Analysis of the Effectiveness of Magnetic Filters of Shovel Hydraulic System and Successive Hydraulic Failure Prediction using Data-Driven Approaches

by

Prerita Odeyar

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

In

Mining Engineering

Department of Civil and Environmental Engineering

University of Alberta

ABSTRACT

Surface mining dominates the world's production of minerals. Currently, almost all non-metallic minerals (more than 95 %), most metallic minerals (more than 90%), and a significant fraction of coal (more than 60%) are mined by surface mining methods. The hydraulic shovel-truck system forms the backbone of the surface mining industry. With the complexity of the shovel and the tough working conditions, hydraulic system components are more susceptible to failures and are often the most expensive system to repair. The failures are highly unpredictable, associated with high follow-on failure rates, and cause lengthy downtime. Wear and debris-related failures are the most common cause of hydraulic failures. The most common contaminant of wear and debris failures is iron contamination which is a constant issue as it is the by-product of machine operation and component failures. As the component wear gradually increases in the hydraulic system, it leads to debris accumulation in the oil that might trigger multiple component failures allowing the contamination to spread rapidly, resulting in catastrophic damage. Hence, companies are constantly aiming to improve the hydraulic system condition and prevent major failures by using different methods like condition monitoring methods, increasing hydraulic filter capacities, and implementing new methods like introducing new filters and using statistical data-driven techniques to predict and stop failures.

Magnetic filters are the newly introduced hydraulic filters that use the most advanced magnetic technology to prevent contamination and system wear failures. The first part of the research aims to quantitively evaluate magnetic filter performance on hydraulic system components and test their effectiveness using failure data of hydraulic shovels.

The second part of the research focuses to further enhance the mitigation of catastrophic hydraulic failures and increase component life with the use of data-driven techniques. The aim is to assess

the likelihood of successive failures of hydraulic components in the next 1000 hours of operation after a component failure. Historical failures are studied using different machine learning algorithms and probability of successive failures are predicted based on the failure patterns identified.

PREFACE

This dissertation is an original and independent work by Prerita Odeyar under the supervision of Professor Robert Hall and Professor Derek Apel at the University of Alberta. No part of this thesis has been previously published in any form.

ACKNOWLEDGEMENTS

I would like to take this opportunity to acknowledge all the individuals who gave me their supervision, guidance, dedication, and support to make this research successful.

First, I would like to express my greatest gratitude to my supervisor Prof. Robert Hall for providing me with this opportunity to accomplish one of my biggest goals. His support, time, guidance, and advice throughout the research have been very valuable. It was a great privilege and an honor to work and study under his guidance. I would also like to express my deep and sincere gratitude to my supervisor Prof. Derek Apel who has been a constant support in successfully completing my research and for the invaluable guidance and advice, he has provided to me in my research and thesis work. I am extremely grateful for all his assistance.

I would also like to express my gratitude and appreciation to the thesis examining committee members, Prof. Hooman Askari and Dr. Zhang, for their invaluable time and feedback.

I am especially thankful to my husband, Gautam Punit, for his constant support and being instrumental in achieving my dreams. He has been my greatest strength since we met. I am very grateful for his love and understanding which helped in completing this research work. I am forever in debt to my parents for all their sacrifices to give me the best in life. Without their wishes and support, nothing would have been possible. I am also indebted to the rest of my family members especially my sister, Prarthana, who has been encouraging and motivating me throughout this journey. She has been important in every step of my life. Lastly, I am also grateful for my friends and the support that I am privileged to have in life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

This chapter provides an overview of this research. It includes a general background of the research topic, the objectives of this research, methodology adopted to achieve the objectives and describes the organization of this thesis.

1.1 General Background

Large mining equipment such as hydraulic shovels are critical in a mine's production system, and unplanned equipment failures negatively affect their ability to meet production targets and generate revenue. Hydraulic systems of the large shovels working in mines are often subjected to contamination that results from abrasive wear particles generation from the system components and dust particles from air drawn into the hydraulic tank. Tiny particles break off hydraulic components as normal wear occurs. These particles build up in the hydraulic system often leading to catastrophic failures that need the whole system to be repaired and replaced. These failures are proven to be very costly and often lead to extended downtime resulting in production loss. Proper maintenance of oil and prior knowledge about potential failures can prevent catastrophic failures by strategizing maintenance activities that in turn can reduce failure costs and downtime of the equipment.

Several filters have been introduced in the hydraulic system that trap particles of different sizes to prevent system contamination, enhance component life and mitigate catastrophic failures. The introduction of new magnetic filters (Mag-filters) uses the most advanced magnetic technology and can trap metal contaminants up to 4µ sizes in the system and prevent failures. The introduction of Mag-filters is widely regarded as positive but there is no existing framework to analyze and quantify the impacts of filter effectiveness. Hence, the first part of this analysis tries to establish a framework for pre-post analysis of implementing a solution for mine equipment failure mitigation.

Monitoring and recording equipment conditions/failures regularly and making failure predictions based on collected data will help minimize maintenance costs and further prevent catastrophic failures of hydraulic systems. Several data-driven techniques are now used in predictive maintenance areas to know failures ahead of time so that maintenances interventions can be planned accordingly to prevent unplanned downtime of a run-to-failure strategy. From the equipment failure perspective, machine learning can be useful to identify and replace or repair a component before a fault happens and restore the original condition of the equipment to maintain reliability especially in an application where avoiding downtime is likely more valuable than extracting all the usable life out of a component. The algorithms use historical failure data or vibration/ condition monitoring data of the equipment to study failures and make predictions. This directly would lead to decreased downtime and achieving expected levels of production. In addition, machine learning helps predict future failures to accurately schedule maintenance operations. ML techniques are designed to derive knowledge out of existing data.

Although machine learning techniques have gained a lot of popularity in various engineering domains and have been successfully implemented in failure analysis and failure predictions, the application is still not widespread in mining engineering to detect faults or predict failures. Hence the second part of this research tries to use machine learning methods in assessing the likelihood of successive failures that are impacted by other failures which can help reduce maintenance costs, downtime costs, labor costs and increase production time leading to an overall increase in profits.

1.2 Research Objectives

Addition of magnetic filters in the hydraulic system of the shovel fleet is generally regarded as positive by the company that manages shovel hydraulic failures, but the benefits have been anecdotal. Hence, the first part of this research focuses on understanding the effectiveness and quantitative impacts of Mag-filters. Different methods are used to compare the effectiveness of Mag-filters and a pre-post analysis is done using different key performance metrics to quantify impact. Hydraulic failures are inevitable. In order to enhance maintenances practices, improve hydraulic system conditions and to further reduce sudden and unexpected hydraulic failures, the second part of the research aims to predict successive hydraulic system failures in the next 1000 hours of operation, given that a hydraulic component has already failed.

In order to understand successive failures, data-driven techniques are applied that make use of historical data to predict future failures. Using feature variables influencing failure and the lag variables, data-driven approaches are used that help identify and understand failures and predict the probabilities of successive failures. Using historical failure data, successive hydraulic failures that are most probable to occur in the next 1000 hours of operation after a hydraulic component failure are identified. To be more precise on the failure analysis, no fabricated data or data generated by simulations is used in this research.

In summary, the objective of the research can be divided into two parts:

- The first part aims to quantify the performance and effectiveness of hydraulic magnetic filters introduced in a company's shovel fleet
- The second objective of the study is to predict successive hydraulic failures given a hydraulic component failure by developing various machine learning algorithms that use historical shovel failure data

1.3 Research Methodology

To achieve the objectives of the analysis, a complete framework was developed to quantify the impacts of magnetic filters using historical data of giant hydraulic shovels that are working in the mines. The conditions of the components before installation of Mags are compared to the component conditions post-installation and a weighted analysis method is established to quantify the overall efficiency of the system before and after installation of Mag-filters. Different machine learning algorithms are also used in the analysis to identify successive failure probability.

This thesis demonstrates the use of statistical approaches like I.I.D Tests, ANOVA Tests, Bayes theorem, probability distributions, and data-driven approaches like data collection, data processing, feature engineering, hyperparameter tuning, and implementation of artificial intelligence models to analyze hydraulic system failures. Popular software and platforms such as Tableau, R, and Python were used for data visualization and statistical computing. Toolkits and packages such as pandas, Matplotlib, Seaborn, and Scikit-Learn were employed in this research.

To achieve the objectives of this research, the following tasks have been completed:

- **Literature Review**: The literature review required for the analysis was divided into 4 sections and an extensive study in each of the following sections was reviewed:
    - Hydraulic system overview, filter types, hydraulic failures, and criticality levels
    - Oil analysis methods for system condition monitoring
    - Statistical tests and reliability assessment methods of mining equipment
    - Machine learning approaches for failure predictions and predictive maintenance
- **Data Collection**: Information regarding historical failures and failure conditions was collected from three different data sources for three units of giant shovels working in different mines. In order to accurately model and analyze real scenarios, no fabricated data or simulated data was used in this research. Component changeout data recorded for each failure of the hydraulic system for different units of hydraulic shovels were collected from the company database. Failure details and conditions in which these failures occurred were also used to evaluate different metrics and also to analyze failure patterns. Data

transformation and filtration were also done in this step according to the analysis requirements.

- **Developing framework for assessment of quantitative impacts of Mag-filters:** Different methods such as component life analysis, reliability assessment, oil analysis, particle count methods, comeback failure rate and cost analysis methods were used to quantify the effects of Mag-filters on the hydraulic system performance. In addition, critical components of the system were identified, and their most common failure causes, and effects were studied.

- **Developing ML models for prediction of the probability of downstream failures in the hydraulic system**: Different ML-based algorithms were used to predict the chance of a successive failure given a component has failed and probability of downstream component failures in the hydraulic system using historical data of failure events. Algorithms were evaluated based on performance criteria identified and the algorithm that best describes the failure data was chosen.

1.4 Organization of Thesis

To achieve the objective of this research, the thesis is divided into five chapters.

Chapter 1 (Introduction) – provides a general overview of the project, an introduction to the area of research, the problem statement, the objective of the research, and the methodology adopted to achieve the objective.

Chapter 2 (Literature Review) – provides an extensive review of the already existing work in this area of research. The chapter is divided into four sections. The first part provides an overview of the hydraulic system of giant shovels used in mining, the components that constitute the hydraulic system, hydraulic circuits, and filter types, hydraulic failure causes and modes, and particle contamination causes and effects in the hydraulic system. The second section provides an overview of the condition-based monitoring of the hydraulic system using hydraulic oil samples. The section discusses existing oil analysis methods to detect failures, factors influencing oil analysis procedures, ISO cleanliness code to measure system contamination, ICP particle count method, and the importance of oil particle count in failure analysis. The third section focuses on different statistical techniques and reliability analysis methods used in equipment performance measurements. The fourth section provides a review based on different ML approaches used in fault detection and predictive maintenance methods.

Chapter 3 (Quantitative Impacts of Hydraulic Mag-filters) presents different methods used to compare the effectiveness of magnetic filters and discusses the criticality of different component failures. The chapter also provides iron content oil analysis basis. The quantitative impacts of the filters are assessed by component life analysis, reliability methods, oil and particle count analysis, comeback failure analysis, and cost analysis methods. The analysis helps in understanding the influence of Mag-filters on the life cycle of hydraulic components and valuates if the system performance has improved post-installation of the filters.

Chapter 4 (Machine Learning algorithms for failure predictions) introduces a method to predict successive failures given a component failure and predict the probability of downstream failures in the hydraulic circuits. The chapter establishes the benefits of using different machine learning algorithms to comprehensively understand failures and predict future failures more accurately. Both supervised and unsupervised learning methods were studied to detect faults using data-driven techniques. Different ML algorithms are explored to find failure patterns using historical hydraulic component changeout data. The model that best describes the data is selected.

Chapter 5 (Conclusion) presents the summary and conclusions of this research. This chapter also discusses the significance of this research and recommends future work in this area of failure detection and predictive maintenance of the hydraulic system and in general.

2. LITERATURE REVIEW

This chapter provides an overview of literature related to this research. The first part of the chapter presents an overview of hydraulic system design, different hydraulic filters, types of hydraulic failures and their effects. The second part of the chapter presents overview of oil analysis methods for condition monitoring including particle count lab analysis methods and ISO cleanliness code for contamination detection. The third part of this chapter presents different statistical tests and reliability methods and their application for time to failure analysis. The fourth section presents an overview of machine-learning based algorithms for failure analysis, predictive maintenance, and their application.

2.1 Hydraulic Shovels Overview and Working

2.1.1 Introduction

"Surface mining methods dominate the world production of minerals. Surface mining accounts for over 30 billion tonnes of ore and waste materialized each year, that accounts for nearly 25 billion tonnes" (V.Ramani, 2012).

Surface mining today is not feasible without using large excavation equipment that is an essential part of the mining process. "In surface mining operations, truck and shovel systems are the most prominent type of haulage system" (Yaghini, Hall, & Apel, 2020). A single-bucket mining shovel is one of the most used machines in surface mining for digging and loading material. Their cycle includes digging, moving the filled bucket to the truck, unloading the excavated material from the bucket, and returning to the digging face (Andreev, 2015).

Shovels are heavy earth-moving machinery that are popularly used in construction and mining industries. They are made up of a moving body, a swing body, and a front digging manipulator to conduct digging operations. The digging manipulator comprises many moving parts, the most important of which are the boom, stick, and bucket, all connected by rotary joints. (Mitrev et al., 2011). Rope and hydraulic shovels are the two most important types of heavy machinery. They are found at almost all modern large-scale surface mining sites.

2.1.2 Hydraulic Vs Rope Shovel

Equipment selection is regarded as one of the most important decisions that affect the mine design, annual production, and other economic parameters in mining (Snetkov & Kosolapov, 2019). The advantages and disadvantages of various established factors are weighed to assess the feasibility of using a certain type of excavation equipment. The production capacity of the mine; properties of overburden and minerals, their state of occurrence; slope angles, and bench heights are some other considerations that will influence the choice (Burt & Caccetta, 2018).

Rope shovels and hydraulic excavators have advantages over one another, depending on the type of mineral, and the geotechnical and climatic conditions of the region. Electrical rope shovels are typically thought to be more reliable and long-lasting and are often simpler and less expensive to maintain. They are more resistant to temperature extremes and, as a result, are capable of successfully offering higher digging force. Hydraulic shovels, on the other hand, are typically preferred where complicated geological conditions exist as they are more technologically robust, weigh less, and can be equipped as a "front shovel" or a "backhoe." It offers qualitative selective excavation and has a higher power delivery for excavation and has much higher mobility. The hydraulic shovels' lower static ground bearing pressure makes them ideally suited to soft mine floor conditions (Ozdogan, 2003) (Kelsh, 2008).

Andreev (2017) carried out investigations and data analysis for evaluating hydraulic shovel and rope shovel major costs in operation and discussed that the rope shovel's main advantages over hydraulic shovels are lower cost of excavation and longer expected life. On the other hand, hydraulic shovels are considerably less expensive, technologically more flexible, and capable of higher production, although they require higher-quality maintenance and servicing. Hydraulic shovels have a higher physical availability compared to electric cable shovels.

Hydraulic shovels are the most common excavation machine used in surface mines. "Although the electrification of their powertrains has already begun, hydraulic systems continue to power the main actuators, even in a hybrid s excavator" (Park, Yoo, Ahn, Kim, & Shin, 2020).

It is safe to claim that the hydraulic shovel excavator generation represents a true breakthrough in loading technology and a significant turning point in the global machinery industry. Renowned companies including Caterpillar, Hitachi, Komatsu, Liebherr, Terex Mining, Demag, and others are manufacturing hydraulic shovels in various styles, bucket capacities, operating weights, and other features. They range in size from mini shovels with bucket capacities of 0.008 m^3 and working weights of 6.7t to super large excavators with bucket capacities of 52 m^3 and working weights of 900t. This provides a relatively open market for hydraulic shovels, allowing all consumers to pick and invest in the appropriate equipment based on their working conditions and financial capabilities (Bui & Drebenstedt, 2009).

2.1.3 Hydraulic Shovel Design and Working

The hydraulic shovels consist of three major assemblies, an upper structure, a lower structure, and an attachment. The upper structure consists of a machinery dwelling, an operator's cab, and a counterweight. The propel drive and crawler mechanism are housed in the lower frame, which also serves as a stable base for the machine (Andreev, 2015). A hydraulic cylinder powers these elements through a transmission mechanism, which is connected to the excavator links and working tools. Extension and retraction of hydraulic cylinders as commanded by the operator, actuate these elements (Mitrev, Gruychev, & Pobegailo, 2011). Figure 1 represents hydraulic shovel front design and Figure 2 represents a simplified diagram of hydraulic system of shovels.

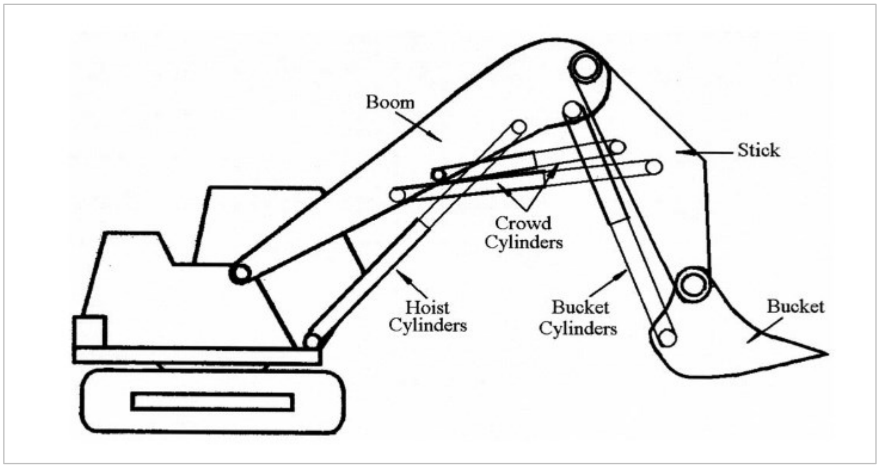


Figure 1.Hydraulic front shovel design (Frimpong, Hu, & Inyang, 2008)

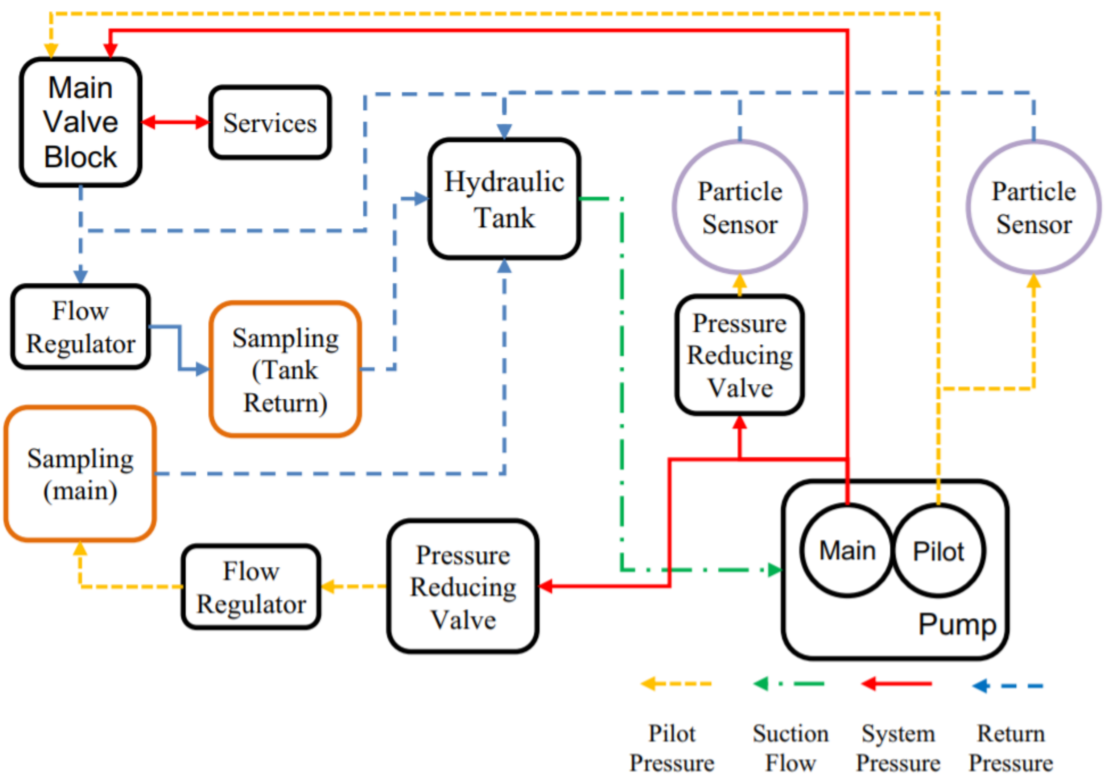


Figure 2. Simplified diagram of shovel hydraulic system (Felix Ng, Jennifer A. Harding, and Jacqueline Glass 2016)

The main components of a hydraulic system are a reservoir, pump, control valve, motors, actuators, hoses and pipes, hydraulic fluid, oil cooler, and filters (Brain, 2021). Every hydraulic system has a tank to supply hydraulic fluid to the pump and store fluid returning from the hydraulic circuit. Before the fluid re-enters the pump, the reservoir has enough volume to store the returning fluid giving it enough time to cool and allow air to escape. Hydraulic pumps generate fluid flow and transform mechanical energy into hydraulic energy (Khayal, 2014). The system includes both main and pilot pumps. The main pump is a variable displacement axial plunger pump, and the engine drives the main pump through a coupler. The pilot gear pump is directly attached to the main pump's driveshaft in the pilot circuit. In hydraulic circuits, valves are used to regulate pressure, volume flow rate, and flow direction which is in turn used to maintain a safe pressure level in a hydraulic circuit. A hydraulic actuator is a system that transforms hydraulic energy into mechanical energy. There are two types of actuators, rotary and linear (Khayal, 2014) (Brain, 2021).

The actuators in a hydraulic system are powered by pressurized fluid. Hydraulic oil is used to actuate power to the shovel operations. The cylinder press exerts pressure on small quantities of oil to generate a large amount of power. The energy is transferred to the piston, which will then be pushed upward to perform the shovel operation (Strelnikov, Markov, Rattmann, & Weber, 2018).

2.1.4 Hydraulic System Failures

Hydraulic system components are often susceptible to failures. With the complexity of the shovel hydraulic circuits and the tough working conditions they endure, the reliability of such systems remains a major concern (An & Sepehri, 2005). Around 80% of equipment shortages and component failures are caused by hydraulic breakdown.

Hydraulic failure studies include broad-based surveys of the hydraulic failures of the mobile equipment working on-site. Hydraulic issues are often caused by faulty pumps, system breaches, or temperature issues (Baker, 2020). Air contamination into a hydraulic system often causes the problems of aeration and cavitation, both of which can lead to severe damage of the hydraulic system over time by wearing down the pump and surrounding components and contaminating hydraulic fluids, and even overheating the system.

Aeration occurs when air enters the pump cavity from an outside source. Aeration is usually caused by loose connections or leaks in the system. Aeration problems can lead to constant noise in the hydraulic system when the pump runs (Baker, 2020) (Smiley, 2021). Water contamination is also a common issue in hydraulic systems, and it is often caused by system leaks or condensation caused by temperature changes. Hydraulic components can deteriorate over time due to oxidation and freeze damage (Hu & Men, 2020).

Hydraulic systems that run too hot or too cold can cause severe problems over time. Hydraulic fluids can thin due to heat, preventing lubrication and increasing the likelihood of leakage (Smiley, 2021). Hydraulic fluid can oxidize and thicken in extreme heat. This fluid thickening can cause buildups in the system, which restricts flow and restricts flow and reduces the system's ability to dissipate heat. Low temperatures increase hydraulic oil viscosity, making it more difficult for the oil to reach the pump. Cavitation may be caused by putting systems under load until the oil reaches 70 degrees or more.(Baker, 2020) (Smiley, 2021) (Hu & Men, 2020). All these factors can greatly enhance component wear, which results in the contamination of hydraulic oil.

Hydraulic equipment has improved significantly in complexity and operating pressures over the last 30 years. As a result, not only is it more costly to repair modern hydraulic equipment when it fails, but proactive maintenance is imperative to increase system life and reduce operating costs

(Khayal, 2014). Oil contamination is the major source of failure and wear of hydraulic system components. According to a literature survey, approximately 70 % of hydraulic system failures are caused by oil contamination. Therefore, in order to operate the hydraulic system reliably, the contamination management should be managed effectively (Singh, Lathkar, & Basu, 2012). The National Research Council of Canada also reported that particle-induced failures such as abrasion, corrosion, and fatigue account to around 82 % of wear problems. (Garvey, 2003).

(Singh, Lathkar, & Basu, 2012) studied the major causes of failure of pumps, valves, actuators and reported that the contamination once developed/ingressed in the system, while circulating in the system, damages the surface of other components resulting in failure of the entire system. (Khayal, 2014) also conducted similar studies in the gold mine and identified major causes of equipment breakdown. They concluded that most hydraulic system failures in gold mining equipment resulted from contamination of hydraulic fluid and suggested good filtration systems to ensure long life and proper operation of the system.

