and food processing, and so on. This thesis project aims to propose a study on a Remaining Useful Life (RUL) estimation for predictive maintenance by using of a data-driven approach.

This thesis analyzes the present maintenance practices that are being followed in Volvo Cars Torslanda and introduce data-driven based maintenance planning to optimize the available resources. For this reason, the suitable predictive maintenance strategies are examined for gluing workstations in the automobile production factory. The purpose of the project is to transfer from time-based maintenance to predict equipment failure and RUL using state-of-the-art machine learning algorithms. In order to perform a data driven approach for predictive maintenance, the CRoss Industry Standard Process for data mining (CRISP-DM) is followed as a reference methodology in this thesis. Different data sources are analyzed and the most relevant sources for the study are selected. An exploratory data analysis is done with the available data from specified workstation and the most suitable parameters are selected for the study. Due to data from different sources, many data pre-processing techniques are utilized in order to merge and make the data suitable for the Machine Learning (ML) algorithms used for prediction. Two of the most common ML algorithms such as Auto Regressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) are utilized for the prediction and LSTM model provides a better prediction on the test data. In conclusion, this thesis provides a set of recommendations for the company, which would enable them to conduct future predictive maintenance projects and to help Volvo Cars in their advancements with Smart maintenance road map.

Keywords: Smart Maintenance, Predictive Maintenance, Machine Learning, Data-driven Decision Making, Exploratory Data Analysis, Industrial Robots, Manufacturing

Acknowledgements

We would like to express our gratitude to Ebru Turanoglu Bekar, academic supervisor at Chalmers for her continues guidance and support. She was a constant source of encouragement even during the hardest times of the thesis work. Her valuable inputs and detailed reviewing of our thesis helped us to come up with presentable thesis report. We would like the thank Anders Skoogh, Examiner at Chalmers University for his valuable feedback and the motivation to strive hard.

We would like thank Volvo Cars AB for entrusting us for this thesis work and providing all the necessary resources required for the completion this thesis work. We would like to express our sincere gratitude for Catarina Dudas, supervisor at Volvo cars, for constantly guiding us through out the project. Her availability and knowledgeable inputs where quintessential for our learning and for the successful completion of the project. We would like to thank Ronnie Jansson, Head of Advanced Analytics and AI, Volvo cars, for selecting us for this project and providing us the advanced facilities, which were crucial for the project. We would like to thank Per Andersson, Daniel Thomson, Mickael Svensson, from the maintenance department of Volvo cars for sharing their domain knowledge and giving us the freedom to explore and learn the different data sources for the thesis project. We would also like to thank Praffful Mishra, Data engineer at volvo cars, for providing all the technical supports and helping us to easily adapt to the work flow.

Finally express my deepest gratitude to our parents and friends who have constantly supported and motivated us throughout the time of our masters studies. We would like to thank all our professors, who inspired us through out our masters program at Chalmers University.

Doyel Joseph & Tilani Gallege
Gothenburg, September 2021

Contents

List of Figures

List of Tables

1
Introduction

This chapter gives a picture about the background of the thesis project and discusses the aim, problem description, research questions, the scope and limitations of this thesis work.

1.1 Background

A phenomenal boom during the industrial revolution gifted notable innovations in fields such as automation, rapid manufacturing and big data analytics. The fourth industrial revolution which is also termed as Industry 4.0 aims at digital transformation of manufacturing and production industries . This digital transformations creates value to companies as well as to its customers by improving productivity, quality and sustainability. Highly complex, dynamic, and integrated information systems, and very high computational powers are the key enablers of Industry 4.0 [48].

When it comes to adapting to changes, the automotive industry has always been the pioneer. This was mainly due to the enormous amount of competition and incriminating challenges to meet the customer expectations. The scenario was not different when it comes to application of data science for improving the standard and quality of production. This adaptability of the industry is visible since the first industrial revolution to the state-of-the-art Industry 4.0 techniques, and this has enabled the industry to improve their production sustainably [69]. Out of the key enablers of Industry 4.0, smart maintenance plays a crucial role [69]. The integration of Information and Communication Technology (ICT) to the manufacturing process gave a breakthrough for gigabytes of data in to the manufacturing industries [71]. This big data and advanced computing algorithms enables us to use a relevant data to generate precise operational decisions.

Volvo Cars is one of the worlds' leading manufacturer and supplier of safest generation of cars. Being a company which is always known for the state-of-the-art safety standards and sustainability, Volvo Cars emerged out to be one of the most promising brand name when it comes to automobile sector. The organisation has been working towards circular economy and climate neutrality for decades and has always been a role model for its competitors.

Volvo Cars has their state-of-the-art manufacturing plant with a very high level of automation at Torslanda, Sweden. When the the level of automation increases it becomes hard for the maintenance team to schedule and execute the maintenance activities with least possible impact on the production process. Conducting maintenance at right time is very crucial for the life of the machine. Unpredicted breakdown leads to an economical wastage is also crucial, because these breakdowns at critical moments lead to disruption of production in an entire production line. This leads to a huge financial issues which might result in losses such as eviction from the market and reputation loss. When it comes to big production lines, like that of Volvo Cars, the loss would be very huge. Even a marginal downtime can cost significantly, when it comes to mega factories.

In order to avoid unwanted breakdowns there are different types of maintenance strategies that are being used in industries. Presently Volvo Cars follow a scheduled maintenance strategy in all the robots in the assembly line. This strategy involve conducting preventive maintenance at fixed time intervals, irrespective of the condition of the machines. Even though this strategy mostly solve the issue of unpredicted breakdowns, but this approach has two major drawbacks. By conducting scheduled maintenance a phenomenal part of the useful life of a machine will not be utilised for a beneficial production and this increases the overall maintenance cost. Also, the disruptions in the production lines for maintenance activities will also consume some of the useful production hours, which reduces the productivity of the plant.

1.2 Aim

The current maintenance practices at Volvo cars are based on time-based preventive maintenance strategy. One of the major problem with this strategy is that, a good amount of useful life of the machine is left unutilized. The main objective of this work is to apply machine learning (ML) methods to predict failure of a gluing robot which is being utilized in Volvo Car's production facility. The final outcome of the thesis work would facilitate the gluing machine system with data driven decision making for maintenance work, which would prevent unprediced failures and also the remaining useful life estimation (RUL) estimation. The thesis work also aims to explore the different data sources at the company and provides a road map for the company to facilitating data driven decisions in the future.

1.3 Research Questions

The following research questions are concluded based on our objectives of the thesis as follows:

> *RQ1 - Is the quantity and quality of available data sufficient to build the ML model for predictions?*

RQ2 - Which is the best ML model in predicting RUL of the machine based on performance metrics?

RQ3 - How can we improve the predictive maintenance system considering data acquisition and create a future road map?

1.4 Problem Description

Volvo Cars being one of the worlds' leading manufacturer of automobiles. Efficiency and productivity in manufacturing is very important for meeting the very high customer demand of the automobile industry. In order to achieve this efficiency a company should have a well defined strategy for the maintenance activities to ensure maximum up-time for all of its machines. Volvo Cars currently make use of value driven maintenance strategy for maintaining the robots and machines in good working conditions in their body shop factory. Value driven maintenance is always improving based on the learning from the previous feedback Rosqvist et al. [63]. Currently the scheduling and prioritisation is done only from the experience of the workers and the company wants to make maintenance decisions based on data driven approaches.

Volvo Cars is currently outlining a smart maintenance strategy, specifically describing the future development of data driven decision making. The aim is to move from calendar based, or some times even reactive maintenance to being able to predict equipment failure and planning maintenance on demand. Our project is conducted at the body shop of the company's production plant at Torslanda, Sweden. There are several robots in the plant for manufacturing activities like welding, gluing etc. In our project we are working with a gluing robot, which applies glue beads on work piece before spot-welding the work pieces together. The gluing robot was chosen for the study as it was acting like a bottle neck in the system. There were sudden, unpredicted break-downs in the gluing machine which led to stoppage of the production line several times. This was causing many problems for the maintenance team as they often needed to do reactive maintenance. On interviewing the maintenance experts from the company it was clear that the companies strategy was also to first change in to predictive maintenance based on the cost of each process.

The gluing station has an adhesive unit consisting of gluing gun, doser, docking station, pump, controller and media connection. The Robot Unit consist of the controller, media panel and manipulator. The aim of this project is to make use of the log data from the sensors attached to different parts of the gluing station and analyse whether the data available is sufficient for a ML approach to predict the next failure of this gluing robot . The aim is to use ML models and analyse how good they are performing with the current data available from the gluing machine. From experience learned from modeling the next step will be to lay down a road map for the future data driven plan at Volvo Cars which would enable them to achieve their goal of attaining smart maintenance capability.

1.5 Scope and limitation

As mentioned in problem description the scope of the project is to conduct an exploratory data analysis to see the relevance of data available and to make use of State-of-the-art ML algorithms for future prediction of failure in a gluing robot at Volvo Cars. The predictions would help the maintenance department to plan the maintenance activities in advance and also make use of the valuable RUL of the gluing machine.

The project concentrates on a gluing robot cell from a station consisting of many different robots. In this project we are analysing data from different sources, including high resolution alarm data, teamster log data which contains log readings from different temperature, pressure and volume sensors. All these data were extracted in the beginning of the project. There were no historical data available from the past before the starting date of the project. This scarcity of relevant historical data can be considered as one of the limitations of the project.

The aim of the project was to make a ML model which would predict the future failure of the gluing machine. In order to conduct supervised ML it is necessary to label the data with relevant data. The best possible data that could be used for labelling is the maximo data (maintenance log), but the time when the maintenance took place was manually entered by the technicians in the plant. On detailed discussions with the maintenance team it was made clear that the data was not accurate as each technician enter the time of maintenance and also the comments for the cause of failure and actions taken to rectify the failure, in different ways. Because of these reasons the maximo data was not used for labelling. Instead of using maximo data the alarm log was used for this purpose. Alarm log is automatically logged and the severe alarms from the log were assumed as failures. This is possible because whenever a class 'A' alarm (severe alarm category) occurs the machine is stopped immediately. Predictions of the occurrence of severe alarms are almost same as predicting the failures of the gluing machine.

The study was conducted during the pandemic COVID-19, due to which most parts of the work and discussions were conducted remotely. The remote working environment caused some delays in the data acquisition. Since it was the first time a similar project was conducted with specific data sources in the body shop of Volvo Cars, the data acquisition process was very iterative and time consuming. There were many sources of data which were not relevant, understanding and acquiring the relevant data also consumed majority of the time of the project. Also the historic data was not readily available, so at some points we had to wait for the data to be generated. These were the major limitations of the project.

2
Literature Study

For the master thesis work, an extensive literature study was carried out to gain knowledge and also to validate the results obtained. Literature study was carried out on various topics like Maintenance strategies, strategy used at Volvo Cars, smart maintenance, related works, ML models used for time series prediction and finally the evaluation techniques for assessing the models.

2.1 Maintenance strategies

In the past years, we could see a considerable increase in automation in almost all fields. The story of manufacturing industry is not so different. In order to meet the high customer demand, manufacturing industries are trying to automate most of their operations. As a byproduct to this, the importance of maintenance activities are also increasing in order to keep these machines up and running. Different maintenance strategies have been evolved and is being used by manufacturing plants all across the globe. From the literature study, it was clear that the suitability of each method depends on the severity of the repercussions caused by the failure of the machine and depends on the type of production activity [6]. For example if a product is continuously manufactured then, a breakdown in the production should be avoided. So for such a production strategy corrective or breakdown maintenance is not the correct fit. Maintenance strategies are classified differently with different standards and depending on their application. Al-Turki et al. [6] broadly classifies maintenance in to corrective maintenance and preventive maintenance, this classification is shown in Figure 2.1.

Corrective maintenance

In corrective maintenance strategy the maintenance activity is carried out after a breakdown or failure has taken place. This strategy is mainly implemented in less critical machines, whose breakdown has very less impact on the production line. But one of the disadvantage of running machine to failure is that, sometimes the maintenance cost are huge, if a critical and expensive component is effected. Another downside of this method is that it is not possible to predict the maintenance activities as the time of failure of the machine is unknown [51].

Preventive maintenance

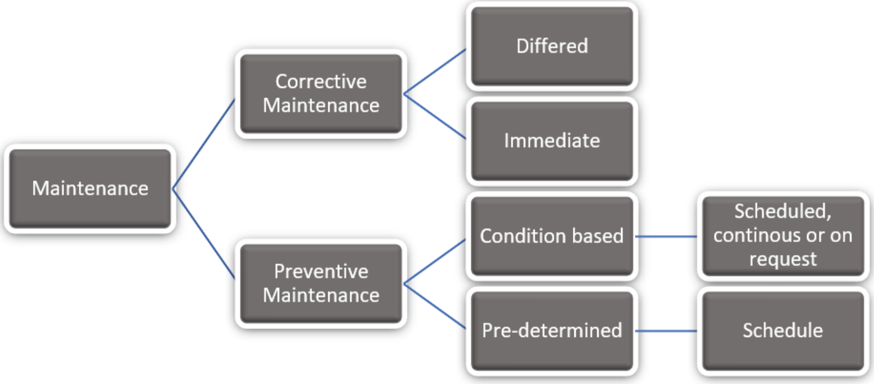


Figure 2.1: Types of maintenance [6]

In preventive maintenance strategy, maintenance activities are carried out on scheduled time intervals to prevent the breakdowns of the machine. This strategy is being used in critical machines. The advantage of preventive maintenance method compared to corrective maintenance activities is that, in preventive maintenance strategy the maintenance activities can be planned ahead. The draw back of this method is loss in useful life of the machine. Löfsten et al. [51] state that 15% to 40% of total production cost is taken by maintenance activities.

Predictive maintenance

In predictive maintenance strategy, the failure of the machine is predicted using condition monitoring tools or by data driven approach making use of Artificial Intelligence (AI)/ ML. As mentioned above when using preventive maintenance strategy a good amount of useful life of the machine is lost. By employing predictive maintenance this useful life of the machine can be utilized [51].

One of the most common way to achieve predictive maintenance is by calculating the RUL of an equipment. It is usually calculated from condition monitoring and health monitoring techniques, by physical experiments and also by data analysis. The data driven approach is based on the historical data available form the equipment and statistical models according to [70]. Si et al. [70] discuss the applicability, cost and implementation flexibility as the advantages of data driven approach for RUL compared to the traditional condition based approaches.

2.2 Current maintenance strategy at Volvo Cars

Volvo Cars currently make use of value-driven maintenance planning as the maintenance strategy in their production plant. Value driven maintenance planning is a maintenance program for the equipment in a plant system. One of the major element that makes it different from other maintenance approaches is that, their use of the experience feedback to optimise the current method. This approach is known as EBRCM (Experience Based Reliability Centered Maintenance) as discussed by Rosqvist et al. [63]. In value driven maintenance planning method, the equipment's locations are classified into maintenance classes and on the basis of the risk matrices, the critical equipments are identified [63]. An expert panel would then review the key performance indicators (KPI) and maintenance performance indicators (MPI) to monitor the goals and, investigate the cause of deviation for the goals [63].

The classification of the plant location is done on the basis of the consequence the machine will cause on the functionality of the plant. The most critical machines are classified as maintenance class 1 (MC1) and the least critical ones as maintenance class 3 (MC3) [63]. The KPIs and MPIs are used as conditions for this classifications [63]. This is because failure events are random and only the risks of failures are measurable. One of the important things to be noted that these events are continuously improved and changes takes place with the accumulation of experiences. After the classification of the maintenance task into MC1, MC2 and MC3 different maintenance strategies are to be selected such as condition based, predetermined and breakdown [63].

In this era of industry 4.0, making use of digital technology for improving the production and quality if inevitable for any company [48]. Volvo Cars being one of the most innovative companies in the world is also aiming to implement smart manufacturing.Implementation of smart maintenance is important for implementation of smart manufacturing, as smart maintenance act as the core of smart manufacturing.Prediction of the failures and performing maintenance activities just before the failure would help the company to avoid the wastage of RUL available from the equipment.This thesis project is a first step towards understanding the data and digital resource the company has to perform smart maintenance.

2.3 Smart maintenance

Smart manufacturing utilizes analysis of real-time data for creating intelligence for manufacturing and across the entire supply chain of a factory. The fourth industrial revolution and digitized machinery have produced huge amount of data which is the key enabler of smart manufacturing. The digitization in the manufacturing domain has introduced many new data sources including sensors reading, data logs and disturbance logs etc integrated by Industrial Internet of things (IIoT). The data collected from the machines can be used to visualise the operating conditions of the equipment [18]. A wide variety of strategies are being utilized to perform condition monitoring, diagnostics, prognostics and maintenance of a product or machine [17]. Smart maintenance is a broader term and there are many key enablers for this, as mentioned above the sources of quality and relevant data is a key enabler for smart maintenance and along with this the capability to perform analysis data driven decision making is also important [18].

2.3.1 Data-driven decision making

Data driven decision making is what differentiates traditional maintenance and smart maintenance. Data driven decision making is deciding scheduling maintenance activities on the basis of analysis of historical data rather than from the intuitions and experience of humans. The advancements in high computation and the extensive researches in the field of ML has enabled the capability to prediction, classification and anomalies detection which would support decision making. Researches have proved that data driven decision making is capable of providing better accuracy in decision making compared to traditional methods [61]. Some of the key enablers of data driven decision making are briefly discussed bellow.

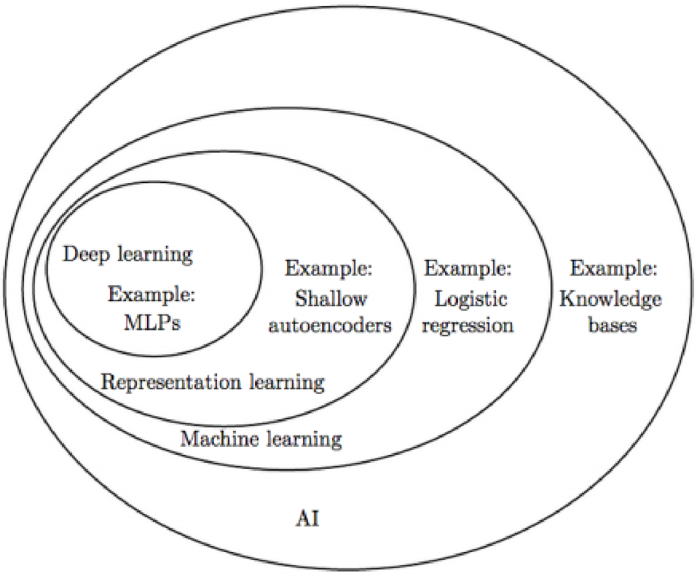


Figure 2.2: Venn diagram showing the correlation between AI, ML and Deep learning[5]

Artificial Intelligence

AI is a broad terminology that can be used for any entity which can analyse data and recognise the patterns. It is a broad team and has many subdivisions like ML, deep learning etc. Ahmed et al.[5] describes the correlation between AI, ML and Deep learning and it is as shown in Figure 2.2. Waltz wt al.[74] points out that the development in parallel computing and massively parallel programs that has the capability to learn things has enabled human level intelligence in modern day machines.

Machine learning

According to Burkov et al.[19] ML is a tool for finding the hidden patterns and trend from the available data. ML techniques make use of mathematical algorithms for optimisation, prediction, classification, anomaly detection etc [19]. The major role of a data scientist is to prepare the data set in the best possible way which is suitable for the algorithms or statistical model chosen based on the business requirement and application.

2.4 Related work

As discussed in the above sections predictive maintenance is a key enabler for smart maintenance is a key enabler for smart maintenance. High computational power and highly accurate ML algorithms makes it possible to perform advanced predictions for implementing different Predictive Maintenance (PdM) applications such as health indicators construction, anomaly detection, and RUL estimations. The related part section concentrates on the research works conducted on RUL estimations and its applications in different fields.
Cline et al. [26] implies that the maintenance actions has to be taken at the right time otherwise it will not only cause waste of usable resources but also unnecessary downtime. The research team has collected all the related historical data from a very long period of time. It is proven that application of ML improves the estimation/ prediction the failure of risky assets. Predictive maintenance using AI / ML algorithms could aid in taking up maintenance activities just in time, when they are really crucial. Scientific studies have proved that predictive maintenance reduces the breakdown time at a rate of 70 - 75%, reduces the maintenance cost by 25 - 35% and increases production by 25 to 35% [52]. These statistics motivates industry to move forward in implementing predictive maintenance using AI/ ML algorithms.

Baptista et al. [12] compares the different AI and statistical approaches for PdM, and concludes that AI approach produces a better result than statistical approach and also discusses about the ability of ML algorithms to handle high dimensional multivariate data for predictive maintenance applications in industries. Hashemian et al. [34] discusses that the biggest hurdles for implementation of a PdM project is the absence of complete database and the expertise of the professionals in the company to implement of PdM strategy.

Wu et al. [76] develops an accurate prediction model based on MS-RUL which integrates the classification-regression algorithms expecting to accurately diagnose the failures and reduce the maintenance costs. This work has mainly three steps: feature extraction from raw data,modelling and evaluation. The system's health condition is categorized based on the model and raw data, multi-stage RUL is implemented for estimation along with regression method. The difference between currently available RUL prediction methods and multi-stage RUL estimation methods is, combining ML classification and regression methods to improve the performance in terms of accuracy.

Wang et al. [75], proposed a multi-task learning approach to jointly select common features in predicting the RUL and failure types of wheel sets. To come up with a common solution for both the problems such that RUL and failure type prediction, it proposes a convex optimization formulation method to apply least square loss and negative maximum likelihood and performing feature selection over both problems.

To perform feature selection, it merges several loss functions such as L2/L1 regularization, linear and logistic regression loss. The L2/L1 regularization also helps to identify the common input variables in both tasks. This performs much better than with a single learning method in terms of prediction accuracy.

Bekar et al. [14] introduces an intelligent approach for data pre-processing and analysis in PdM based on an industrial case study. The authors used a real-world industrial data collection method and presented the results after preliminary analysis. This work was based on feature space dimension reduction and clustering of data-points with the idea of understanding the outliers in anomaly clusters. The first step of PdM implementation was the use of unsupervised ML and this formulated approach really helps to collect, analyze, describe, visualize, prepare and understand high-dimensional industrial big data. Apart from that, this study guided the transformation of the domain-expert knowledge to the ML work-flow in the data preparation phase.

Zhang et al. [80] investigated the challenges of data-shortage in run-to-failure time series by introducing the concept of time-series generation to improve RUL estimation. The speciality in this work is combining multivariate time-series and convolutional conditional recurrent - generative adversarial network (CR-GAN) there by introducing the concept of time-series generation. By this methodology, it helps to produce loads amount of data to improve the training with current RUL models. Further, the difference between cyclic and non-cyclic patterns of degradation is identified by studying the degradation types of various systems. The key factors that can be developed about this work are; suitable methodology must be introduced to preserve statistical properties of the time-series generated based on the Run-to-Failure data, experiment about Gaussian distribution is important because it affects the performance of time-series generation and lastly the unreasonable time-series should be excluded in a more knowledgeable way.

Zhao et al. [81] designed a different way of predicting RUL, based on gradient descent that builds a neural architecture search method. The suggested methodology works better than the traditional RUL method which finds more suitability in terms of cost and migration capability. Firstly, the neural architecture search (NAS) comes up as a partial remedy for automatic model construction issue. But it is a bit time consuming process as NAS is reinforcement learning and it considers the objective function as some kind of black box. They have introduced a gradient-based neural architecture search method to overcome this problem.

Liu et al. [50] discussed about a novel method, it is a generalized cauchy method in predicting RUL which has long-range dependence and fractal. It is very important to do RUL prediction considering fractal and Long-Range dependence as RUL deals with long range of time series data. The proposed RUL estimation based on Generalised Cauchy (GC) method as stochastic process as the current RUL methods are not possible to handle both factual and Long-Range dependence stochastic characteristics at the same time.

According to Saigiridharan et al. [64], the statistical models are much better than the experienced-based predictions. It presents a statistical modelling approach instead of experience-based model as the author states that it is hard to apply experience-based prediction approach in a reference population. The hurdle Gamma model and hierarchical Gamma model is used to perform model training and also to handle the hierarchical structure of the system. In regard with the feature selection process, the data from multiple sources undergoes Akaike Information Criterion (AIC). The performance of this model is evaluated using two standard statistical strategies such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Again the Akaike Information Criterion (AIC) is applied to find the best model based on the prediction performance and identified that hierarchical hurdle Gamma model performs better.