Solid contamination is known to be the leading cause of hydraulic system failures and early decay; it is impossible to eliminate solid contamination completely, but it can be kept under control. The most popular standard for Contamination Classes in hydraulic systems is ISO 4406:1999.

In a hydraulic system, there are a surprising number of different sources of contamination including new oil that is unsuitable for hydraulic systems; particles that are added during routine system maintenance or service; wear contamination that is caused by the pumps, actuators, valves, and hydraulic motors; and failure to clean the system thoroughly after failure. Surface degradation contaminants are responsible for more than 70% of all hydraulic system downtime (McCloy & Martin, 1980) (Khayal, 2014).

A contaminant is defined as any material foreign to a hydraulic fluid that has a detritus effect on the fluid performance in a system (Singh, Lathkar, & Basu, 2012). Contaminant particles come in all shapes and sizes and are made up of a wide variety of materials (S.Cundiff, 2001). The majority are abrasive, so they plough and cut fragments from critical surfaces in the components when they interact with surfaces. This abrasive wear and surface fatigue is responsible for about 90 % of degradation failures (Singh, Lathkar, & Basu, 2012).

‘Wear rates in successfully operating industrial equipment can vary enormously, from very high values under particularly aggressive conditions to very low values in more benign circumstances” (Williams, 2005). Wear metals include iron, chromium, nickel, aluminum, copper, lead, and tin. The contaminant metals include silicon, sodium, and potassium. The phenomenon of wear was given a formal definition in 1968 by the OECD as ‘the progressive loss of material from the operating surface of a body occurring due to relative motion at its surface’.

Hydraulic circuit contaminants affect the performance and life of hydraulic equipment, leading to one of three types of system failure:

- Degradation: “clearance-sized particles interact with both faces, often causing abrasive wear, corrosion, and aeration issues” (Babcock & Battat, 2006).
- Intermittent: “solid particles cause damage according to their size. High concentrations of small particles (£ 10 mm) form silt, eroding components' interior mating surfaces, rendering them inoperable. Contaminating solid particles equal in size to the clearance between two moving surfaces can cause both jamming and wear” (Babcock & Battat, 2006).

- Catastrophic: “this happens suddenly when a few large particles or a large number of small particles cause complete seizure of moving parts” (S.Cundiff, 2001).

In mining, wear has a significant role in energy losses too. The total energy consumption of global mining activities, including mineral and rock mining, is estimated to be 6.2% of the total global energy consumption. (Holmberg, 2017) stated that about 16.6 % of the energy consumed in mining industry, equalling 2 EJ annually on a global scale, is used to remanufacture and replace worn out parts and reserve spare parts and components needed due to wear failure. Hence, any hydraulic system, no matter how simple or complex, requires contamination control. Unfortunately, filters are often viewed as a necessary evil and are installed in a system as an afterthought rather than an advantage (Khayal, 2014).

2.1.5 Hydraulic filters and prevention of system contamination

Protecting expensive internal hydraulic system components from premature wear and catastrophic failure is critical. Longer component life means less system downtime. In a typical hydraulic system, filters function together to preserve these components. Filters are the first line of defense to reduce the number of particles in the oil. There are various types of filters situated in the hydraulic system depending on their use and filtration capacity (Gannon, 2018). Figure 3 shows a typical hydraulic filtration system.

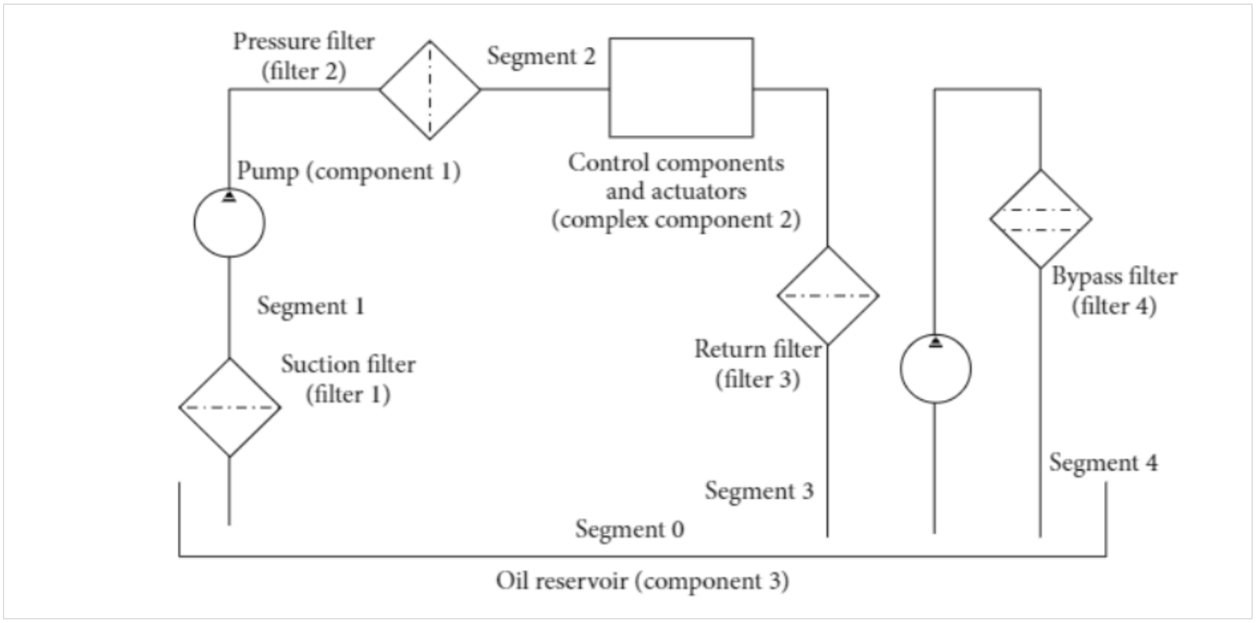


Figure 3. Typical hydraulic system filtration(Huang, Nie, & Ji, 2014)

There are four basic types of filters in the hydraulic circuit:

- Suction filters: Suction filters are usually installed at the pump inlet to protect the pump from large contaminants that can cause catastrophic failures. They also protect the pump from ingesting any debris that is built up in the oil. Suction strainers normally filter particles from 75 to 150 microns (Gannon, 2018).
- Return-line filter: Another common location for filters is in the return line. The return-line filter prevents contaminants caused by part wear from getting into the tank. The filter ratings range from 3 to 25 microns (Reik & Oberli, 2014).
- Bypass filter: The flow through the return-filter constantly keeps changing even if it is correctly sized. A steady flow through the element gives the most efficient filtration. If the

return- filter fills up with contaminant, then the bypass filter valve opens (Reik & Oberli, 2014).

- In-line filters: Control valves normally require pressure-line filters as they have low contamination tolerance. These valves have small, very close-tolerance fits and must shift rapidly at a low pilot-pressure differential (Heney, 2014).

(Huang, Nie, & Ji, 2014) studied identification of contamination control for fluid power system and discussed several disadvantages of improper maintenance framework. Although several studies were undertaken on hydraulic contamination sensitivity analysis, they had difficulties reflecting the filtration systems from the perspective of economic effects. They conducted a detailed analysis of the filter debris and used the inexact chance-constrained integer programming (ICIP) method to control contaminants of the hydraulic system. The method used improved upon the existing interval-integer and chance-constrained programming approaches of filter replacement. They concluded that effective filter management and control of hydraulic wear particles in the system play a crucial role in ensuring reliability and increasing the service life of components. The incorrect methods of installing or replacing or prolonging system maintenance posed serious contamination threats to the entire system.

2.2 Oil Analysis for CBM

2.2.1 Maintenance types and its evolution

(Dhillon, 2002) defined maintenance as a set of activities performed to restore a component or machine to a state where it can perform its intended functions. Improving a maintenance management program is a never-ending task that necessitates a progressive approach and active participation. In the mining industry, equipment maintenance accounts for a large portion of overall operating costs (Murthy, Atrens, & Eccleston, 2002). “Equipment maintenance costs range from 20% to over 35% of total mine operating costs” (Unger & Conway, 1994). “Approximately 10% of production time is lost by unplanned maintenance in Australian underground coal mining industry” [Clark, D 1990]. Selecting a proper maintenance strategy to prevent failures is significant in mining due to its fallbacks in the safety and economics of operation (Azadeh & Zadeh, 2015). The approach towards maintenance has changed throughout the years, from merely being a part of the production to an essential strategic element in mining operations. Figure 4 represents different maintenance strategies used in mining industry.

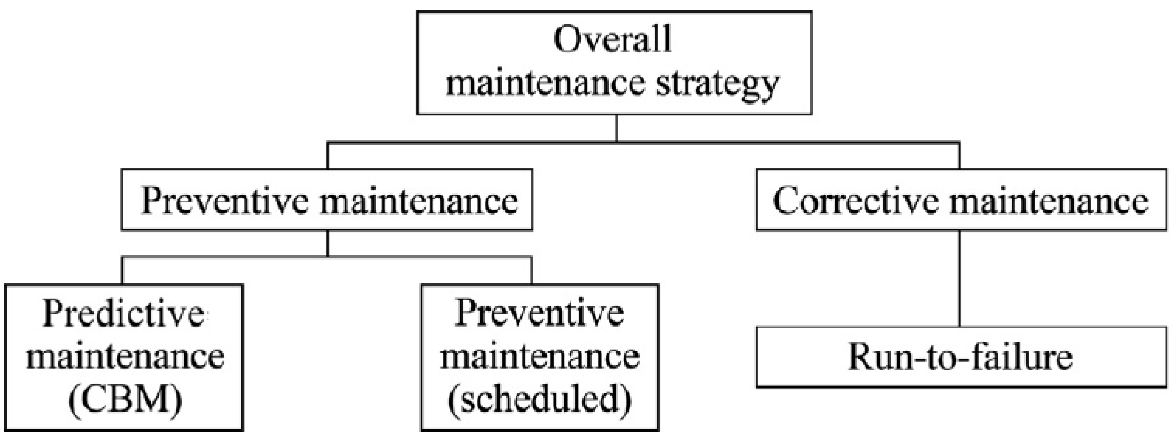


Figure 4. Different maintenance strategies (Alla, Hall, & Apel, 2020)

Maintenance strategies over the years have evolved from reactive (corrective) actions to ongoing predictive activities with an aim to optimize time, costs, and quality. Nowadays, maintenance management aims to decrease both unscheduled and scheduled downtime. Available time, production quality, reduced costs and performance are the basic key performance indicators (KPI), which combined give the overall equipment effectiveness (OEE).

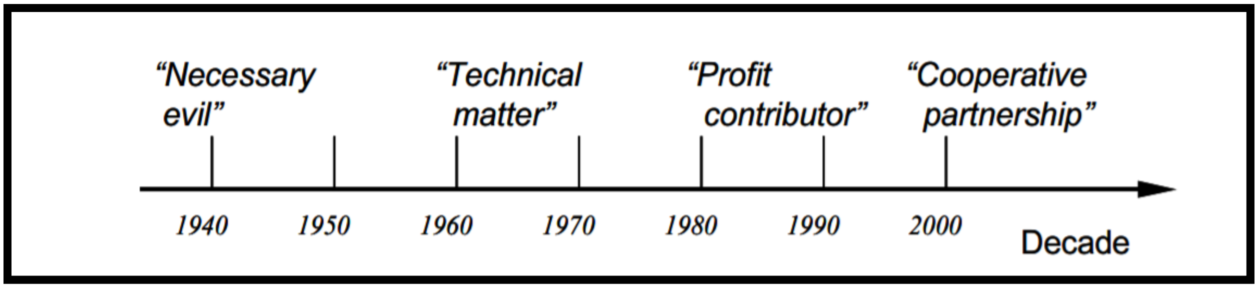


Figure 5. Maintenance Operations in a time perspective (Pintelon & Parodi-Herz, 2008)

Maintenance was once thought to be nothing more than an essential part of production; it was merely a necessary evil. Repairs and replacements were made as needed, and no concerns about optimization were posed. Later, maintenance was a technical issue. Maintenance became a full-fledged function rather than a production sub-function as time passed. Maintenance management has clearly evolved into a complex role requiring both technical and managerial expertise and the ability to adapt to a changing business climate. Top management recognizes that having a well thought out maintenance plan along with a detailed practice of that strategy could have a significant financial impact (Pintelon & Parodi-Herz, 2008). Figure 5 represents evolution of maintenance practices over the years in mining industry.

As depicted in Figure 4, maintenance actions or interventions can be of two types. They are either corrective maintenance (CM) or precautionary/preventive maintenance (PM) actions. Corrective maintenance implies replacing or repairing the equipment after it has failed. When equipment fails, CM tasks locate the problem and fix it so that the equipment can be repaired and thus the facility's production can resume. The key drawback of this maintenance technique is the inherent level of complexity involved (Dhillon, 2002). Similarly, the technique is highly reactive, capable of shutting down an entire operation just because of a single system failure. As a result of its disadvantages, a more proactive maintenance method of recognizing the equipment needs periodic maintenance to function smoothly before a breakdown occurred as developed (Deighton, 2016).

“Preventive Maintenance is carried out at predetermined intervals and according to a prescribed criterion; it is intended to reduce the probability of failure or the degradation of an item” (EN 13306 2001). All preventive management programs are time driven. The component to be maintained can either be replaced or reconditioned depending on its condition. PM can be further categorized into preventive-based and predicted maintenance (Coetzee, 2004).

The predictive maintenance strategy detects issues that can be overlooked by preventive maintenance. Predictive maintenance monitors the performance and condition of equipment during normal operations to prevent failures. The goal of predictive maintenance is the ability to predict when the equipment failure could occur based on certain conditions, followed by preventing failures through appropriate maintenance strategies. Condition-directed maintenance can detect the onset of an equipment failure mechanism. Many businesses are on the brink of a second digital revolution known as 'Industry 4.0,' owing to enhanced connectivity and significantly expanded access to low-cost computing power (Short & Twiddle, 2019) (Deighton, 2016).

Condition-based monitoring (CBM) is a form of scheduled maintenance that repairs a system before it fails by looking for signs of fatigue or other failure precursors. CBM creates an optimum maintenance period by extending the time between preventive maintenance and reducing the expenses of unnecessary excessive maintenance and downtime. CBM is based on the study of maintenance of gathered data (such as vibration, crack propagation, oil, pressure, and viscosity) (A, Correa, & Guzman, 2020). Operators collect, store, and analyze appropriate data and observations to make accurate and timely decisions. The key benefit of CBM is that it aims to optimize maintenance cycles. Operators can predict when a system will fail by continuously tracking key elements of machine failure over time. This helps the system continue running until it is about to fail, with the failure scheduled ahead of time to save on manpower, expedited shipping of parts, and rescue operations (Sellathamby, Moore, & Slupsky, 2010).

Reliability-centered maintenance (RCM) approach includes a structured framework to analyze equipment components' functions and potential failures, such as pumps, compressors, motors, etc. The strategy of this analysis is to preserve system function instead of focusing on preserving the actual equipment. The next step in the process is to determine the function or functions that the equipment or systems are intended to perform (Deighton, 2016).

(Thomas, 2018) in their work mentioned that a case study conducted by (Feldman et al., 2008) estimated a return-on-investment ratio of 3.5:1 for moving from reactive maintenance to predictive maintenance on an electronic multifunctional display system within a Boeing 737. (Drummond & Yang, 2008) mentioned of examination of train car wheel failures showed a potential cost savings of up to 56 % of the associated costs when switching from a reactive maintenance approach to a predictive maintenance approach. Another work mentioned by (Thomas, 2018) estimates that for pumps, reactive maintenance costs $18 per horsepower per year while preventive maintenance was $13, predictive was $9, and reliability-centered maintenance was $6.

2.2.2 Condition-based Oil Analysis Monitoring to Predict Hydraulic Wear Failures

Nonstationary and time-varying load conditions are common in mining machinery. Condition monitoring (CM) is the result of the evolution of diagnostic and prognostic systems. The aim of monitoring mining machinery condition is to determine whether it is in good working order. Real-time functional health assessment is useful for predicting component failures and thus improving equipment efficiency (Marcus, 2004). CM is the method of continuously monitoring various condition parameters to detect any noticeable changes that may indicate the onset of a fault. CM techniques are normally used on rotating equipment and other machinery such as pumps, motors, and internal combustion engines, while stationary equipment is subjected to periodic inspection using non-destructive testing techniques and fit-for-service assessment (Chaulya & Prasad, 2016). "Diagnostic and prognostic are two important components in a CBM program, where diagnostic deals with fault detection and prognostic deals with fault and degradation prevention before they occur" (Madenas, Tiwari, Turner, & Woodward, 2014).

Oil analysis is a central part of any condition-based maintenance program. The right type of oil, the right grade of oil, and changing the lubricant and filter at prescribed intervals are critical for a machine's reliability and efficiency (Junior, et al., 2020) (Kumar & Kumar, 2018). Establishing warning limits is a critical part of an oil-analysis program from a predictive and proactive basis (Thibault, 2013). Oil analysis majorly helps to identify changes in oil conditions and detect problems related to wear debris, elevations of metal levels, oil consumption and leakage, filter problems and deteriorating properties of additives.

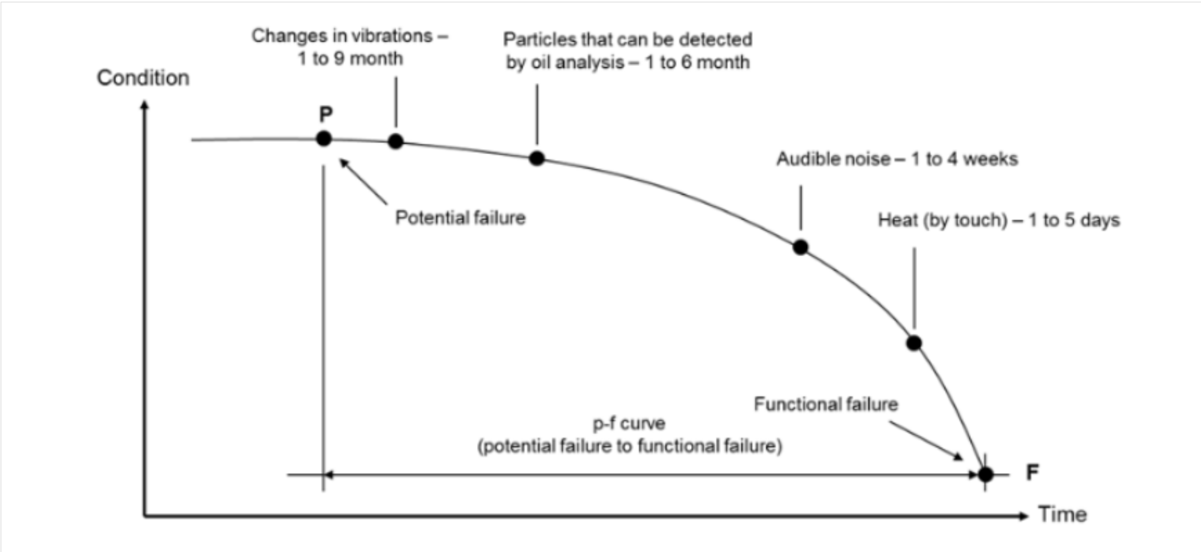


Figure 6. PF Curve for hydraulic system component (Bengtsson & Lundstrom, 2018)

Figure 6 shows the potential failure curve for the wear failures of the hydraulic component. It indicates the progress of wear failures with time. P (in the graph) indicates measurable potential failure and F indicates functional failure. Particles in oil can be monitored and can be detected anywhere between 1-6 months prior to failure. For condition monitoring to be feasible, a failure progress time of significance must exist to serve as a warning period. The National Research Council of Canada stated that 82% of wear problems are attributable to particle-induced failures such as abrasion, erosion, and fatigue. These particles can range from 1 micron to 200 microns (Fitch, 2013). CAT Ltd maintains that the concentration of wear particles in oil is a key indicator of potential component problems. Hence, oil analysis techniques for condition monitoring offer significant potential benefits to the operator (Ng, Harding, & Glass, 2016).

In recent decades, the aerospace industry has become an expert in using real-time data for product tracking and maintenance scheduling. Significant amounts of real-time usage data from product monitoring are generated and transmitted back to OEMs for diagnostic and prognostic purposes. Other sectors, such as mining, construction and automotive, have recently begun to develop capabilities in these fields, and condition-based maintenance (CBM) is gaining traction as a method of meeting customer demands. CBM necessitates continuous product data tracking in real-time (Fitch, 2013) (Ng, Harding, & Glass, 2016).

(Ng, Harding, & Glass, 2016) did research that focused on current dynamic data acquisition techniques for mobile hydraulic systems, using inline particle contamination sensor; the aim was to assess suitability to achieve both diagnostic and prognostic requirements of condition-based maintenance. They could conclude that hydraulic oil contamination analysis, namely the detection of metallic particulates, offers a reliable way to measure real time wear of hydraulic components.

(Ron LeBlanc, 2019) cited a case study of Dominion Diamond Mines, including oil analysis-based condition monitoring across the Dominion fleet in conjunction with changing the oil. They mentioned that in over two years, Dominion Diamond Mines was able to extend drain intervals across its vehicle and equipment fleet to result in $900,000 of potential savings. In addition to achieving considerable cost savings, Dominion achieved an overall reduction in idle time for its mobile equipment and drastically improved fleet reliability during harsh Arctic winters delivering cold-start capabilities.

2.2.3 Oil Analysis Tests for Particle Count Detection

Routine tests of oil analysis vary based on the originating component and environmental conditions but almost always include tests for viscosity, elemental (spectrometric) analysis, moisture levels, particle counts, Fourier transform infrared (FTIR) spectroscopy, and acid number. Other tests based on the originating equipment include analytical ferrography, ferrous density, demulsibility and base number testing (Almasi, 2014).

Elemental analysis works on the principles of atomic emission spectroscopy (AES), which is sometimes called wear metal analysis. This technology detects the concentration of wear metals, contaminants, or additive elements within the oil. The two most common types of atomic emission spectroscopy are rotating disc electrodes (RDE) and inductively coupled plasma (ICP). The methods have limitations in analyzing particle sizes, with RDE limited to particles less than 8 to 10 microns and ICP limited to particles less than 3 microns. Particle counting method measures the size and quantity of particles in the oil. Many techniques can be used to assess this data, which is reported based on ISO 4406:99. This standard designates three numbers separated by a forward slash providing a range number that correlates to the particle counts of particles greater than 4, 6, and 14 microns (Bennett, 2013) (Macuzic, 2010).

ISO 4406:99 cleanliness codes are referred on a Renard series table in conjunction with particle counts of specific micron size. Each code is associated with a specific ISO code. Every time the number of particles in the oil gets doubled, the ISO cleanliness code increases by 1. Figure 7 depicts the ISO code for the range of particles per milliter presents in the oil.

ISO 4406	Number of particles per ml	
ISO Code Number	More than	Up to and including
24	80,000	160,000
23	40,000	80,000
22	20,000	40,000
21	10,000	20,000
20	5,000	10,000
19	2,500	5,000
18	1,300	2,500
17	640	1,300
16	320	640
15	160	320
14	80	160
13	40	80
12	20	40
11	10	20
10	5	10
9	2.5	5
8	1.3	2.5
7	0.64	1.3
6	0.32	0.64

Figure 7. ISO 4406:99 code table for particle cleanliness

2.2.4 Oil Analysis Drawbacks

(Mayer, 2009) analyzed drawbacks of oil analysis and summarized the following points on how CBM fails to detect failures at early stages.

- The common pitfall is noted as performing analysis too infrequently.
- Poor sampling techniques are the weakest links in the oil analysis chain, and representative samples have many data disturbances
- Poor interpretations of the fault diagnostics result and making poor decisions on the supplied information

- Oil analysis condition monitoring technologies fail to overlap with other condition monitoring technologies.

(Fitch, 2013) mentions that an integrated approach of oil analysis with other predictive maintenance technologies would improve equipment reliability and reduce catastrophic failure. With the advent of Industry 4.0, the possibility of increased integration of the real-time monitoring and operational aspects of equipment can help strategically plan the maintenance activities (Short & Twiddle, 2019) (Deighton, 2016).

2. 3 Reliability Analysis of Hydraulic System

Reliability refers to the probability that the system meets its desired performance standards in yielding output for a specific time duration when used under specific conditions (Dhillon, 2008). The component's reliability is a function of time and is always measured at a specific operating time. Reliable operation is interrupted or terminated by failures. A failure is an event that results in the inability to complete the required duties and meet the requirements. The theoretical definition of reliability is (Reliability = 1 – Probability of Failure), given by R(t). Availability and maintenance are related to reliability and are defined as essential components of it (Mencik, 2016).

If a component's reliability value is near 1, the chances of it being susceptible to interference or failure are low. If the unit is run according to procedures and maintenance is performed regularly, these conditions will be achieved. The reliability importance of a component is a measure of how important it is to the system's overall reliability. This can be useful information for improving the overall system reliability as the efforts will be to concentrate on improving the reliability of those components that have the greatest effect on the system reliability (Suryo & Bayuseno, 2018) (Weibull, 2013).

Understanding the complexities of mining equipment, their efficiency, and failures can help achieve better results and reduce unexpected and unneeded costs. Mines can maintain consistent levels of productivity by conducting regular reliability assessments. Performance measurement is significant because it identifies existing performance gaps between existing and desired performance and shows how far the gaps have been closed (Troy, 2018) (Amy, 2020).

Many KPIs are used to monitor the long-term trends in reliability and maintenance performance. These KPIs help in understanding if all the small and large modifications in maintenance practices and system changes are having the desired effect over time. The Mean Time Between Failure (MTBF) and Mean Time to Failure (MTTF) are two essential KPIs for determining the system's reliability (Kendon, 2019).

The output of a healthy reliability process is optimal asset reliability at optimal cost. Result measures include maintenance cost (as a contributor to total operating cost), asset downtime due to planned and unplanned maintenance (as a contributor to availability) and a number of failures on assets (the measure of reliability: this can then be translated into the mean time between failures). All these results measure lag (Weber, 2005). With the ever-increasing worldwide industrial competitiveness, examining the reliability of every piece of equipment (equipment reliability) is an important maintenance policy. The most widely applied system reliability models are based on graphical techniques, lifetime distribution models, fault tree analysis, and Markov models. Each of these approaches has advantages and limitations. (Chatterjee & Bandopadhyay, 2012).

2.3.1 Graphical Techniques

To examine failure trends of repairable systems, graphical approaches using Total Time on Test (TTT) plots, first developed by Barlow and Campo, and later advocated by Bergman and Klefso, are useful (K. R. M. Rao, 2001). TTT plots can be used to track the health of equipment by looking at the failure rate, whether it is constant, growing, or decreasing (Kumar, Klefsjo, & Granholm, 1989). After each failure-repair process, it is assumed that these components are as good as new, and the data follows an independent and identical distribution. Kaplan-Meier, Piecewise Exponential, and Maximum Likelihood Estimators can be used to create TTT plots. Non-homogenous Poisson process models, which describe repairable equipment with minimum repair, can be applied for data that does not satisfy the i.i.d assumption and when a trend is seen. Furthermore, proportional hazard models can be used to describe repairable equipment whose performance is influenced by concurrent variables (Ascher & Feingold, 1984).

2.3.2 TTT Plotting Method

The graph plots number of failures per unit versus the total time on test per unit. The time between failures (TBFs) is assumed to be independently and identically distributed in this method, therefore, the actual chronological orderings of the TBFs can be ignored. Thus, it is not useful to evaluate failure data that has structures or is positive to the serial correlation test by using a TTT-plot. However, a significant aspect of these plots is that they can be used to analyze incomplete data. The failure rate of the equipment can be inferred from the shape of the plot. If the plot is concave downwards, the equipment is deteriorating (increasing failure rate), but if it is concave upwards, the equipment improves over time. If the plot crosses diagonal multiple times, the equipment has a constant failure rate. (Kumar, Klefsjo, & Granholm, 1989). The procedure for making the plot is presented below:

Suppose that $0 = t_{(0)} < t_{(1)} < t_{(2)} \ldots < t_{(n)}$ is an ordered sample of times to failure of a unit and let

(And $S_0 = 0$) denote the total time on test (TTT) at time $t_{(i)}$, j = 1, 2, ... n. The TTT-plot is obtained by plotting (j/n, u_j) where $u_j = S_j / S_n$. A TTT plot lies within the unit square starting at (0,0) and ending at (1,1).

The double TTT-plot proposed by Akersten is a generalization of the TTT-plot. This generalization is particularly useful for studying systems with non-constant failure rates. If the successive failure times are independent and exponentially distributed, the points plotted in the double TTT-plot are randomly spread in the unit square. The points pattern will tend to cluster in the upper section of the unit square in cases of increasing failure rate (IFR). There is a tendency for decreasing failure rates (DFR) and related instances to cluster in the lower section of the square (Akersten, 1986) (Akersten, 1987). Maintenance intervals can also be determined using graphical approaches. For repairable equipment with an increasing failure rate, maintenance planning with optimum maintenance periods based on minimum cost per unit time can be derived graphically (K. R. M. Rao, 2001).

2.3.3 Test for Trend and Independence

Before fitting the data with a lifespan distribution, the failure and repair data should be tested for the presence of trend and serial correlation. The IID assumption is that the observations in the samples are distributed independently and identically. Graphical approaches can be used to verify the presence of trends in failure and repair data by plotting the cumulative number of failures against the cumulative time. A straight line shows a lack of trend. A convex or concave curve indicates a system with a decreasing and increasing failure rate respectively (Ascher & Feingold, 1984). Before modeling the reliability data, it should also be tested for mutual independence by testing it for the presence of serial correlation. The presence of serial correlation can be tested by plotting the i^{th} TBF X_i against $(i-1)^{th}$ TBF, X_{i-1}. If the plotted points are plotted without any pattern, it can be interpreted that the TBFs are free from serial correlation. In case the plot reveals serial correlation, then the TBFs are plotted at greater lags, such as X_i against X_{i-2}, X_{i-3}, X_{i-4} …. etc. to search for serial correlation over greater lags (Ahmadi, Hajihassani, Moosazadeh, & Moomivand, 2020).

When the data are free from the presence of a trend and serial correlation, the next step is to choose a best-fit probability distribution model using goodness-of-fit tests to study its statistical characteristics. If the TBF shows the presence of any trend, it should be analyzed by nonstationary models such as the power law process. The prediction from statistical distributions also serves as a measure of the success of the design process (Smith, 2005).

2.3.4 Weibull distribution

The Weibull distribution is one of the best-known life-time distributions. It adequately describes observed failures of many different types of components and phenomena. The Weibull distribution is one of the most widely used lifetime distributions in reliability engineering. They are widely used in reliability and survival analysis. In addition to the traditional two-parameter and three-parameter Weibull distributions in the reliability or statistics literature, many other Weibull-related distributions are also available (Lai, Murthy, & Xie, 2006) (Hall R. , 1997).

The 3 parameter Weibull distribution is given by

$$f(t)=\frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}} \tag{1}$$

Where *f(t) >= 0, t >= Υ; β > 0; η > 0 and - ∞ < Υ < + ∞.*

And η = scale parameter, or characteristic life; β= shape parameter (or slope); γ= location parameter (or failure free life)

The 2 parameter Weibull distribution is given by

$$f(t)=\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}} \tag{2}$$

and the 1 parameter Weibull distribution is given by

$$f(t)=\frac{C}{\eta}\left(\frac{t}{\eta}\right)^{C-1} e^{-\left(\frac{t}{\eta}\right)^{C}} \tag{3}$$

where the only unknown parameter is the scale parameter, η.

The equation for the 3-parameter Weibull cumulative density function, *CDF,* is given by:

$$F(t) = 1- e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}} \tag{4}$$

This is the failure rate which is also referred to as unreliability. The reliability function of distribution is simply 1- CDF (cumulative density function); the reliability function for the 3-parameter Weibull distribution is then given by:

$$R(t) = e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}} \tag{5}$$

The Weibull distribution is widely used in reliability and life data analysis due to its versatility. It allows for the identification of which components, or locations, in a system contribute to most of the unreliability, as well as the calculation of life-cycle costs. The prediction also serves as a metric for the system performance. Depending on the values of the parameters, the Weibull distribution can be used to model a variety of life behaviors. The values of the shape parameter, β, and the scale parameter, η, affect distribution characteristics such as the shape of the curve, the reliability, and the failure rate. β is the shape parameter, η scaling parameter and ϒ is a location parameter. These equations represent the 3 parameter Weibull. For the 2 parameter Weibull, ϒ is set to zero. The shape parameter β indicates whether the failure rate is increasing, decreasing or constant (Lazzari, 2017) (Hall R. , 1997).

2.3.5 Exponential Distribution

The exponential distribution is the simplest and most widely used reliability distribution. The exponential distribution is a continuous distribution that measures the expected time for a failure event to occur. It can be used to measure the likelihood of incurring a specified number of failures within a specified time. Systems whose failures follow the exponential distribution exhibit a constant failure rate. In other words, the failure process has no memory, which means that if the device is still functioning at time *t*, it is as good as new, and the remaining life has the same distribution (Kissell & Poserina, 2017). The exponential distribution is given by:

$$F\,(t;\lambda) = \lambda e^{-\lambda t} \tag{6}$$

$$t \geq 0;\ \lambda > 0$$

where λ is the failure rate, the failure rate function $h\ (t;\ \lambda) = \lambda$, is constant over time. The exponential model is thus uniquely identified as the constant failure rate model.

The reliability function for the exponential model is

$$R\,(t;\lambda) = exp\,(-\lambda t) \tag{7}$$

and the reliable life becomes

$$t_R = -lnR/\lambda \tag{8}$$

The MTTF for the exponential model is λ^{-1}, the reciprocal of the failure rate. The CDF for this distribution with parameter α can be written as

$$F(x) = \int_0^x \propto e^{-\propto x'} dx' = 1 - e^{-\alpha x} \tag{9}$$

The exponential distribution is a skewed continuous distribution. Its rise is vertical at 0, on the left, and it descends gradually, with a long tail on the right.

2.3.6 Log-Normal Distribution

Lognormal distribution plays an important role in probabilistic design. Typical uses of lognormal distribution are found in descriptions of fatigue failure, failure rates, and other phenomena involving a large range of data (Chang, 2015).

The lognormal distribution of a random variable X with expected value μ_X and standard deviation σ_X is denoted LN (μ_X, σ_X) (Hall R. , 1997) and is defined as

$$F(t) = \frac{1}{t\sigma\sqrt{2\pi}} exp\left[\frac{-(lnt - \mu)^2}{2\sigma^2}\right] \tag{10}$$

f(t)$\geq 0, t \geq 0, -\infty < \mu, \sigma > 0$

in which *f(t)* is the PDF of the failure distribution, and

$$R(t) = 1 - F(t) = \int_t^\infty \frac{1}{t\sigma\sqrt{2\pi}} exp\left[\frac{-(lnt - \mu)^2}{2\sigma^2}\right] \tag{11}$$

$$\lambda(t) = \frac{f(t)}{R(t)} \tag{12}$$

The parameters μ and σ represent the mean and standard deviation of the natural logarithms of the data. An increase in μ indicates an increase in the meantime between failures, and an increase in σ indicates that there is more variation in the TBF. Additionally, μ is the scale parameter, and σ is the shape parameter (Robert & Jim, 2017).

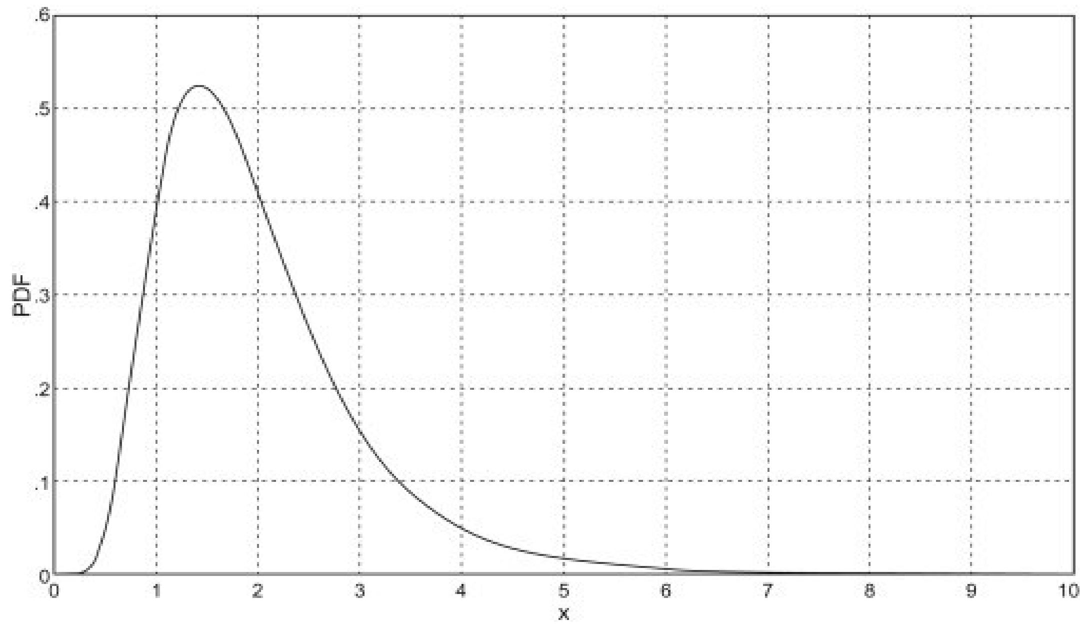


Figure 8. PDF of Log-Normal distribution(Maymon, 2018)

2.3.7 Gumbel Distribution

Gumbel distribution is used to model the distribution of the maximum (or the minimum) of a number of samples of various distributions. It is also known as the log-Weibull distribution and the double exponential distribution. The Gumbel distribution's pdf is skewed to the left, unlike the Weibull distribution's pdf, which is skewed to the right. The Gumbel distribution could also be appropriate for modeling the life of products that experience very quick wear-out after reaching a certain age (Chang K. H., 2015). The pdf of Gumbel distribution is given by

$$f(t) = \frac{1}{\sigma}\, e^{z - e^{z}} \tag{13}$$

$$f(t) \geq 0, \sigma > 0$$

where $z = \frac{t - \mu}{\sigma}$ and μ is the location parameter and σ is the scale parameter.

The reliability at time t is given by:

$$R(t) = e^{e^{-z}} \tag{14}$$

2.3.8 Parameter Estimation

The term parameter estimation refers to the process of using TTF data to estimate the parameters of the selected distribution. Several parameter estimation methods are available. Rank regression (Least squares), Maximum Likelihood Estimate (MLE), and Bayesian estimation are some of the commonly used methods for estimating parameters (Hall R. , 1997). The rank regression is a simple technique which engages replacing the data with their corresponding ranks. The calculation for the rank correlation coefficient is the same as that for the Pearson correlation coefficient but is calculated using the ranks of the observations and not their numerical values. This method is useful when the data are not available in numerical form, but information is sufficient to rank the data. The basic idea behind MLE is to obtain the most likely values of the parameters for a given distribution that will best describe the data (Almazah & Ismail, 2021). It is mathematically formulated as follows:

Let X1, X2, ⋯, Xn be a random sample from a distribution that depends on one or more unknown parameters θ1, θ2, ⋯, θm with probability density function f (x_i; θ1, θ2, ⋯, θm). Suppose that (θ1, θ2, ⋯, θm) is restricted to a given parameter space Ω. Then:

$$L\ (\theta_1, \theta_2, \dots, \theta_m) = \prod_{i=1}^{n} f\,(x_i\,; \theta_{1,}\theta_2, \dots, \theta_m) \tag{15}$$

is called the likelihood function.

The log likelihood function is given by

$$l\,(\theta|x) = log \prod_{i=1}^{n} f\,(x_i|\theta) \tag{16}$$

and $\alpha = \sum_{i=1}^{n} l(\theta|x_i) = \sum_{i=1}^{n} logf\,(x_i|\theta)$

The maximum likelihood estimators (MLE) of the unknown parameters θ are obtained by maximizing *l*.

Maximization can be accomplished by taking the derivative of α with respect to each θ, setting the equations equal to zero, and solving them simultaneously.

2.3.9 Goodness of fit tests

A goodness-of-fit test indicates whether it is reasonable to assume that a random sample comes from a specific distribution. Goodness-of-Fit (GOF) tests include (Donadio et at., 2006):

- Chi-square test
- Kolmogorov-Smimov test
- Cramer-von Mises test
- Anderson-Darling test
- R-squared test
- Residuals

2.3.10 Chi-square Test

The Chi Square test involves comparing the number of data that fall into selected classes with the number that would be expected to fall in those classes from the assumed distribution. The test compares the empirical histogram against the theoretical histogram (Hall R. , 1997).

$$X^2 = \sum_{1}^{N} \frac{(xi - \mathrm{Ei})^2)}{\mathrm{Ei}} \tag{17}$$

where,

x_i is the observed quantity in the i_{th} class

E_i is the expected value from the given distribution

χ^2 is the calculated value of Chi Square

N is the number of classes

The above equation can be used to calculate the chi-square value. This value can be compared to the Chi-Square value for N-P (where P is the number of parameters estimated) degrees of freedom at a given confidence level. If the calculated value of χ^2 is greater than the tabulated value, then the assumed distribution of the data is not supported at the chosen confidence level.

2.3.11 KS Test

The Kolmogorov-Smirnov test is a non-parametric test used specifically for a continuous distribution. It is an exact test based on the empirical distribution function of observed data. In contrast to chi-square test, KS test compares the empirical cumulative distribution function against theoretical CDF. KS test is more powerful for small-sized datasets. There is a comparison of the ranked value of the data with what the expected value of the ranks would be from the assumed distribution. Test statistics are simply the maximum absolute difference between the observed and expected rank value of the distribution. If the calculated value is greater than the tabulated value, then the assumed distribution of the data is not supported at the chosen confidence level.

The null hypothesis of the test states that the data comes from the specified distribution, and the alternate hypothesis states that at least one value doesn't match the specified distribution (Hall R. , 1997). The test statistics D is given by:

$$D = \sup | F_0 (x) - F_{data} (x)| \tag{18}$$

where

- $F_0 (x)$ = the cdf of the hypothesized distribution,
- $F_{data}(x)$ = *the* empirical distribution function of your observed data

If D is greater than the critical value, the null hypothesis is rejected.

2.3.12 Residuals

A residual plot is a graph that is used to examine the goodness-of-fit in probability distributions. The residuals from a fitted distribution are the differences between the response data and the fit to the response data at each value. It can be defined as the total error in describing the data by a probability distribution plot.

2.3.13 Pareto Analysis

Pareto analysis is a technique for determining which components cause the most significant failures. The Pareto principle states that "80% of problems are caused by 20% of events." The first step in a Pareto analysis of equipment is to break down it into appropriate subsystems or to identify distinct subsystem components. The cumulative percentage cost or downtime or failure frequencies are plotted as a function of the cumulative percentage of failures using recorded failure data and repair and replacement costs for the associated failures for each of these.

It is common to find that about 20% of components are responsible for 80% of failures, emphasizing that maintenance activities should be concentrated on these components (Hall, Knights, & Daneshmend, 2000), (Kumar U. , 1990).

A graphical depiction of the findings of Pareto analysis is a Pareto histogram. It displays a cumulative percentage curve through the right side of the first bar and lists data (failure frequency or repair or cost) in descending order of value. The Pareto histogram analysis focuses solely on one of downtime, cost, or failure frequency, and it cannot determine which aspects are more important. All the failures that occur frequently or that have longer downtime and higher costs have an influence on production and are crucial aspects for improving reliability. However, a Pareto histogram miss events such as impacts of failures with high downtime or maintenance costs and low failure frequency (Du, 2008) (Knights, 2001).

2.4 Machine Learning Algorithms for Failure Predictions

Early diagnosis of potential failures will improve component and system reliability. The overall repair and restoration costs of equipment are determined by scheduling the maintenance operations. Maintenance costs account for a major amount of a hydraulic system's total operating costs. Wear problems can necessitate unanticipated repair and replacement procedures, which can be avoided through preventive maintenance. Early failure prediction may aid in the scheduling of routine maintenance, reducing downtime caused by component failures or machinery breakdowns. Machine learning is a new technology that is expected to grow in the next years. Machine learning methods have used in a variety of applications, including prediction of failures and preventive maintenance of systems (Din, Guizani, Rodrigues, Hassan, & Korotaev, 2019). The key element of machine learning and the decision-making system is data preparation. It organizes the data so that it may be used to make decisions. The decision results impact future forecasts, failure events, and equipment availability (Jan, Farman, Khan, Talha, & Din, 2019). Data mining is a technique for classifying and converting data into usable information. It extracts useful models from a large amount of data and uses a variety of methods to uncover stored data. Data mining is the process of extracting knowledge from raw data.

Feature selection could be a fundamental issue in data processing and machine learning algorithms that target the features which are the most relevant to the intended prediction (Kamepalli & Mothukuri, 2014). Features collected from the observation of a circumstance aren't all equivalently significant. Normally, operational data tends to be incomplete, partially meaningful, or not meaningful. Some of the data may be noisy, redundant, or irrelevant. Choosing a feature set relevant to a certain duty is the goal of feature selection. This is a multidimensional and complicated situation (Rosario & Thangadurai, 2015). Based on the correlation coefficient clustering method, Hsu suggested a unique feature selection methodology. It aimed to reduce aspects that were noisy, repetitive, or unnecessary. By removing irrelevant features, performance in terms of computational speed and classification accuracy can be enhanced (Hsu & Hsieh, 2010). Data processing methods aid in the improvement of data quality and the accuracy of data mining, making it more efficient. Data quality is critical for the finding of information, the detection of anomalies, and the prediction and analysis of data for decision-making. Predicting equipment failures is critical for minimizing repair and maintenance costs and evaluating equipment availability (Fan & Fan, 2015).

Businesses can benefit from “big data”, which helps guide systems along the proper pathways. It is vital to obtain useful data from the dataset to improve the performance of machine learning algorithms (Fahad & Mahbub, 2016). Depending on the availability of labelled data, ML-based data-driven methods can be further classified as supervised, semi-supervised or unsupervised approaches. Machine learning algorithms are classified into taxonomies based on the algorithm's expected outcome. The following is a list of common algorithm types:

- Supervised learning: The algorithm creates a function that maps inputs to outputs. Output variables are known. The classification problem is a common supervised learning challenge in which the learner must learn (or estimate the behaviours of) a function that maps a vector into one of many classes by studying multiple input-output samples of the function.
- Unsupervised learning: In this method, there is no target or outcome variable to predict/estimate. It is used for clustering population in different groups and when there is a lack of sufficiently labelled data (Abdi, 2016).
- Semi-supervised learning: Combines both labelled and unlabeled examples to generate an appropriate function or classifier (Ayodele, 2010).
- Reinforcement learning: Using this algorithm, the machine is taught to make a certain decision. It works like this: the machine is placed in an environment where it would constantly train itself through trial and error. This system learns from its previous experiences and seeks to capture as much information as possible to make accurate decisions (Abdi, 2016).

2.4.1 Logistic Regression

In binary classification, logistic regression analysis performs exceptionally well, particularly with categorical variables with [0,1] classes. Based on the values of predictor variables, either categorical or numerical, logistic regression models can estimate the likelihood of a failure occurrence (Kleinbaum & Klein, 2002). In logistic regression, the dependent variable has a Bernoulli distribution. Thus, for any given linear combination of independent variables, an unknown probability, P, is estimated for the response variable is estimated. To do so, a link function must be used to link the independent variables to Bernoulli's distribution, with the natural log of the odds ratio or the logit acting as the link function. This function converts a linear combination of explanatory variables to Bernoulli's probability distribution, which has a domain of 0 to 1. In turn, the natural log of odds is a linear function of the input variables. As a result, as illustrated in Equation, logistic regression models the logit-transformed probability as a linear relationship with the predictor variables (Bhattacharjee, Dey, & Mandal, 2020) (Hildreth & Dewitt, 2016).

$$Logit(P) = log\left(\frac{P}{1-P}\right) = y \tag{19}$$

where log $\left(\frac{P}{1-P}\right)$ is called the logarithm of the odd, also known as log-odd or logit.

The probability of the component failing is represented by the odds. It is the ratio of "successes" to "non-successes," i.e., odds are the likelihood of a component failing divided by the probability of a component not failing. From above,

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots\ldots\ldots\ldots + \beta_P X_P \tag{20}$$

where y is the probability of failure, and β_0, β_1, β_2, ……. β_P are regression beta coefficients of explanatory variables X_1, X_2…. X_P (Bhattacharjee, Dey, & Mandal, 2020) . Thus, the logistic regression equation can be written in terms of an odds ratio.

$$p = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}} \tag{21}$$

Parameter Estimations of Logistic Regression

Maximum likelihood estimation (MLE) is used in logistic regression to provide model coefficients that link predictors to the target. The method is repeated until the LL (Log-Likelihood) does not change considerably after the initial function is estimated. The probability mass function is given by the Bernoulli equation

$$P(Y=y|X=x) = \beta(\theta^T x)^y . [1 - \beta(\theta\hat{} T\ x)]^{(1-y)} \tag{22}$$

From the probability mass function, the likelihood of the data can be written as

$$L(\theta) = \prod_{i=1}^{n} P(Y = y^{(i)}|X = x^{(i)}) \tag{23}$$

$$=\prod_{i=1}^{n} \beta(\theta^T x^{(i)})^{y(i)} . [1-\beta(\theta^T x^{(i)})]^{(1-y^{(i)})} \tag{24}$$

Taking log on both sides of the equation

$$LL(\theta) = \sum_{i=1}^{n} y^{(i)} log\beta(\theta^T x^{(i)}) + (1-y^{(i)})\, log\,[1-\beta(\theta^T x^{(i)}] \tag{25}$$

The parameter values are obtained by maximizing the log-likelihood function (Monroe, 2017)

Goodness of fit for Logistic Models

Goodness of fit in logistic regression attempts to get at how well a model fits the data. It is usually applied after a final model has been selected.