Deng et al. [27] develops an effective technique to extract forward differential features such that demonstrates the actual degradation trend in the entire life time. This paper discusses about three main issues that can experience in predicting RUL such as finding the actual degradation trend, addressing the temporal correlation among raw data and handling the highly non-linear data. After they identify these issues, the authors introduces a new concept with three steps as a remedy for above mentioned issues. First step is to proposing a long-term differential method for differential feature extraction, then introducing Fibonacci window to extend short-term features and as the final step CatBoost algorithm utilize the non-linear data to perform better RUL prediction. As a conclusion, the aircraft engine data from NASA is used to confirm this methodology experimentally.

Susto et al. [72] presents a new multiple classifier for PdM. The prediction decisions are improved by these multiple classifiers working parallel to utilize the knowledge of logistic variables in each process cycle. As we already discussed, PdM is very important concept in smart maintenance system as it helps to optimize production resources. Monitoring the status of any system is critical in implementing PdM applications. The numerical results are called "health factors" which describes any maintenance problems associated with the system. The introduced methodology has speciality which can be used to manage high-dimensional as well as censored data set. The result will be to train through multiple classification modules that has different prediction capabilities. The selection of these different training modules is done by the frequency of failure and downtime of machine parts. The results which obtained from simulation and real maintenance problem in semiconductor production are the verification for this methodology.

Amruthnath et al. [8] discusses about building an unsupervised learning model when there are less amount of historical data available. In a critical production environment, there can be situations where it is needed to build a model with very less amount of data. This paper recommends that unsupervised learning can be a better suggestion and it has experimented with different unsupervised learning methods such as Hierarchical clustering, Fuzzy C-Means clustering, K-Means and PCA T^2

statistic method and proposes an algorithm to choose the best-fit model in the end.

Paolanti et al. [57] discusses the importance if condition monitoring with predictive maintenance of machine parts concerning the optimized use of resources and continuous production. Random Forest method is the ML approach which is used in this study to implement in the predictive maintenance. The model which is applied on a real industrial application and compares the results with simulation analysis results. Data analysis tool available on Azure cloud architecture is used to perform the data analysis and the results of this experiment verifies the accuracy of this methodology.

Jahnke et al. [38] covers an entire process from data collection, feature extraction to ML model design and evaluation. This thesis work compares the components of failure detection in advance and PdM approaches. The raw data has to be cleaned, pre-processed and important features are extracted through the process called "Data acquisition and feature engineering". Following this, labelling of data has to be performed for data analysis using two strategies such as "State Driven Data Labelling" and "Residual Useful Lifetime Labelling". The right data labelling strategy must be selected to proceed for correct predictions. Later, ANN and SVM techniques are utilized in detecting failure and predicting maintenance decisions. The evaluation of the ML model performance are carried out based on true positives, false negatives, true negatives and false positives. This work focuses on Prognostic Horizon, $\alpha - \lambda$ performance, Relative accuracy and convergence evaluation techniques.

Prytz et al. [62] implements predictive maintenance solution in automobile industry by using both supervised and unsupervised learning methods. The project works with on-board data as well as off-board data in order to predict the failures of the vehicles. The on-board data are gathered with the use of telematics communication platform. The off-board data received from recorded databases such that vehicle repair history data represents the statistics of vehicle usage. With more attention given to off-board data handling, a supervised learning method is considered. The prior recorded repair data was used to label the vehicle usage patterns and statistics. The developed ML model is able to predict the vehicle maintenance based on historical data related with repair occurrences.

The research conducted by Kroll et al. [46], involves the developments in industrial plants by anomaly detection. The suggested approach here is to use timed hybrid model based on the normal working condition of machine parts in predictive maintenance. The difference of this model compared with other type of models is it converts both discrete and continuous data into individual states that corresponds with the present situation of the machine part.

Kaparthi et al. [43], refers to performing predictive maintenance based on machine parts manufacturing industry. However this system can be applied to any industry because of its scalability. And it provides an introduction about decision tree-based ML strategies as well as an experiment based on real case study. Decision tree-based learning method follows the relationship identified between the input variables. Main focus in this research is conditional inference tree statistical methodology. A confusion matrix is used to evaluate the model performance and it is done by comparing the prediction data with real data.

Fink et al. [29] discusses about the five different level of condition based and predictive maintenance and also arranges them in their level of complexity. Fink et al. [29] also states that availability of labelled data set is one of the biggest challenge faced on conducting supervised learning projects. It also puts forward two solutions for this issue. The first one is obtaining labels directly or indirectly from the health indicator and the second one is to simulating the machine in a virtual environment and getting the required data set from the digital twin. This is a great solution to start with initially and the data analysis can gradually be done with the real time data which would increase the accuracy of the predictions [29].

Abu-Samah et al. [4] make use of Bayesian based methodology for predictive maintenance application. The occurrence of faults are first determined using the Bayesian network and it is then used to derive the conclusions on the critical regions and derive patterns of failures. In this approach the predictions of failures and faults were made on the basis of these predictability index and patterns. Kumar et al. [47] Make use of polynomial regression models for autonomous diagnostics of preventive maintenance. For the diagnostics of the machines health, the different health states of the machines were Identified and then the current state of the machine was mapped to one of these states.

According to Carvalho et al. [21] the Random Forest (30%) is the most used ML method when a detailed literature review was conducted. Other popular methods were Neural network based networks which includes ANN and CNN (27%), Support vector machines (25%), and k-means (13%). During the study it was also found out that PdM were applied in wide variety of machines and there were no preferences.It was also found that the vibration signal data was commonly used in most of the PdM projects for anomalies detection [21].

Baptista et al. [13] gives an insight on combining different methods for better accuracy. In this literature they make use of auto regressive moving average (ARMA) modeling along with data driven prediction methods. The time series of failure and past schedule is fed in to the ARMA model output is fed in to the Data driven model. The data driven model make use of the features given to it after principle component analysis (PCA) to provide the final model predictions [12]. The findings showed that the integration of ARMA model with data driven methods for the prediction of fault events, was able to provide much more accurate predictions compared to the traditional life usage method.

From the extensive literature study it was clear that there were not many research works on data quality issues in real time industrial scenario and the ways to solve them and overcoming these data challenges for performing RUL estimation. Different ML model give different accuracy values depending on the fitness of data which is input in to the ML model. From the extensive literature survey it was clear that recurrent neural networks with a memory gave the best accuracy when it come to time series prediction.

2.5 ML models for time series Prediction

2.5.1 Neural networks

Neural networks (NN) are made up of layers of neurons. These neurons are the core processing unit of a neural network. The first layer of a neural network is known as the input layer, and it is the layer which receives the input. The final layer of the neural network predicts the final output. All the layers which are in between the input and the output layers are known as hidden layers and they perform most of the computations required by the network. Neurons of one layer is connected to neurons of another network through channels. The results from each node is adjusted by multiplying it by weights which assigned to each channel [66]. The inputs are multiplied to its corresponding weights and their sum is send as inputs to the neurons of the next hidden layer. Each of these neurons is associated with a value known as bias which is then added to the input sum. This value is then passed though a function threshold function called as the activation function. The result of the activation function determines if a particular neuron will get activated or not. The activated neuron transmits data to the neurons of the next layer. In this manner data is propagated through the network. In the output layer the neuron with the highest value gets fired and determines the output. This is known as forward propagation as in the as stated by Schmidhuber et al. [66].

2.5.2 Recurrent neural networks

Standard neural networks, work independently with data were as recurrent neural networks (RNN) has a memory. This enables them with the capability to work with process information like the human brain. Human's make decision based on their previous memory of events. This is not possible for normal neural networks. Where as an RNN has loops in them which enable them to pass information from one step of the network. According to Zaremba et al. this capability of having memory enables them to predict future data points and because of this they are widely used in applications like prediction of next word in a sentence and predictions made understanding the context [77]. In RNNs the modules usually has a very simple structure as depicted in Figure 2.3. These modules are repeated like a chain to from the RNN network and has a single network layer [54].

Long term dependencies

Long term dependencies is a phenomenon which is one of the major disadvantage of a normal recurrent neural network where the neural network needs to have a memory of the context. RNN has memory but when it comes to problems which has context it is important for the RNN to have a memory from the beginning to the end. According to Gers et al [32], this can be achieved by optimising the parameters of the neural network manually. According to Gers et al [32], this tiring optimisation

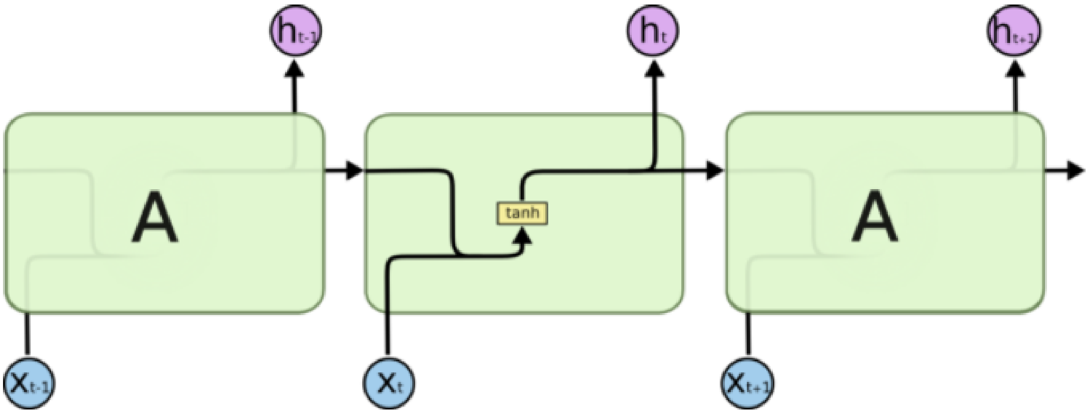


Figure 2.3: Standard RNN module [54]

can be avoided just by using LSTM [32]. Bengio et al. [15] discuss about the reasons why the optimisation is difficult.

2.5.3 Long Short-Term Memory (LSTM)

Theory

LSTM was introduced by Hochreiter Schmidhuber (1997) [36]. What made them popular was the capability to overcome long term dependencies. LSTM has four network layers and are more complex than standard RNN [33]. LSTM modules has the capability to forget and add information to a cell state and LSTM make use of gates to achieve this. The gates used in a standard LSTM module is sigmoid gates which gives output value between zero and one. A standard LSTM module is depicted in Figure 2.4.

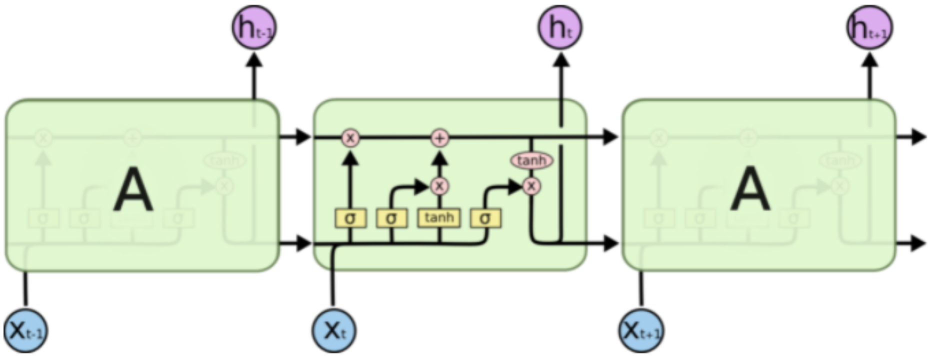
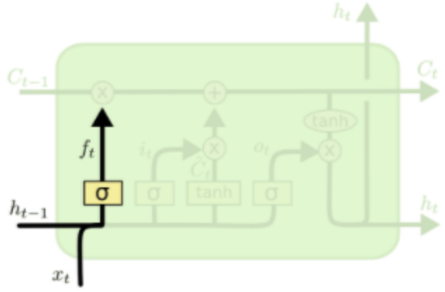


Figure 2.4: LSTM module [54]

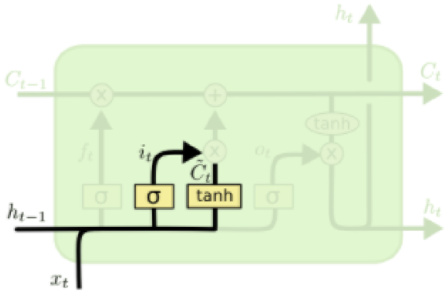
The first layer of LSTM is a sigmoid layer which decides whether to forget or take the cell state. h_{t-1} and x_t are evaluated by the sigmoid function to give a value between 0 and 1 for the values of cell state C_{t-1} [33]. This is depicted in Figure 2.5 The second layer comprise of two parts, the sigmoid function decides whether the information is stored in the cell state. New candidate value is created by the tanh



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

Figure 2.5: LSTM module First layer [54]

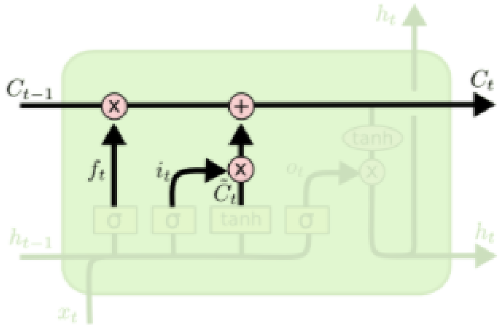
layer. This is depicted in Figure2.6.



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

Figure 2.6: LSTM module second layer [54]

Now the old state C_{t-1} is multiplied by f_t and is added to $i_t * C_t$. This is being depicted by Figure 2.7

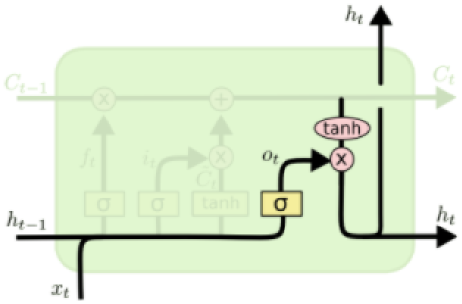


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Figure 2.7: LSTM module third layer [54]

The final layer of the LSTM decides the output of the module. The sigmoid function is used to filter out the part of the cells and then its result is multiplied by tanh function to make the values between -1 and 1. This is being depicted in Figure 2.8.
Related work - LSTM for time series prediction

LSTM is one of the most popular deep learning method for time series prediction. Hua et al. [37] applies LSTM for investigating issues traffic prediction and user



$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Figure 2.8: LSTM module fourth layer [54]

mobility forecasting. From the results it was seen that a reduction in 1% neural connections reduced the computing cost by 30%. According to Hua et al. [37] LSTM model outperform conventional models like SVR, ARIMA and FFNN .

Lim et al. [49]) make use of LSTM for prediction of future cargo transport at port for port authorities. It was found that the LSTM model is able to handle the missing values from liquid cargo techniques up on usage of various missing value handling techniques to replace the missing values. The prediction results were tested on real time data and when compared with ARIMA based and VAR prediction models. It was found that the LSTM model was having higher accuracy compared to the traditional methods [49].

Zangh et al. [79] uses LSTM prediction model for multidimensional time series prediction of gas concentration for coal mine safety management. The literature states that Adam optimization algorithm to continuously update weights and hyper parameter tuning to get the optimum number of layers and batch size was crucial for attaining the best results. The model was made to predict the gas concentration in the next time period and was compared with the real time data. The results showed that the LSTM model was much more accurate than Bidirectional recurrent neural network (BidirectionalRNN) model and the gated recurrent unit (GRU) models. The average mean square error of the prediction model was 0.003 and the root mean square error of prediction was 0.015. These matrices was highly reliable for gas concentration prediction [79].

Heidari et al. [35] make use of LSTM model for the prediction of hot water energy use in modern buildings. The literature states that the prediction is challenging for cases that include stochastic environment. The literature uses different input data sets with short span and long spans. It was observed that for prediction of highly stochastic cases, long term time-series data is required [35].

Cao et al. [20] make use of LSTM for prediction of maximum connection in telecommunication. The telecommunication networks become slow when many connections are established in same base station. So the aim of time series predicting the maximum number of connections in a base station is to tune the capacity of these stations at the peak hours to address the issue of telecommunication network slowing down. The three main evaluation matrices used in the literature was mean squared error(MSE), Mean absolute error (MAE), Mean absolute percentage error (MAPE). The Literature states that LSTM was able to forecast satisfactorily on practical cases and clearly outperforms traditional models like bagging, random forest, XG boost and TCN [20].

2.5.4 Auto Regressive Integrated Moving Average (ARIMA)

Theory

Moving Averages or Regression Analysis technique can be used for predicting series data. ARIMA model technique is used to work with data of having a trend or non-stationary as discussed by Devi et al.[28].

In 1970, this model was introduced by George E.P.Box and Gwilym M.Jenkins to be used [65] in predicting economical, sociological and industrial applications. One main advantage of predicting using ARIMA is accuracy in short-term time series data.

Based on the trend of the data series, the two types of data such as seasonal and non-seasonal can be fed into the model [23]. Non-seasonal or stationary data are defined as its statistical properties such as mean, co-variance does not change with time as opposed to seasonal time series has changing statistical properties over the time. The model predicts the future values considering the past values using linear regression.

The main components of ARIMA are as follows according to Almasarweh et al.[7];

- AR: stands for ‘auto regression’, means that the model uses the dependent relationship between current data and its past values.
- I: stands for ‘integrated’, which means that the data is stationary.
- MA: stands for ‘moving average model’, indicating that the forecast of the model depends linearly on the past values.

AR, **MA** components are commonly used in linear models related with time series stationary data and **I** component is used to make the non-stationary time series data transformed into stationary data.

These above components are included as parameters in the model. The specific integer values are assigned on this p, d and q parameters depends on the model type as shown in Figure 2.9.

ARIMA (p, d, q)

- **p** is the number of lag observations used in the model and also called as ’lag order’.
- **d** is the number of occurrences that raw observations are subtracted in making stationary data. It is also called ’degree of differencing’.

- **q** is the order of moving average or can be called as moving average window size.

Becoming zero for any of the above parameter implies the unavailability in the model.

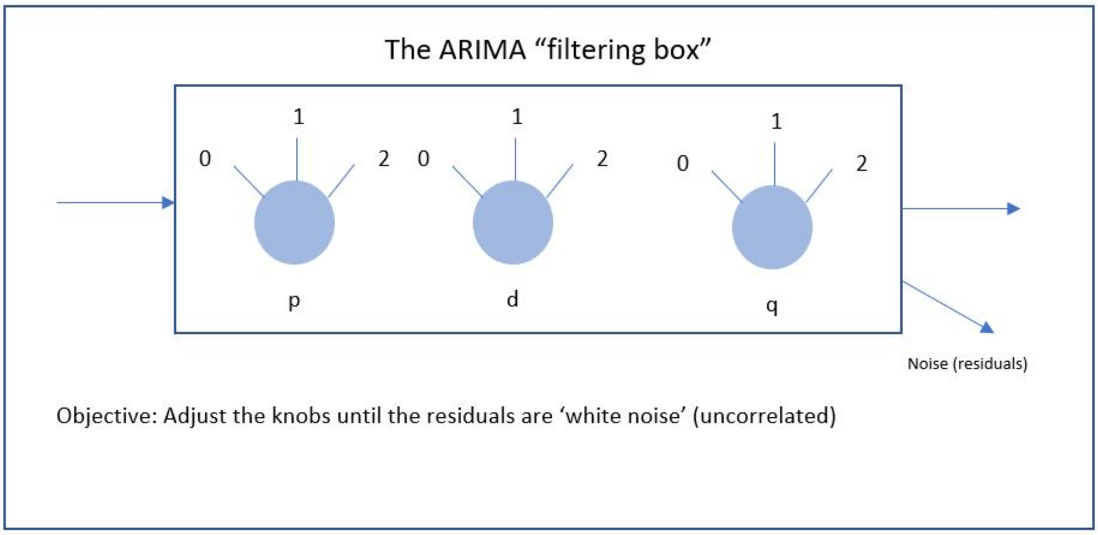

The ARIMA "filtering box"
0
1
2
p
0
1
2
d
0
1
2
q
Noise (residuals)
Objective: Adjust the knobs until the residuals are 'white noise' (uncorrelated)

Figure 2.9: ARIMA model structure

The general equations [3] in Equation 2.1 for ARIMA model is as follows. By convention, the AR terms are positive and MA terms are considered negative.

$$\begin{aligned}\hat{y}_t &= \mu + AR - MA \\ \hat{y}_t &= \mu + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 e_{t-1} ... - \theta_q e_{t-q}\end{aligned} \tag{2.1}$$

The formulation of ARIMA model can be described in five steps [24];

1. Visualise the time series data
2. Checking the stationarity of the data series if not, perform the differencing to make it stationary
3. Plot the Auto correlation (ACF) and Partial autocorrelation (PACF) statistical graphs to identify the basic parameters in the ARIMA model (p, d, q)
4. Build the model with the estimated model parameters
5. Perform the prediction and evaluate the model for future time instances

Related work - ARIMA for time series prediction

The work done in Francis et al. [30] is related with fault prediction while performing real time data analysis. Hence, it can be considered a predictive maintenance application utilizing ARIMA model to analyse the trend and predict faulty scenarios in advance. The proposed system uses sensors to record real time data for predictions. These time series data is entered to ARIMA model to trend analysis and then it passes to PCA for feature reduction. And then the new data with reduced features are fed into random forest algorithm as their ML model. The result from the training is fed into the support vector regression model for prediction phase to execute predictions of the maintenance.

Jakaša et al. [40] discusses the application of ARIMA model in predicting the day-ahead electricity prices in EPEX power exchange. The hypotheses test shows that ARIMA models is good enough to forecast day-ahead electricity prices. ARIMA models have been already applied for price forecasting but usually simple, on smaller number of observations, usually three weeks data up to one year. In this paper, the original dataset has 3836 observations (10 years). The expert modeler is used to find the best fitted ARIMA model.

The ARIMA models have been explored in literature for time series prediction. The work done by Ariyo et al.[9] presents extensive process of building stock price predictive model using the ARIMA model. Results obtained revealed that the ARIMA model has a strong potential for short-term prediction and can compete favourably with existing techniques for stock price prediction. The results obtained from real-life data demonstrated the potential strength of ARIMA model to provide investors short-term prediction that could aid in investment decision making process.

Chen et al. [24] applies ARIMA model in forecasting property crime situations and compared the results with two other model fitting techniques such as SES (Simple Exponential Smoothing) and HES (Holt two parameter Exponential Smoothing). It was a trending research field to use time series model in short term crime forecasting applications. And the results found to be more fit with the series and able to identify the turning points. However the one downside of using ARIMA was not being able to identify the fluctuations in the series data.

Peiris et al. [60] work is about forecasting the tourist arrival using seasonal series data. Seasonality of the data is verified by the HEGY test. It mentions that seasonal ARIMA model is popular in predicting tourism arrival than regular ARIMA model since tourist arrivals are highly dependent on seasonal changes. The accuracy of the model is represented by measurement criteria such as Mean Absolute Error(MAE), Root Mean Squared Error(RMSE) and Mean Absolute Percentage Error(MAPE).

The research based on forecasting of stock market performance is done by P Jansson [41] involved modeling and comparing different ARIMA models' results. It has been done for two indices in the stock exchange. They have used Expert Modeler in SPSS software to find the best-fit model for each index. This model is verified by Box-Jenkins method and AIC (Akaike Information Criterion) is used to indicate that this model is the best-fit compared with another predicted model. Ultimately, the performances of these models are evaluated by MPE and MAPE measurements using actual values and predictions. '

2.6 Evaluation of ML models

The success of the model performance can be defined by different evaluation metrics such that comparing the predicted values with the actual values. The actual values are chosen from the test data which we have from the split data set as train and test. The commonly used accuracy metrics in evaluation towards regression modelling are;

- Min-Max Error (minmax)
- Mean Error (ME)
- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Lag 1 Autocorrelation of Error (ACF1)
- Mean Absolute Percentage Error (MAPE)
- Mean Percentage Error (MPE)
- Correlation between the Actual and the Forecast (corr)

There are other evaluation criteria in classification modeling are;

- Confusion matrix
- Classification accuracy
- Precision
- Recall
- AUC

However, we concentrate on two evaluation metrics in our thesis, which are:

1. Mean Absolute Error (MAE)
2. Root Mean Squared Error (RMSE)

Mean Absolute Error (MAE)

Mean absolute error (MAE) is the simplest error metric and considers only absolute value to avoid canceling out of negative and positive residuals (errors). The average of all the residuals is also taken. The equation for calculation of mean absolute error [22] is shown in Equation 2.2.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y - y'| \tag{2.2}$$

Root Mean Squared Error (RMSE)

RMSE can be described as the standard deviation of the errors in the predictions. This error is the difference between data points with the regression line and how well it is distributed around the best fit line. The equation for calculation of RMSE [22] is given in Equation 2.3.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y - y')^2} \tag{2.3}$$

3

Methodology

This section focuses on the scientific concepts, selected approach and the data exploratory analysis that has been used. The initial planning were started to follow the standard CRISP-DM methodology which is having Business understanding, Data understanding, Data preparation, Modeling, Evaluation and Deployment as the main phases, Schröer et al.[67]. However for practical reasons, this work represents the modified CRISP-DM methodology by putting more attention in data acquisition phase.

3.1 Enhanced CRISP-DM methodology

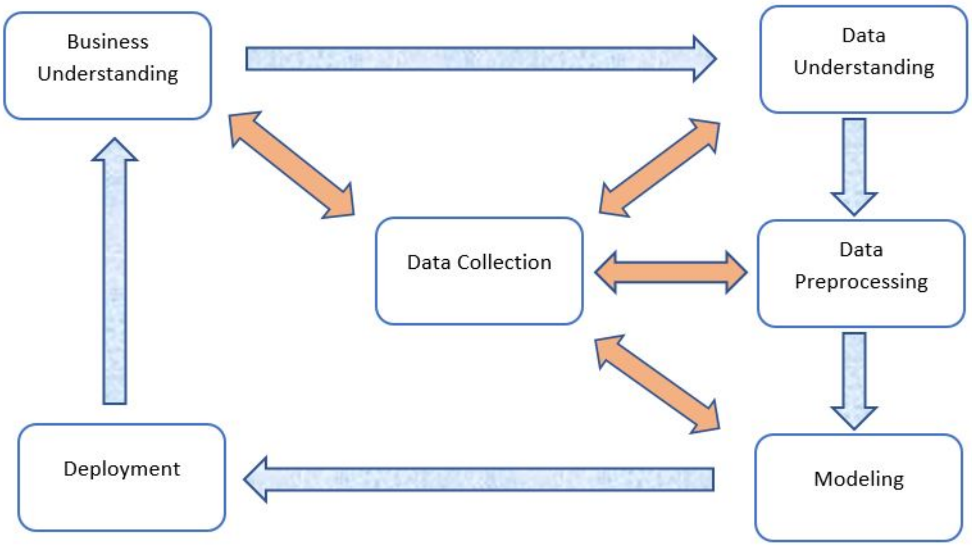


Figure 3.1: Enhanced Methodology - adopted from the original CRISP-DM [67]

The project starts with business understanding phase. It is essential for any project to understand customer needs. This is kind phenomenal because all the company objectives and goals have to be formulated. The data understanding part to identify the relevant data sources, collect data and evaluation of data comes secondly. The third main step would be data preparation and it takes the most of the time out of all the phases in the methodology. Once the sufficient amount of quality data is ready, the next phase is modeling and it is the shortest phase. The trial and error method should be performed to select the best possible model. In the training step, the model performance with the business requirements will be assessed. In the final step, the deployment is performed to conclude the work. The following content discuss about the main phases of CRISP-DM structure and how it is modified in the present investigation.

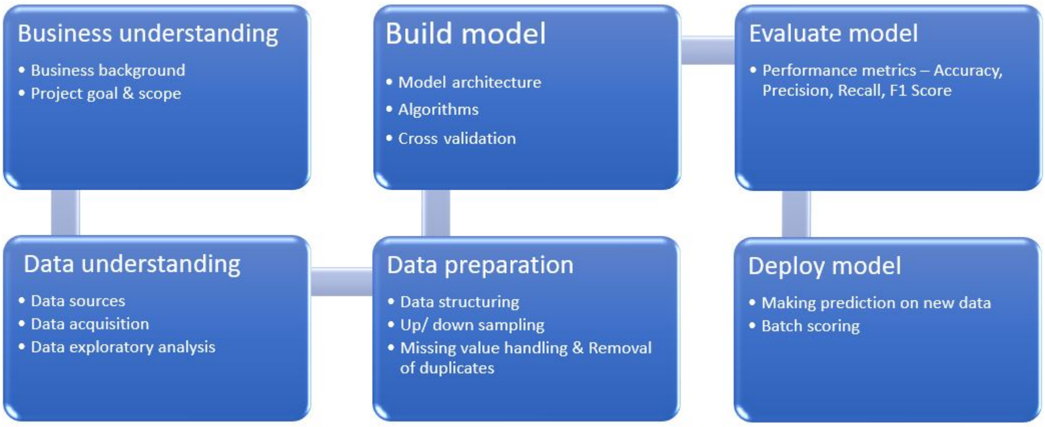


Figure 3.2: Overall methodology of the project

As portrayed in the figure, all the phases from business understanding to modelling is connected to data collection phase since historical data was not available and it has to be collected during the project duration.

3.2 Business Understanding

The first phase of the methodology concentrates on the company requirements and objectives from the business point of view. It is necessary to learn the company goals, objectives and what are their expectations clearly before we set up the project scope and plan. The initial plan and the scope has to be decided based on the available resources and time. The time limitation is about six months and few discussion sessions were carried out to understand the data availability. Hence, understanding the business objectives, assessment of available resources, and finalising the project plan is done in this phase.

The software tools incorporated in the project are identified to handle the work for data analysis, visualization and data modelling with the help from supervisor and team.

Eventually this define phase, represents the fundamental work in succeeding any data science project.

3.3 Data Understanding

The data understanding phase starts with data collection to perform data analysis. The main functions are data familiarization, data quality identification and data visualisation to gain insights. The objective of data understanding is to find hidden patterns and parameters that might affect the results. The exploratory data study includes handling missing data, finding relations among data and creating synthetic data. Apart from that, data quality issues must be properly addressed in this phase. Since this project is at the initial stages and historic data was not readily available, we were able to involve in the data collection phase as well. In this project, the discussions with the maintenance team was performed in order to record new data. Therefore data understanding part was divided into two sections, data collection and data analysis. The software tools such as Power BI and Microsoft Excel, and Jupiter notebook (python programming language) were used to perform visualisations and finding patterns.

3.3.1 Data Collection

The relevant data sources and corrected data has to be identified to proceed with the project. It takes considerable amount of time to scrutinize the data before the main phase. It is required to discuss with the maintenance team and also suppliers in data collection to verify the data sources. The maintenance team has the relevant experience and knowledge on robot as well as machine parts of the system. Therefore it is necessary to contact the data company before accessing the suitable data sources. Collected data comes in the form of csv or excel files and all of them are referenced with timestamps. The issues can occur when the timestamps from different data sources differs, and it has to be resolved by a sampling procedure. It is possible to compare and visualise the data from different sources.

3.3.2 Exploratory Data Analysis

'Without domain knowledge Exploratory Data Analysis (EDA) has no meaning, without EDA a problem has no soul' [45].The EDA can be performed once the data is available. It is necessary to check whether the data set is balanced and if not relevant actions such as K-fold cross validation, re-sampling with different ratios, choosing appropriate evaluation metrics are to be taken [44]. The main tasks of exploratory data analysis are data cleaning by handling missing values, duplicate values and then data visualisation in terms of graphs or statistical parameters. The next step is understanding the relationships between data distributions through the trends and patterns. The focus should be on data distribution understanding and identify the patterns[45]. The most commonly used EDA tools are median, box plot, residuals and running median.

3.4 Data Preparation

Data preparation is the basic step in data analysis and it requires most of the time in data engineering process. Preparing data to feed into the model will be the major task in this phase. After collecting all the relevant data a new data set must be prepared by using strategies like synthetic data generation and transformation. Data integration from different sources, dimensionality reduction and transformation of data are part of the preparation process.

The reasons for data preparation is to alter the issues due to missing data and for formatting the data into a suitable ML format [78]. The dataset obtained after preparation will be clear, quality ensured, error-free, complete and obviously smaller in size compared with the original dataset.

3.5 Predictive Modeling

The Modeling phase is about trying out different modeling techniques and comparing the outcomes for selecting the best model with optimum hyper parameters. There is a connection between modeling and data preparation such that data issues can be identified during the modeling stage. A correct model type can be chosen depending on the available data in preparation phase. The expected outcome of the model is also considered during the model selection. Once the data set is finalised, the selection of model design and correct algorithm is chosen based on the data available and also the output requirement of work. LSTM and ARIMA models are suitable for this work as it can handle the time series data efficiently. An introduction and theory about these models are included in the literature review. The software tools used in this phase are Jupyter Notebook with Python programming.

3.6 Evaluation

Building the model is not the final phase of the project. The generated model is improvised until the model is best suited. The Evaluation phase includes comparing the model results with the actual data and suggest evaluation metrics such as F1 score, Accuracy. Also, the outcomes of the model has to be evaluated with the business objectives of Volvo Cars. The accuracy of the model can be represented using different evaluation metrics and those error values should be minimized. It is important to discuss the utility of prediction results to achieve the business goals which was defined in the beginning stage.

3.7 Deployment

The best generated model and optimal results obtained should be presented to the organisation in order to ensure the benefits of the results to the organisation. The deployment phase includes implementation of the results in the company and making it a positive impact. The results will be formulated in to a scientific report.

4

Results

This chapter discusses the insights and results gained during the different phases of the thesis work, following CRISP-DM methodology. Business understanding section portrays the business objectives of the thesis work. The data collection process is describes in data collection section. The different sources of data, data quality analysis and EDA is described in data understanding section. Data preparation section describes all the data preprocessing performed in order to make the data suitable for the ML model. Model design section discusses the test designing and all the models which were used for prediction. The evaluation of the ML models and the results are presented in the evaluation section.

4.1 Business Understanding

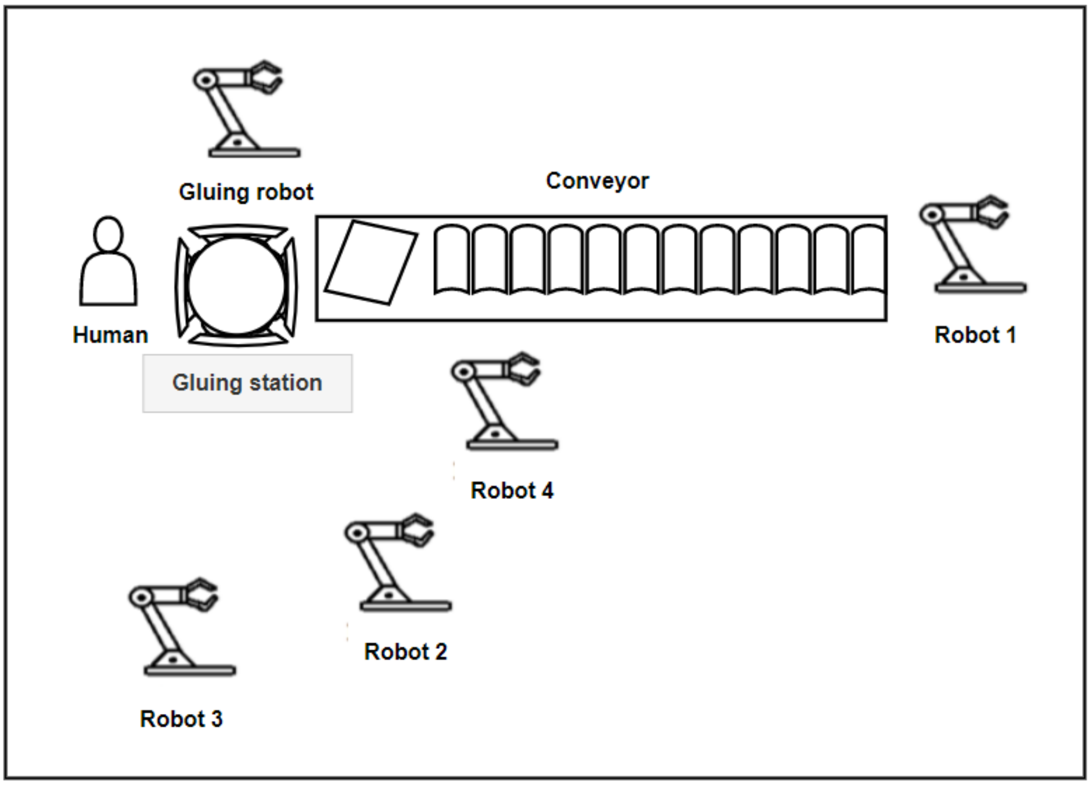


Figure 4.1: Workspace

The project was based on the gluing station in an assembly line of Volvo cars Torslanda factory. As it is shown in the Figure 4.1, the gluing robot is connected to two other sections in the station. The work pieces comes from a rotating conveyor which involves a human to locate the parts. After the gluing is done, the parts are kept on a separate rack/ conveyor which proceeds to the welding operation.

The main components of the gluing station are composed of two main sections: Robot and the Adhesive part. The robot comprised of the robot manipulator, controller and media panel. The robot established in the station belongs to the family of ABB IRB6700.

- *Robot manipulator*

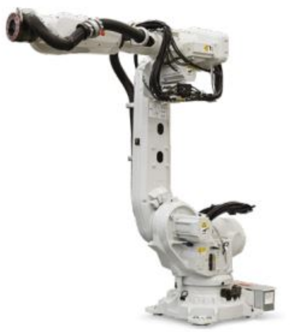


Figure 4.2: Robot Manipulator [2]

Parameter	Value
Reach (m)	2.6-3.2
Handling Capacity (kg)	150-300
Center of gravity (kg)	300
Wrist Torque (Nm)	900-1900
Supply voltage	200-600 V, 50/60 Hz
Energy consumption (kW)	2.8
Robot base (mm)	1004 x 720
Temperature in operation (C)	5 - 50
Relative humidity	Max. 95%
Noise level (dB)	Max. 71
Axis 1 rotation	Default: ±170° Option: ±220°
Axis 2 arm	-65°/+85°
Axis 3 arm	-180°/+70°
Axis 4 wrist	Default: ±300°
Axis 5 bend	±130°
Axis 6 turn	Default: ±360° Max. rev: ±93.7

Table 4.1: Specifications of IRB6700 [2]

This particular robot family has payloads from 150 to 300 kg and 2.6 to 3.2 meters, also floor mounted. And adhesive system comprised of glue gun, doser, docking station, Mesamoll pump, controller, pump and media connections.

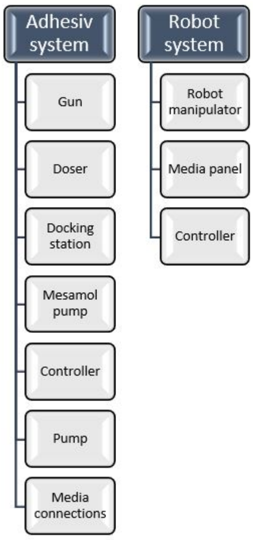


Figure 4.3: Components of the Gluing Station

The project focuses on the adhesive components as it requires higher rate of maintenance services out of all the parts in the gluing robot system. Hence we only consider the data from doser and gun. The entire material application system is called T2X2 BiW. Below Figure 4.4 shows the doser and gun as a combined equipment.

- *Doser & Gun*

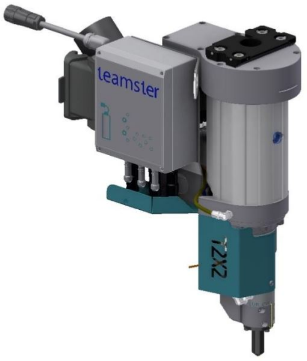
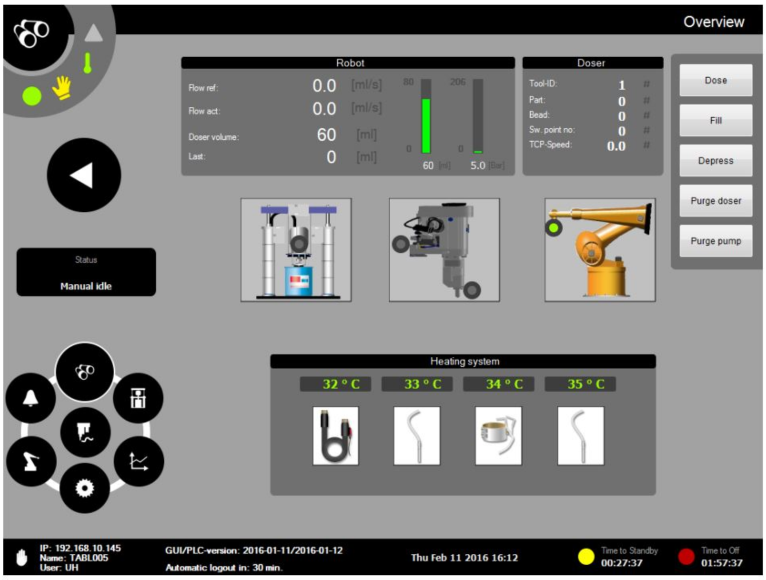


Figure 4.4: Doser and Gun [1]

Doser and Gun combined tool are used to apply the sealing material on the work piece. The doser feeds the sealing material to the gun with required pressure and flow rate. There are pressure sensors mounted on the gun and also on the doser. And temperature sensor is mounted on the gun. The doser is robot mounted and material filled by a docking station. This system even has a controlling computer (PLC - Programmable Logic Controller) and GUI (Graphical User Interface). T2X Host is the software tool to manage the settings and parameters of T2X2 system and the overview can be seen in Figure 4.5.



Flow ref: 0.0 [ml/s]
Flow act: 0.0 [ml/s]
Doser volume: 60 [ml]
Last: 0 [ml]
60 [ml] 5.0 [Bar]

Tool-ID: 1 #
Part: 0 #
Bead: 0 #
Sw. point no: 0 #
TCP-Speed: 0.0 #

Status
Manual idle

32 ° C 33 ° C 34 ° C 35 ° C

IP: 192.168.10.145
Name: TABL005
User: UH
GUI/PLC-version: 2016-01-11/2016-01-12
Automatic logout in: 30 min.
Thu Feb 11 2016 16:12
Time to Standby 00:27:37
Time to Off 01:57:37

Figure 4.5: Graphical User Interface of T2X2 [1]

T2X-Log software is used to record the temperature and production log data and the alarm event data (alarm log) can also extracted from this same software. There are three alarm and warning categories such as 'A', 'B' and 'D'. Category 'A' refers to immediate stopping alarms, category 'B' for critical error alarms and category 'D' refers warning and information.
It is considered that the occurrences of category 'A', 'B' or severe alarms implies tendency to a failure. Therefore, one task of the project is to predict these severe alarms beforehand. The other task is to predict the parameters of the doser considering the doser volume and pressure.

4.2 Data Collection

The process of data collection occurs from various data sources mainly from PLC, Teamster logs, Maximo, Axxos as well as SQL database.

Teamster Logs

The temperature, production, doser and alarm logs were extracted from Teamster system. Temperature logs contains sensor data from chamber, gun, hose, dock valve, plate and body with the timestamp. Doser log contains doser volume, servomotor moment, flow reference and pressure. Production log is about the production details such as cycle time and volume per bead, per cycle. The alarm event data can be extracted from alarm log. Alarm category, alarm ID, timestamp and description are included in the alarm log csv file.

SQL (Structured Query Language) Data

We had the opportunity to access their database server of PLC records during the initial phase of the project and found out that there is a similarity between these data with other data sources.

Maximo Data

Maximo is the current maintenance plan which is used in the factory to perform their maintenance work. It is really valuable resource to understand about the present situation in terms of maintenance.

Axxos Data

Axxos data provides information about the disturbances in production of the facility.

4.3 Data Understanding

As mentioned the data was collected from different sources, but all of the data collected from different sources were not suitable for our study. Since there were no historical data available the data collection and understanding process was iterative. As mentioned in section 4.2 data were collected from the teamster log, Axxos data set, Maximo data set etc. A data quality analysis was done and it was found that all the numbers were in float format, but the date and time was in object format which had to be converted. This section is focused on understanding the contents and parameters of the extracted data.

Teamster Log

The data from the teamster Log are the logging's from the temperature, pressure, volume sensors and the alarm log. There logging are extracted as four different csv files. Each of them are explained briefly in the following sections. For this project we consider data from 26 August 2021 to 23 June 2021. The general information regarding the teamster data log is described in Table 4.2

SI no	Data source	Rows	Columns
1	Temperature Log	83549	7
2	Doser Log	8274705	6
3	Alarm Log	169964	4
4	Production	385575	9

Table 4.2: Teamster data sources

Temperature log

The attributes of temperature log are the temperature measurements from the chamber, gun, hose, dock valve, plate and the body. The temperatures are measured in Celsius scale (°C). The logging are recorded in every minute and mentioned in Table 4.2 there are 83,549 rows of data points. On a discussion with the machine experts from the OEM it was clear that one log entry in one minute was the most suitable sampling size for the temperature log as temperature variations are minimal under a minute. A sample temperature log is being shown in Table 4.3.

Timestamp	Chamber	Gun	Hose	Dock Valve	Plate	Body
2021-04-26 00:00:31	39.8	43.0	45.1	48.0	45.0	45.1
2021-04-26 00:01:31	39.9	42.9	45.0	47.9	45.0	45.1
2021-04-26 00:02:31	40.0	43.0	45.0	48.0	45.0	45.1
2021-04-26 00:03:31	40.1	43.1	45.1	48.0	45.0	45.0
....	...	...	...	...	...	...
....	...	...	...	...	...	...
2021-06-23 00:29:51	46.8	43.0	45.0	48.0	45.0	45.0
2021-06-23 00:30:51	46.8	43.0	44.9	48.0	45.0	45.0
2021-06-23 00:31:51	46.8	43.1	45.1	48.0	45.0	45.0
2021-06-23 00:31:51	46.8	42.9	45.0	47.9	45.0	45.0

Table 4.3: Sample temperature log

Doser log

Doser log contains the parametric data regarding the flow of glue, the pressure exerted and also the toque of the motor. Table 4.4 depicts the unit in which each of these attributes are measured. Looking in to the doser log data it was seen that some time there are 500 and more logging under a minute and some other time it is 180. From Table 4.2 it is clear that for the same time span doser log has 8274705, which is significantly higher than temperature, alarm and production logs. From Table 4.5 we can see that the resolution of timestamp of the doser log is very high as it is of millisecond level. A logging take place whenever there is an activity and the activities varies with time. This is the reason why the doser log has uneven sampling and a very high number of logging compared to all other logs.

Attribute	**Unit**
Flow reference	ml/s
Actual flow	ml/s
Pressure	bar
doser volume	ml
Torque	Nm

Table 4.4: Units of the attributes in doser Log

Timestamp	Flow referens	Acctual flow	Pressure	Doser voluem	Torque
2021-04-27 00:00:07.840	1.5	0.0	20.0	41.6	15.3
2021-04-27 00:00:07.890	1.2	-0.4	18.3	41.6	31.8
2021-04-27 00:00:07.940	3.6	2.1	15.2	41.6	53.2
2021-04-27 00:00:07.990	4.6	4.8	29.8	41.3	45.0
....	...	...	...	...	...
....	...	...	...	...	...
2021-06-23 00:33:20.287	1.8	0.2	41.5	64.4	54.5
2021-06-23 00:33:20.337	1.7	0.8	41.5	64.5	52.3
2021-06-23 00:33:20.387	1.8	1.4	41.7	64.2	51.6
2021-06-23 00:33:20.437	1.6	1.7	41.8	64.1	49.1

Table 4.5: Sample doser log

Alarm log

The alarm log from the teamster data log contains all the alarms that has occurred in the gluing machine. From Table 4.2 we can see that there are 385575 alarms occurring in the time span. But most of this alarms were depicting the completion of actions and also resetting of alarms. The attribute Priority from Table 4.6 depicts the severity to the alarms. Prepossessing of data is required to delete all the non relevant alarms from the log in order to achieve the required severe alarms for predictive maintenance.