The following measures of fit are available for the Logistic Model:

- Chi-square goodness of fit tests and deviance
- Hosmer-Lemeshow tests
- Likelihood estimate
- ROC curves
- Logistic regression Pseudo R^2

The likelihood of the estimated model fitness of the logistic regression analysis is checked, and the lower the value, the more appropriate it is and therefore the logistic regression analysis can be an evaluation scale that confirms the fitness of the data analysis model. The fitness method statistic is χ^2, and if the significance probability is less than 0.05, the null hypothesis that the statistical model is valid can be chosen. The receiver operating characteristic (ROC) curve is also used to evaluate the classification results that predict two categories of the predictor variable. For the ROC curve, when the value of the x-axis is 0, the value of the y-axis becomes 0, and as the x-axis increases, the y-axis increases. The ROC curve is a straight line with a slope of 1 across the origin if the data is totally randomized. The ROC curve of the classifier is located above the straight line with a slope of 1 going through the origin if the classification model's performance is better than the random prediction. The ROC curve having a large y-axis value if x points are fixed can be said to have a good performance. The AUC (area under this curve) is 1 for a model that totally predicts the result of all cases and 0.5, if the completely predicted model and the randomly predicted model are identical. The higher the AUC, the better the model distinguishes between two categories (Jin-Hee, 2018) (Nakamura, 2007) (Narkhede, 2018).

The Wald test is usually used to assess the significance of the prediction of each predictor. Another indicator of the contribution of a predictor is odds-ratio of the coefficient, which is the amount the logit (log-odds) changes, with a one-unit change in the predictor variable (Forthofer, Lee, & Hernandez, 2007).

Logistic regression is often used in failure predictions and preventive maintenance strategies. (Hildreth & Dewitt, 2016) used logistic regression models based on cost and use metrics to accurately predict economic success or failure using the fleet data for the 378 single axle dump trucks. (Bhattacharjee, Dey, & Mandal, 2020) proposed a systematic approach for developing a standard equation for Risk Priority Number (RPN) measure, using the methodology of interval number-based logistic regression. The aim was to reduce risks of failure, using FMEA in terms of the Risk Priority Number (RPN). The developed logit model helped in identifying probability of risk of failure of high-capacity submersible pumps in the power plant. In another study the aim was to propose a model for predicting mechanical equipment failure from various sensor data collected in the manufacturing process. This study constructed a Hadoop-based big data platform to distribute many datasets for research and performed logistic regression modelling to predict the main variables causing the failure from various collected variables. As a result of the study, the main variables in the manufacturing process that cause equipment failure were derived from the collected sensor data, and the fitness and performance evaluations for the prediction model were made using the ROC curve (Ku, 2018). (Battifarano, DeSmet, Madabhushi, & Nabar, 2018) applied logistic regression to predict machine state 24 hours in the future given the current machine state. A confusion matrix was used to evaluate model performance.

2.4.2 Decision Trees

The decision tree is a supervised machine learning method for constructing classification systems based on multiple parameters or generating prediction algorithms for a target variable. In this method, a population is divided into branch-like segments that form an inverted tree with a root node, internal nodes, and leaf nodes. The algorithm is non-parametric, and it can handle huge, complex datasets without imposing a complex parametric framework (Lu, 2015). Decision trees are mainly effective in handling non-linear datasets. Like stepwise selection in regression analysis, decision tree methods can be used to pick the most relevant predictor variables from a large number of features in datasets and to assess the relative importance of these variables on the decision variable. Moreover, decision trees can also handle missing data very well. It is also easy in handling a variety of input data: nominal, numeric, and textual (Rokach & Maimon, 2015).

Decision trees can be divided into two types: categorical variable and continuous variable decision trees (Sharma, 2020). Figure 9 represents a general decision tree structure of models.

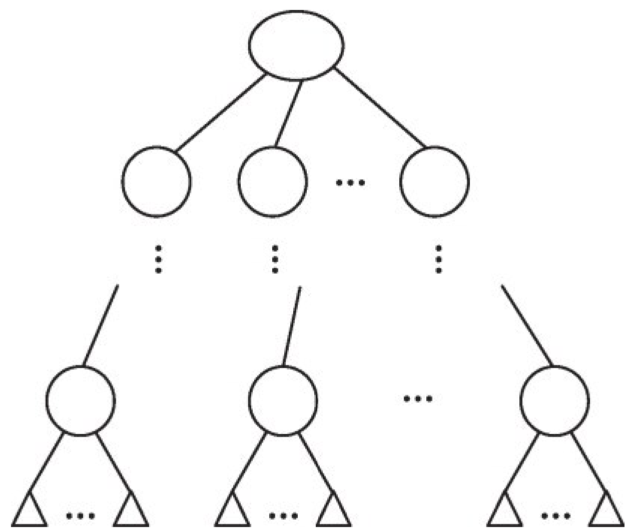


Figure 9. General decision tree structure(Du & Sun, 2008)

Attribute Selection Measures (ASM)

The main aim is to find the optimal decision tree by minimizing the generalization error. However, other target functions of the decision tree can also include minimizing the number of nodes or minimizing the average depth to find the most important predictors. Heuristics methods are required for solving the problem. The decision node and the leaf node are the two nodes that make up a decision tree. Decision nodes have been used to make any decision and have multiple branches, and leaf nodes are the output of those decisions and do not contain any more branches. The tree's algorithms are greedy by nature, and they build the decision tree in a top-down, recursive manner. The algorithm considers the partition of the training set using the result of a discrete function of the input characteristics in each iteration. Some splitting measures are used to determine which function is the most appropriate (Lu, 2015). After choosing an appropriate split, each node divides the training set into smaller subsets until no split achieves a sufficient splitting measure or a stopping criterion is reached. In most cases, the discrete splitting functions are univariate, meaning that an internal node is split according to the value of a single attribute. Consequently, the algorithm searches for the best attribute upon which to split. There are various univariate criteria. These criteria can be characterized in different ways, such as

- According to the origin of the measure: information theory, dependence, and distance.
- According to the measure structure: impurity-based criteria, normalized impurity-based criteria, and Binary criteria [for complete details (Rokach & Maimon, Decision Trees, 2005)]. The most popular technique for ASM is Information Gain.

Information gain is the measurement of changes in entropy after the segmentation of a dataset based on an attribute to calculate how much information a feature provides about a class. According to the value of information gain, the decision tree is built by splitting the node. A decision tree algorithm always tries to maximize the value of information gain, and a node/attribute having the highest information gain is split first. The entropy of an attribute is given by,

Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no)

*Information Gain= Entropy(S)- [(Weighted Avg) * Entropy (each feature)*

where, S= Total number of samples, P(yes)= probability of yes of the split, P(no)= probability of no of the split.

Pruning is the practice of removing redundant nodes from a tree to obtain the best decision tree possible. Using strict stopping criteria results in decision trees that are poorly fitted. On the other hand, using loosely stopping criteria results in huge decision trees that overfit to the training set.

Pruning techniques were created to address this issue. This method provides a weakly stopping criterion, allowing the decision tree to overfit the training set. The over-fitted tree is then pruned down to a smaller tree by deleting sub-branches that do not contribute to improving accuracy. It has been indicated in many research studies that pruning methods can improve the generalization performance of the decision tree, especially when there is noise in the data. (Patil, Wadhai, & Gokhale, 2010) (Rokach & Maimon, Decision Trees, 2005).

Decision trees are used in equipment reliability analysis mainly to decide the most important factors that influence failure rates, equipment downtime and safety incidents. (Kohli, 2021) in their work proposed an equipment reliability model for pumps, designed by applying a data extraction algorithm on equipment maintenance records residing in SAP application. The author has initially applied unsupervised learning to perform cluster evaluation. After that, the data from the finalized model was applied to a supervised learning algorithm where the classifier was trained to predict equipment breakdown. The classifier was tested on test data sets where it was observed that Support Vector Machine (SVM) and Decision Tree (DT) algorithms were able to classify and predict equipment breakdown with high accuracy and a True Positive Rate (TPR) of more than 95%. (Jiarula, Gao, Gao, Jiang, & Wang, 2016) developed a model of fault mode prediction based on the decision tree, C4.5 algorithm. (Doostan & Chowdhury, 2017) studied an approach for identifying equipment failure faults in power distribution systems. The output variable was a binary classification problem in which outages were classified as equipment failure and non-equipment failure types. Actual outage data was collected and variables that contributed to equipment failure were identified, and their relationships were examined using a decision tree, logistic regression, and naïve Bayes classifier, and their performances were evaluated.

2.4.3 Naive Bayes Classification

The Bayesian Classification is a supervised learning and statistical classification method. The Bayes Theorem was proposed by Thomas Bayes, and this classification is named after him. The algorithm assumes an underlying probability distribution and captures uncertainty about the model in a logical manner by calculating probabilities of occurrences. It is used to solve diagnostic and predictive issues and calculates explicit hypothesis probabilities and is robust to noise in the input data (Zhang, Wu, Yang, & Guan, 2018).

The Naive Bayes algorithm is a straightforward probability classifier that derives a set of probabilities by counting the frequency and combinations of values in a data set. When assessing the value of the class variable, the method applies Bayes' theorem and assumes that all variables are independent. In a range of controlled categorization challenges, the algorithm learns quickly (Saritas & Yasar, 2019). The Bayes theorem is a mathematical formula for calculating conditional probability, is

$$P\,(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{26}$$

Here;

- P(A|B) is the probability of the occurrence of event A when event B occurs,
- P(A) is the probability of the occurrence of A
- P(B|A) is the probability of the occurrence of event B when event A occurs
- P(B) is the probability of the occurrence of B

If there are X = (x_1, x_2, x_3 ...., x_n) are the feature variables then

$$P(y|x_1, \dots, x_n) = \frac{P(x_1|y)P(x_2\,|y) \dots P(x_n|y)P(y)}{P(x_1)P(x_2) \dots P(x_n)} \tag{27}$$

The presence of one feature does not affect the other and hence it is called naive (Moghaddass & Zuo, 2012). Assuming the prior class probabilities are equal, and the features are independent, this can be written as:

$$P(y\backslash x_1, \dots, x_n) \propto P(y) \prod_{i=1}^{n} P(x_i|y) \tag{28}$$

And Naïve Bayes classifier is obtained as

$$y = argmax_y P(y) \prod_{i=1}^{n} P(x_i|y) \tag{29}$$

There are different types of Naive Bayes classifiers. When characteristic values are continuous, it is assumed that the values associated with each class are spread according to the gaussian distribution, which is the normal distribution. On multinomial distributed data, multinomial Naive Bayes is preferred. It is commonly used in natural language programming (NLP) for text classification. Bernoulli Naive Bayes is employed when data is distributed according to multivariate Bernoulli distributions that is, multiple features exist, but each one is considered to have a binary value. As a result, binary values are required for features (Gandhi, 2018) (Prabhakaran, 2018).

Data-based fault diagnostics of mechanical components has become a new hotspot. (Zhang, Wu, Yang, & Guan, 2018) used Naïve Bayes for bearing fault diagnosis on enhanced independent data. Their approach was based on processing the data vector (attribute feature and sample dimension) to reduce the limitations of Naive Bayes by independence hypothesis. The statical characteristics of the original signal of the bearings were extracted, and decision trees were used to select important features of the signal, and low correlation features were selected. The authors used SVM models in the next step to prune redundant vectors and in the last step used Naïve Bayes to the processed data to diagnose faults. (Moghaddass & Zuo, 2012) studied non-repairable equipment with multiple and independent failure modes where only incomplete information about the failure mode was obtained through condition monitoring. The study focused on obtaining a probability matrix representing the relationship between actual health and condition monitoring information of the equipment, and Naïve Bayes was used as a classifier to classify each failure mode based on the degree of damage. (Yi, Chen, & Hou, 2017) applied Naïve Bayes classifier for diagnosing faults of rolling element bearings and indicated that Naïve Bayes classifier presented higher levels of accuracy without any feature engineering requirement.

2.4.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification and regression problems. In the SVM algorithm, each data item is plotted as a point in n-dimensional space where n is the number of features considered, with the value of each feature being the value of a particular coordinate (Ray, 2017). The aim is to perform classification by finding the hyper-plane that differentiates the two classes very well. SVMs maximize the margin around the separating plane, and the decision function is fully specified by a subset of training samples called the support vectors. This becomes a standard quadratic problem that can be easily solved by standard methods. Support vectors are the data points that lie closest to the hyperplane. They are the data points that are more difficult to classify. The support vectors have a direct decision on optimizing the hyperplane location (Zisserman, 2015) (Gordon, 2004).

A linear classifier of SVM has the form

$$y = f(x) = w^{\mathrm{T}}\,\mathrm{x} + \mathrm{b} \tag{30}$$

Where w is the n-dimensional input vector and b is the bias.

Learning the SVM can be formulated as an optimization:

$\min_{\mathrm{w}} \| w^2 \|$ subject to $y_{\mathrm{i}} \; (w^{\mathrm{T}}\,\mathrm{x}_{\mathrm{i}} + \mathrm{b}) \geq 1$ for $i = 1 \ldots\ldots \mathrm{N}$

This is a quadratic optimization problem subject to linear constraints, and there is a unique minimum. Not all the points are correctly classified into their respective groups. Choosing the hyperplane that accurately divides the data into two groups can lead to overfitting. Hence a regularization parameter is required to trade of between the margin and the number of mistakes (wrongly classified points) on the training data (Zisserman, 2015). Figure 10 represents the SVM hyperplane of a binary machine learning model.

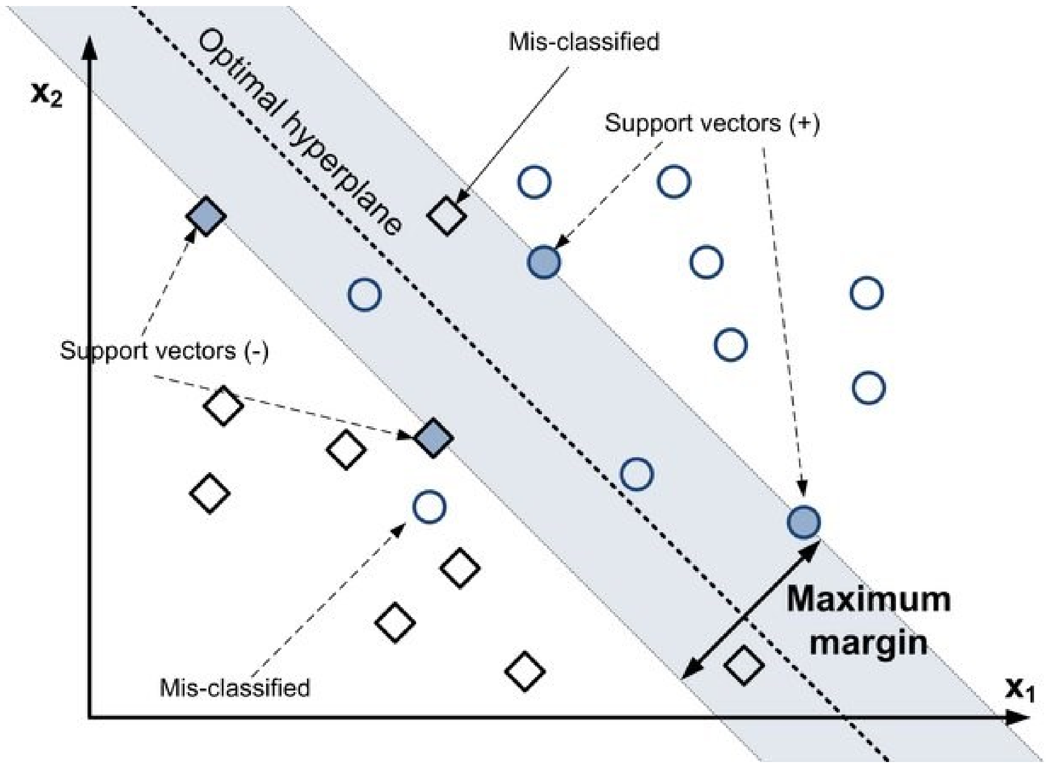


Figure 10. Optimal hyperplane for a binary classification(Duc, Kamwa, Dessaint, & Cao-Duc, 2017)

SVM Kernel Function

A separating hyperplane can be used to divide data that is linear. However, the data is frequently non-linear, and the datasets are closely linked. The input data is non-linearly mapped to a high-dimensional space to account for this. After that, the new mapping is linearly separable. Kernel trick allows SVM's to form nonlinear boundaries. The kernel function aims to allow operations to be conducted in the input space instead of the possibly high-dimensional feature space. As a result, the two classes can be separated in the feature space. Different kernel functions exist, such as polynomial, radial basis function (RBF), and sigmoid function, and the choice of a kernel function is determined by the application (Jakkula, 2011).

(Nirmal, Agarwal, & Reddy, 2019) developed SVM model-based approach to detect faults of equipment in oil industry. Temperature and pressure data were collected from sensors every minute and failure occurrences were recorded. The noise in recordings was fine-tuned using Exponentially Weighted Moving Average (EWMA) method and tabulation of the mean and peak measurements every 15 mins were tabulated. These features were used for fault diagnosis in the system using SVM K-means algorithm. (Celikmih, Inan, & Uguz, 2020) used aircraft failure and maintenance data with nine input and one output variables across two periods. Then a hybrid data model was prepared in two steps. In the first step, using a feature selection method the important features were extracted and in the next step a K- means algorithm is used to eliminate noise. Then different ML methods like SVR (Support Vector Regression), ANN (Artificial Neural Network), LR (Linear Regression) are used for performance evaluation and to improve the success of failure count prediction. (Hwang & Jeong, 2018) used SVM to detect defects and fault patterns of unexpected sudden equipment failures. SVM classifier was used to divide data as normal and abnormal, and only normal data was used for learning using the Restricted Boltzmann Machine (RBM) and then based on patterns faults in the system were identified.

2.4.5 k-NN (k Nearest Neighbour)

The k-nearest neighbours (k-NN) method is a supervised machine learning algorithm that can be used to address classification and regression problems (Harrison, 2018) . k-NN is a kind of instance-based learning (also known as lazy learning), in which the function is only estimated locally, and all computation is deferred until classification. When there is very little prior knowledge about the data distribution, the k-NN is the most basic and simplest classification algorithm. The data points are categorized based on how their neighbours are classified. The algorithm's idea is that all data points with similar characteristics will be found near together. As a result, the data points are initially split based on their similarity in attributes. Given a K value, the nearest K neighbours are chosen for any new point, and the class containing the most points out of the k points is allocated to the new point. The choice of K, as well as the distance measure used to pick the nearest K points, determine the performance of a k-NN classifier. In the case of k-NN, a small training sample size can have a significant impact on the selection of the optimal neighbourhood size K, and the sensitivity of K selection can significantly decrease k-NN classification performance. In general, k-NN is susceptible to data sparsity, noisy mislabeled points, and outliers from other classes if K value chosen is too small or too large (Imandoust & Bolandraftar, 2013) (Jabbar, Deekshatulu, & Chandra, 2013) (Zhang Z. , 2016).

(Vahed, Ghodrati, & Hosseinie, 2018) studied a historical failure dataset of a dragline to conduct predictive maintenance. The authors used the k-Nearest Neighbors algorithm to predict the failure mode but there was a chance of overfitting in the methodology. Hence, a combination of the genetic algorithm and k-Nearest Neighbor algorithm was applied for the failure dataset. This enhanced the model performance, and the results were better predicted. In another study (Moosavian, Ahmadi, & Khazaee, 2013) collected vibration signals of main journal-bearings of IC engine from condition monitoring methods. The vibration signals were classified on three conditions: normal, oil starvation condition and extreme wear fault. Thirty features were extracted from the processing of signals and k-NN and ANN were applied to train the dataset and later for diagnostic use. Variable K ranging from 1 to 20 with the step size of 1 was used to get better classification results. The experimental results showed diagnostic methods were reliable to separate fault conditions in the bearings. (Sharma, Jigyasu, Mathew, & Chatterji, 2018) proposed a new methodology of Weighted k-Nearest Neighbor Classifier where a square inverse weighting technique is used to improve the accuracy of k-NN model for fault diagnosis of rolling element bearings. Three bearing conditions, healthy, inner race fault and outer race fault were classified. The algorithm indicated that this method enables the fault detection in bearings with high accuracy.

2.4.6 K-Means Algorithm

K-Means Clustering is an unsupervised learning approach that is used in machine learning to handle clustering problems. It divides the unlabeled data into many clusters. The K-Means Clustering method is easy and accurate, flexible to handle large data, has a good speed of convergence, and has the adaptability to sparse data. K-Means clusters the data into different groups and provides a simple technique to determine the categories of groups in an unlabeled dataset without any training. It is a centroid-based approach, where each cluster has its own centroid. The goal of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters. The k-means Clustering algorithm finds the best value for K center points or centroids by an iterative process and assigns each data point to its closest k-center. Those points which are near to the k-center create a cluster. The distance of the point from the centroid in each step is calculated using the Euclidean method. Hence data points from each cluster are similar in some way and are far from other clusters. The K value is user-defined for the algorithm that is generated. The Elbow method is the most popular way that helps in selecting the optimal K value. The method is based on minimizing within-cluster Sum-of-square-values (WCSS) that defines total variation in the data. The WCSS method formula is as follows:

$$WCSS=\sum P_{i\ inCluster1} distance(P_i C_1)^2 + \sum P_{i\ in\ Cluster2} distance(P_i C_2)^2 + \sum P_{i\ in\ Cluster\ 3}\ distance(P_i C_3)^2 \tag{31}$$

where

$$\sum P_{i\ in\ Cluster1} distance(P_i C_1)^2$$

is the sum-of-the-square of the distances between each data point and its centroid within a cluster. The WCSS values are plotted against K value starting from 1. The sharp point where the graph drops abruptly is selected as the best K value.

(Zheng, Dai, & Zhou, 2019) applied K-means clustering to process the high non-linear wind turbine equipment failure data and group the data into different clusters then analyze the fault list for useful information using Long-Short Term Memory (LSTM) model. (Wei, Luo, & Yu, 2019) used historical maintenance data of ship's equipment failure and proposed a data mining method based on K-means clustering and analyzed fault phenomenon in the equipment to establish different failure modes. (Abdelhadi, 2017) tried to implement a clustering method to group maintainable equipment based on their need for maintenance according to time to failure and the location of these machines. The main aim was to reduce the scheduling process and time and a standard maintenance procedure for the machines in each cell. (Riantama, Prasanto, Kurniati, & Anggrahini, 2020) examined the condition-based equipment data using data analytics approach to develop a predictive maintenance program. K-means for clustering the failure characteristic, Support Vector Regression (SVR) model used for predicting equipment failure are the two models used in their study.

2.4.7 ANN Algorithm

The Neural Network (NN) plays a vital part in the human brain, and ANN is an unsupervised learning technique created from biology. ANN stands for Artificial Neural Networks, and it was inspired by biological neurons. It is a massive parallel computing system made up of many basic processors connected by many interconnections. ANNs learn the basic rules from a series of given symbolic circumstances in instances, rather than following a set of laws specified by human experts. They are organized in three or more layer, (i.e., input layer, several hidden layers, and an output layer). Furthermore, the relationships between the network processing units are the source of the ANNs' analytical activity. In comparison to other classic machine learning techniques, ANNs models have significant advantages in dealing with random, fuzzy, and nonlinear data. ANNs are best suited for systems with a complicated, large-scale structure and ambiguous data. ANNs are the most extensively used machine learning algorithms. Multilayer Perceptron (MLPs) with backpropagation learning are based on a supervised technique and have three layers: input, hidden, and output ((Park & Lek, 2016) (Walczak & Cerpa, 2001). They are commonly employed for a wide range of issues. The sample neural network architecture is shown in Figure 11.

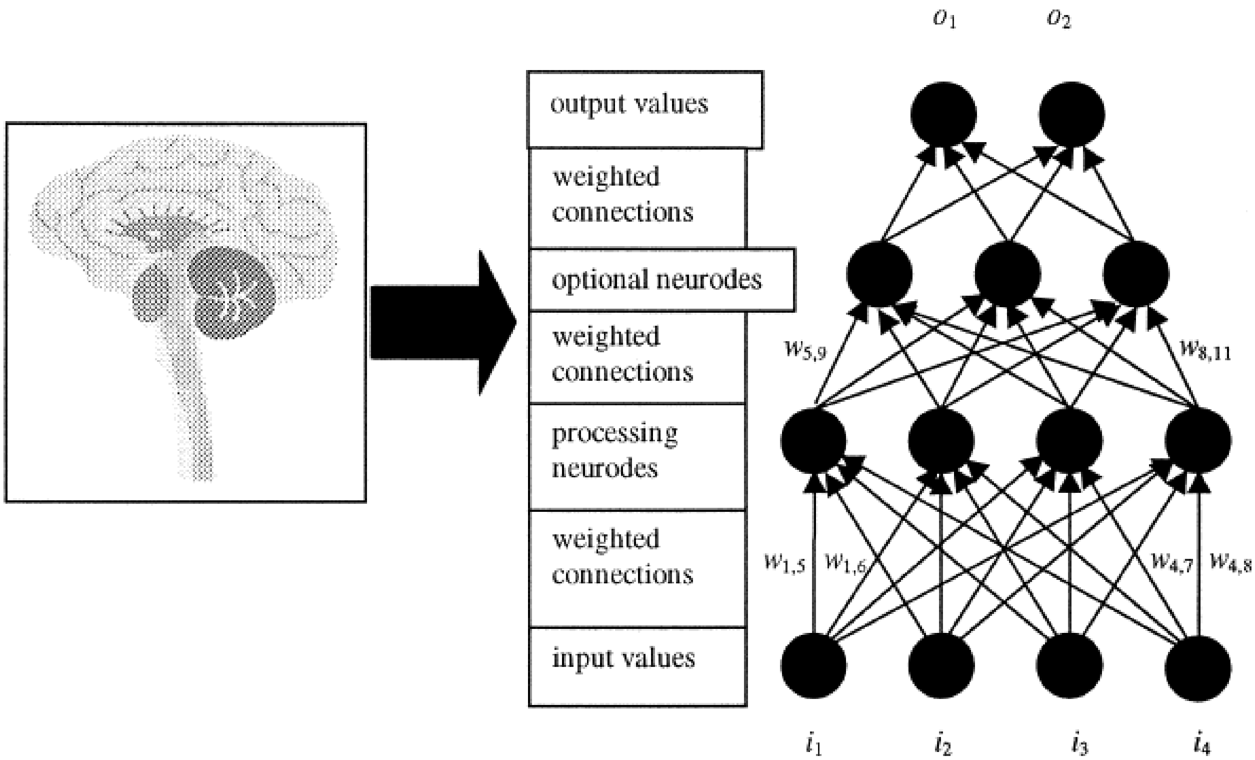


Figure 11. Sample neural network architecture (Walczak & Cerpa, 2001)

ANN is widely used in reliability and fault analysis of mining machines. Several literature works can be found using ANN for analysis. ANN is used for fault diagnostics of numerous rotating machinery that use signal processing techniques to extract features and further input these to ANN model to classify faults. (Jia, Lei, Lin, Zhou, & Lu, 2016) (Jonathan P Peck, 1994) (Liu, Yang, Zio, & Chen, 2018) (Bin, Bin, Gao, Li, & Dhillon, 2012). (Li, Mechefske, & Li, 2004) studied electric motor faults with ANN feedforward networks and self-organizing maps. Data was taken from stator current and mechanical vibration signals for major motor faults. The study showed the effectiveness of both algorithms and feedforward network looked more promising for electric motor analysis. (Zhu, et al., 2019) proposed a rotor vibration fault diagnosis approach that transforms multiple vibration signals into Symmetrized Dot Pattern (SDP) images, and then identifies the SDP graphical feature characteristic of different vibration states using a convolutional neural network (CNN). A CNN can reliably and accurately identify vibration faults by extracting the feature information of SDP images adaptively through deep learning. Proposed approach was tested experimentally using a rotor vibration test bed, and the results obtained were compared to those obtained with an equivalent CNN-based image recognition approach using orbit plot images. The rotor fault diagnosis precision was improved from 92% to 96.5%.

2.4.8 Performance Evaluation Metrics

Model evaluation is a process through which the quality of the system's prediction can be quantified. The trained model measurement is tested on validation and test datasets and the labeled data (dependent variable) in the validation or test dataset is compared with its own predictions (Mishra, 2019). There are different evaluation metrics for classification models like accuracy, precision, recall, confusion matrix, AUC score, specificity, and sensitivity. Though empirical studies have shown that choosing a metric to employ for a given problem might be difficult, each of them has specific attributes that measure different aspects of the evaluated algorithms. Machine learning performance evaluations involve certain level of trade-off between true positive and true negative rate. Confusion matrix, precision, recall and F-score are commonly used in the information retrieval as performance measures. Accuracy is used to define the overall efficiency of the model in predicting output. Receiver Operating Characteristic (ROC) curve serves as graphical representation of the trade-off between the false negative and false positive rates. The efficiency of any machine learning model is also determined using measures such as True Positive Rate, False Positive Rate, True Negative Rate and False Negative Rate (Danjuma, 2020).

This study mainly uses accuracy, confusion matrix and AUC score to estimate the performance of test dataset using machine learning models. The standard definitions for the metrics are as follows:

- **Accuracy** measures the proportion of true results to total cases. Accuracy is presented as

$$accuracy = \frac{True\ Positive + True\ Negative}{True\ Postive + True\ Negative + False\ Positive + False\ Negative} \quad (32)$$

- **Confusion Matrix** assesses the relationship between the label and the classification of the model. A confusion matrix has two axes: one for the predicted label and the other for the actual label. N represents the number of classes. Confusion Matrix is represented as

True Positive	False Positive
False Negative	True Negative

- The area under the curve (AUC) is calculated by plotting true positives on the y axis and false positives on the x axis. This metric is useful since it gives you a single value to compare different sorts of models.

2.4.9 Hyperparameter Tuning

The process of selecting a set of ideal hyperparameters for a learning algorithm is known as hyperparameter tuning. A hyperparameter is a model argument whose value is determined prior to the start of the learning process. The method entails objectively searching for different values for model hyperparameters and selecting a subset that produces the best model on a given dataset. A hyperparameter optimization yields a single set of high-performing hyperparameters with which the model can be configured. The developer specifies hyperparameters to guide the learning process for a specific dataset. Each hyperparameter value is represented by a vector at each point in the search space. The purpose of optimization is to find a vector that gives the best performing model, such as maximum accuracy or least error. Random search and grid search are two of the simplest and often applied optimization algorithms. Grid search is great for spot-checking combinations that are known to perform well generally. Random search is great for discovery and getting hyperparameter combinations that the user would not have guessed intuitively, although it often requires more time to execute (Brownlee.J, 2020).

2.5 Summary of Literature Review

The above literature review helped in understanding in detail the hydraulic system of giant shovels, the design of the hydraulic system, the importance of hydraulic system filters, different failure modes, and failure causes in the hydraulic system, condition monitoring of the hydraulic system based on oil samples, and crucial factors affecting oil contamination and hydraulic system wear. The above literature review also summarized various statistical techniques that can be applied in different analyses and the major contributions of previous researchers seeking to better understand and implement various machine learning techniques for failure analysis and predictions in various mining equipment. Researchers have implemented numerous machine learning models and achieved satisfactory results for fault diagnostics and predictive maintenance analysis in mining and other industries.

Although numerous statistical techniques are applied in the analysis of failures of mining equipment, there is no framework described to quantify the pre-post effects of implementing a solution to prevent hydraulic failures. The first part of this study aims to apply different statistical techniques to quantify the effects of magnetic filters installed in the shovel hydraulic system. Several machine learning techniques are already used in the analysis of mine equipment failures. Despite the popularity and application of machine learning techniques, no work is previously found related to the prediction of the probability of successive failure of the mine equipment components that are influenced by the previous failure. The second part of this study is an attempt in implementing machine learning models to find the probability of successive failure of different hydraulic components in the next 1000 hours of operation after a component has failed.

3. QUANTITATIVE ANALYSIS OF MAGNETIC -FILTERS

This chapter presents an approach to analyze the impacts of magnetic filters on hydraulic system of shovels and provides a framework to quantify effectiveness of a solution to prevent mine equipment failures. The chapter provides information on iron content baselines in oil analysis and presents a detailed analysis of impacts of magnetic filters in preventing contamination by quantifying different key performance metrics using statistical techniques.

3.1 Background Information

The objective of this phase is to understand the need for magnetic filters in the hydraulic system of shovels and evaluate equipment performance before and after installation of Mag-filters in the shovels. The hydraulic shovels used in this analysis are 518t machine with a Cummins engine which powers this machine with up to 1007kW. The shovels are categorized in the biggest machine segment in the crawler excavator's category. The heaped bucket capacity of the shovels is $27m^3$, which is above average for this type of machine. The system is powered by two engines that supply energy for the function of hydraulic actuators. The hydraulic system of the shovels is very complex in design and is often subject to failures. The most common cause of hydraulic failures in the shovels are related to wear and debris contamination. The two most common causes of wear-related system failures are poor fluid cleanliness and external system contamination. Contamination causes degradation of oil performance and results in hydraulic system failures for numerous reasons.

Hydraulic system contamination results from abrasive wear particles generation from the system components, externally drawn in past cylinder seals and dust particles from air drawn into the hydraulic tank during servicing. The component surface asperities can break off either by external contamination 3rd body abrasive wear or through normal wear, at times of low lubrication or high load, and thereby lead to subsequent 3 rd body abrasive wear. Copper alloy particles and prevent smooth component operation whereas relatively harder metal alloy particles result in more aggressive scoring wear. cause minor abrasive damage. As the number of wear particles increase, it can lead to catastrophic component failure, generating many abrasive particles which will be required to be cleaned even if they don't result in the functional failure of the component they contaminate. Neglecting to properly clean the hydraulic system after a failure will very likely result in further cause repetitive failure.

At the same time, external particles also enter the system because of the working environment and where the equipment is operated and stored. Leaky or improperly installed seals and caps can allow abrasive dirt and particles into the hydraulic system. Water ingress throughout the system will increase the formation of corrosion, oxidization, acids, and rust in the hydraulic fluid, which can severely impact a hydraulic system's performance and lifespan.

Hydraulic oil should be maintained at certain cleanliness levels to minimize wear particles and prevent debris accumulation. To achieve this, different methods are used to maintain hydraulic system cleanliness. Hydraulic screens and filters are installed at different locations in the system to continuously remove contaminants in the hydraulic oil. The other popular techniques are condition monitoring techniques where equipment parameters indicating failure are measured. Oil based condition monitoring method is helpful in detecting high levels of iron and other contaminants. This study uses historical hydraulic failure data and oil analysis results of shovels to assess the impacts of the newly installed magnetic filters on the shovel hydraulic system. The shovels considered in this study are currently working at different mine sites in North America.

3.2 Oil Contamination Analysis

Oil samples are taken every 600 hours from the return circuit of the hydraulic system of shovels used in this analysis, and the oil contaminant elements are monitored using the ICP particle count. ISO4406 code is used as a standard method for expressing fluid cleanliness. This section is documented in detail in Chapter 2 of the thesis. The code is expressed in a three-part cleanliness rating such as 18/16/13, representing 4/6/14 micron-sized particles per milliliter of oil. Twentythree different metals are identified in the oil including iron, chromium, nickel, aluminum, copper, lead, and tin. The contaminant metals include silicon, sodium, and potassium. A system contamination failure is usually followed by a hydraulic tank flushing, replacing the failed or contaminated components, lines and filters.

An excel based oil contamination analysis tool based on oil sample results of the hydraulic shovels used in the analysis was built to analyze the effects of variation in ICP measurements and its impacts on predicting failures. The tool also helps to understand the criticality levels of different elements present in the oil for a given volume of particles present in the oil. The elements that are included in the analysis involve wear metals and contaminants mentioned above. The ICP particle count per milliliter in oil that is measured in the laboratory has a variation of +/- 15%. The impact of variation increases with the increasing particle count in oil. Hence the lower and upper bound ranges are added to the analysis to know the range of variation for a particular metal particle count and its effects on the criticality index. The following metrics are calculated for each of the wear and contaminant metals for a given volume of any metal (V_M) with metal density (D_M), the hydraulic tank capacity (HT_C) i.e., 2200 liters for the shovels considered in the analysis, and hydraulic oil density (D_O) which is approximately 0.847 kg/lt. Different parameters of oil particle count analysis are assessed as follows:

Quantity of metal present in the oil (kg),

$$Q_M = V_M * D_M \tag{33}$$

The minimum quantity of metal that could be present in oil (kg)

$$L_B = Q_M - (0.15 * Q_M) \tag{34}$$

The maximum quantity of metal that could be present (kg)

$$M_B = Q_M + (0.15 * Q_M) \tag{35}$$

Resultant ICP Particle Count of metal (M_1)

$$ICP_{M1} = \frac{Q_M * 10^6}{HT_C * D_O} \tag{36}$$

If there are n metals present in the oil with particle count $ICP_{M(n)}$, then the total ICP Count is

$$ICP = \sum_{i=1}^{n} ICP_{M_i} \tag{37}$$

The ISO cleanliness code can then be calculated from the ISO:4406 table based on the total ICP count. If a certain volume of metal is present in the oil, then the tool gives information about different metrics such as mass of metal present, ICP count, the hydraulic acceptable, moderate, and critical limits for each metal, resultant ISO code, and error range by volume, mass, and ppm. Different colours indicate the criticality of metal concentration in the ISO particle count cell. An example of the results can be found in Appendix A. The tool was used to find the relation between the metal volume present in the oil and the ISO cleanliness code.

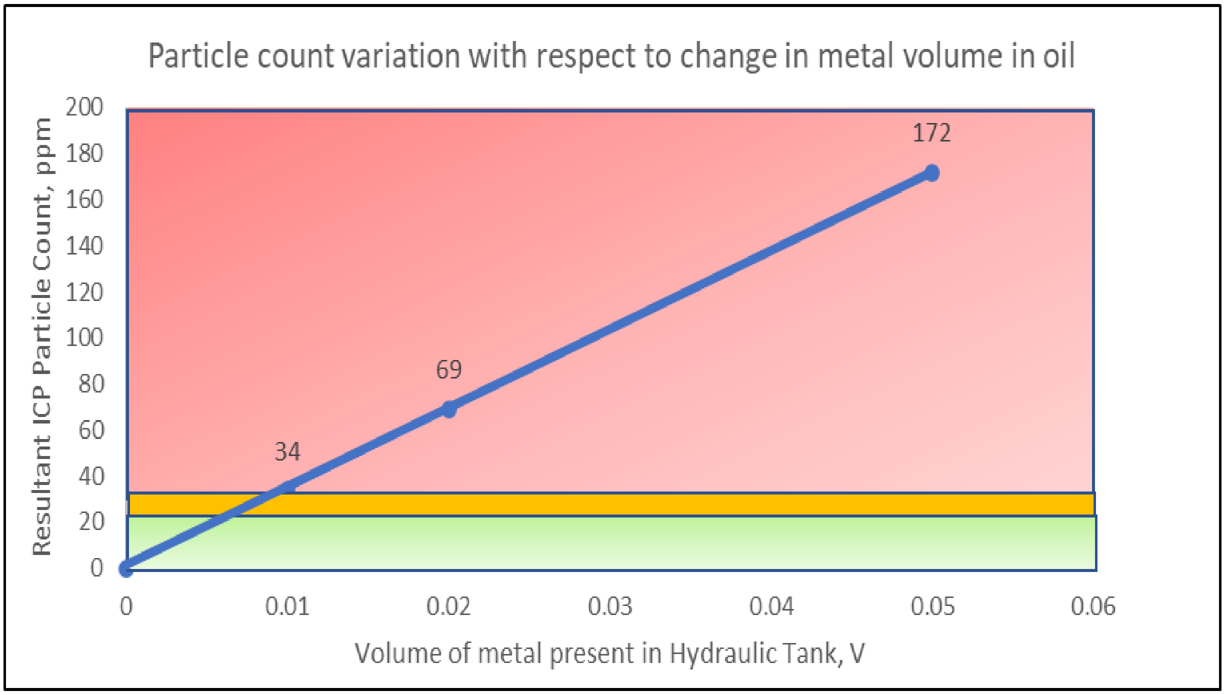


Figure 12. Particle count variation with change in metal volume in hydraulic oil

The iron particle variation in oil with respect to change in metal volume is depicted in the graph represented in Figure 12. The green shade in the graph indicates acceptable range of particles, the orange shade indicates reportable range, and the red shade indicates critical levels of particles present in the oil for the resultant ICP count. It can be inferred that if the iron volume in the tank is less than 0.007 L, the ICP particle count is in the acceptable range, i.e., below 22 ppm. The ICP particle count is at a reportable range when the iron volume in the tank is between 0.007 Land 0.01 L. When the iron volume in the hydraulic tank exceeds 0.01 L, the iron count present in the oil is at critical levels which are to be immediately reported. From the graph, it can be noted that the number of particles per milliliter in oil doubles every time the volume doubles. The analysis tool is also used to calculate the relation of ISO code and ICP count with the number of tablespoons of metal present in the oil. Using this information, the minimum amount of material (in tablespoons and liters) that can impact the performance of main pumps, motors, and cylinders are calculated. The analysis gives a benchmark for oil sample analysis to track pump, motors, and cylinder performance and wear. The following tables provide information about the minimum quantity of material wear from the pumps and cylinders of hydraulic system that can lead to their performance issues in the shovels considered.

Particle count analysis for pump performance issues	*Piston to bore clearance should not exceed exceeds 0.002" as per the manual*
Diameter of Piston (in.)	1.25
Height of piston (in.)	3.25
Approximate wear amount to performance reduction per piston (in^3)	0.013
Number of Pistons per pump	9
Total wear per pump section (in^3)	0.115
Total wear per pump section leading to performance change (liters)	**0.002**

Table 1. Particle count analysis for performance issues in pumps

The minimum amount of material wear for performance change in cylinders					
Cylinder Type	Boom	Level	Arm	Bucket	Dump
From Tube (lt)	NA	0.26	0.20	0.14	0.04
From Rod (lt)	0.08	0.08	0.07	0.04	0.01
Tube (tablespoons)	NA	17.31	13.62	9.64	2.41
Rod (tablespoons)	5.16	5.16	4.47	2.87	0.61

Table 2. Particle count analysis for cylinder performance issues

Although oil analysis can be used as a baseline to track the hydraulic system failures, there are some limitations associated with oil analysis. The analysis noted that particles per million double every time the volume doubles. It is challenging to predict failures at early stages as particles present per milliliter (particles/ml) hardly show a significant increase with the change in the volume of particles. The ICP count measurement in the lab has a variation of +/- 15%. Hence, a substantial difference in the number of metal particles present in the oil results in a significant variation in the particle count. For example, the ICP particle count of 150 ppm has variations ranging from 127 ppm to 172 ppm. This means that when the resultant ppm value is 150 ppm, it can range anywhere from 127 to 172 ppm. Hence it is difficult to predict failures only by considering the oil analysis reports.

3.3 Problem Statement

In order to remove wear particles that are induced in the oil during normal shovel operations, different types of filters are used in different locations of the hydraulic system. The basic types of filters in the hydraulic system include suction filters, high-pressure strainer filters, full-flow filters, bypass filters, drain filters, and pilot filters. Hydraulic filters protect hydraulic system components from damage due to contamination of oils or other hydraulic fluid caused by particles. The following table gives information about the different original equipment manufacturer (OEM) filters, their location, and the filtration capacity size of the filters present in the hydraulic system of shovels that considered in this analysis.

Filter	Location	Filtration Capacity (micron-meter)
Suction Filter	Hydraulic Tank	177 µm
High-Pressure Strainer	Behind Main Pumps	120 µm
Full flow filter	Return Line	10 µm
Bypass Filter	Return line	5 µm
Drain Filter	Off-line	10 µm
Pilot Filter	Pilot circuit behind the pilot pump	10 µm

Table 3. Different filters in the hydraulic system and their filtration capacity

In addition to the existing filters, magnetic filters (Mags) are also newly introduced in the hydraulic system that use magnetic technology to filter contaminant particles. Mags remove iron particles of 5 microns and smaller that pass through OEM filters, resulting in cleaner oil. These particles are removed before they enter any pump suction ports. Unlike the OEM filters, Mags are placed in the hydraulic reservoir and are able to operate with clearances that functionally produce no pressure drop and as a result do not “plug” even at high collection amounts.do. Using this location means all reservoir flow will be cleaned before returning to any pump in the system. Mags remove 95% of all iron particles during each pass without restricting the flow.

Magnetic filters are designed to help prevent OEM filters from plugging, allowing them to continue doing their job and remove particles in the reservoir before they can cause damage. During normal operation, Mags should remove around 30% more iron particles ($\leq$ 5 microns). Magnetic filters are said to have a 15+ year life span.

While the addition of the magnetic filters has been widely regarded as positive by the company, the benefits have generally been anecdotal. The study in this section aims to quantify the effects of magnetic filters that are installed on the giant hydraulic-shovel fleet presently working in the mines. Figure 13 presents a flowchart detailing the steps involved in identifying the impact of Mags that use historical failure data and oil analysis results collected for different hydraulic components, along with using troubleshooting information (initial failure reports (IFR)) and detailed disassembly component inspection reports known as component condition reports (CR) that are tabulated for each failure. The details of data collection are presented in the next section.

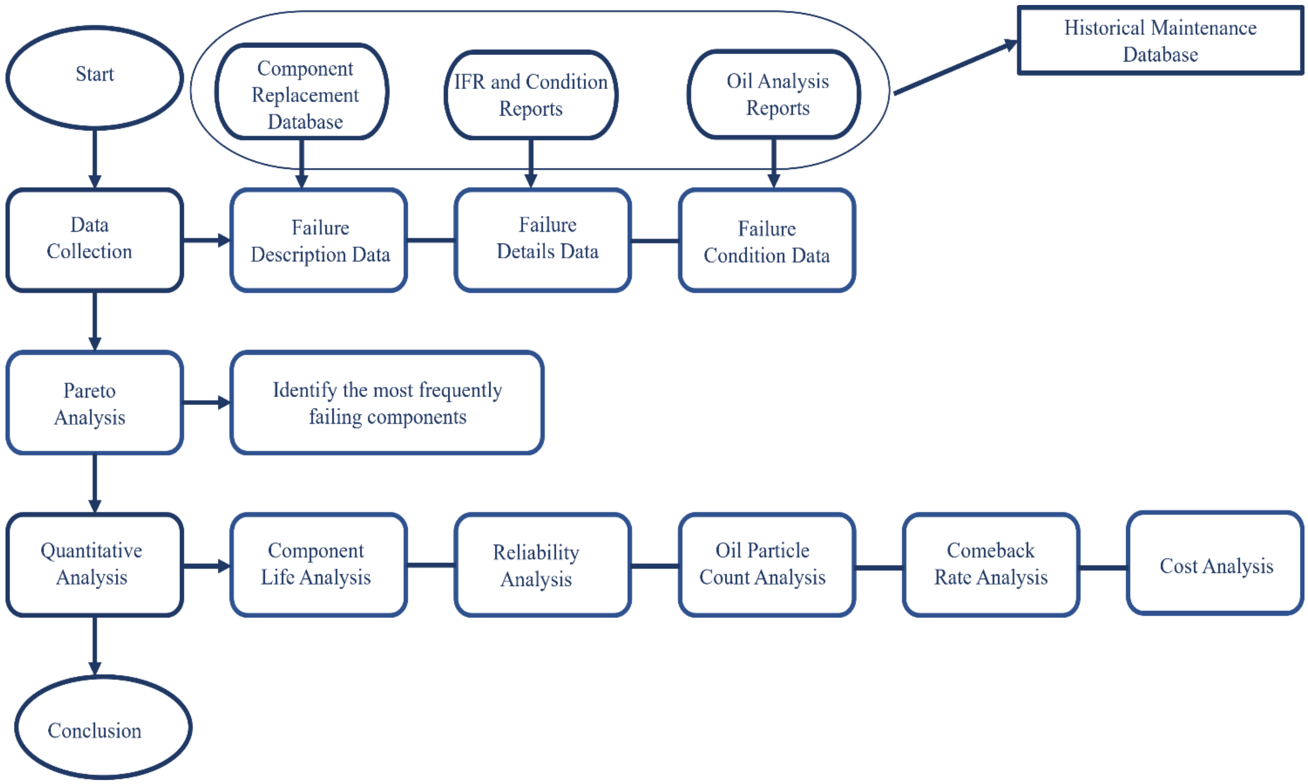


Figure 13. Flowchart detailing the steps involved in identifying magnetic filter impact

Only major component replacement failures of the hydraulic system of shovels are used in this analysis. Different metrics using component average life, reliability analysis, oil particle count analysis, cost analysis, and comeback failure rate analysis are calculated to analyze the quantitative impacts of the filters. The details of each of these methods are explained in the later sections. The main purpose of magnetic filters is to prevent wear and debris related failures of the hydraulic system. As the analysis evaluates the effects of magnetic filters, only wear and debris-related failures are considered in this study. Hence, there is a loss of wear-related information on the components that have functioned and failed for other reasons. The study also does not include time-series data and does not track any component wear rate information and hydraulic system condition and oil conditions over time. Only historical data of the component replacement failures and oil conditions monitored at the time of failure are used. The following sections provide detailed analysis of methods used in identifying the impacts of hydraulic in tank Mag-filters.

3.4 Data Collection

Historical hydraulic component replacement data was collected for three giant hydraulic shovel fleet units working at different mine sites in North America. The three machines are named S1, S2, and S3 for reference. The failure data for the first shovel, S1, is collected over 15 years from 2006. The shovel has been in service for 75,198 meter unit hours (referred to as SMU hours or SMU unit hours throughout the document) as of 2020/08/01. Hydraulic failures within 75,000 hours of S1 operation were used in the analysis. The magnetic filter in this unit was installed at 57,193 SMU hours. The failure data for the second unit, S2, is collected over ten years from 2011. The shovel has been working for 55,972 SMU hours as of 2020/10/06. Hydraulic failures within 48,225 hours of S2 operation were used in the analysis. The magnetic filter in this unit was installed at 29,783 SMU hours. The failure data for the third shovel, S3, is collected for over four years from 2017. The shovel has been working for 19,125 SMU hours as of 2020/11/18. Magnetic filter was installed before the unit started working in the mine. Hydraulic failures within 19,000 hours of S3 operation were used in the analysis.