Timestamp	ID	Name	Priority
2021-04-26 00:00:42.336	10	Doser1 volume is low	D
2021-04-26 00:00:44.355	-1	Alarm reset	Ack
2021-04-26 00:00:47.995	-1	Alarm reset	Ack
2021-04-26 00:00:50.438	-1	Alarm reset	Ack
....	...	...	...
....	...	...	...
2021-06-23 00:29:41.612	-1	Alarm reset	Ack
2021-06-23 00:29:43.052	-1	Alarm reset	Ack
2021-06-23 00:29:45.952	-1	Alarm reset	Ack
2021-06-23 00:29:48.694	-1	Alarm reset	Ack

Table 4.6: Sample Alarm log

Production log

The production log enter logging when each bead of the gluing takes place. This log seems to have use in productivity analysis and seems to have little importance for our study.

Maximo data

Maximo data log is the Maintenance logging of production log. At present Volvo cars make use of scheduled maintenance in their body shop. Maximo data log contains the scheduled time and also the time in which the maintenance activity took place. It had both scheduled maintenance activities and also unplanned maintenance activities. But the problem with making use of this data for supervised learning is that the logging were manually done by the workers in the factory and there were variations added due to the manual input. It was advised by the machine expert from the company as they doubted the quality and accuracy of this data set, and many mismatches were found during the data understanding phase.

Axxos data

Axxos data set had all the disturbances that took place in the factory but it doesn't had any attributes logged which is of interest for this project.

So in conclusion the useful data set for this project was the temperature, doser and alarm logging from the teamster logging.

High resolution alarm

The high resolution alarm log was recorded and stored in the data base. The data has specifications of alarms which was triggered in the entire station. But this data set doesn't had the attributes required for the project. The alarm log from the teamster log was more relevant to the robot cell which we are analysing in this project.

4.3.1 Data quality analysis

In this section we try to analyse the quality of the data that is to be used for failure prediction. There were many quality issues with the data available. Many data prepossessing methods were essential to make the data suitable to be fitted in to a ML model. Historical data was not readily available but the data sources which were required was accessible during the thesis work. But there were some time delay in accruing and mainly it was the waiting time for the data to be produced from the machine, so there wasn't a punctual availability of data. Authorisation for the data required for the study was given. So in total the availability of the data was limited.

From the available data sources most of them were usable. Some of the unusable data were axxos data and the maximo data which were not utilised during the study. Most of the data sources were automatically recorded and hence had a high credibility. In conclusion the data usability was good. The data received from the sources were accurate, and there weren't any missing values or errors in the data which was available, which made it consistent and complete. The data set was very big which will make it difficult to manually audit it, but it is easily auditable visually. Since all the necessary criteria for a reliable data, we could conclude that the data available was reliable for a ML project.

When it comes to relevance of data available the temperature and doser data available was relevant for predictive maintenance ML analysis. But at the same time relevant data for labelling the data-set for a supervised learning was missing. This was because the maintenance of the less accuracy of the data from the maintenance log (maximo data). All the data was available as csv files and was easily readable using python and also with power BI. For files from some data sources the headings were missing which was then added at preprocessing stage. But overall the structure and readability of data was really good.

From the data quality analysis and data understanding sections it is clear that quantity and quality of data available is sufficient to build an ML model for prediction of failures. This answers the first research question (*RQ1*) . There is some unbalances in the data, but which is due to the nature of the data produced from the machines and is reasonable. Table 4.7 gives a summery of this data quality analysis.

Availability	Usability	Reliability	Relevance	Presentation Quality
Accessibility ✓	Definition/ Documentation ✓	Accuracy ✓	Fitness ✓	Readability ✓
Timeliness X	Credibility ✓	Integrity ✓		Structure ✓
Authorization ✓	MetaData X	Consistency ✓		
		Completeness ✓		
		Auditability ✓		

Table 4.7: Data quality analysis summary

4.3.2 Exploratory Data Analysis

In this section we make use of data visualisation techniques to investigate the quintessential patterns hidden inside the data set. The variation of temperature and doser values with change in time is visualised to analyse how these parameters change with respect to time and also to have a deeper understanding about the working of the machine. The severe alarms are plotted against temperature and doser data to analyse the patterns before and after break up and to investigate if these patterns are repeating whenever these alarms are occurring.

4.3.2.1 Teamster data visualisation

Figure 4.6 shows the plot between temperature and time. From this figure we can infer that the temperature when the machine is working normally is between 40 °C to 50 °C. One of the important inference that we get from this plot is that the temperature logging takes place even when the machine is not working. From Figure 4.6 we can see during week ends when the machine is not working the atmospheric temperature is logged. Also it was seen that the chamber temperature was constantly increasing reaching the weekends every week.

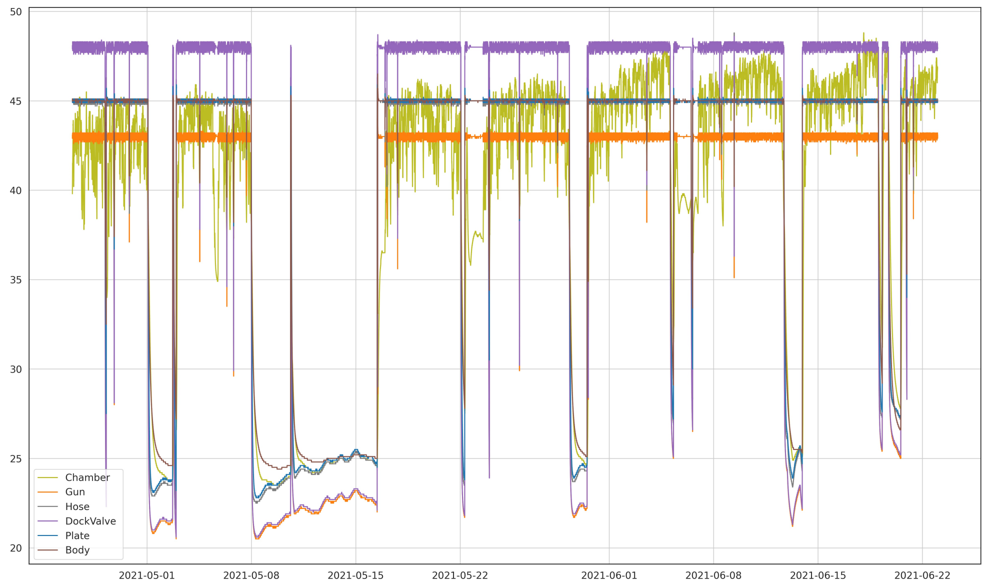


Figure 4.6: Temperature vs time plot

Since there is a huge number of data points doser vs time plot for the entire data set was not readable this is being shown in Figure A.1. so in order to investigate the change in patterns when in doser values during normal working conditions a smaller time frame was visualised. Figure 4.7 depicts the doser parameters from 2021-06-16 12:56:30 to 2021-06-16 12:59:30. From this plot it is seen that the doser parameters are logged only when the machine is working. This is a line plot and there are no values in between the two cycles, although it looks like it, since it is a line plot and not a scatter plot. After each gluing cycle there the doser volume is restored back to the initial level and during the gluing process the doser volume is constantly decreasing. Negative values were observed for actual flow and reference flow which denotes that there is a flow of glue in opposite direction. When consulted with the machine experts it is found that it is a normal working phenomenon of this particular gluing machine.
When looking into a much larger time window of 1 hour it was seen that this pattern is being repeated though out the operating time of this gluing machine. This is depicted in Figure 4.8. From this plot we can clearly see that the doser is being refilled by glue after every second in gluing cycle and the normal working pressure range is from 7 bar to 60 bar.

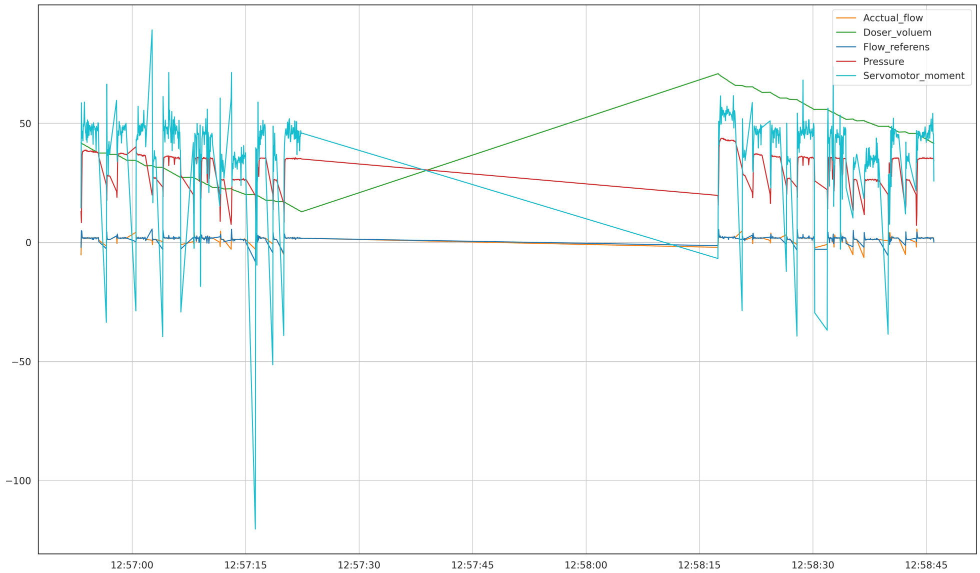


Figure 4.7: Doser vs Time plot for 3 minutes time frame

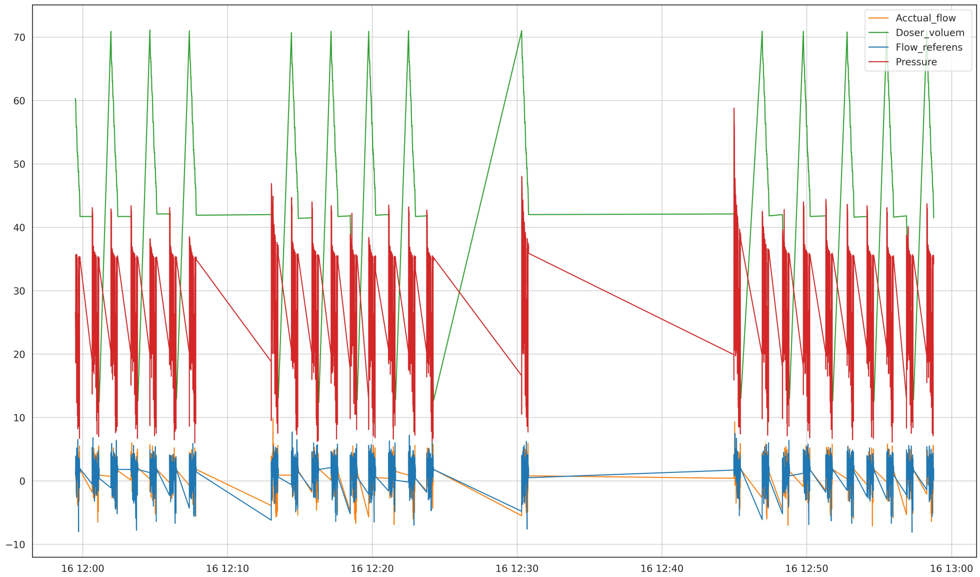


Figure 4.8: Doser vs Time plot for 1 hour time frame

In section 4.4 we discussed about the prepossessing done on the doser data in order to shrink the data to minute time frame. Figure 4.9 depicts the doser plot from 2021-06-16 18:00:00 to 2021-06-16 23:59:30.From this plot we can see that the mean, minimum and maximum of actual flow and reference flow dosent vary much through out the operating range.
But when analysing the entire data visually it was difficult to see anomalies before the real breakdowns at the plant from these visualisations. Further visualisations are required with temperature and doser values and alarm Id to analyse their relations.

The analysis for the Teamster data needs to visualise the entire series data con-

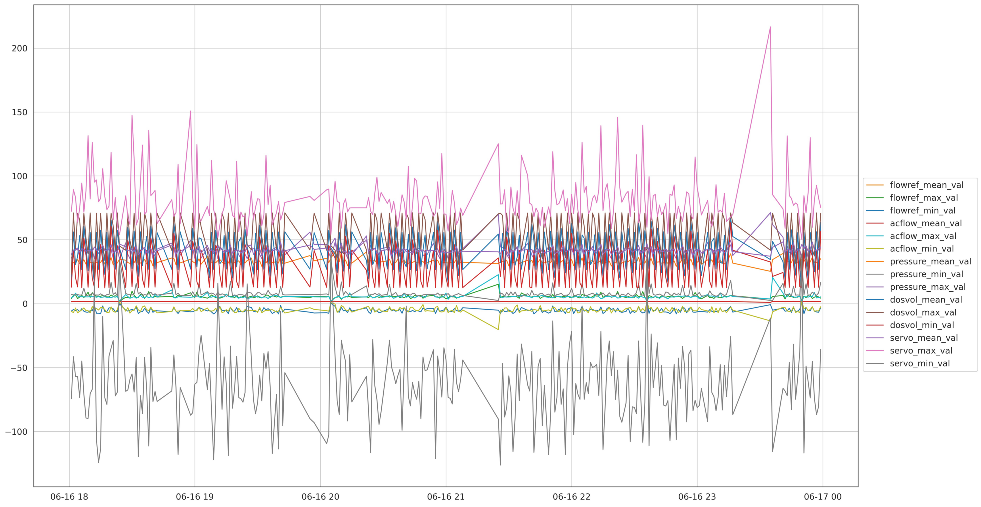


Figure 4.9: Doser plot after prepossessing

sidering their minimum, maximum and median values. The following Figures 4.10, 4.11, 4.12, 4.13 and 4.14 represents the minimum, maximum and median values of each parameter respectively. These plots are a good contribution in identifying the functional misbehaviour when we share this with the maintenance team and found out some parameter variations are not due to failure but manual changes done by production team.

The Figure 4.10 is the plot regarding flow reference parameter in the doser log. And x-axis, y-axis represents the timestamp and parameter value respectively. The parameter value can be either median, minimum or maximum value. All the values were calculated for every minute since it is the common method of analysing in this study. Therefore, we calculated the median, minimum and maximum per every minute for all the parameters.

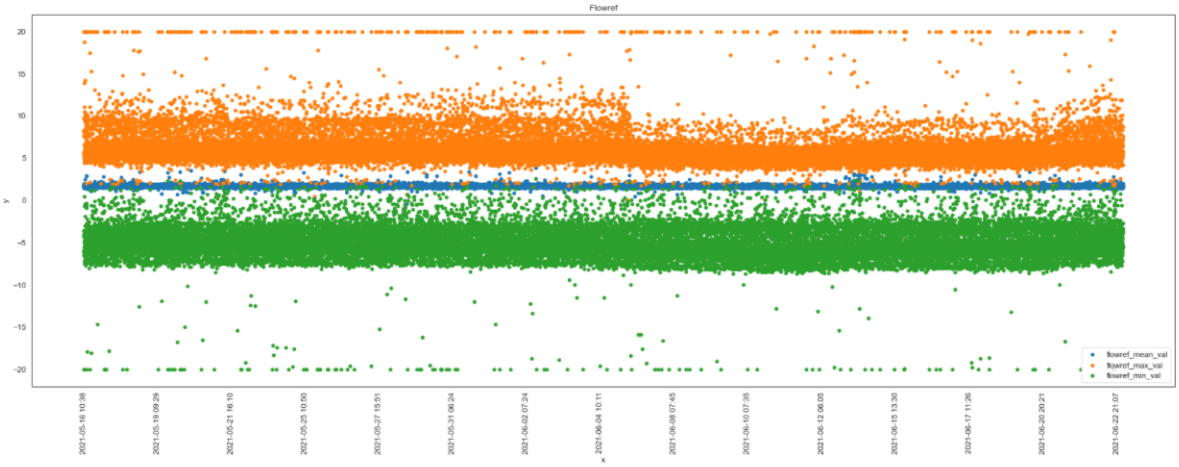


Figure 4.10: Flow Reference from Doser log

The following Figure 4.11 show the actual flow statistical data plot from the doser log throughout the time period. And it can be seen that it is always constant and no variation occurred. Hence it will not add much contribution to our study.

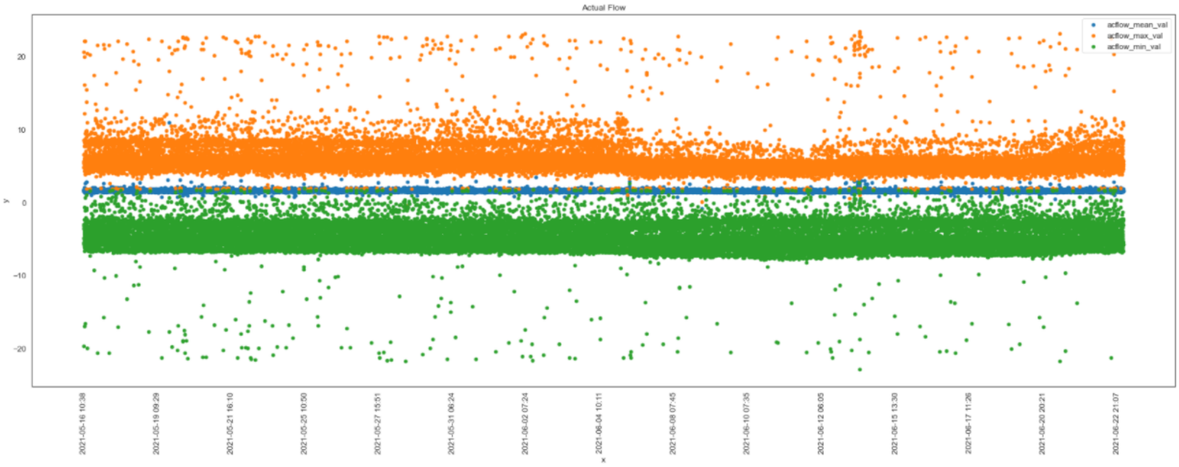


Figure 4.11: Actual Flow from Doser log

Following Figure 4.12 shows the pressure plot variation and the it is possible to see a slight decrease in the value. Apparently this is due to some manual change of machine part done to mitigate another issue not because of failure.

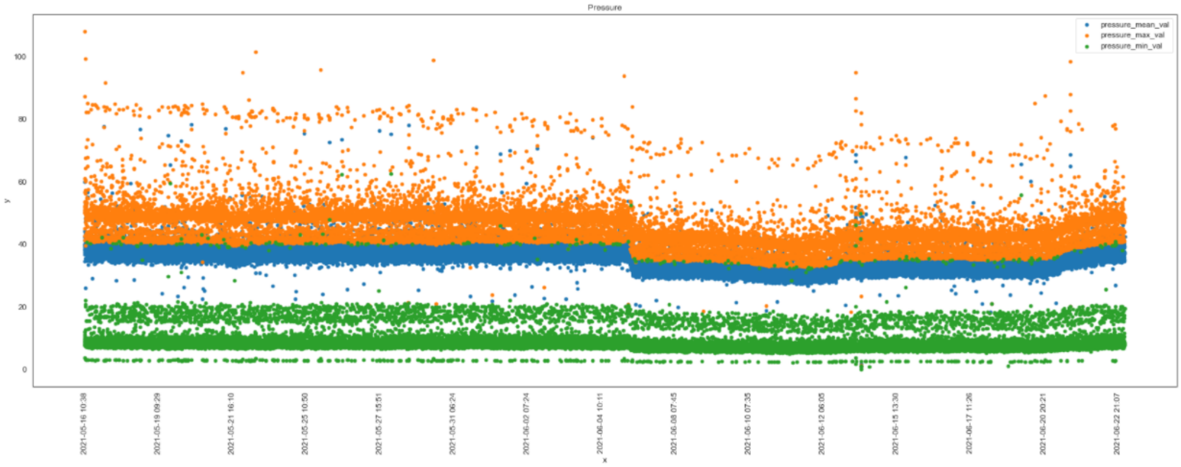


Figure 4.12: Pressure from Doser log

The doser volume variation is depicted in the Figure 4.13 and mean value is seen to be widely spread out rather than densely populated.

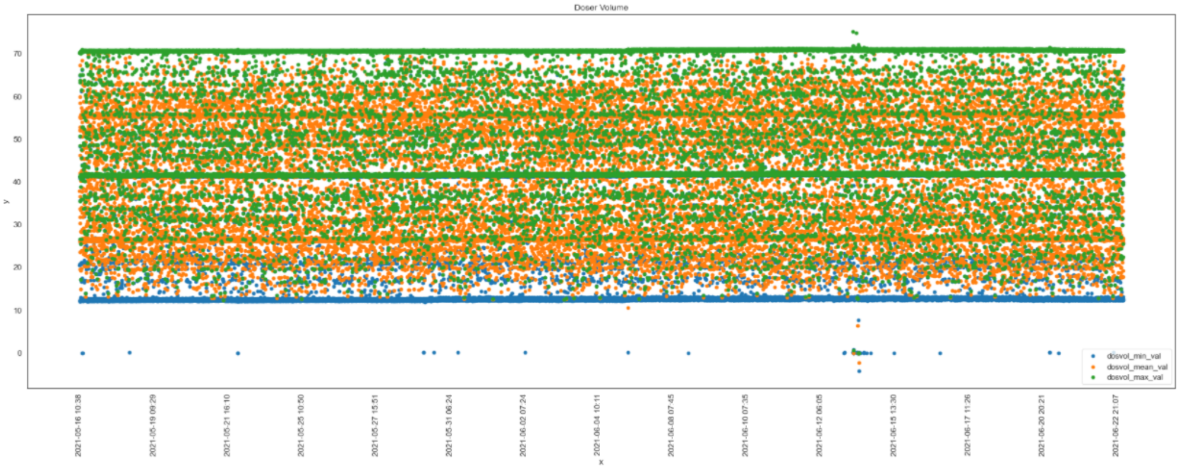


Figure 4.13: Doser Volume from doser log

The torque of the servo motor is represented by Figure 4.14 and is quite similar to the behaviour of actual flow and flow reference. It also shows almost the constant value throughout the time period except a slight change in the mid way and we are not completely certain about the cause for this indication.

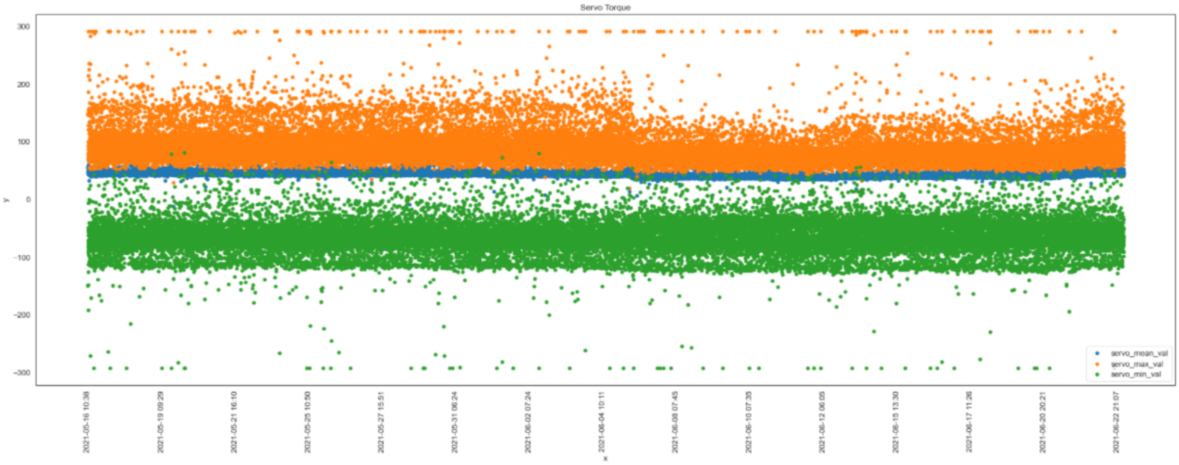


Figure 4.14: Torque from doser log

4.3.2.2 Correlation matrices

Correlation coefficients

The statistical relationships or the correlation among the features in dataset can be identified by the correlation matrices. Finding the correlation between parameters gives an insight about the dependencies between these parameters. Correlation coefficient is a statistical parameters which indicates the relationship between two features as discussed by Taylor et al.[73]. Different correlation coefficients can be calculated based on the information about relationship and distribution of the variable. Pearson and Spearman are two important correlation coefficient which can be used in Python programming.

- Pearson Correlation Coefficient
  This parameter is a measure of linear relationship between two variables and results a value between +1 and −1 in which +1 represents 100% positive linear correlation, −1 represents 100% negative linear correlation and 0 represents having no correlation between the two variables. Pearson coefficient can be only used to evaluate linear relationship such that exists a proportional difference between variables. If it is not possible to detect a linear relationship, Pearson coefficient results a zero value as discussed by Nahler et al.[53].

- Spearman Correlation Coefficient
  Spearman correlation is the statistical measure of dependence in rank correlation of variables. And evaluates the monotonic correlation which based on ranks of variables more than raw data. The rate of change in monotonic relationship is not constant as it is available in a linear relationship as discussed by Artusi et al.[10].

It is important fact to understand that any of these coefficient cannot be used in detecting non-linear relationships. Nonetheless, this project is focused in understanding the direct relationships between the input and output variable. Therefore, these coefficient criterion were used in evaluating the correlation matrices for different features in this project.

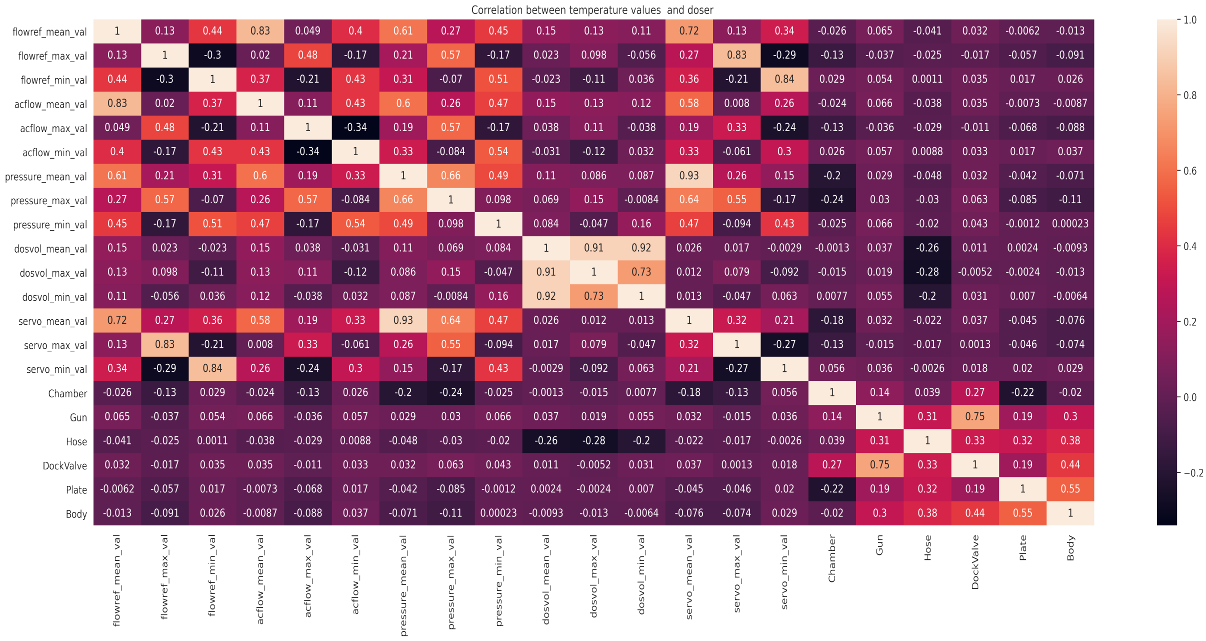


Figure 4.15: Correlation between temperature and doser logs

Figure 4.15 depicts the co-relation between temperature and doser log values. From the co-relation matrix it is clearly visible that there is no co-relation between the preprocessed doser and temperature values. But strong corelations was visible between some of the attributes with in the doser and temperature logs. Some of these attributes are shown in Table 4.8 and the most prominent co-relation was between mean pressure and and servo moment mean (93%). This relation is reasonable as pressure of the glue and the torque in which the motor is pumping the glue are co-related.

Attribute 1	Attribute 2	Correlation (%)
Reference flow mean	Servo moment mean	72%
Reference flow maximum	Servo moment max	83%
Reference flow minimum	Servo moment min	84%
Mean pressure	mean servo moment	93%
Hose	Gun	75%
Dock valve	Gun	75%

Table 4.8: Strong Correlation between temperature and doser data

Figure 4.16 and Figure 4.17 depicts the co-relation between temperature and doser log data with severe alarm ID. From the co relation matrix between pre-processed doser data and severe alarm Id , Figure 4.16 there is no strong relation with any of the severe alarms occurring and the doser data. But in Figure 4.17 we can see a considerably strong relation between chamber temperature and Severe alarms.

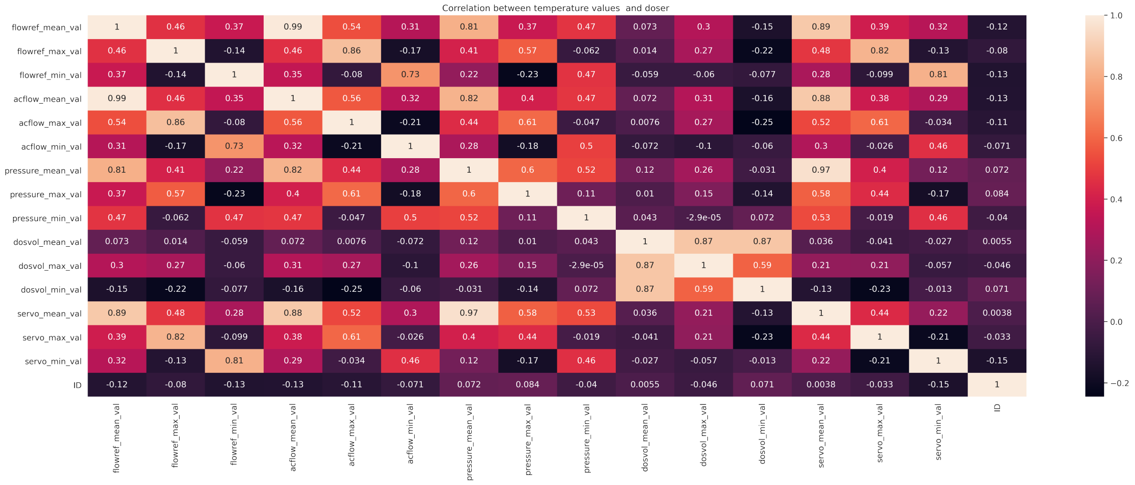


Figure 4.16: Correlation between doser logs and severe alarm alarm ID

4.3.2.3 Relationship between temperature, doser data and the alarms

In this Project we make use of alarms log for labelling the dataset for supervised learning. Figure 4.18 shows the severe alarms (A & B) that got triggered during the observed span. From Figure 4.18 we can infer that the highest number of occurrences are emergency stops and it is the most frequent alarm. However there is no evidence to track the actual reason for emergency stop alarms. It can be manual pressings by the technical team due to necessary work in the production or real failure condition. The ambiguity of the alarm occurrences leads to unclear situation in predicting

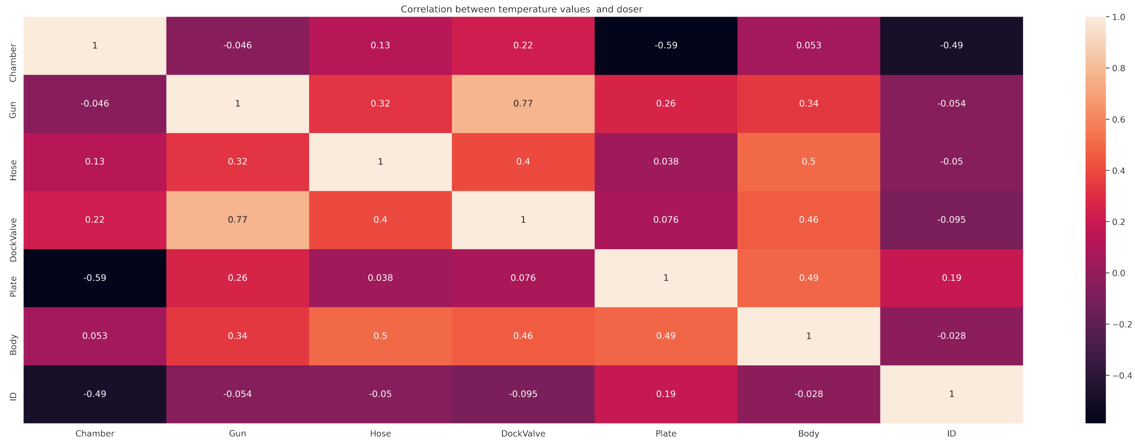


Figure 4.17: Co-relation between temperature and severe alarm ID

actual failures. However, this visualisations became really good insights in data analysis.

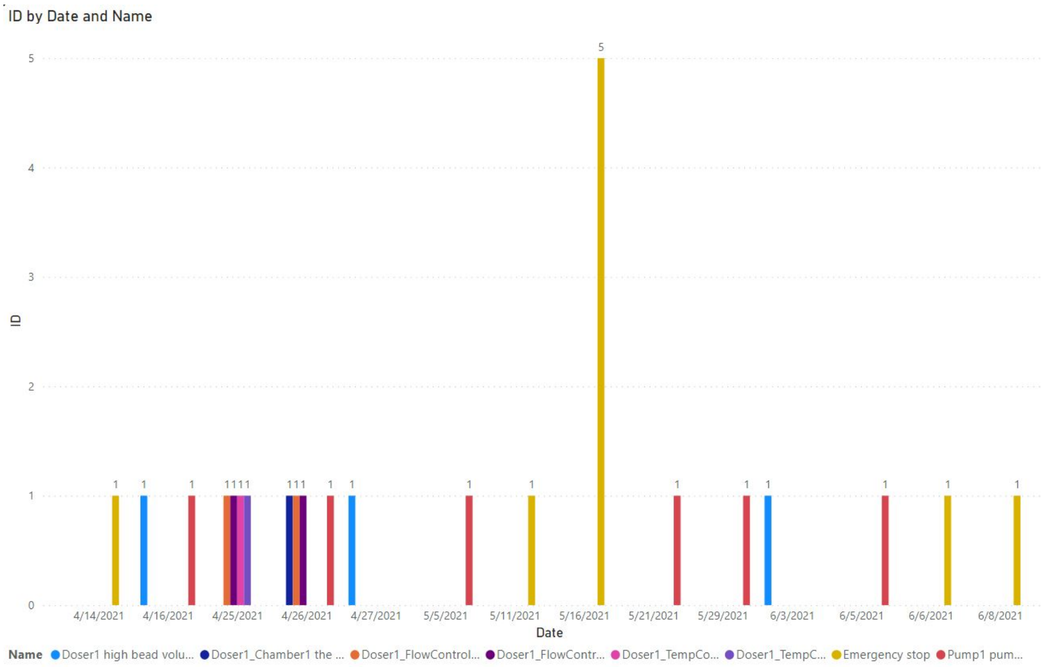


Figure 4.18: Category A and B alarm count

In order to investigate the relation between the severe alarms and temperature and pressure attributes the visualisations shown in Figure 4.19 and Figure 4.20 are combined plots of alarm with temperature and doser. From these plots we could see that most of the time when a severe alarm is occurring a breakdown is occurring as mentioned above in this section some of the emergency alarms are pressed by the employees and does not indicate a failure so we can expect some errors in the ML predictions as well. When compared to the normal working data the failure data or the severe alarms are very less. So we haven't removed emergency alarms from the final data set.

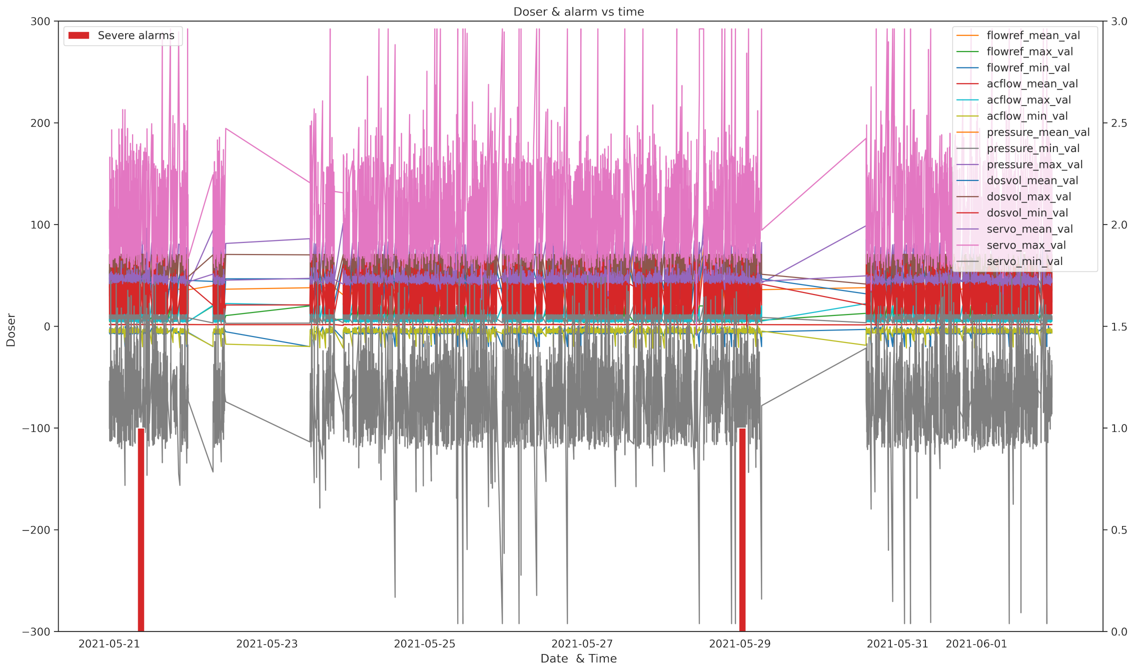


Figure 4.19: Doser and alarm plot

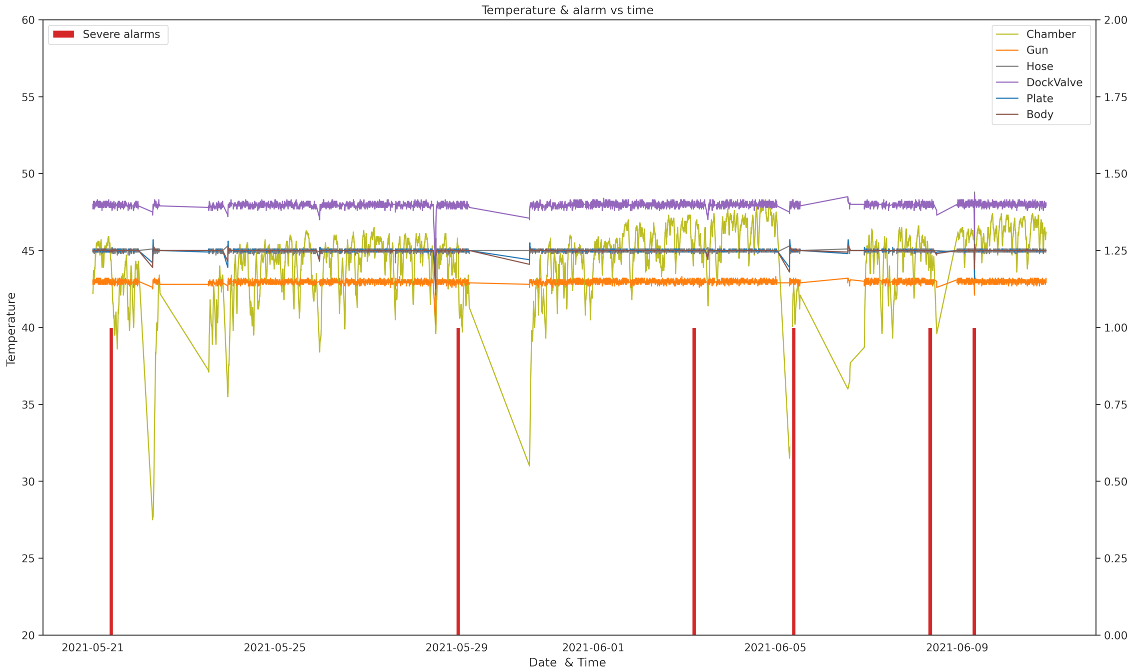
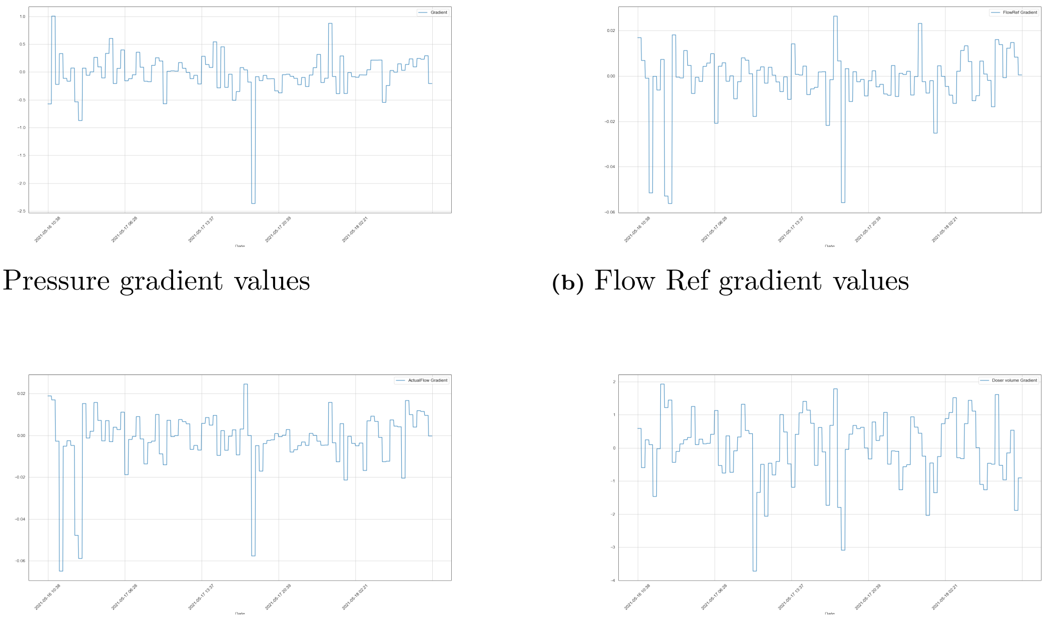


Figure 4.20: Temperature alarm plot

4.3.2.4 Pattern detection - Gradient Analysis

During the analysis of doser data, the variations of the parameters such as Flow Reference, Actual Flow, Pressure, Doser Volume and Torque were analysed in order to predict future malfunctions in the machine. The usual working conditions were learned from the maintenance team and the threshold of the abnormal conditions

were identified based on these background research of machines.
The time duration was chosen for this analysis from 2021-05-16 to 2021-06-23. X-axis represents the time axis and y-axis represent the gradient levels. The smaller sets of rows were selected i.e., it was ten consecutive rows to calculate the gradients (slops). The gradient calculation was performed on the each selected set. Then these calculated gradients were plotted against the relevant time frame. These plots are shown in the below graphs and level of gradients can be detected which might affects the usual working condition. Any parameter that goes below or higher the threshold value, can consider as unusual behaviour. The maintenance team can pay attention for the machine parts related with the parameter.
The following figures corresponds to the gradient plots for Pressure parameter in the doser log. It is possible to see some higher gradient values that are out of safer boundary.



(a) Pressure gradient values

(b) Flow Ref gradient values

(c) Actual Ref values

(d) Doser volume values

Figure 4.21: Gradient vs Time for doser parameters

4.4 Data Preparation

In order to conduct a supervised ML to predict the maintenance the data should be created with depended and independent variables. In our case we have data extracted from different sources and has to be formatted and prepossessed in order to come to that required data format.

The first data prepossessing is to be carried out is the merging of all the files together. The temperature data was received throughout the duration of the project in the format of 1 csv file per week. So all the 9 temperature files were merged together in to a data frame. Similarly there were 9 alarm logs and 42 doser logs which were merged together in to a data frame for alarm and another data frame for doser log.

The doser log was received without headers which were added later while reading the files to python using pandas library. Some of the headings in the temperature and alarm dataset were also changed in order to improve the readability. Date and time was the only attribute which was common to all three different data sets and therefore in order to merge all these files together we should have it in date-time format. The time stamp in object format was converted in to date-time format.

Most of the alarm logging were reset alarms and notification on completion of some actions. As we don't have a usable maintenance log for labelling the failures in this project the severe alarms are approximated to failures. All the severe alarms which occurred during the time-span of the data are shown in Table A.2. The teamster operating manual classifies alarms with priority 'A' and 'B' are the most severe alarm. When alarm 'A' is triggered, the machine stops immediately. When alarm 'B' is triggered the ongoing cycle is completed and the machine stops. The severe alarms were separated by creating a new data frame which only had the rows with priority 'A' and 'B'. The severe alarms that were during the period of study are depicted in Table A.2.

As discussed in the section 4.3 the doser log has a varying sampling frequency. From Table 4.2, we can see that the doser log has almost 100 times more data points than temperature log. One of the options was to merge the temperature data in to the doser data by filling out the temperate value near to the doser time stamp. In this case as the machines are not working on week ends those dates will also get deleted automatically. But when doing it this way the major draw back was over fitting. Bilbao et al. [16] states that over fitting is one of the main problem from artificial neural network. Over fitted models give very good results in train data and poor results on test data . The aim was to shrink the data points from micro seconds to minute scale. But the major challenge in shrinking data is the chance of loosing valuable information.

The best way to do this was to use different strategies like taking mean, minimum and maximum value of all the values inside a minute time frame. It was a trial and error approach and the best results were obtained when mean, minimum and maximum values were used. Every attribute in the doser Table 4.5 were replaced by the mean median and mode of all the values inside a minute time frame. A sample of this processed table is being depicted in Table 4.9. The time stamp we can see that each is in minutes time frame now. Each row has the mean, minimum and maximum values of each of the attributes from the doser log. All the processed attributes except timestamp is in float format and the timestamp is in date-time format.

After this step, the data frames of temperature, processed doser and severe alarms were merged together in to a new data frame. The data frames of doser and temperature were merged together up on timestamp column with a tolerance of 10 minutes. This new data frame is then merged with the severe alarm data frame up on timestamp and a tolerance of 30 minutes.

The data received has some scrambled historical data without continuation in the beginning. Continues data was available from all the data sources from 6th may 2021, so all the rows with a timestamp previous to this date was removed from the merged data frame. This prepossessing was done after the merging to avoid doing the same process for all the data sources separately.

As mentioned above the merging of severe alarm was done with a tolerance in time of 30 minutes. So all the severe alarms were merged to rows in the doser and temperature data frame. As there are only a few severe alarms compared to the entire data frame, all the other values were termed as null values. These null values were replaced with Zeros. All the severe alarms, that is with priority 'A' and 'B' were converted to float value 1.0. So the binary digit 1.0 zero represent an occurrence of a severe alarm and zero represents a normal working condition of the gluing machine. The final table after all the prepossessing is depicted in Table 4.10.