Historical failure data of the three shovels is maintained in a component replacement database. For every component replacement, a work order (WO) is generated with a unique ID, the date on which the repair started, and the date on which the WO was completed. Additional details w.r.t each WO were collected from different sources for the analysis. The data collected from these sources were categorized into three parts: Failure Description, Failure Details, and Failure Conditions. Figure 14 shows the failure information collected for Mags's analysis and different sources used to obtain this information. A total of 750 failure records of hydraulic component failure replacements of the three units of the hydraulic shovel fleet were collected.

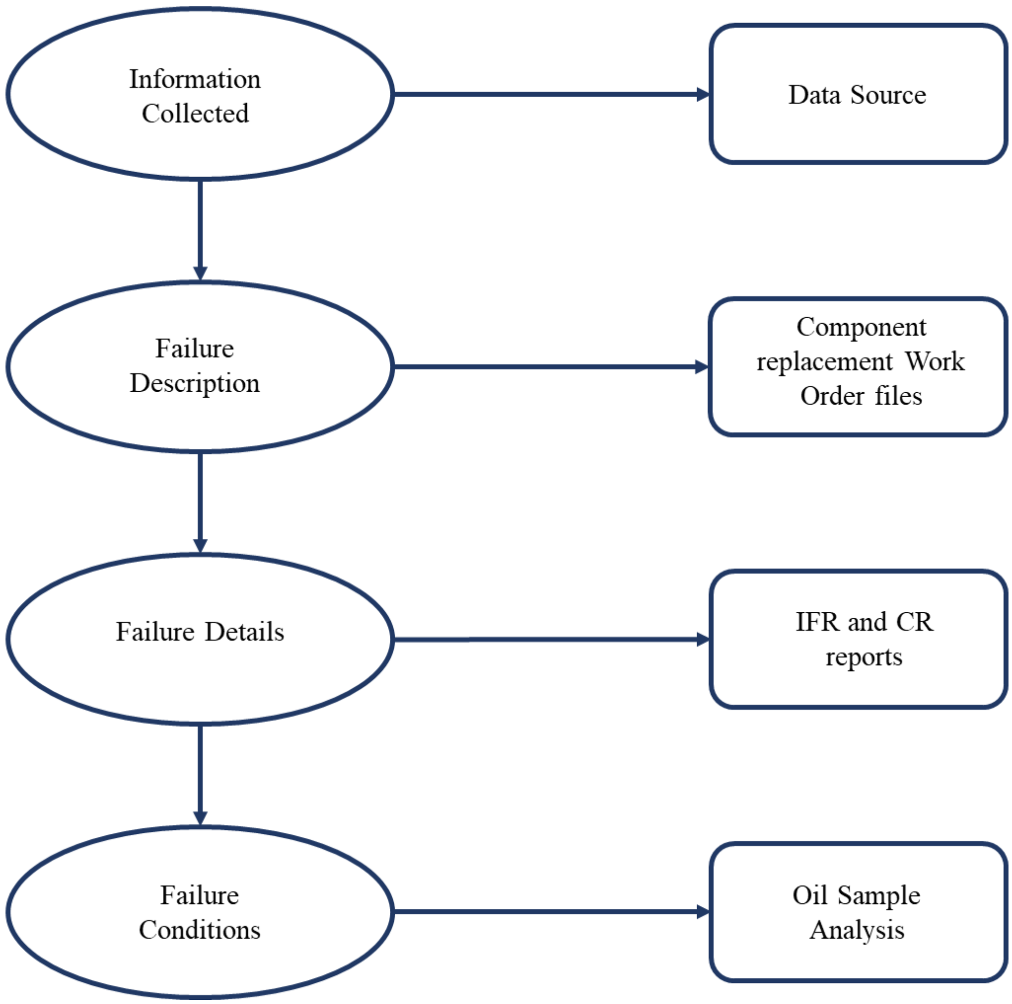


Figure 14. Details on hydraulic failure data collection

The following information is collected in each of these three categories of data collected:

- Failure Description: Data on work order, time of failure, unit number, fleet information, SMU unit at the time of failure, component description, and information about magnetic filter installation are collected in this category.

- Failure Details: Details of the failures such as initial complaint, cause of failure, mode of failure, failure location, if the warranty is applicable on the failed component, and if the failure was caused by damage (operator abuse) are available in this set of information.
- Failure Conditions: Conditions of the oil at the time of failure and prior to failure, such as wear particles generation at the time of failure; if the oil was changed before failure, meter reading since the last oil analysis sample was taken; if the oil analysis indicated any impending failure, particle count at the time of failure, particle count 600 hours before the failure, particle count 1200 hours prior to the failure and H_2O level in the oil at the time of failure are recorded from the oil analysis results. Particle count measurement measures the number of 4μ, 6μ, 14μ, 25μ, 35μ, and 70μ particles (ppm) in the oil. The following table represents the percentage of the information available for each type of data collected in the three different units of shovels considered for analysis.

		Percentage of Data Available		
		Failure Data Category		
Unit	Time Period	Failure Description	Failure Details	Failure Conditions
S1	2005-12-19 to 2020-05-12	100%	40%	85%
S2	2011-02-06 to 2020-03-09	100%	57%	83%
S3	2016-12-20 to 2020-06-29	100%	32%	66%

Table 4. Percentage of Information available for each unit

Initial failure reports contain information recorded on-site at the time of failures. A typical IFR has a unique work order and consists of information regarding the unit-id, hydraulic failure images, description of failure (example: severe damage to the cylinder eye bearing bore and snap ring broken), and details regarding wear of the component. A sample work order history is presented in Appendix A.

Condition reports contain all component rebuild (maintenance/ repair work) data carried out on equipment. A typical condition report consists of essential information such as the associated work order replacement id, equipment id, the date on which repair work has started and ended, a brief description of the failure, and other information about warranty and damage. A sample condition report is presented in Appendix A.

3.5 Data Processing

The initial stage in evaluating Mags' performance is to filter for wear and contamination-related failures and eliminate component failures caused by physical damage and warranty-covered component failures. Failures related to design flaws and discrete components that do not satisfy manufacturer's standards are typically covered under warranty. As wear-related failures are not covered under warranty, the warranty failures have been eliminated. A few examples of wear-related failures in hydraulic shovels include:

- Scorn, worn, and cavitated cylinder and swash plates
- Boom cylinder seal failure – internally bypassing
- Clam Cylinder Leaking
- Main Pump - Seal Failure - Hardened and Flattened – Leaking oil

A few examples of warranty-related hydraulic failures are:

- Rotary Group Free Spinning
- Servo Spools Backed Out
- Poor Rebuild - Did Not Lap Cylinder Block and Valve Plate Surfaces Together -Leaking Internally and Plate Separation

Some examples of hydraulic damage failures are as follows:

- Overloading - Delivery tube hit hard and bent; damage sustained by overloading
- Third Party Damage to Chrome - External Leakage

Figure 15 represents the percentage of wear-related, warranty, and damage failures. 86 % of the total hydraulic component replacements were due to wear and contamination related problems, 5% of the component failures were due to physical damage, and 9% were warranty covered failures.

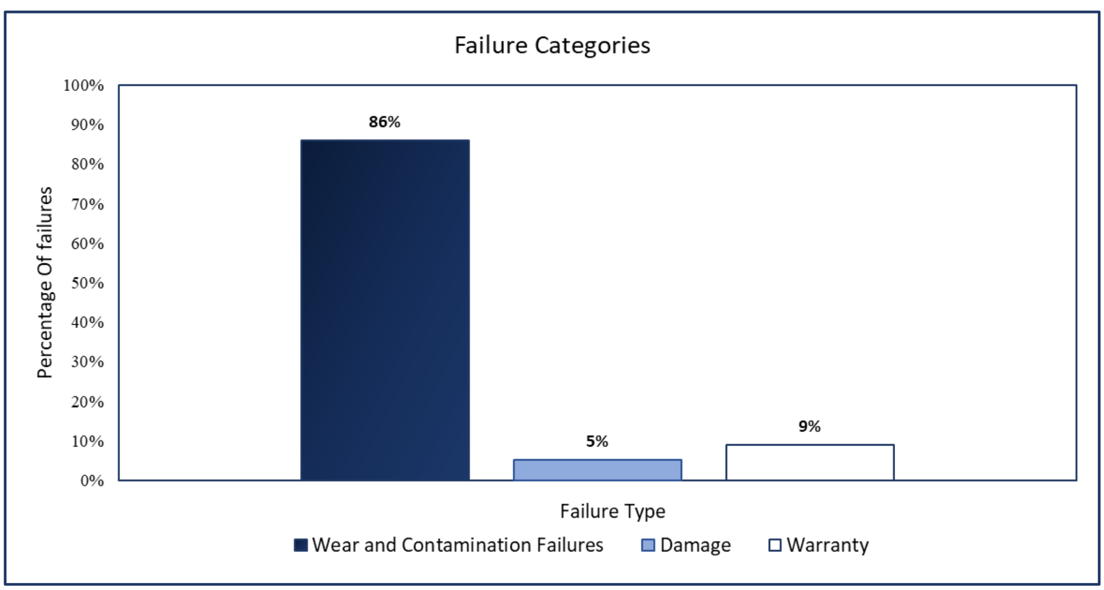


Figure 15. Different types of hydraulic failures and their proportions

After the exclusion of damage and warranty failures, the wear contamination failures were used to analyze the quantitative impacts of Mag-filters. Different criteria such as average component life, reliability estimations, particle count analysis, comeback failure rates, and cost analysis were used to assess the performance of filters. These indicators were used to assess Mag's performance throughout various aspects and potential benefits/effects. The component life analysis is a primary key performance indicator that provides an overview of the equipment failures. The reliability analysis describes component behavior over time. The oil particle count analysis aids in determining the hydraulic conditions at the time of failure and how they have changed since Mag-filters were installed. The measures for comeback failure rate will aid in determining how component failure affects other hydraulic components. The cost analysis metrics would indicate the difference in cost per failure before and after magnetic filter installation. Figure 16 describes different steps used in the quantitative analysis of Mag-filters.

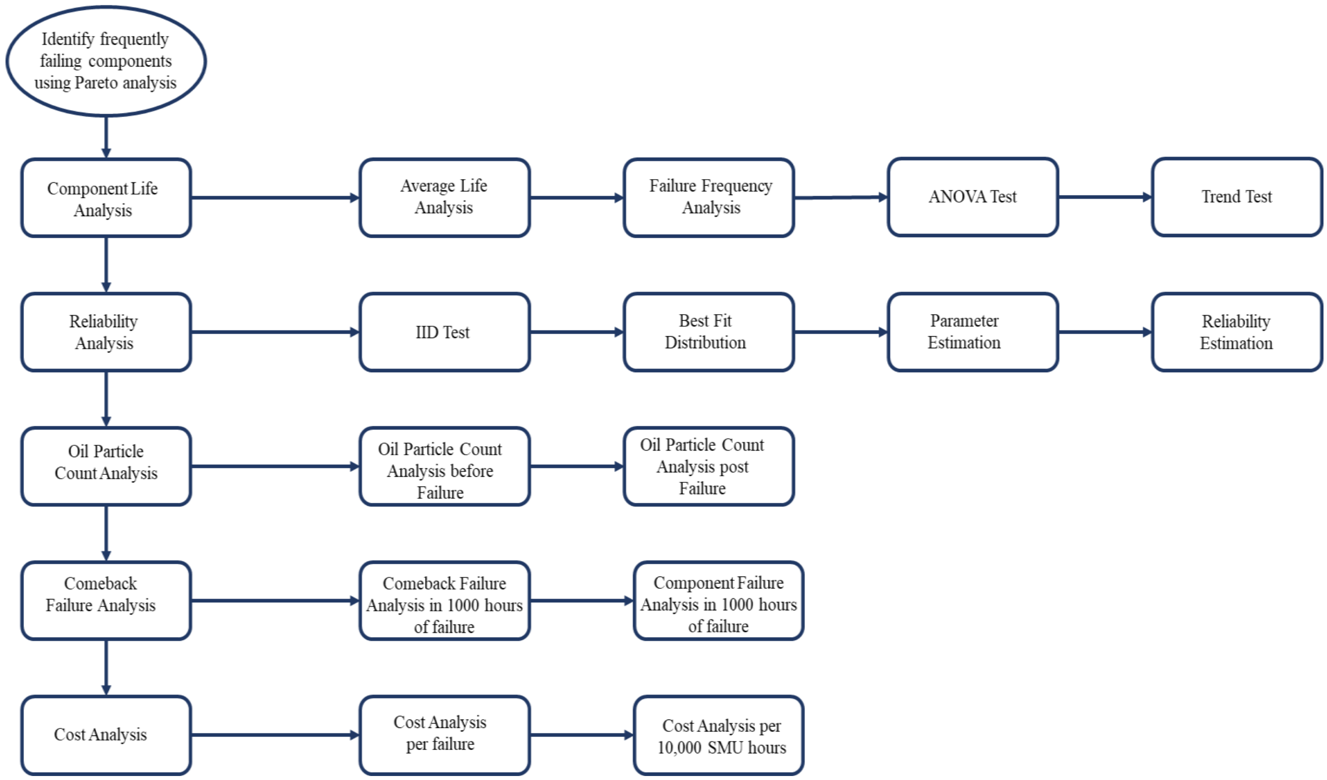


Figure 16. Different methods used in measuring efficiency of magnetic filters

3.6 Hydraulic System of Shovels

The hydraulic system consists of the pilot circuit, main circuit, pump transmission oil cooling circuit, oil cooler fan motor circuit, compressor motor circuit, and the travel shock damper/travel stop circuit. There are 45 different components located on either side of the system. The left engine provides power to the components on the left side of the system, and the right engine provides power to the right-side hydraulic components. There is one hydraulic tank in the system with 2200L storage capacity. The following circuits constitute the hydraulic system:

- Pilot Circuit supplies the pressure oil from the pilot pump to the operation control circuit, the pump control circuit, the travel mode control circuit, the travel parking brake release circuit, the swing parking brake release circuit, the oil cooler fan motor speed control circuit, the Fast-Filling panel control circuit, and the auto lubrication circuit.

- Main Circuit Controls the pressure oil from the main pump at the control valve to drive the hydraulic actuators such as the cylinders and the motors.
- Oil Cooler Fan Motor Circuit drives the oil cooler fan motor using the pressure oil from the oil cooler fan motor drive pump.
- Compressor Motor Circuit drives the air compressor motor using the pressure oil from the compressor motor drive pump.
- Pump Transmission Oil Cooling Circuit sends the transmission oil by means of the pump transmission oil pump to the oil cooler to lubricate and cool the transmission.
- Travel Shock Damper / Travel Stop Circuit damps shock loads applied to the front idlers and stops traveling if excessive loads are applied.

3.7 Quantitative Analysis of Mag-filters

The top eight components of the hydraulic system with the highest failure frequency were identified and the behavior of these components before and after the installation of Mag-filters were used to quantify the effects of the filters. Pareto analysis is performed to determine the components that contributed to the highest failure frequency, as represented in Figure 17 and Figure 19. The X-axis on the graphs represents the failure component, and the Y-axis represents the percentage of failures for each component.

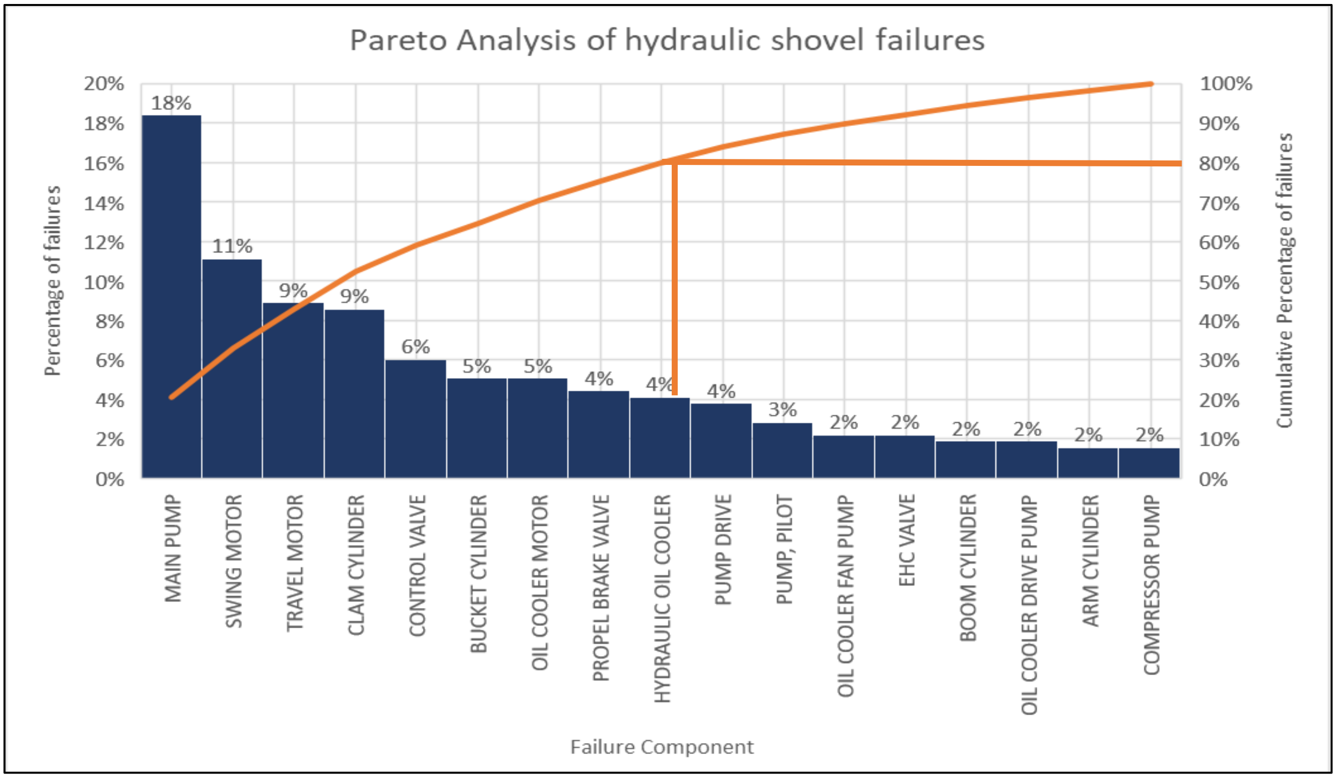


Figure 17. Pareto Charts of frequency of hydraulic failures

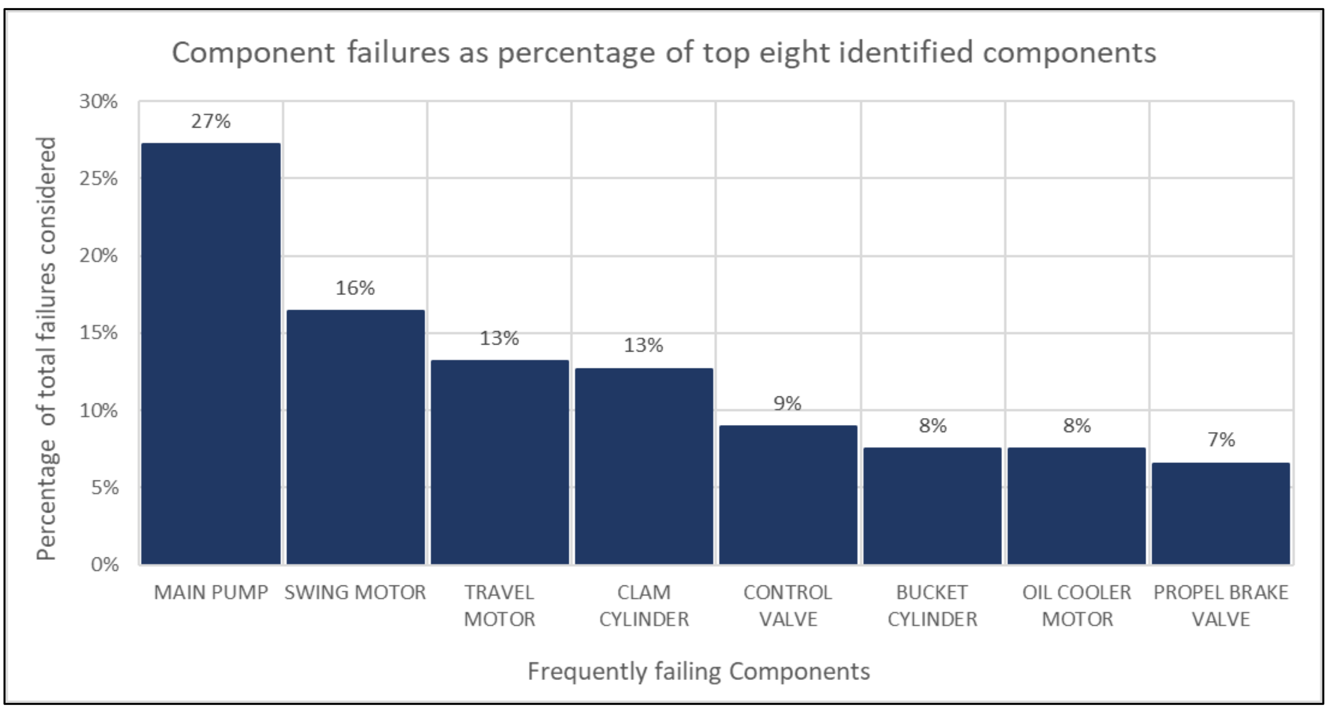


Figure 18. Percentage of component failures of top eight identified components

Figure 17 indicates that main pumps have the highest failure frequency followed by the swing motors, travel motors, clam cylinders, control valves, bucket cylinders, oil cooler fan motors, and propel brake valves. The graph indicates that main pumps account for 20% of all failures. As a result, it can be deduced that main pumps fail more often than other components. According to the graph, the top eight components are responsible for 80% of failures. As a result, these components were included in the research to better understand the effects of Mag- filters. Figure 18 shows component failures as a percentage of the top eight components identified. The graph represents the distribution of data among the top eight identified components considered for further analysis. Main pumps, which have the highest failure frequency in the system, account for 27% of total failures considered in the analysis, whereas propel brakes, which have the eighth highest failure frequency in the system, account for roughly 7% of total failures considered in this study. Different approaches represented in Figure 16 were used to analyze and quantify the effects of Mag-filters. The five methods used in this study include component life analysis, reliability estimations, oil particle count analysis, cost analysis, and study of comeback failures. The data for each of these identified components were classified into two groups (before the installation of Mag-filters and post-installation of Mag-filters). These methods were used to evaluate the performance of the two groups. The details of the analysis are presented in the following sections.

After identifying the top eight components, Mag-filters are analyzed on five different metrics as represented in Figure 16. For all the eight identified components, performance was measured in terms of component life, reliability at different given time, oil analysis methods, cost analysis methods, and comeback failure method.

In the component life analysis, performance is measured mainly based on the average life of hydraulic components before and after the installation of Mag-filters. The ANOVA test is used to see if there is a significant difference in component life before and after Mag-filter installation. The frequency of failures is calculated before and after installation of filters. The variation in component life before and after installation is visualized using trend and histogram charts.

Reliability analysis estimates the variation of probability of failure over time. Reliability of the components before and after the installation of Mag-filters is compared at 5000 hours, 10,000 hours, and 15,000 hours. This gives an idea about the changes in wear rate and particle contamination failures in the system after the installation of Mag-filters.

Particle count Analysis method is used to understand the change in particle count behavior before and after installation of Mag-filters. Pie charts are used to analyze percentage of failures in acceptable range, reportable and critical range.

Comeback failure analysis is used to analyze the impacts of component failure on the performance of other components. The main idea of the analysis is that magnetic filters trap particles when a component fails. Because the particles are trapped in the filter, there is no spike in particles in the hydraulic oil, which helps to prevent failures of other components that might otherwise be influenced by prior failures. Hence, comeback failures in the next 1000 hours of operation following a component failure are investigated, as well as the change in behavior following the installation of Mag-filters is studied.

Cost analysis helps in identifying the difference in cost spent on each failure before and after installation of Mag-filters across different components. This gives an idea about the cost saved per failure after the installation of Mag-filters. The following sections represent in detail on the analysis of impacts of Magnetic filters.

3.8 Component Life Analysis

For component life analysis of hydraulic system, failure data is initially treated for outliers and missing values. The performance of magnetic filters is measured mainly based on the average life of hydraulic components before and after the installation of Mag-filters. ANOVA is used to check if there is significant variation in the average life in the two groups. The frequency of failures for every 10,000 hours before and after installation is calculated. Trend charts, box plots, and histograms are used to study the variation of component life before and after installation.

3.8.1 Outlier and Missing value treatment

Outliers in the collected data are the points that are distant from other similar points. Outlier data represents component life beyond manufacturers' standards. The cause of outliers is due to variability and measurement errors in manually recording the failure SMU hours. They are detected using a z-score. Z-score indicates how much a given value differs from the standard deviation. The z-score of a value is provided by:

$$\boldsymbol{Z = \frac{x - \mu}{\sigma}} \tag{38}$$

The value of z represents the standard score of the observed value. x represents the observed value, μ represents the mean of the component life, and σ represents the standard deviation. Data points that were 1.5 standard scores above the mean were eliminated.