In the final data set the fundamental units of the attributes are different. For example the pressure values are in bars, temperature values are in degree Celsius (°C) and flow values are in ml/s. In this thesis we made use of Sklearn standard scaler from the pre-processing library. If the sample is termed as x, mean as u and standard deviation as s [59], the standard score is calculated by the Equation 4.1.

$$z = (x - u)/s \quad [59] \tag{4.1}$$

The scaling is done on individual attributes after statistical computation on the training dataset. The transformation on data is done using the stored mean and standard deviation [68].

flow_ min_val	pressure_ mean_val	pressure_ max_val	pressure_ min_val	dosvol_ mean_val	...	Time stamp	Chamber	Gun	Hose	DockValve	Plate	Body	ID	Name	priority
-19.6	42.480833	108.0	3.7	37.933056	...	2021-05-16 10:38:00	24.5	43.0	45.3	47.6	44.1	44.3	6222.0	Emergency stop	1
-5.5	42.376190	51.7	9.6	52.251701	...	2021-05-16 10:39:00	24.5	42.7	44.7	47.5	44.5	44.4	6222.0	Emergency stop	1
-4.8	44.223077	60.9	12.0	30.287854	...	2021-05-16 10:43:00	24.6	43.2	45.1	48.1	45.5	44.9	6222.0	Emergency stop	1
-3.5	41.347340	52.3	9.9	36.573404	...	2021-05-16 10:44:00	24.6	42.9	45.1	47.9	45.6	45.0	6222.0	Emergency stop	1
-5.5	37.987805	44.9	8.7	52.000697	...	2021-05-16 10:45:00	24.6	43.0	44.8	47.8	45.8	45.1	6222.0	Emergency stop	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
-5.7	32.718130	42.2	7.0	55.423796	...	2021-06-21 00:45:00	43.5	42.9	45.0	47.8	45.0	45.0	4006.0	Pump1 pump A frame air pressure is low	1
-3.7	33.099167	43.6	6.0	26.611944	...	2021-06-21 00:46:00	43.5	43.0	45.0	47.9	45.0	45.0	4006.0	Pump1 pump A frame air pressure is low	1
-5.7	33.867373	39.2	7.7	60.890254	...	2021-06-21 00:47:00	43.6	43.0	44.9	47.9	45.0	45.0	4006.0	Pump1 pump A frame air pressure is low	1
-3.7	30.217829	36.0	6.2	47.023256	...	2021-06-21 00:48:00	43.6	43.0	45.1	47.9	45.0	45.0	4006.0	Pump1 pump A frame air pressure is low	1

Table 4.9: Doser volume after prepossessing

flowref_ mean_val	flowref_ max_val	flowref_ min_val	acflow_ mean_val	acflow_ max_val	acflow_ min_val	pressure_ mean_val	pressure_ max_val	pressure_ min_val	dosvol_ mean_val	...	Timestamp	Chamber	Gun	Hose	DockValve	Plate	Body	ID	Name	priority
1.491667	20	-20	1.527778	20.7	-19.6	42.480833	108	3.7	37.933056	...	2021-05-16 10:38:00	24.5	43	45.3	47.6	44.1	44.3	6222	Emergency stop	1
1.660544	8.2	-6.1	1.609184	6.6	-5.5	42.37619	51.7	9.6	52.251701	...	2021-05-16 10:38:00	24.5	42.7	44.7	47.5	44.5	44.4	6222	Emergency stop	1
1.801215	8.3	-4.1	1.697976	4.8	-4.8	44.223077	60.9	12	30.287854	...	2021-05-16 10:43:00	24.6	43.2	45.1	48.1	45.5	44.9	6222	Emergency stop	1
1.734574	9.8	-6.1	1.721809	4.9	-3.5	41.34734	52.3	9.9	36.573404	...	2021-05-16 10:44:00	24.6	42.9	45.1	47.9	45.6	45	6222	Emergency stop	1
1.618467	5.5	-6.5	1.610105	5.6	-5.5	37.987805	44.9	8.7	52.000697	...	2021-05-16 10:45:00	24.6	43	44.8	47.8	45.8	45.1	6222	Emergency stop	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1.666856	6.8	-2.9	1.57762	5.7	-5.7	32.71813	42.2	7	55.423796	...	2021-06-21 00:45:00	43.5	42.9	45	47.8	45	45	4006	Pump1 pump A frame air pressure is low	1
1.689722	6.7	-3.3	1.661944	5.4	-3.7	33.099167	43.6	6	26.611944	...	2021-06-21 00:46:00	43.5	43	45	47.9	45	45	4006	Pump1 pump A frame air pressure is low	1
1.699576	5.3	-3.7	1.691102	5.4	-5.7	33.867373	39.2	7.7	60.890254	...	2021-06-21 00:47:00	43.6	43	44.9	47.9	45	45	4006	Pump1 pump A frame air pressure is low	1
1.668217	5.5	-0.2	1.555814	5.6	-3.7	30.217829	36	6.2	47.023256	...	2021-06-21 00:48:00	43.6	43	45.1	47.9	45	45	4006	Pump1 pump A frame air pressure is low	1
1.879651	9.7	-2.9	1.838372	6.4	-3.8	42.934302	56.9	18.9	34.211047	...	2021-06-21 01:01:00	43.6	42.9	45	48	45	45	4006	Pump1 pump A frame air pressure is low	1

Table 4.10: Final table

4.5 Model Design

This section deals with the ML modeling and the parameters used for the models. From the extensive literature survey it was clear that ARIMA and LSTM are the two best models for time series prediction. Modeling part is one of the crucial step for achieving the desired output and as mentioned in the Methodology section, Figure 3.1 the modeling process was iterative in nature. Selecting, designing, building and evaluation of the model are the four steps followed for modelling.

4.5.1 Selecting modelling techniques

Selecting the most suitable model is the key to achieve the business goals. Every ML model is not suitable for all the application and dataset. Awan et al. states that the best ML model is the the one which give the most accurate results with lesser resources and it is very important to consider the format and nature of the data set before selecting a ML model [11].

- **Business goals:** Our desired goal is to predict future severe alarm and thus the failure of the gluing machine. So the ML model should be able to perform a multivariate time series prediction. The model should be able to predict the day in which the failure is going to occur so that the maintenance department can plan and schedule the maintenance before this failure is occurring.

- **Available dataset:** The final dataset after prepossessing contains multivariate time series data. Since the attributes are of different fundamental units, it is necessary to do a scaling operation before fitting it in to a ML model. The dependent variable severe alarms and this attribute is converted in to a binary category, 1 when a severe alarm is occurring and 0 when the machine is in normal working condition.

Taking reference from the literature and analysing the available data set and business goals it was clear that ARIMA and LSTM are the two best methods for predicting the severe alarms.

4.5.2 Designing the testing

Cross validation is quintessential for any data science project. It is the stage were the accuracy of the model and prediction is evaluated. In this thesis project cross validation is done using the classic hold out method as shown in Figure 4.22. The entire dataset is divided in to train set and test set. The ML model is trained using the training set and the other portion of the data is hold out for validation which is known commonly known as the test set. The majority of the data is used for training and a the small portion left out is not seen by the model.

In this thesis project 20% of the total dataset is holdout for cross validation and 80% of the data is used for training the models. Since it is a time series prediction

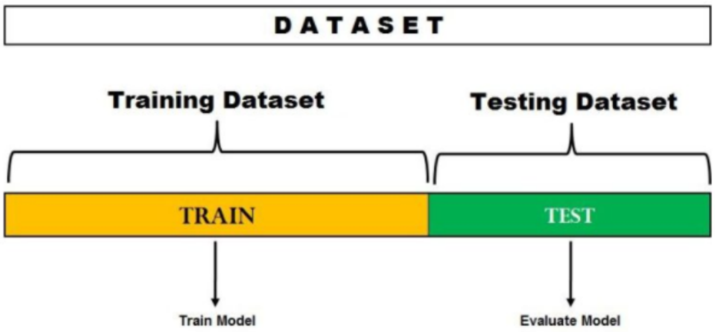


Figure 4.22: Sketch depicting the holdout method used for cross validation

problem the future is dependent on previous data points random selection of data points can not be bone. Data from 15th May to 14 th June is used for training and data from 15th June to 19 th June, 2021 is hold out for testing.

4.5.3 Building ARIMA model

ARIMA is a linear regression type of model means that predictors are own lag values of the model and the best performance can be achieved when those predictors are independent with each other.
This section is about the ARIMA model design to predict time series parameters in doser log such as pressure and doser volume. Since we use past values in predicting parameters it is called 'Univarate time series forecasting'. And this study works based on the fact that information on historical data can only be used in predicting future values. The ARIMA model works with lags of past values, lagged values of prediction errors. ARIMA model is capable of handling 'non-stationary' data having a pattern.
The main steps of the ARIMA model design;

- Visualisation of the time series data
- Checking whether if the data series is stationary or not, transform if it is non-stationary
- Graphical representation of the correlation and auto-correlation plots
- Build the ARIMA model using ARIMA() function
- Train the model with train dataset, using fit() function
- Perform the predictions using predict() function

4.5.3.1 Data preparation

The ARIMA model was tested on the doser log data and tried to predict a week ahead doser parameters. In the preparation process, we imported doser log data as .csv files and concatenated all of them into a single file. The mean values of all the parameters (flow reference, actual flow, pressure, doser volume and servo torque) were used to feed to the model.

4.5.3.2 Check the seasonality of the data

As mentioned above, there are two primitive ways of checking the status of stationary. i.e., rolling statistics and Augmented Dickey-Fuller Test. Regarding the rolling statistics, rolling mean and standard deviation has to be constant in order to have a stationary series.

The Figure 4.23 shows the series data, 12-rolling mean and 12-rolling standard deviation plot. Even though there are small changes in the value of rolling mean and standard deviation, it is not possible to see a considerable increase or decrease.

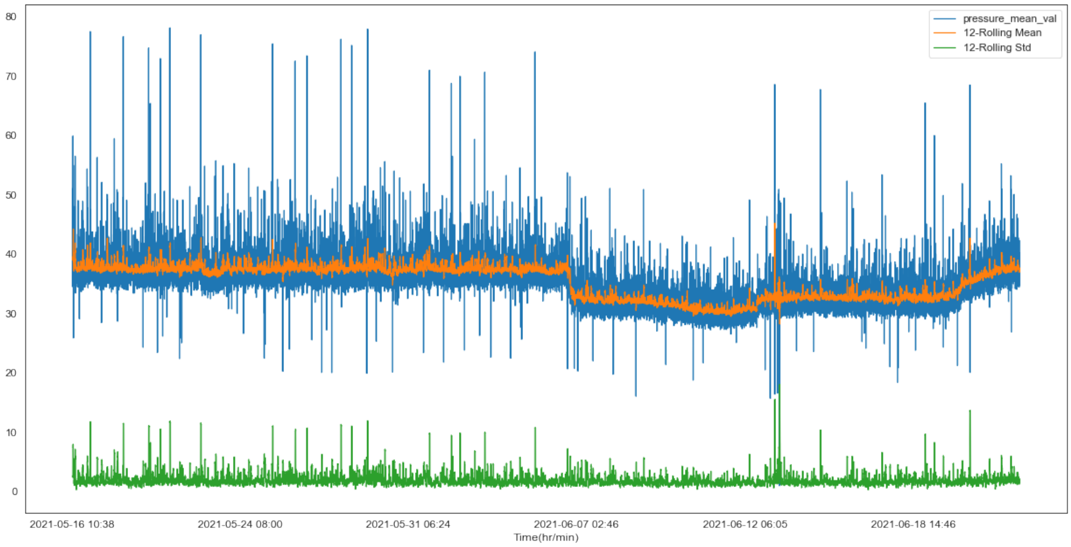
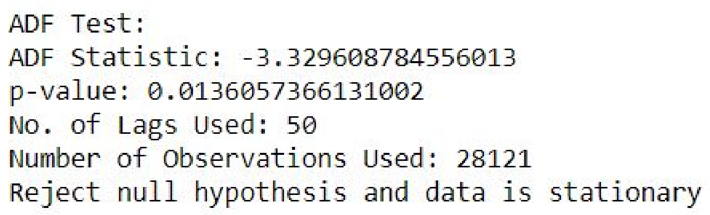


Figure 4.23: Rolling statistics

The basic principle of ARIMA model is to decide the correct values for p,d and q parameter.

4.5.3.3 Finding the d parameter

This is about finding the order of differencing or the d parameter. The order of differdincing means the minimum differencing we need to get to a stationary series. However this only needs if the series data is non-stationary and we used the Augmented Dickey Fuller Test i.e., adfuller() function in the statsmodel package. In contrast, the d value become zero when there is non-seasonal series data. If the test becomes null hypothesis or p value is higher than 0.05 threshold level, the time series is non-stationary. Anyhow, the ADF result for this study became $p < 0.05$, so reject null hypothesis resulted the time series is stationary. And d value is equal to be zero. Otherwise, the order of differencing (d) has to be found.



```
ADF Test:
ADF Statistic: -3.329608784556013
p-value: 0.0136057366131002
No. of Lags Used: 50
Number of Observations Used: 28121
Reject null hypothesis and data is stationary
```

Figure 4.24: ADF statistics

4.5.3.4 Selecting the p parameter (AR term)

The next step of the model design is to decide the p parameter or the order of AR term. This can be chosen by analysing the partial correlation plot. In simple terms, partial correlation gives an idea about the inter-connection between series and its lagged data.
After observing the partial correlation plot (Figure 4.25) we can see that PACF lag 1 is quite above the significance area and therefore consider p value as 1.

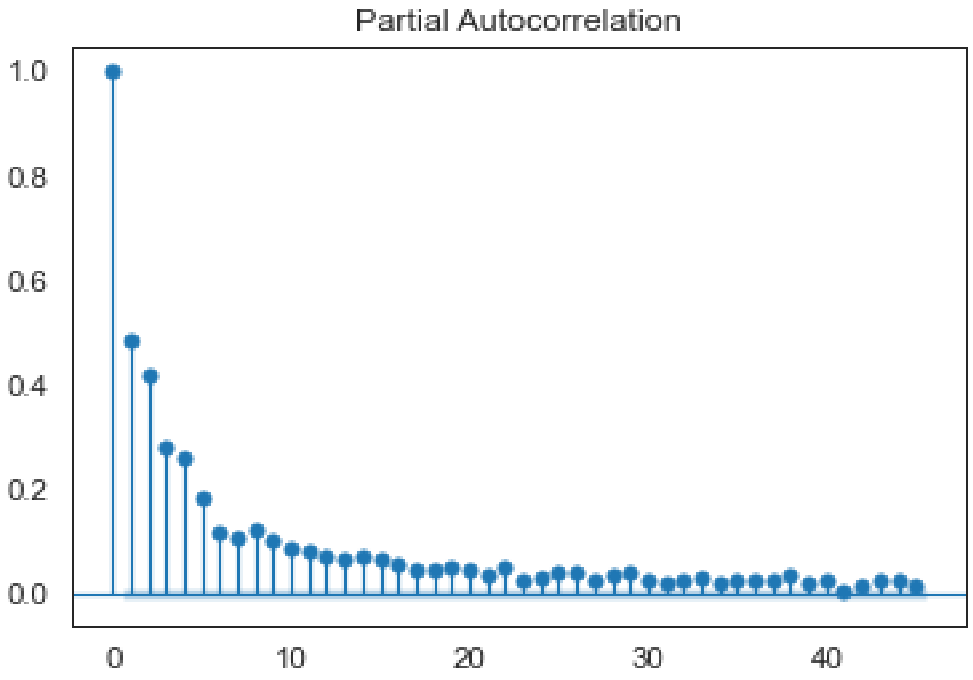


Figure 4.25: Partial Autocorrelation

4.5.3.5 Selecting the q parameter (MA term)

The correlation graph represents the correlation between the present time observations and all the observations in the past. The auto correlation function can be used to find the optimal number of MA term in the model as well as the q parameter. It is the number of MA terms needed to take out the auto correlation in the data series. This MA term represents the lagged forecast error in the ARIMA model function. It can be seen that around three lags are above the significance line and conclude the value of q as 3.

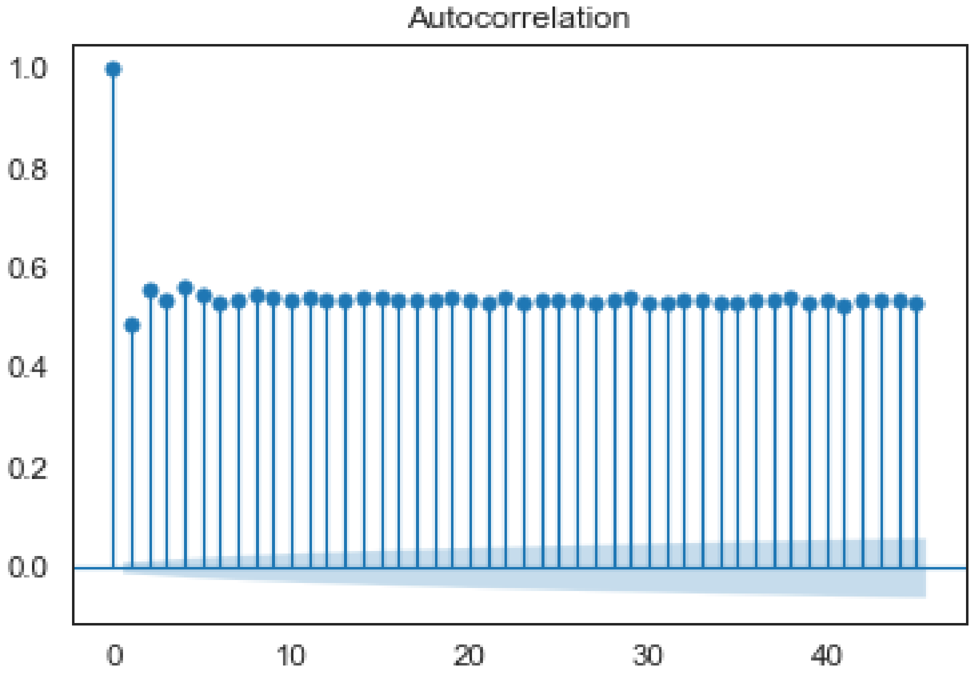


Figure 4.26: Auto correlation

The conclusion we can draw from the Figure 4.26 is correlation is less between data points therefore randomness is high.

4.5.3.6 Build the ARIMA model

After choosing the best possible values for p, d and q parameters, it is the time to implement the ARIMA model by using ARIMA() function in statsmodels library. The predicted result of ARIMA model with order (1,0,3) is depicted in the Figure 4.27 in orange color and the actual data is drawn in blue color.

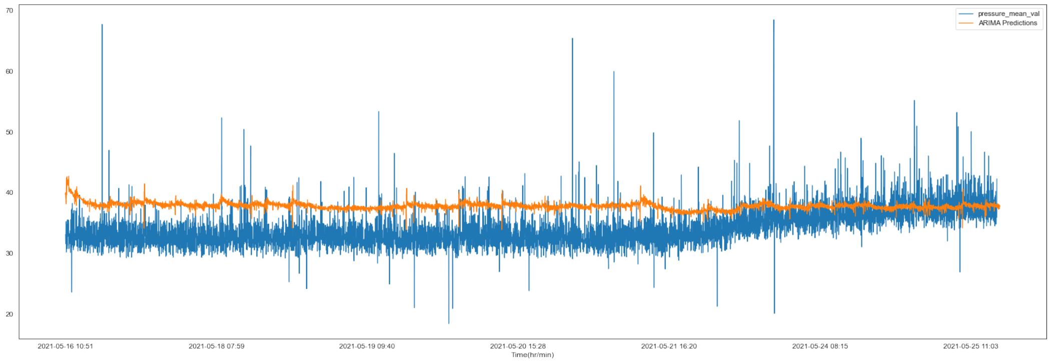
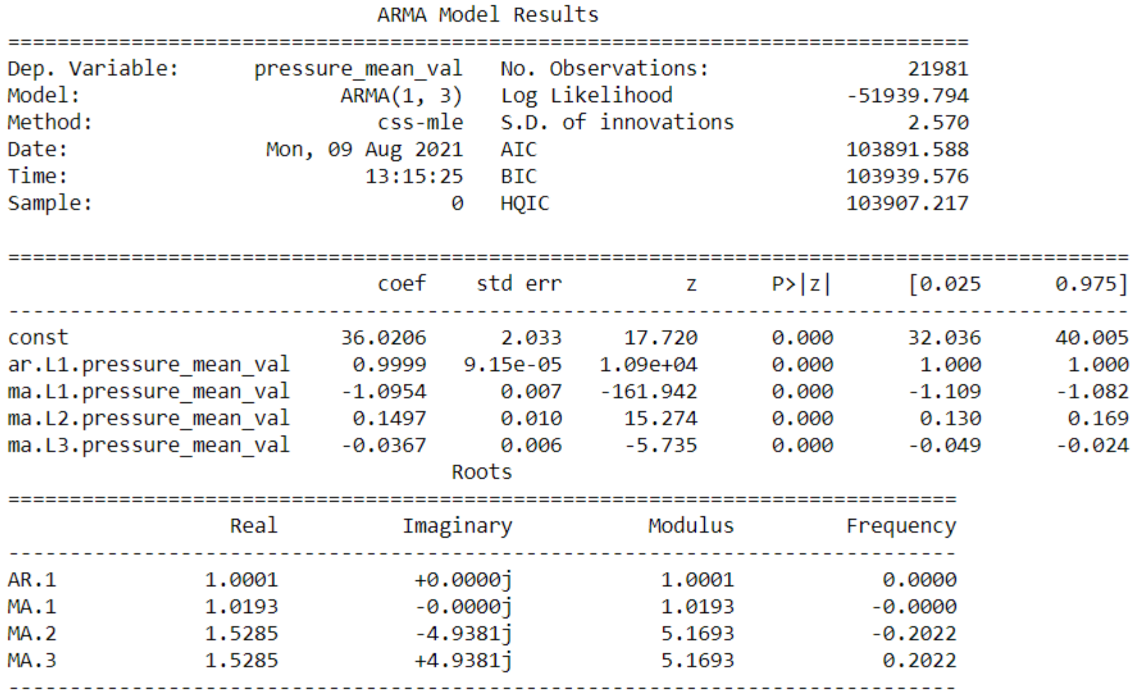


Figure 4.27: Predicted Result Plot



```
                              ARMA Model Results
==============================================================================
Dep. Variable:      pressure_mean_val   No. Observations:                21981
Model:                     ARMA(1, 3)   Log Likelihood              -51939.794
Method:                       css-mle   S.D. of innovations              2.570
Date:                Mon, 09 Aug 2021   AIC                         103891.588
Time:                        13:15:25   BIC                         103939.576
Sample:                             0   HQIC                        103907.217

============================================================================================
                              coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                      36.0206      2.033     17.720      0.000      32.036      40.005
ar.L1.pressure_mean_val     0.9999   9.15e-05   1.09e+04      0.000       1.000       1.000
ma.L1.pressure_mean_val    -1.0954      0.007   -161.942      0.000      -1.109      -1.082
ma.L2.pressure_mean_val     0.1497      0.010     15.274      0.000       0.130       0.169
ma.L3.pressure_mean_val    -0.0367      0.006     -5.735      0.000      -0.049      -0.024
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.0001           +0.0000j            1.0001            0.0000
MA.1            1.0193           -0.0000j            1.0193           -0.0000
MA.2            1.5285           -4.9381j            5.1693           -0.2022
MA.3            1.5285           +4.9381j            5.1693            0.2022
-----------------------------------------------------------------------------
```

Figure 4.28: ARIMA model summary

The Figure 4.28 is the representation of the ARIMA model summary of what each criterion it is associated with. It is important to understand what these parameters are which is related with the ARIMA model. The 'Dep. variable' and 'Model' parameters are the variable name and the model with it's respective orders. The 'method' is mentioned as css-mle and stands for 'conditional-sum-of-squares & maximum-likelihood-estimation', referring that maximum-likelihood-estimation will determine the parameter value when the model result having the highest probability and the model output will be close to the actual data.

The 'log-likelihood' parameter represents the maximum-likelihood-estimation in a simpler way such that this is generated by considering the logged values of prior values. This is beneficial in comparing between other similar models to find the best. The higher the lag value, better the model will be.

The parameter 'AIC' will be introduced next and it stands for Akaike's Information Criterion. This will be an indication of strength of the model. It utilizes the maximum-likelihood-estimation result and number of parameter as the inputs. It is always better to use less number of parameters in the model as having more parameters affects the maximum-likelihood-estimation to be increased. The model with the lowest 'AIC' value will have the best performance. Another criterion similar to 'AIC' is 'BIC' which stands for Bayesian Information Criterion is mentioned in Devi et al.[28]. The model performs well when this criterion is less and related with number of rows in the data set. In the understanding of model performance, both of these AIC and BIC are important in feature selection as well as helps to decide the best model with reliable output.

The least frequently used criterion for feature selection is 'HQIC' or Hannan-Quinn Information Criterion. The 'coef' column gives an idea about feature significance. The ar.L1 term is about lag 1 in auto regressive. Similarly, ma.L1, ma.L2 and ma.L3 represents the lag1, lag2 and lag3 values of moving average parameters. $AR(1): Y_t = \mu + \phi_1 * Y_{t-1} + \epsilon_t$ and $MA(3): Y_t = \mu + \epsilon_t + \theta_1 * \epsilon_{t-1} + \theta_2 * \epsilon_{t-2} + \theta_3 * \epsilon_{t-3}$ are the general equations for ARIMA model such that ϕ stands for auto regressive terms and ϵ stands for moving average terms in the equation (Padhan et al.[56]).

The 'std err' parameter is about estimation errors pf prediction values. This indicates the importance of the residual errors on the predicted values. The 'z' parameter is calculated from the division of 'coef' value by 'std err' value and it is called as standardised coefficient. The next parameter would be '$P > |z|$' and it corresponds with the p value of coefficient. This p value has to be less than the threshold i.e., 0.05. The model selection has to be based on this condition since otherwise it will result in unrealistic outcome. The two columns from the end related with confidence intervals set according to the error margin.

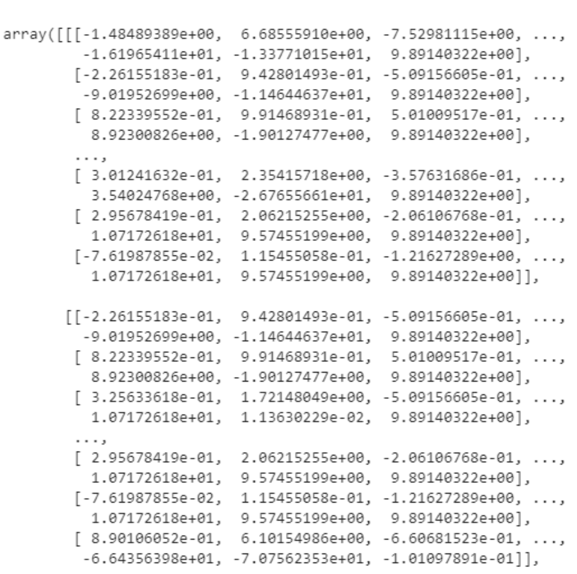
To sum up, the result of the ARIMA model was not practical, that it did not meet our proposed expectation in the project and it can be seen in Figure 4.27. It is not much helpful in forecasting failure time because it requires some threshold values to decide critical parameters and that information was not available at the moment. However, the purpose of using this model is to experiment the capacity to predict the future trends and this can be developed to detect the future failures with more information such as threshold values of the relevant parameters.

4.5.4 Building LSTM model

4.5.4.1 Data preparation for LSTM

One of the most important step in handling multivariate LSTM prediction is preparing the data in a shape that the LSTM model is able to process. This data preparation is to be done on the training data set as well as the holdout dataset. LSTM networks need data to be reshaped in the format of N number of samples * time steps. In our case the time step is one minute. The pre-processing is done so that, if n number of rows are utilized for training the n+1 th row is predicted, and this sliding window is moved by one row. For example, if one to 14 th value is used for training in the first sliding window the 15th value is used for prediction and then the sliding window moves by one row , and uses second row to 15 th row for training and 16 th row for prediction. Like ways all the attributes are reshaped in two lists (trainX and trainY) using for loop.

In our thesis work the sliding window looks back on previous 14 minutes and prediction is done on the 15th minute. After this step both the lists are converted in to arrays. The shape of the train X array is (19427, 14, 22). In this matrices of shapes, 14 is the number of minutes that looked back, 22 is the number of attributes, and 19427 is the total number of data points. The train X array after pre-processing is show in Figure 4.29. The shape of train Y is (19427, 1). There is only one attribute for train Y because the prediction is only to be done on the alarm column, as 0 or 1, where 0 depicts the normal working condition of the gluing machine and 1 represents the occurrence of a severe alarm. The same data pre-processing is done on the test set as well.



```
array([[[-1.48489389e+00,  6.68555910e+00, -7.52981115e+00, ...,
         -1.61965411e+01, -1.33771015e+01,  9.89140322e+00],
        [-2.26155183e-01,  9.42801493e-01, -5.09156605e-01, ...,
         -9.01952699e+00, -1.14644637e+01,  9.89140322e+00],
        [ 8.22339552e-01,  9.91468931e-01,  5.01009517e-01, ...,
          8.92300826e+00, -1.90127477e+00,  9.89140322e+00],
        ...,
        [ 3.01241632e-01,  2.35415718e+00, -3.57631686e-01, ...,
          3.54024768e+00, -2.67655661e+01,  9.89140322e+00],
        [ 2.95678419e-01,  2.06215255e+00, -2.06106768e-01, ...,
          1.07172618e+01,  9.57455199e+00,  9.89140322e+00],
        [-7.61987855e-02,  1.15455058e-01, -1.21627289e+00, ...,
          1.07172618e+01,  9.57455199e+00,  9.89140322e+00]],

       [[-2.26155183e-01,  9.42801493e-01, -5.09156605e-01, ...,
         -9.01952699e+00, -1.14644637e+01,  9.89140322e+00],
        [ 8.22339552e-01,  9.91468931e-01,  5.01009517e-01, ...,
          8.92300826e+00, -1.90127477e+00,  9.89140322e+00],
        [ 3.25633618e-01,  1.72148049e+00, -5.09156605e-01, ...,
          1.07172618e+01,  1.13630229e-02,  9.89140322e+00],
        ...,
        [ 2.95678419e-01,  2.06215255e+00, -2.06106768e-01, ...,
          1.07172618e+01,  9.57455199e+00,  9.89140322e+00],
        [-7.61987855e-02,  1.15455058e-01, -1.21627289e+00, ...,
          1.07172618e+01,  9.57455199e+00,  9.89140322e+00],
        [ 8.90106052e-01,  6.10154986e+00, -6.60681523e-01, ...,
         -6.64356398e+01, -7.07562353e+01, -1.01097891e-01]],
```

Figure 4.29: Pre-processed training set for LSTM

4.5.4.2 LSTM model architecture

After the data pre-processing the next important step is designing the architecture for the LSTM model. As mentioned in section 2.5.3 LSTM networks is a combination of number of LSTM modules which have the capability to store, delete and pass information from one module to another. If the number of modules is less that optimum number there is a chance of under fitting and if it is more than the optimum number there is high chance of over fitting. In order to avoid these issues hyper-parameter tuning to find the optimum number of modules and layers is quintessential.

In this thesis project we have made use of Keras Tuner to conduct hyper-parameter tuning to find out the optimum parameters. Keras tuner is a scalable hyper parameter optimisation framework that makes the process easier and efficient. Keras tuner has build in search algorithms like Bayesian Optimization, Hyperband, and Random Search, and in this thesis work we have made use of random search algorithm for searching [55].

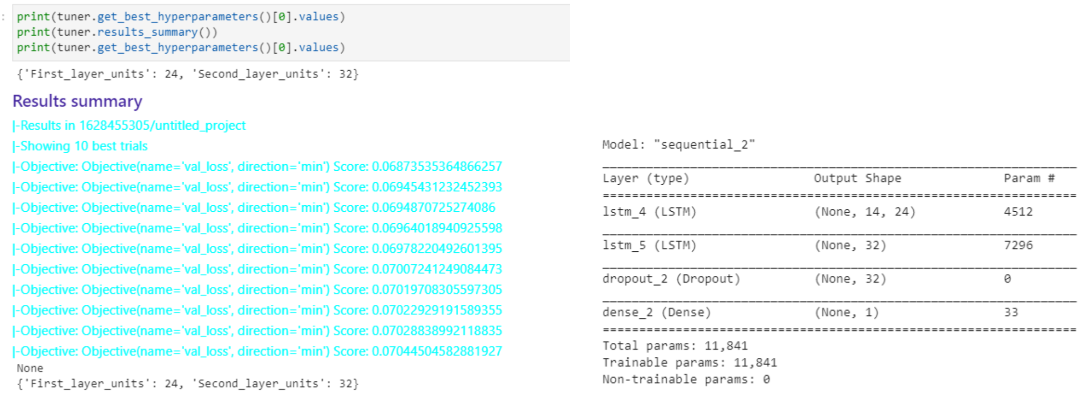
The first objective of the hyper parameter tuning was to find the optimum number of layers of LSTM model and then fine tuning the number of LSTM models inside these sequentially arranged layers.For the each LSTM layer the input parameter for searching was with in the range of 16 LSTM module to 256 modules, and the search algorithm took a step value of 16 modules. All the permutations and combination with in this given range was ran for 5 layers. The algorithm looked on the mean squared error of each epoch for analysing the quality of result from each of the combination.

After the two LSTM a layers a drop out layer, which drop is added in order to avoid over fitting. Since it is an unbalanced data-set, the ML model tend to show some of the alarms as normal working, as the majority of the data points of the dataset was of normal working of the gluing machine. This drop out layer was useful in avoiding over fitting problem by randomly eliminating data from the neural network. The inputs which are not eliminated are scaled up by 1/(1-rate) so that there is no change happening in the overall inputs [25].

Since the output of an LSTM model is not a softmax, there occurs a mismatch in dimensions of the output and the dimension of the target.In order to avoid this issue of mismatch a dense layer with the shape In this thesis the number of nodes in the dense layer is one, as the shape of the prediction (train Y) is one. Adam optimiser is used for first order gradient based optimisation. Adam optimiser was chosen because it gives the best results for problems with noise and sparse gradient, and it gave the best results up on trial and error with other optimizer.

After performing the hyper-parameter tuning it was found that the best model was a two layers LSTM model with 24 modules in first layer and 32 modules in second layer. The obtained result of hyper parameter tuning using keras tuner is shown in Figure 4.30a and the model architecture of the best model is depicted in Figure

4.30b.



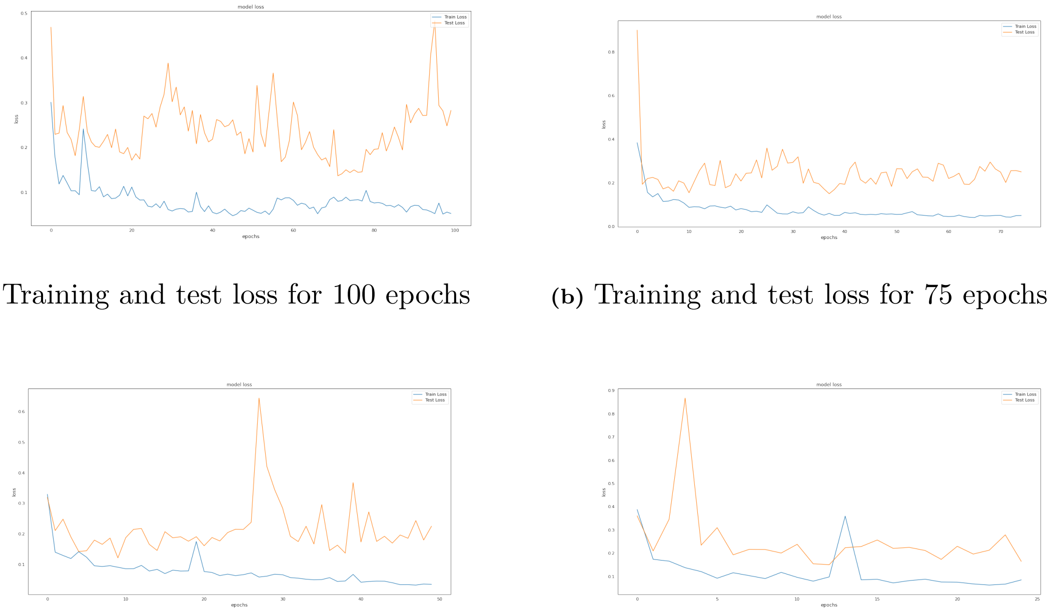
```
print(tuner.get_best_hyperparameters()[0].values)
print(tuner.results_summary())
print(tuner.get_best_hyperparameters()[0].values)

{'First_layer_units': 24, 'Second_layer_units': 32}
Results summary
|-Results in 1628455305/untitled_project
|-Showing 10 best trials
|-Objective: Objective(name='val_loss', direction='min') Score: 0.06873535364866257
|-Objective: Objective(name='val_loss', direction='min') Score: 0.06945431232452393
|-Objective: Objective(name='val_loss', direction='min') Score: 0.0694870725274086
|-Objective: Objective(name='val_loss', direction='min') Score: 0.06964018940925598
|-Objective: Objective(name='val_loss', direction='min') Score: 0.06978220492601395
|-Objective: Objective(name='val_loss', direction='min') Score: 0.07007241249084473
|-Objective: Objective(name='val_loss', direction='min') Score: 0.07019708305597305
|-Objective: Objective(name='val_loss', direction='min') Score: 0.07022929191589355
|-Objective: Objective(name='val_loss', direction='min') Score: 0.07028838992118835
|-Objective: Objective(name='val_loss', direction='min') Score: 0.07044504582881927
None
{'First_layer_units': 24, 'Second_layer_units': 32}
```