3.8.2 Average Life Analysis

After eliminating the outliers, the average life of the components before and after the installation was compared for each of the top eight identified components. A one-way ANOVA test was used to compare if the average of the component life before and after installation were statistically significant. ANOVA works by comparing the variance of samples within and between the groups. If there is a signification variation (spread of data away from the mean) within the two groups, then there are greater chances that the mean of the samples selected are different due to chance. The F-value and p-value were used to understand whether there is a significant difference in the means of the two samples considered.

The F-value column in the ANOVA test is the test statistic from the F test: the mean square of each independent variable divided by the mean square of the residuals. The larger the F value, the more likely the variation associated with the independent variable is real and not due to chance. The p-value is the probability value of the F-statistic. This shows how likely it is that the F-value calculated from the test would have occurred if the null hypothesis of no difference among group means were true. The null hypothesis of the test states: Means of the population are equal, all population samples have a common variance, and there is no significant difference in the mean of the groups. The alternate hypothesis states that the means of two populations vary significantly. The p-value is the evidence against the null hypothesis. Smaller the p-value (<0.05), the stronger the evidence to reject the null hypothesis. For the component analysis of Mag-filters, the aim is to check if the mean component life before and after the installation of Mag-filters is significantly different or not. The null hypothesis is rejected if there is a significant difference between the two means i.e., if $F > F_{crit}$ (F-critical value) for given degrees of freedom and p-value < 0.05.

Histograms were used to analyze the component life variation after the installation of magnetic filters by grouping TTF data of components into bins of equal width of 5000 hours to detect early failures (component life < 5000 hours), their causes, and to check if component life has increased after the installation of Mag-filters. The number of component failures every 10, 000-SMU hours and performance of components after installation was compared to the average component performance before installation for each of the top eight components. Trend charts were used to study how the component life changed over time. The results of component life analysis for each of the eight components are presented in the following section.

3.9 Results of Component Life Analysis

3.9.1 Main Pump Analysis

Main Pumps are located in the main circuit of the hydraulic system. Main pumps transport fluid from the pilot circuit to hydraulic actuators through control valves. Two-unit main pumps are arranged in tandem, and 12 units in total (six units each on the right and left engines) are mounted on the machine. The right and left suction manifolds are connected to the hydraulic oil tank by two suction pipes. Each main pump draws hydraulic oil from the suction manifold. 6 suction filter units are provided in the hydraulic oil tank. As a control lever is operated, the pilot pressure oil from the pilot valve flows through the main pump regulator at the flow rate control pressure Pi. The pressure oil may not be supplied to the corresponding circuit depending on the operations of the control valves, even if the main pump delivers pressure oil at the maximum flow rate. Hydraulic oil delivered from a total of 12 main pump units is routed to 6 control valve units via the high-pressure strainer. 3 control valve units are located on both right and left sides. Hydraulic oil delivered from main pumps (MP1 to MP6) on the left engine flows to the 3 control valve units on the left side, and hydraulic oil delivered from the main pump (MP7 to MP12) flows to the 3 control valve units on the right side. The following section presents component life analysis results of the main pumps.

The average life of main pumps before the elimination of outliers is 16,231 hours, and the standard deviation (SD) is 9395.9 hours. Data points above 1.5 standard scores, i.e., failures with TTF > 30,000 hours, were eliminated. The following plot represents outliers in the main pump TTF data. The X-axis represents the meter reading (SMU hours) of the shovel fleet and the Y-axis represents the TTF hours (Component Life) of the main pumps. 12% data points that had errors in recordings with pump life of greater than 30,000 hours. These data points were eliminated.

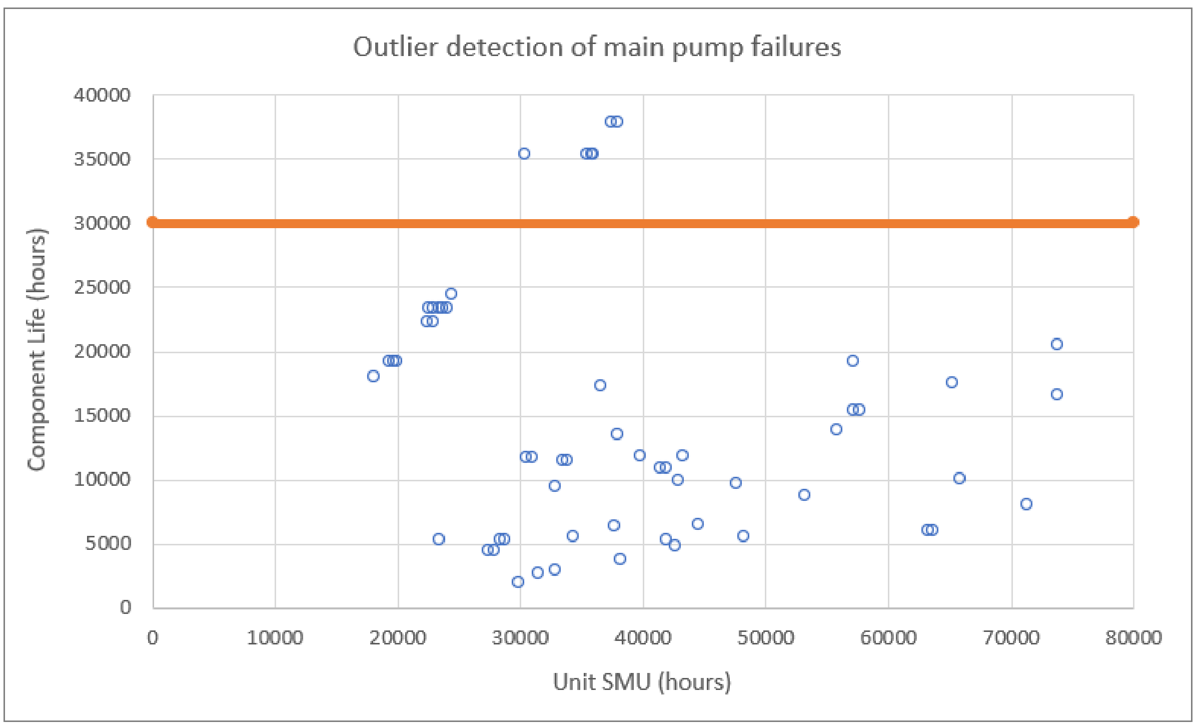


Figure 19. Outlier detection in main pump component analysis

After eliminating the outliers, the main pump average life was calculated. Table.5 through Table.7 shows the component life analysis of main pumps. The average life of pumps before the installation of Mag-filters is 14915 hours, and post-installation is 12197 hours. The average performance of main pumps is 12568 hours. 52% of the pumps performed better than average life before Mags' installation and only 38% of the pumps performed better than average life post-installation. When compared to the average life of main pumps before Mags' installation (i.e., 14915 hours), 32% of the magnetic filter pumps have performed better than the pump average life before installation. On average, in the S1 unit, 3 pumps were replaced every 10,000 hours before installation, and 3 pumps were replaced every 10,000 hours post-Magnetic filter installations. On average, in S2, 1 pump was replaced every 10,000 hours before installation and 3 post-Mag installations. The average life of S1 main pumps has increased post-installation of Mag-filters.

MAG-FILTERS	AVERAGE LIFE (HOURS)	%PUMPS PERFORMING ABOVE AVERAGE LIFE
BEFORE INSTALLATION	14915	52%
AFTER INSTALLATION	12197	38%

Table 5. Average life analysis of main pump failures

X- Average Life of Main Pumps before Mag-filters installation	*Avg Life – 14915*
%Pumps with higher average life than X after Mags installation	32%

Table 6. Percentage of Mags installed main pumps compared to average life pre-installation

	#PUMPS REPLACE PER 10000 BEFORE MAG-FILTER INSTALLATION	#PUMPS REPLACE PER 10000 HOURS AFTER MAG-FILTER INSTALLATION
S1	3	3
S2	1	3

Table 7. Frequency of main pump failures before and after installation of Mag-filters

ANOVA test was used to check if the mean of the two groups, i.e., (the main pump average life before installation of Mags and the main pump average life after installation of Mags) varied significantly. Normality of main pump failure data is checked by comparing mean and median values of the distribution, skewness of the data, and Normal Probability Distribution (NPP plot). The details of the analysis are mentioned in Appendix A. With the above tests, the data is nearly normally distributed and hence ANOVA test is used to check the significance. The ANOVA test results for the main pump average life analysis before and after installation of Mag-filters are represented in Table.8.

The null hypothesis of the test states: Means of the samples are equal, and all populations have a common variance. The alternate hypothesis states that the means of two populations vary significantly. For the main pump component analysis, F > Fcrit and $p < 0.05$. Hence, the two-sample means vary significantly. This infers the component life of main pumps post-installation of Mags and the component life prior to Mag-filters installation vary significantly. The decrease in component life can be largely influenced by the working location and condition of the equipment, physical properties that impact rock and soil hardness and abrasion index.

ANOVA - MAIN PUMP

SOURCE OF VARIATION	*SS*	*Df*	*MS*	*F*	*P-value*	*F crit*
BETWEEN GROUPS	81115623	1	81115623	1.901863	0.017517	0.072654
WITHIN GROUPS	1.79E+09	42	42650607			
TOTAL	1.87E+09	43				

Table 8. ANOVA results for average life analysis of main pumps

Figure 20 and Figure 21 represent histograms for TTF data of the main pumps. TTF data is grouped into different bins with equal widths of 5000 component life hours and the number of failures is plotted on the y axis corresponding to each bin. Figure 19 indicates that overall, less than 10% of the pumps have failed in less than 5000 hours. The pumps (with TTF < 5000 hours) are from unit S2 and have failed after the Mag-filters were installed. Most pumps have component life in the range of 5000 – 10,000 hours and 15,000 – 20,000 hours intervals. The number of failure before and after installment is randomly distributed among all intervals. Figure 20 indicates the number of failures in different TTF intervals before and after the installation of Mag-filters. Around 4% of pumps have less than 5,000 hours of component life before the installation of magnetic filters and 10% of pumps have failed in less than 5,000 hours after installation of Mag-filters. Around 26% of pumps have achieved a pump life of greater than 20,000 hours before the installation of Mag-filters and around 14% of pumps have achieved a pump life of greater than 20,000 hours after the installation of Mag-filters. The histogram charts indicate that there is same distribution of pumps having component life of <10,000 hours before and after the installation of Mag-filters whereas a higher number of pumps have achieved greater than 20,000 hours of life before the installation of Mags compared to post-installation of filters.

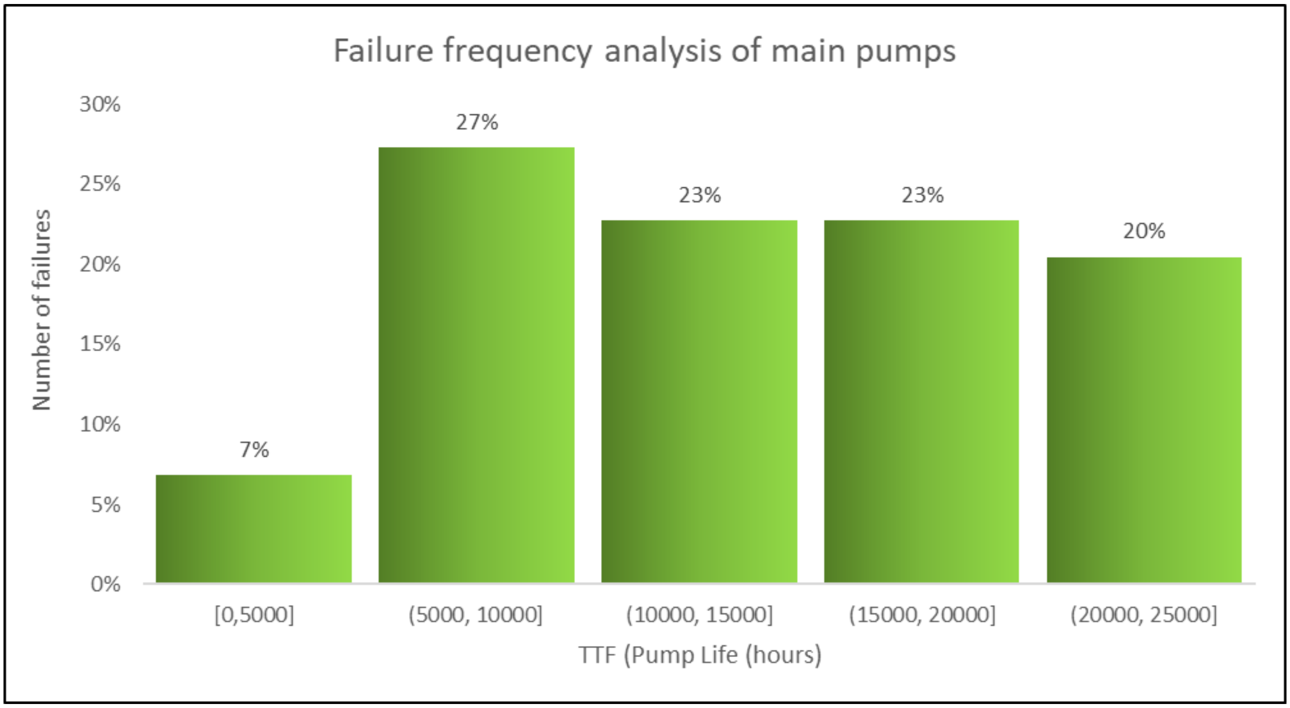


Figure 20. Failure frequency of main pumps in different TTF intervals

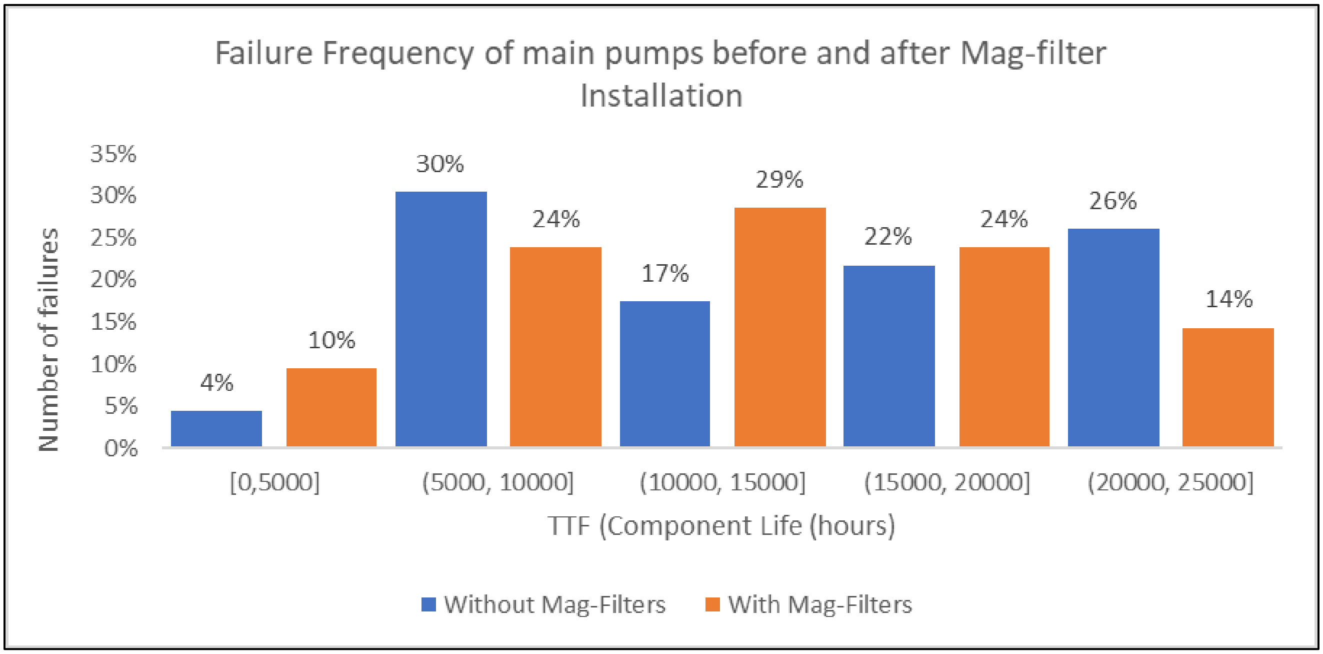


Figure 21. Failure frequency of main pumps in different TTF intervals before and after Mag-filters installation

In order to study the details of performance difference before and after installation of Mag-filters, trend charts were plotted to study the change of component life over time. Figure 22 and Figure 23 represent the component life trend before and after the installation of Mag-filters in both S1and S2 units. Unit SMU hours VS Component Life graphs were plotted to check for the variation in pump performance before and after the installation of magnetic filters. The red line denotes the Mag-filter installation SMU hours. From Figures. 22 and Figure 23, it can be inferred that the initially installed pumps when the machine was newer have greater average life (>20,000 hours) and the performance of pumps has steadily been decreasing as the equipment is aging. From Figure 22, in the S1 unit, the pump performance was observed to be varying randomly post-installation of magnetic filters whereas, in the S2 unit (Figure 23), the component life of pumps has improved steadily after installation of Mag-filters, but the pumps are not able to achieve their full life of (>20000 hours). The reasons might be the location of the equipment in the mine site or the previous failures and wear generation in the hydraulic oil over time are having greater impact on pump average life. However, after the installation of Mag-filters, the pump life seems to have improved.

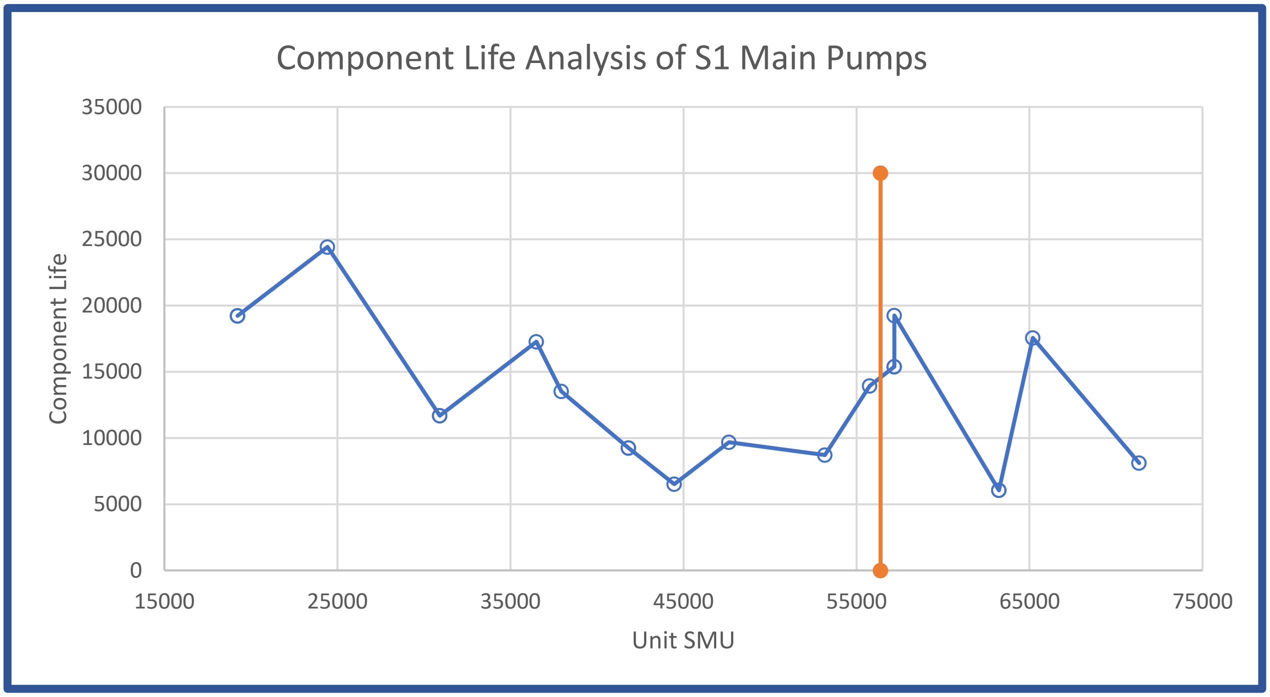


Figure 22. Component Life Analysis of S1 Main Pumps

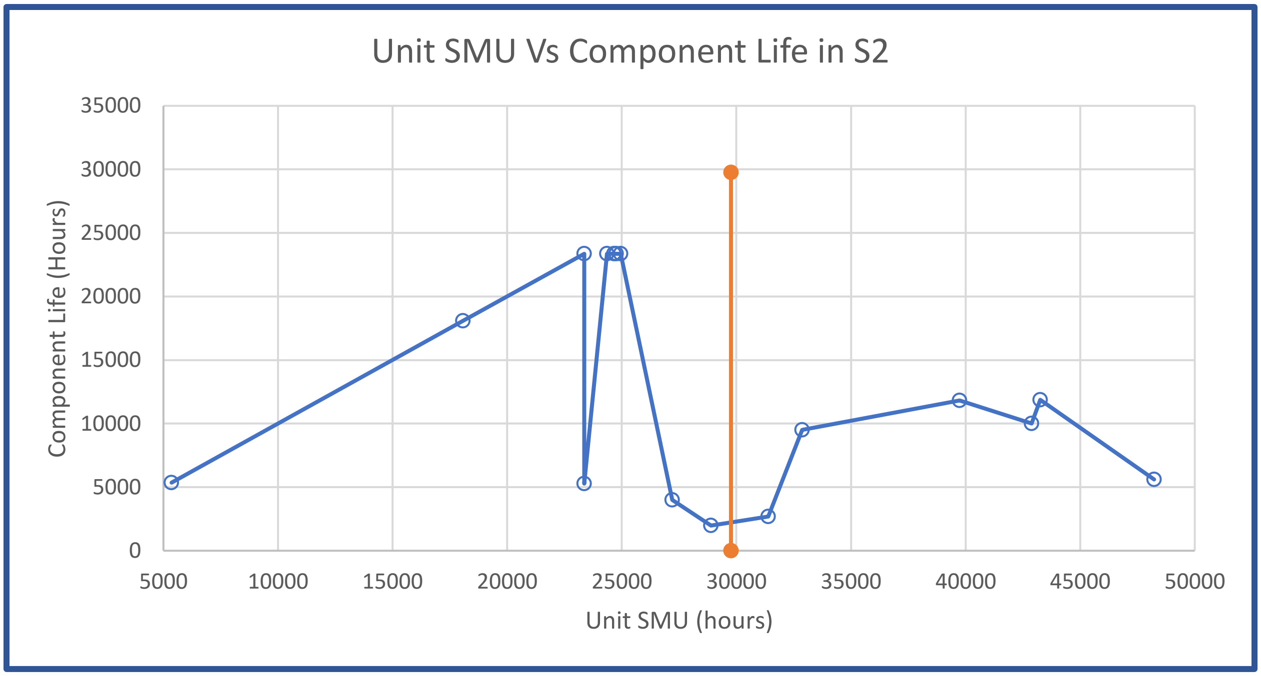


Figure 23. Component Life Analysis of S2 Main Pumps

3.9.2 Swing Motor Analysis

Swing motors are located in the main circuit of the hydraulic system. Swing motors are hydraulic motors that form the proper turning force of excavator booms using oil pressure, shift rotation either to the right or left and convey or cut off turning power. Swing motors are supplied oil from upper left and lower right control valves. There are 4 swing motors, two of which are situated in the front and two in the rear of the hydraulic system. The swing motor is driven by the pressure oil from the pump and transmits its output to the reduction gear. The following section presents component life analysis results of swing motors.

The average life of swing motors before eliminating outliers is 13,273 hours and the SD is 8786.12. Data points above 1.5 standard scores, i.e., failures with TTF > 30,000 hours, were eliminated. The following plot represents the outliers in TTF data of swing motors. The X-axis represents the meter reading (SMU hours) of the shovel fleet and the Y-axis represents the TTF hours of the swing motors. One swing motor failure had errors in recordings with motor life greater than 30,000 hours. This data point was eliminated.

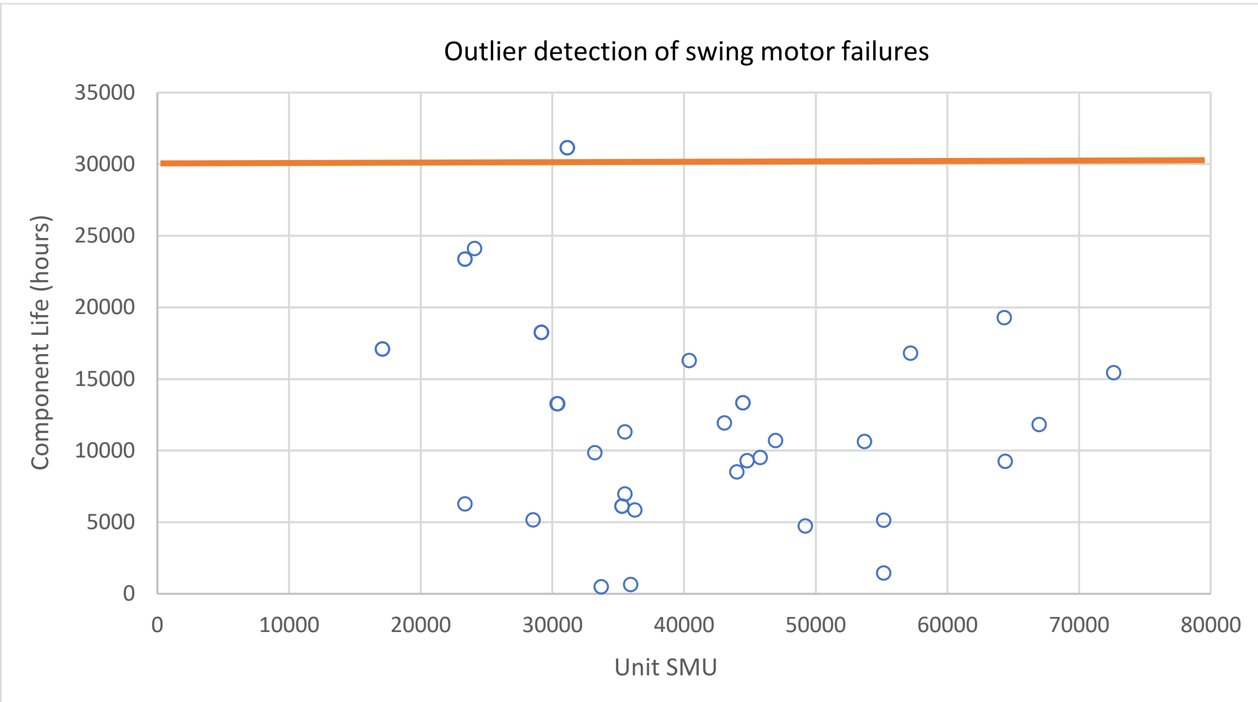


Figure 24. Outlier detection of swing motor failures

After eliminating the outlier, the swing motor average life was calculated. Table.9 through Table. 11 shows the component life analysis of swing motors. The average life of motors before installation of Mag-filters is 12885 hours and post-installation is 11577 hours. The average performance of swing motors is 12230 hours. 53% of the motors performed better than average before Mag installation and 39% of the motors performed better than average after installation. When compared to the average life of swing motors before Mag installation (i.e., 12885 hours), 39% of the Mag installed swing motors have performed better than the motor average life before installation. On average, in the S1 unit, 3 motors were replaced every 10,000 hours before installation and 3 motors were replaced every 10,000 hours post-Magnetic filter installations. On average, in S2, 3 motors were replaced every 10,000 hours before installation and 6 post-filter installations.

MAG-FILTERS	AVERAGE LIFE	#MOTORS PERFORMING ABOVE AVERAGE LIFE
BEFORE INSTALLATION	12885	53%
AFTER INSTALLATION	11577	39%

Table 9. Average Life analysis of swing motor failures

X- Average life of swing motors before Mag installation	*Avg Life – 12885*
Percentage of Mag swing motors performing above average life X	*39%*

Table 10. Performance of Mag installed swing motors compared to average life pre-installation

	TOTAL LIFE	#MOTORS REPLACED PER 10000 HOURS BEFORE FILTER INSTALLATION	#MOTORS REPLACED PER 10000 HOURS AFTER FILTER INSTALLATION
S1	72987	3	3
S2	46963	3	6

Table 11. Frequency of swing motor failures before and after installation of Mag-filters

ANOVA test was used to check if the mean of the two groups, i.e., (the swing motor average life before installation of Mag-filters and the swing motor average life after installation of Mag-filters) varied significantly. Normality of the data is checked by comparing mean and median values of the distribution, and Normal Probability Distribution (NPP plot). The details of the analysis are mentioned in Appendix A. With the above tests, the data nearly followed a normal distribution, and hence ANOVA test was used to check the difference in mean. The following table displays the ANOVA results for swing motor average life before and after installation of Mag-filters.

The null hypothesis of the test states: Mean of swing motor life before installation and after installation of filters are equal, and all both populations have a common variance. The alternate hypothesis states that the means of two populations vary significantly. For the swing motor component analysis, $F < F_{crit}$ and $p > 0.05$. Hence, the two means do not vary significantly. This means there is no significant difference in the component life before and after installation of Mag-filters.

ANOVA TEST - SWING MOTORS						
SOURCE OF VARIATION	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
BETWEEN GROUPS	14984015	1	14984015	0.36733	0.548489	4.130018
WITHIN GROUPS	1.39E+09	34	40791727			
TOTAL	1.4E+09	35				

Table 12. ANOVA test for swing motor average life analysis

Figure 25 and Figure 26 represent histograms for TTF data of the swing motors. TTF data is grouped into different bins with equal widths of 5000 component life hours and the number of failures is plotted on the y axis corresponding to each bin. Figure 25 indicates that around 10% of the motors have failed in less than 5000 hours. Failures are evenly distributed in all intervals from 5000 to 20,000. This means most motors fail randomly anywhere between 5000 to 15,000 hours. Only 8% of swing motors in total have achieved full life. There are 10% more failures in less than 10,000 hours after the installation of Mag-filters. 20% of swing motors have achieved more than 20,000 hours of life before installation, but none of the motors have achieved greater than 20,000 hours of component life after Mag installation. Figure 26 indicates that the performance of swing motors before and after installation is almost similar.

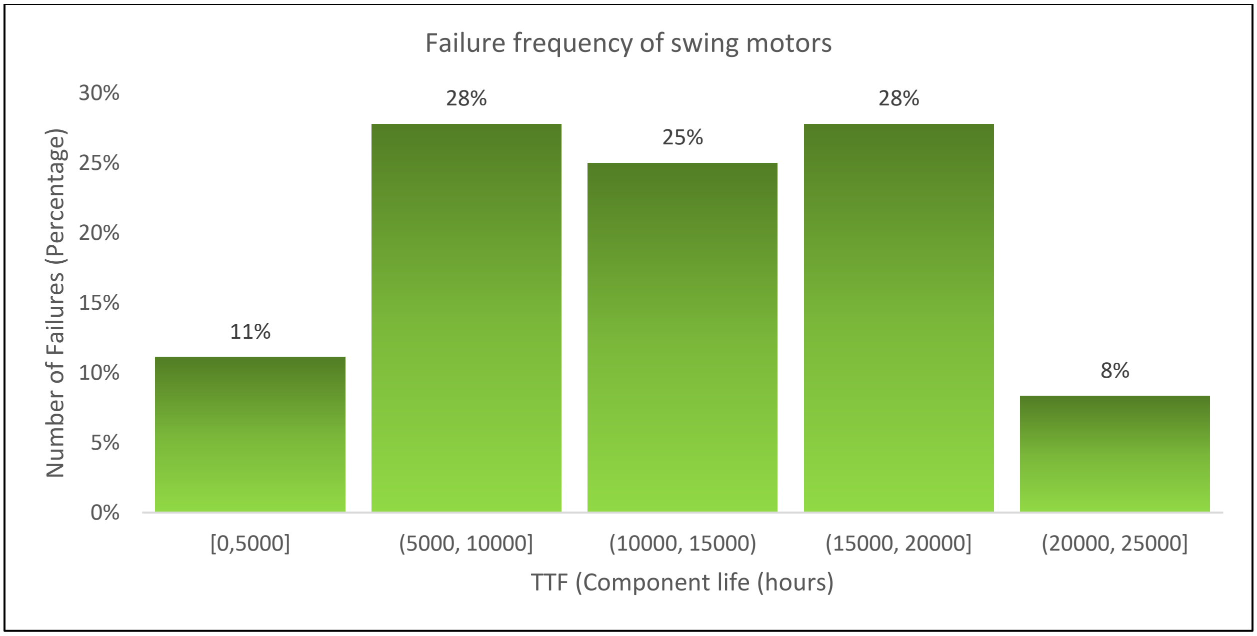


Figure 25. Failure frequency of Swing Motors

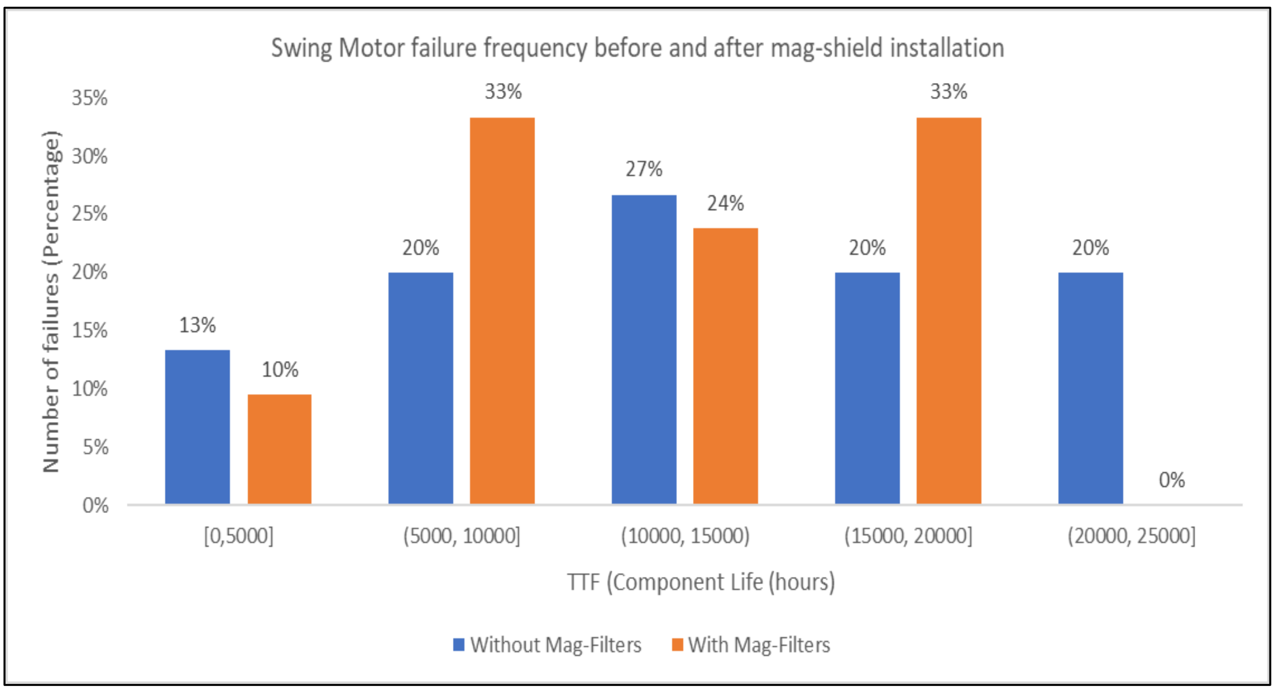


Figure 26. Failure Frequency of Swing motors before and after Mag installation

Figure 27 and Figure 28 represent the component life trend before and after the installation of Mag-filters in both S1 and S2 units. Unit SMU hours VS Component Life graphs were plotted to analyze the variation in swing motor performance before and after the installation of magnetic filters. The red line denotes the SMU hour at which the filter was installed. From Figure 27 and Figure 28, it can be inferred that the initially installed motors when the machine was newer have greater average life (>20,000 hours) and the performance of motors has steadily been decreasing as the equipment is aging. In S1, the first motor has achieved above 20,000 hours life and then the performance is decreasing constantly. The red line denotes Mag installed SMU hour. The performance after the installation has somewhat increased. In S2, there is no trend in component life and failures are random. The initial motor has achieved high hours whereas there is a significant decline in the second motor. Swing motor life has not increased post-installation of Mags. Instead, the failure rate has increased. The reasons might be the location of the equipment in the mine site, or the previous failures and component wear generation in the oil influencing the average life.

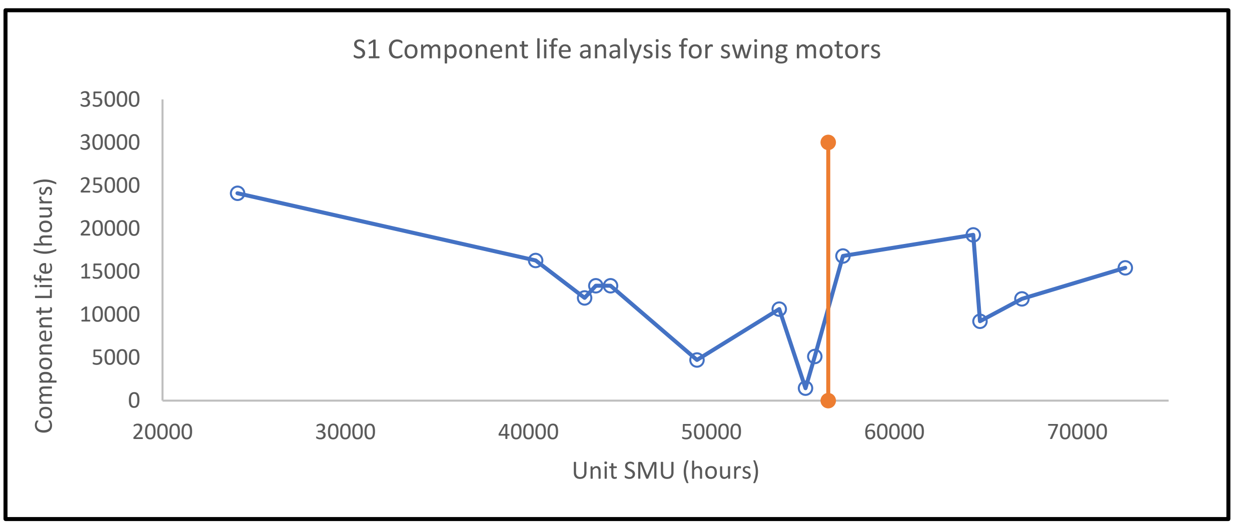


Figure 27. Component Life Analysis of S1 swing motors

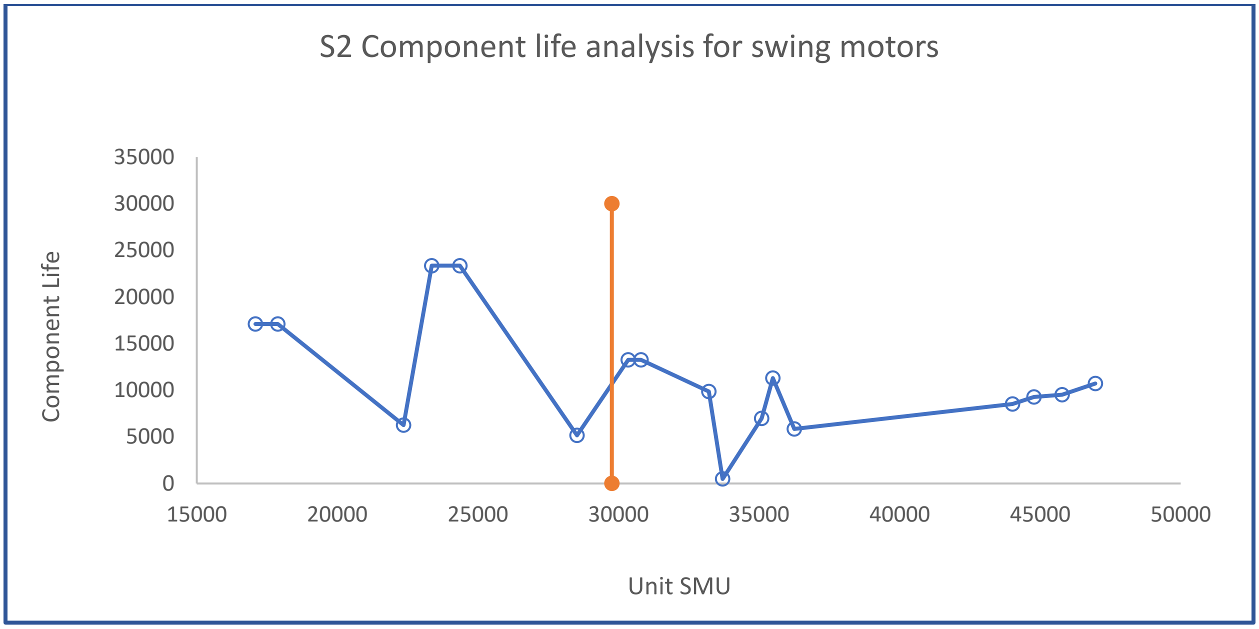


Figure 28. Component Life Analysis of S2 swing motors

3.9.3 Travel Motor Analysis

Travel motor is a part of main circuit of the hydraulic system. Travel motors are hydraulic motors that increase driving power properly for the travel conditions of shovels using oil pressure, maintain adequate speed, shift the gears to move forward or backward, and convey or cut off power. Travel motors are supplied oil from center left and upper right control valves when operating the left track and are supplied oil from lower left and center control valves when operating the right track. There are four travel motors, two of which are located on the left side, and two are located on the right side of the hydraulic system. The travel motor is driven by the pressure oil from the main pump and rotates the travel reduction gear.

The average life of travel motors is 20,410 hours, and the SD is 5440.25 hours. There were no outliers in the dataset. The following plot represents TTF data of travel motors. The X-axis represents the meter reading (SMU hours) of the shovel fleet, and the Y-axis represents the TTF hours of the travel motors.

47% of the travel motor failures were damage-related failures. The component life analysis of travel motors is presented from Table. 13 through Table. 15.

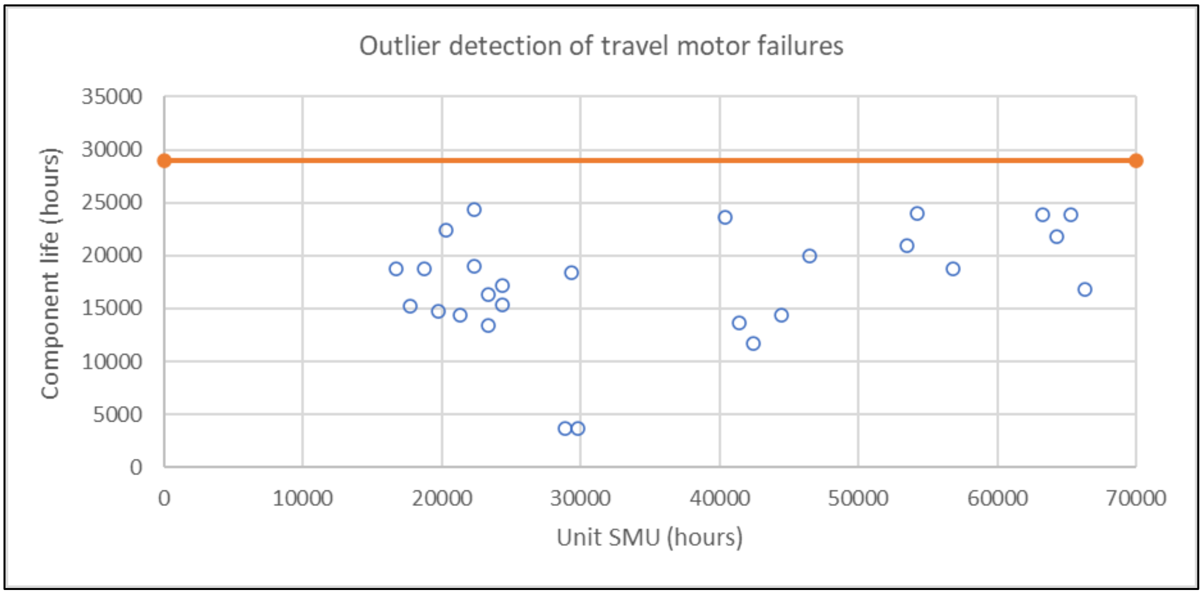


Figure 29. Outlier detection of travel motors

The average life of travel motors before installation of Mags is 18,734 hours, and post-installation is 23,017 hours. The average performance of travel motors is 20,410 hours. 53% of the motors performed better than average life of motors before Mag installation and 100% of the motors performed better than average life after installation. When compared to the average life of travel motors before Mag installation (i.e., 18,734 hours), all the Mag-filter installed travel motors have performed better than the motor's average life before installation. On average, in the S1 unit, two motors are replaced every 10,000 hours before installation and one motor is replaced every 10,000 hours post-Mag installations. On average, in S2, one motor was replaced every 10,000 hours before installation, and the machine has run for 19,375 hours until 2020, and none of the travel motors have failed.

MAG-FILTERS	AVERAGE LIFE	#MOTORS PERFORMING ABOVE AVERAGE LIFE
BEFORE INSTALLATION	18734	57%
AFTER INSTALLATION	23017	100%

Table 13. Average Life analysis of travel motor failures

X- Average life of motors before Mag Installation	*Avg Life – 18734*
#Motors Performing Above Average Life X after installing Mag-filters	100%

Table 14. Performance of Mag installed travel motors compared to average life pre-installation

	#MOTORS REPLACE PER 10000 HOURS BEFORE MAG INSTALLATION	#MOTORS REPLACE PER 1000 HOURS AFTER MAG INSTALLATION
S1	2	2
S2	1	None

Table 15. Frequency of travel motor failures before and after installation of Mag-filters

ANOVA test was used to check if the mean of the two groups, i.e., (the travel motor average life before installation of Mag-filters and the travel motor average life after installation of Mag-filters) varied significantly or not. Normality of the data is checked by comparing mean and median values of the distribution, and Normal Probability Distribution (NPP plot). The details of the analysis are mentioned in Appendix A. With the above tests, the data is nearly normally distributed, and hence ANOVA test can be used to check the significance. The following table displays the ANOVA results for swing motor average life before and after installation of Mag-filters.

ANOVA – TRAVEL MOTORS

SOURCE OF VARIATION	*SS*	*Df*	*MS*	*F*	*P-value*	*F crit*
BETWEEN GROUPS	1.01E+08	1	1.01E+08	3.638638	0.070223	4.324794
WITHIN GROUPS	5.8E+08	21	27627915			
TOTAL	6.81E+08	22				

Table 16. ANOVA test for travel motor average life analysis

The null hypothesis of the test states: Means of travel motor life before and after the installation of Mags are equal, and both populations have a common variance. The alternate hypothesis states that the means of two populations vary significantly. For the swing motor component analysis, F < Fcrit and p >0.05. Hence, there is a 7% chance that the average of both groups does not vary significantly.

Figure 30 and Figure 31 represent histograms for TTF data of the travel motors. TTF data is grouped into different bins with equal widths of 5000 component life hours, and the number of failures is plotted on the y axis corresponding to each bin. Figure 30 indicates around 10% of the travel motors have failed in less than 5000 hours. Around 17% of the motors have failed between 15,000 and 20,000 hours, and 73% of components have achieved more than 20,000 hours life. From Figure 31, it can be inferred that before Mag- installation, 14% of components failed in less than 5000 hours, around 30% of failures are between 15,000 and 20,000 hours and 57% of components have achieved greater than 20,000 hours. After installing Mag-filters, all the components have achieved greater than 20,000 hours of component life. Component life has significantly after the installation of Mag-filters.

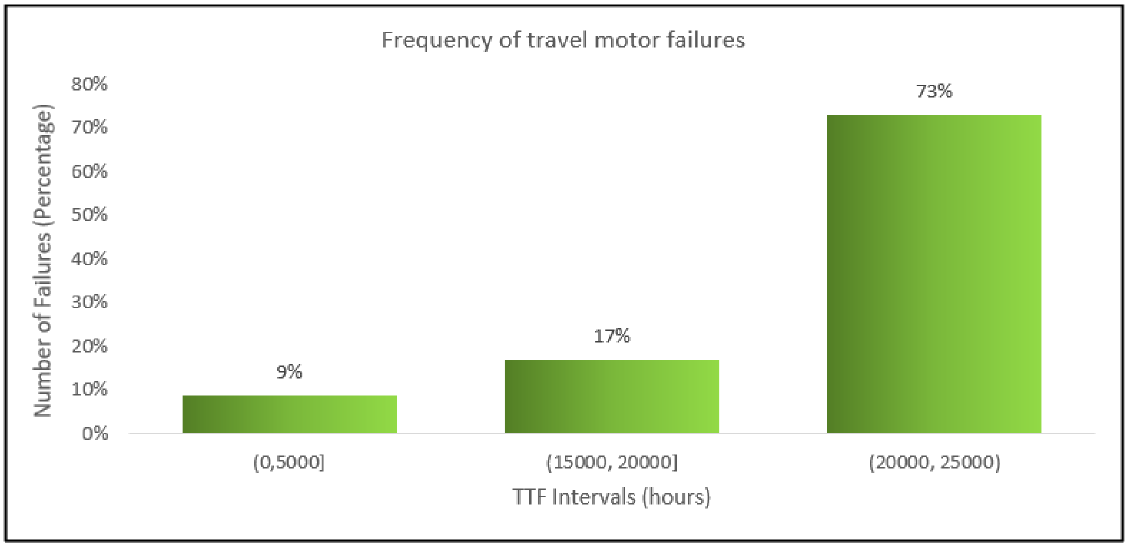


Figure 30. Failure frequency analysis of travel Motors