(a) Results of hyperparameter tuning

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_4 (LSTM)                (None, 14, 24)            4512
_________________________________________________________________
lstm_5 (LSTM)                (None, 32)                7296
_________________________________________________________________
dropout_2 (Dropout)          (None, 32)                0
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 33
=================================================================
Total params: 11,841
Trainable params: 11,841
Non-trainable params: 0
```

(b) Model architecture

In the hyper parameter tuning in order to reduce the computing cost and running time only 10 epochs were used for training. The next step was to find out the optimum number of epochs. This was done using trial and error method. Some of the important results are shown in figures below.



(a) Training and test loss for 100 epochs

(b) Training and test loss for 75 epochs

(c) Training and test loss for 50 epochs

(d) Training and test loss for 25 epochs

From training loss Figures 4.31a, 4.31b, 4.31c,4.31d it is clear that there are spike in test losses when the number of epochs are more. This signifies the model getting over fitted when the number of epochs are more.So in this thesis project we have chosen 25 epochs for training, as it gives the best training results with lower running time and computing cost.

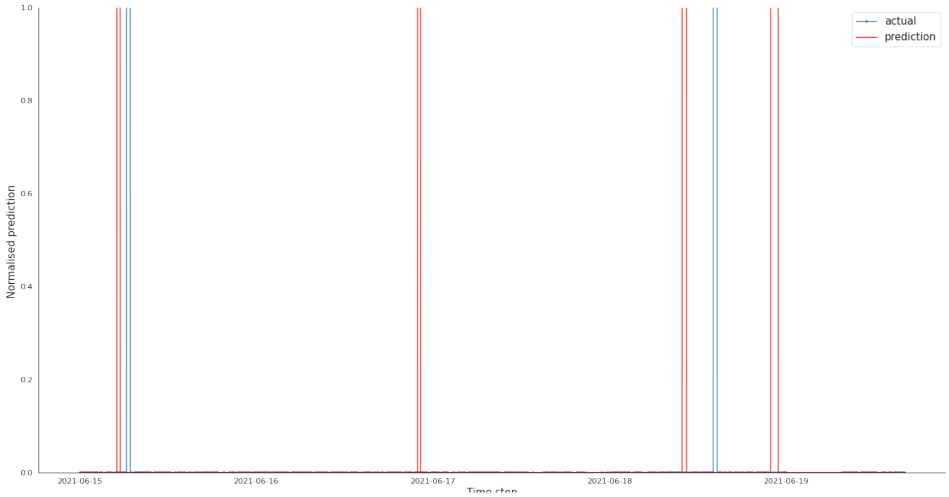


Figure 4.32: Prediction of future alarms with out inputting dependent variables

For cross validation we have performed forecasting of the future alarms without giving any dependent variables from the test data, as in real time the dependent variables from future are unknowns for the model. The result of this future forecast with out giving the dependent variables are shown in Figure 4.32.The x-axis of the graph has future timestamp and the lines in red depicts the prediction of alarms and the blue line depicts the actual alarms. From Figure 4.32 it can be inferred that the ML model is able to predict the day in which the severe alarm is going to occur. The predictions have some false alarms too, but most of the real alarms are predicted correctly. The predictions is not giving the exact minute at which the alarm is occurring, but from Figure 4.32 its clear that the model is able to predict which quarter of the day in which the alarm is going to get triggered.

4.5.5 Assessing the Model

The accuracy of ML models are calculated by calculating the absolute error (MAE),and Root mean squared error (RMSE). The equation and the theory of these calculations are discussed in section 2.6. The ARIMA model was only capable of doing forecasting on continues parameters like pressure temperature etc. The MAE of this prediction was 4.3023 and the RSME was 4.9151. On the other hand the LSTM model was able to predict the occurrence of alarms. The MAE of this prediction was 0.2518 and the RSME of this prediction compared to true alarm values were 1.5539. The lesser the value of these terms the more the accurate the model is, so it is clear that LSTM model is more accurate than ARIMA model. From the manual visualisation of Figure 4.32 it is clear that LSTM model is able to predict the severe alarms but ARIMA model fails to do the prediction on binary forecasting.

4.6 Evaluation

In this project work, the ARIMA model is only focused on the doser log parameters of Teamster data and was used to predict future trends in flow reference, actual flow, pressure, doser volume and servo torque parameters which is related with gluing application process. The objective of utilizing ARIMA model is predicting the future increases or decreases on these targeted parameters by analysing the past trends of these targeted parameters. The main objective of the project is failure prediction and the result from ARIMA does not directly gives the failures, but the parameter predictions from ARIMA model can be used as an input for LSTM model and can be used for training the model and also for prediction.

On the other hand LSTM model is able to predict the future severe alarms and thus the future breakdowns. Even though the model is not able to predict the exact time in which the alarm is occurring but is able to predict the day and in which quarter of the day the alarm is going to be triggered. One of the drawback with this model is that, some false predictions of alarms are occurring. It can also be that the false alarms are just a worker who opens the door to the station or something that has nothing to do with the actual data used. The positive aspect is that none of the real alarms were missed when conducting the cross validation. So it can be concluded that the LSTM model is able to fulfill the business goals by predicting the severe alarms and thus the failure of a machine, which would aid the maintenance department to perform predictive maintenance in the gluing machine.

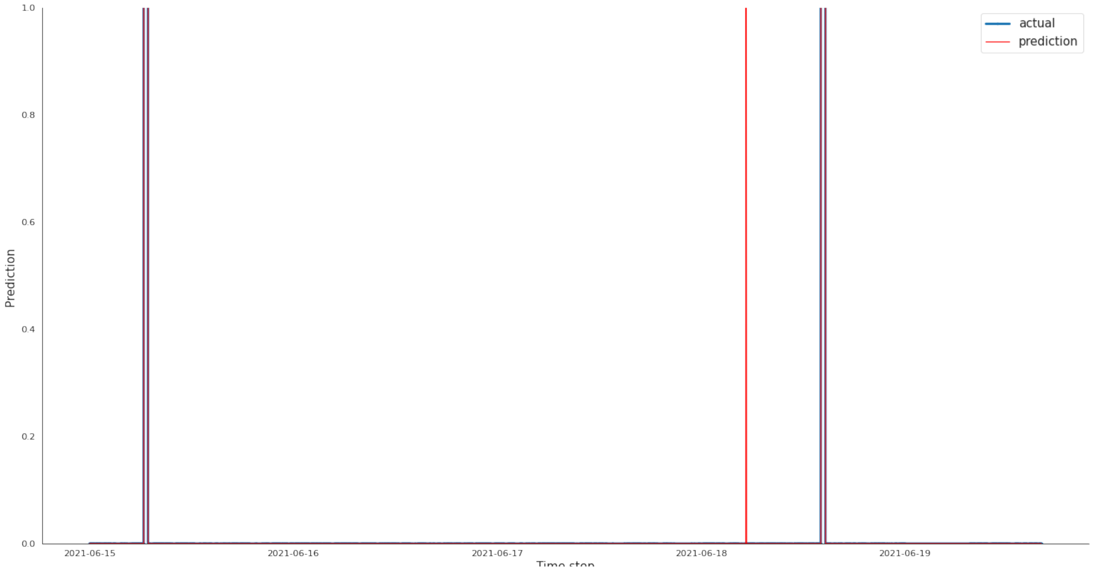


Figure 4.33: Prediction of future alarms after inputting dependent variables

In the result section of LSTM prediction the result of the forecast done by the LSTM model with out inputting the dependent variable (testX) were shown in Figure 4.32. It was done this way because, in real time scenario the dependent variable are unknown for the model as it is going to occur in the future. But if we input the depended variable (testX) values then do the predictions the results obtained are much accurate than forecasting the predictions alone. When dependent variables were given for predictions the prediction of alarms were as shown in Figure 4.33 . From this figure it is clear that inputting dependable parameters makes the prediction more accurate. So by inputting the results of ARIMA as input variables for prediction of LSTM model would also gives a much more accurate result. This can be considered for a future study.

So comparing both the ML models; LSTM was the best ML model for prediction of failures. This answers the second research question (*RQ2*).

5
Discussion

This section sums up the findings and leanings acquired during this data driven predictive maintenance project at Volvo Cars. This thesis project was a success as the project enables the company to predict the severe alarms, thus the future failures that are going to occur in the gluing machine. In this section we portray the challenges faced during the project and the recommendation for the company which would aid the maintenance department to pursue projects which would enable Volvo Cars to perform preventive maintenance in the entire body shop. After this section the academic and industrial contribution of this project is also discussed.

5.1 Challenges

During the course of the project the primary challenges faced was with data collection and acquisition. The maturity of the maintenance department at the body shop of Volvo Cars, for data collection and storage was not advanced. The department is learning about the possibility of incorporating data science for achieving smart maintenance. Since this project was one of the their first try towards this long term goal there were many challenges and limitations in the availability of the data. Most of the data sources used in the project were not from stored data sources, instead was collected after the beginning of the project and this waiting time was the major limitation, as the project was expected to be completed within the deadline date. Many data sources were tried, such as high resolution alarm, axxos data etc and were found to be not relevant for this project. But acquiring new data was a time consuming process, but was a good learning for the students and the maintenance team. This was one of the major hurdle for many other predictive maintenance studies as well [12].

In order to conduct a supervised learning, there is a requirement for labeling the dataset. This would help the ML model to learn when a machine is working perfectly and when the machine is having a breakdown. When it comes to smart manufacturing and predictive maintenance, the maintenance log contains information about the breakdowns and it is the main source of information. In the body shop at Volvo Cars, the maintenance log entries were manually done by the technicians and also the maintenance schedules were done manually. The manual logging had errors and the time entered were mismatching with the reality due to human errors. Because of these reasons the maximo data was not suitable to be used for labelling the data.

Another challenge faced during the project was the lack of a synchronised clock in the plant. Different machines had different clocks and verifying whether all these were synchronised was a challenge. The sampling times of different data sources were different and choosing the most suitable pre-processing methods without missing valuable information was a huge challenge during the thesis work.

Data driven approaches and projects are mostly multi-disciplinary in nature. These projects which are heavily dependent on various domain knowledge such as production lines, systems, machines and data etc, it would be beneficial to conduct a risk assessment strategy to avoid unforeseeable situations.

5.2 Recommendations

5.2.1 Improvisation of data resources

As discussed in section 5.1, one of the major challenge was the process of acquiring relevant data. Even though relevant data source could be figured out during the course of the thesis work, it was a time consuming process. Most of the time we were waiting for data, as the data collection started after the request for the data is being made. So by storing, relevant data in a database could eliminate these unwanted wastage of time. Data storage is also quintessential, for conducting a high quality ML project. This data acquisition was also highlighted as a major hurdle in previous works done in predictive maintenance [12].

In this thesis work the lack of historical data is the prime reason for lower accuracy of prediction. For example, in the alarm log that was used for labelling the data, the severe alarms that got triggered was limited and there were more severe alarms that could be triggered in other circumstances. The ML model is not capable of predicting the severe alarms that were not present in the alarm log used for training the ML model. Jain et al. [39] highlights the importance of the quality of the data and relevance, for attaining highly accurate ML models [39]. So having a data storage and storing the relevant parameters is crucial for future predictive maintenance projects in the body shop of Volvo Cars.

5.2.2 Standardisation of time

For ML problems, data is taken from different sources. As these data are from different sources, a common parameter is quintessential for merging them together. For the gluing machines the data taken from different sources are having different fundamental units and the only common parameter is the timestamp. So having a synchronised clock is necessary for merging these data from different sources together, to perform ML for smart maintenance. So this data standardisation is crucial for obtaining good prediction and classifications using ML algorithms [31].

During the course of the thesis project, it was found that the different data sources and logs available in the body shop of Volvo Cars doesn't have a standard time.

Each of these machines had different clocks and the synchronisation between them could be lost due to human errors. For example, the operator forgetting to change the summer time to winter time can make that particular machine out of sync. Our recommendation is to build a common standard clock for all the machines at the plant to solve this issue.

5.2.3 Automatic breakdown recording system

The maintenance logs at the body shop of Volvo Cars is manually entered. Since there is a high probability for human errors in this manual entry these logging cannot be used for labeling the data. In this thesis work we made use of severe alarms for labeling the data, but the accuracy of the model would have been much better if the breakdown records were made use instead of alarms. Johnson et al. [42] discusses the importance of automatic logging systems for anomalies detection, and states that anomalies are rare events and it is only possible to detect them using automated logging system. In alarm logs the emergency stop alarm can occur when a technician press the emergency stop button also. So such alarms are not a real indicator of a failure. But the code for the program is made in such a way that a data scientist who is going to work on a similar project with automated break down data could just replace the alarm attribute with the automated data, where the breakdowns can be labelled as 1 and the normal working as 0.

An automatic breakdown recording system and storage of the breakdown data is important for data driven smart maintenance project. Volvo Cars currently have axxos system for recording the disturbances occurring at the plant. But this data source did not have recordings from the robot station which we were analysing in this project. Automation of the maintenance logs or adding all the station disturbance to the axxos data logging is quintessential for smart maintenance projects in the future.

5.2.4 Collaboration with suppliers

Volvo Cars have machines from different original equipment manufacturers (OEM) like ABB, Atlas Copco etc. After collaborating with Atlas Copco during the thesis work, it was clear that the suppliers were also interested in data driven approaches for improving the maintenance activities of these machines. The collaboration would enable knowledge transfer from the domain experts from the supplier to the data scientist in Volvo Cars, to conduct further studies which would enable Volvo Cars to attain the capability of data driven predictive maintenance. This collaboration will also be beneficial for the suppliers coming up with additional features in there machines, that predict the failure of the machine. This would help small scale and large scale companies to plan and schedule their maintenance activity before the predicted failure time. Studies have proved that innovation is also positively influenced by collaborations [58].

5.2.5 Improving the scope of study

This thesis work was focusing on a single gluing machine for a short period of time. Better predictions could be made as the model has seen the wide operating condition of the gluing machine. One of the best ways of achieving this is by using data from similar gluing machines from the body shop, this would make the data set more diverse.

The data collected for the study was during the summer of 2021. there can be variations in the operating temperature of the gluing machine, which could influence the failure of the machine and temperature can influence the viscosity of the glue in the gluing machine. This issue can be solved by using a data set with a wider time span, as discussed in section 5.2.1.

This section clearly answers the *RQ3* by pointing out the ways in which the predictions for data driven maintenance can be done and discuss about the future steps that the company can take fulfill its goals of smart maintenance.

5.3 Academic contribution

During the extensive literature survey performed for learning the background and related works, there were very few applications of multivariate time series prediction. LSTM was mostly used for multivariate time series prediction of stock price [37].Lim et.al [49] states that LSTM was better compared to traditional ML algorithms in predictions. So in this thesis project we made use of LSTM ML model for multivariate time series prediction. As predictive maintenance is quintessential for achieving smart maintenance [51] and smart manufacturing, there is a huge potential for many research works in this domain. This thesis work would surely serve as a good industrial research reference for such works in the future.

The methodology used for this thesis project is an enhanced CRISP-DM methodology. This was an enhanced version of the standard CRISP-DM methodology, as standard CRIP-DM was followed by many researchers for data science projects [67]. This methodology is a good reference for future researches dealing with data driven approaches for solving real time problems in industries. In standard CRISP-DM methodology the data storage is readily available and the data understanding, data visualisation, data preprocessing, and modelling is done on the data which is already present. But when it comes to real life scenarios the data collection is an iterative process, as a relevancy of a data source for a particular project can be confirmed only after understanding and visualizing the data-sets from those sources [19]. The enhanced CRISP-DM methodology introduced in this project would give an a better insight about this real time scenario.

5.4 Practical contribution

In this era of Industry 4.0, automation and smart manufacturing are important for the success of any manufacturing company. Predictive maintenance helps to capitalise the RUL of a machine and thus helps to reduce the maintenance cost and avoid unpredicted breakdowns of machines [51]. This thesis project is a data driven approach for finding the remain useful time of a gluing machine, from the body shop of Volvo Cars. The project was a first step towards attaining predictive maintenance for the entire body shop, so the challenges faced and results obtained from the project will be a background for future projects in the company and also would help similar manufacturing industries trying to implement data driven approaches for decision making. The work was formulated using the most commonly used Python language, which is one of the prominently used open source language for data science applications. The formulation of the Python codes for data visualisation, data preprocessing, modelling and evaluation are easily adaptable to a new dataset. This would enable the maintenance department to easily reuse the codes for there future projects.

6
Conclusion

This thesis is the first practical step towards achieving the goal of smart maintenance at Volvo Cars. The final output of the thesis would support the maintenance team for scheduling and planning the maintenance activities at the body shop. The deliverable of this project is a ML architecture and ML model which is capable of predicting the breakdowns in the gluing machine by predicting the severe alarms triggering in the machine. Volvo Cars presently follow value driven maintenance, where optimisation of scheduling and planning is done with the feedback mechanism and individual experience. This project would enable the maintenance team to make data driven decisions for planning and scheduling of maintenance activities in the gluing machine.

The project methodology utilized for this thesis was an enhanced CRISP-DM, as the data collection, data understanding, data preparation, modelling and evaluation was an iterative process due to non-ideal real time scenarios at industries. Several data visualisations were preformed and many insights were obtained about the trends and patterns inside the data. These visualisations were quintessential for data pre-processing and formulating assumptions. The data preparation was carried out to merge relevant data from different data sources and to transform it in to suitable formats for ML. Since the auto recorded breakdown data was unavailable, the triggering of severe alarms were assumed as breakdowns. In this thesis project ARIMA model and LSTM model were used for prediction of multivariate time-series data from multiple sources. The best results was given by the LSTM model. The ML model is able to predict the day in which the breakdown is going to occur. The prediction can be done for the concluding day of training data and the days that can be predicted depends on the amount of historical data used for training.

The formulation of codes were easily adaptable, to a new dataset. When an automated breakdown log is available for the maintenance team, the severe alarms can be easily replaced with the breakdown data in order to achieve a much more efficient prediction of failures. From the insights gained during the thesis project, a set of recommendations for improvements are being discussed. These recommendations aims at providing a road-map for future data driven projects at Volvo Cars.

Bibliography

[1] Teamster AB. *T2X2 BiW Operator manual.*

[2] ABB. Irb 6700 data. `https://new.abb.com/products/robotics/industrial-robots/irb-6700/irb-6700-data`, 2020. [Online; accessed 23-July-2021].

[3] MSK Abhilash, Amrita Thakur, Deepa Gupta, and B Sreevidya. Time series analysis of air pollution in bengaluru using arima model. In *Ambient Communications and Computer Systems*, pages 413–426. Springer, 2018.

[4] A Abu-Samah, MK Shahzad, E Zamai, and A Ben Said. Failure prediction methodology for improved proactive maintenance using bayesian approach. *IFAC-PapersOnLine*, 48(21):844–851, 2015.

[5] War Ahmed and Mehrdad Bahador. The accuracy of the lstm model for predicting the s&p 500 index and the difference between prediction and backtesting, 2018.

[6] Umar M Al-Turki, Tahir Ayar, Bekir Sami Yilbas, and Ahmet Ziyaettin Sahin. Maintenance in manufacturing environment: An overview. *Integrated maintenance planning in manufacturing systems*, pages 5–23, 2014.

[7] Mohammad Almasarweh and Sadam Alwadi. Arima model in predicting banking stock market data. *Modern Applied Science*, 12(11):4, 2018.

[8] Nagdev Amruthnath and Tarun Gupta. A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*, pages 355–361, 2018.

[9] Adebiyi A. Ariyo, Adewumi O. Adewumi, and Charles K. Ayo. Stock price prediction using the arima model. In *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pages 106–112, 2014.

[10] R Artusi, P Verderio, and EJTIjobm Marubini. Bravais-pearson and spearman correlation coefficients: meaning, test of hypothesis and confidence interval. *The International journal of biological markers*, 17(2):148–151, 2002.

[11] Saqib Ejaz Awan, Mohammed Bennamoun, Ferdous Sohel, Frank Mario Sanfilippo, and Girish Dwivedi. Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics. *ESC heart failure*, 6(2):428–435, 2019.

[12] Marcia Baptista, Shankar Sankararaman, Ivo P de Medeiros, Cairo Nascimento Jr, Helmut Prendinger, and Elsa MP Henriques. Forecasting fault events for predictive maintenance using data-driven techniques and arma modeling. *Computers & Industrial Engineering*, 115:41–53, 2018.

[13] Marcia Baptista, Shankar Sankararaman, Ivo P de Medeiros, Cairo Nascimento Jr, Helmut Prendinger, and Elsa MP Henriques. Forecasting fault events for predictive maintenance using data-driven techniques and arma modeling. *Computers & Industrial Engineering*, 115:41–53, 2018.

[14] Ebru Turanoglu Bekar, Per Nyqvist, and Anders Skoogh. An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study. *Advances in Mechanical Engineering*, 12(5):1687814020919207, 2020.

[15] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.

[16] Imanol Bilbao and Javier Bilbao. Overfitting problem and the over-training in the era of data: Particularly for artificial neural networks. In *2017 eighth international conference on intelligent computing and information systems (ICICIS)*, pages 173–177. IEEE, 2017.

[17] Jon Bokrantz, Anders Skoogh, Cecilia Berlin, Thorsten Wuest, and Johan Stahre. Smart maintenance: an empirically grounded conceptualization. *International Journal of Production Economics*, 223:107534, 2020.

[18] Daniel Bumblauskas, Douglas Gemmill, Amy Igou, and Johanna Anzengruber. Smart maintenance decision support systems (smdss) based on corporate big data analytics. *Expert systems with applications*, 90:303–317, 2017.

[19] Andriy Burkov. *The hundred-page machine learning book*, volume 1. Andriy Burkov Canada, 2019.

[20] Kailin Cao, Ting Hu, Zishuo Li, Guoshuai Zhao, and Xueming Qian. Deep multi-task learning model for time series prediction in wireless communication. *Physical Communication*, 44:101251, 2021.

[21] Thyago P Carvalho, Fabrízzio AAMN Soares, Roberto Vita, Roberto da P Francisco, João P Basto, and Symone GS Alcalá. A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137:106024, 2019.

[22] Tianfeng Chai and Roland R Draxler. Root mean square error (rmse) or mean absolute error (mae). *Geoscientific Model Development Discussions*, 7(1):1525–1534, 2014.

[23] Yu-Wei Chang and Meng-Yuan Liao. A seasonal arima model of tourism forecasting: The case of taiwan. *Asia Pacific journal of Tourism research*, 15(2):215–221, 2010.

[24] Peng Chen, Hongyong Yuan, and Xueming Shu. Forecasting crime using the arima model. In *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, volume 5, pages 627–630. IEEE, 2008.

[25] Francois Chollet et al. Keras, 2015.

[26] Brad Cline, Radu Stefan Niculescu, Duane Huffman, and Bob Deckel. Predictive maintenance applications for machine learning. In *2017 Annual Reliability and Maintainability Symposium (RAMS)*, pages 1–7, 2017.

[27] Kunyuan Deng, Xiaoyong Zhang, Yijun Cheng, Zhiyong Zheng, Fu Jiang, Weirong Liu, and Jun Peng. A remaining useful life prediction method with

long-short term feature processing for aircraft engines. *Applied Soft Computing*, 93:106344, 2020.

[28] B Uma Devi, D Sundar, and P Alli. An effective time series analysis for stock trend prediction using arima model for nifty midcap-50. *International Journal of Data Mining & Knowledge Management Process*, 3(1):65, 2013.

[29] Olga Fink. Data-driven intelligent predictive maintenance of industrial assets. In *Women in Industrial and Systems Engineering*, pages 589–605. Springer, 2020.

[30] Freceena Francis and Maya Mohan. Arima model based real time trend analysis for predictive maintenance. In *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 735–739. IEEE, 2019.

[31] Michal S Gal and Daniel L Rubinfeld. Data standardization. *NYUL Rev.*, 94:737, 2019.

[32] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. *Neural computation*, 12(10):2451–2471, 2000.

[33] Alex Graves. Long short-term memory. In *Supervised sequence labelling with recurrent neural networks*, pages 37–45. Springer, 2012.

[34] Hashem M Hashemian. State-of-the-art predictive maintenance techniques. *IEEE Transactions on Instrumentation and measurement*, 60(1):226–236, 2010.

[35] Amirreza Heidari and Dolaana Khovalyg. Short-term energy use prediction of solar-assisted water heating system: Application case of combined attention-based lstm and time-series decomposition. *Solar Energy*, 207:626–639, 2020.

[36] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[37] Yuxiu Hua, Zhifeng Zhao, Rongpeng Li, Xianfu Chen, Zhiming Liu, and Honggang Zhang. Deep learning with long short-term memory for time series prediction. *IEEE Communications Magazine*, 57(6):114–119, 2019.

[38] Patrick Jahnke. Machine learning approaches for failure type detection and predictive maintenance. *Technische Universität Darmstadt*, 19, 2015.

[39] Abhinav Jain, Hima Patel, Lokesh Nagalapatti, Nitin Gupta, Sameep Mehta, Shanmukha Guttula, Shashank Mujumdar, Shazia Afzal, Ruhi Sharma Mittal, and Vitobha Munigala. Overview and importance of data quality for machine learning tasks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3561–3562, 2020.

[40] Tina Jakaša, Ivan Andročec, and Petar Sprčić. Electricity price forecasting — arima model approach. In *2011 8th International Conference on the European Energy Market (EEM)*, pages 222–225, 2011.

[41] Philip Jansson and Hugo Larsson. Arima modeling: Forecasting indices on the stockholm stock exchange, 2020.

[42] Christopher R Johnson, Mirko Montanari, and Roy H Campbell. Automatic management of logging infrastructure. In *CAE Workshop on Insider Threat*, 2010.

[43] Shashidhar Kaparthi and Daniel Bumblauskas. Designing predictive maintenance systems using decision tree-based machine learning techniques. *International Journal of Quality & Reliability Management*, 2020.

[44] KDnuggets. 7 techniques to handle imbalanced data. https://www.kdnuggets.com/2017/06/7-techniques-handle-imbalanced-data.html. [Online; accessed 20-July-2021].

[45] Weiheng Kong, Lili He, and Hailong Wang. Exploratory data analysis of human activity recognition based on smart phone. *IEEE Access*, 9:73355–73364, 2021.

[46] Björn Kroll, David Schaffranek, Sebastian Schriegel, and Oliver Niggemann. System modeling based on machine learning for anomaly detection and predictive maintenance in industrial plants. In *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, pages 1–7, 2014.

[47] Akhilesh Kumar, Ratna Babu Chinnam, and Finn Tseng. An hmm and polynomial regression based approach for remaining useful life and health state estimation of cutting tools. *Computers & Industrial Engineering*, 128:1008–1014, 2019.

[48] Heiner Lasi, Peter Fettke, Hans-Georg Kemper, Thomas Feld, and Michael Hoffmann. Industry 4.0. *Business & information systems engineering*, 6(4):239–242, 2014.

[49] Sunghoon Lim, Sun Jun Kim, YoungJae Park, and Nahyun Kwon. A deep learning-based time series model with missing value handling techniques to predict various types of liquid cargo traffic. *Expert Systems with Applications*, page 115532, 2021.

[50] He Liu, Wanqing Song, Yuhui Niu, and Enrico Zio. A generalized cauchy method for remaining useful life prediction of wind turbine gearboxes. *Mechanical Systems and Signal Processing*, 153:107471, 2021.

[51] Hans Löfsten. Measuring maintenance performance–in search for a maintenance productivity index. *International Journal of Production Economics*, 63(1):47–58, 2000.

[52] Juan José Montero Jimenez, Sébastien Schwartz, Rob Vingerhoeds, Bernard Grabot, and Michel Salaün. Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. *Journal of Manufacturing Systems*, 56:539–557, 2020.

[53] Gerhard Nahler. Pearson correlation coefficient. In *Dictionary of Pharmaceutical Medicine*, pages 132–132. Springer, 2009.

[54] Christopher Olah. Understanding lstm networks. *Neural computation*, 12(10):2451–2471, 2000.

[55] Tom O'Malley, Elie Bursztein, James Long, François Chollet, Haifeng Jin, Luca Invernizzi, et al. Keras tuner. https://github.com/keras-team/keras-tuner, 2019.

[56] Purna Chandra Padhan et al. Application of arima model for forecasting agricultural productivity in india. *Journal of Agriculture and Social Sciences*, 8(2):50–56, 2012.

[57] Marina Paolanti, Luca Romeo, Andrea Felicetti, Adriano Mancini, Emanuele Frontoni, and Jelena Loncarski. Machine learning approach for predictive maintenance in industry 4.0. In *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, pages 1–6, 2018.

[58] Andrea Stefano Patrucco, Davide Luzzini, and Stefano Ronchi. Achieving innovation through supplier collaboration: the role of the purchasing interface. *Business Process Management Journal*, 2017.

[59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[60] H Peiris. A seasonal arima model of tourism forecasting: The case of sri lanka. *Journal of Tourism, Hospitality and Sports*, 22(1):98–109, 2016.

[61] Foster Provost and Tom Fawcett. Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1):51–59, 2013.

[62] Rune Prytz. *Machine learning methods for vehicle predictive maintenance using off-board and on-board data.* PhD thesis, Halmstad University Press, 2014.

[63] Tony Rosqvist, Kari Laakso, and Markku Reunanen. Value-driven maintenance planning for a production plant. *Reliability Engineering & System Safety*, 94(1):97–110, 2009.

[64] Lakshidaa Saigiridharan. Dynamic prediction of repair costs in heavy-duty trucks, 2020.

[65] Renato Cesar Sato. Disease management with arima model in time series. *Einstein (Sao Paulo)*, 11:128–131, 2013.

[66] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[67] Christoph Schröer, Felix Kruse, and Jorge Marx Gómez. A systematic literature review on applying crisp-dm process model. *Procedia Computer Science*, 181:526–534, 2021.

[68] scikit-learn developers. sklearn.preprocessing.StandardScaler. `https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html`, 2020. [Online; accessed 06-August-2021].

[69] Masoud Shayganmehr, Anil Kumar, Jose Arturo Garza-Reyes, and Md. Abdul Moktadir. *Journal of Cleaner Production*, 281:125280, 2021.

[70] Xiao-Sheng Si, Wenbin Wang, Chang-Hua Hu, and Dong-Hua Zhou. Remaining useful life estimation–a review on the statistical data driven approaches. *European journal of operational research*, 213(1):1–14, 2011.

[71] Mukund Subramaniyan, Anders Skoogh, Hans Salomonsson, Pramod Bangalore, and Jon Bokrantz. A data-driven algorithm to predict throughput bottlenecks in a production system based on active periods of the machines. *Computers Industrial Engineering*, 125:533–544, 2018.

[72] Gian Antonio Susto, Andrea Schirru, Simone Pampuri, Seán McLoone, and Alessandro Beghi. Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, 11(3):812–820, 2014.

[73] Richard Taylor. Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, 6(1):35–39, 1990.

[74] David L Waltz. The prospects for building truly intelligent machines. *Daedalus*, pages 191–212, 1988.

[75] Weixin Wang. Joint prediction of remaining useful life and failure type of train wheelsets: A multi-task learning approach. *arXiv preprint arXiv:2101.03497*, 2021.

[76] Ji-Yan Wu, Min Wu, Zhenghua Chen, Xiaoli Li, and Ruqiang Yan. A joint classification-regression method for multi-stage remaining useful life prediction. *Journal of Manufacturing Systems*, 58:109–119, 2021.

[77] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*, 2014.

[78] Shichao Zhang, Chengqi Zhang, and Qiang Yang. Data preparation for data mining. *Applied artificial intelligence*, 17(5-6):375–381, 2003.

[79] Tianjun Zhang, Shuang Song, Shugang Li, Li Ma, Shaobo Pan, and Liyun Han. Research on gas concentration prediction models based on lstm multidimensional time series. *Energies*, 12(1):161, 2019.

[80] Xuewen Zhang, Yan Qin, Chau Yuen, Lahiru Jayasinghe, and Xiang Liu. Time-series regeneration with convolutional recurrent generative adversarial network for remaining useful life estimation. *IEEE Transactions on Industrial Informatics*, 2020.

[81] Jiakun Zhao, Ruifeng Zhang, Zheng Zhou, Si Chen, Ju Jin, and Qingfang Liu. A neural architecture search method based on gradient descent for remaining useful life estimation. *Neurocomputing*, 438:184–194, 2021.

A

Appendix 1

Actual vol	Cycle name	Duration	Num beads	Ref vol	Timestamp	Type name	Type num	cycle num
28.88	C	51.954	12	0.0	2021-04-26 00:00:38.236	0	1.0	485436
4.76	B	3.440	1	4.80	1	6400321,0	2.371	485436
0.63	B	6.883	1	0.64	2	6400319,0	0.536	485436
2.32	B	8.251	1	2.35	3	6400315,0	1.361	485436
2.34	B	10.684	1	2.35	4	6400317,0	1.352	485436
....	...	...	...	....	...	...	...	...
....	...	...	...	....	...	...	...	...
....	...	...	...	....	...	...	...	...
0.61	B	21.515	1	0.65	8	6400324,0	0.560	511153
2.33	B	22.967	1	2.35	9	6400318,0	2.003	511153
2.31	B	26.362	1	2.35	10	6400316,0	1.436	511153
0.63	B	28.678	1	0.65	11	6400320,0	0.565	511153
4.08	B	30.119	1	4.10	12	6400322,0	2.356	511153

Table A.1: Sample Production log

SI no	Alarm ID	Priority
1	4006	B
2	50	B
3	174	B
4	172	B
5	0	B
6	6222	A

Table A.2: Severe alarms

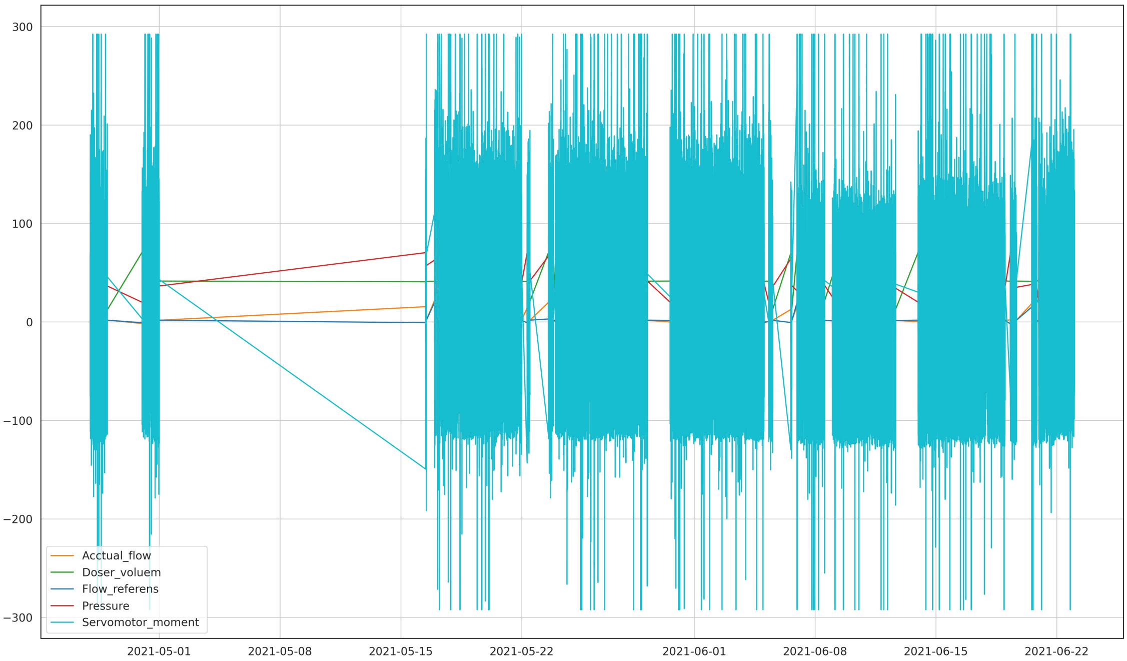


Figure A.1: Doser vs time plot

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY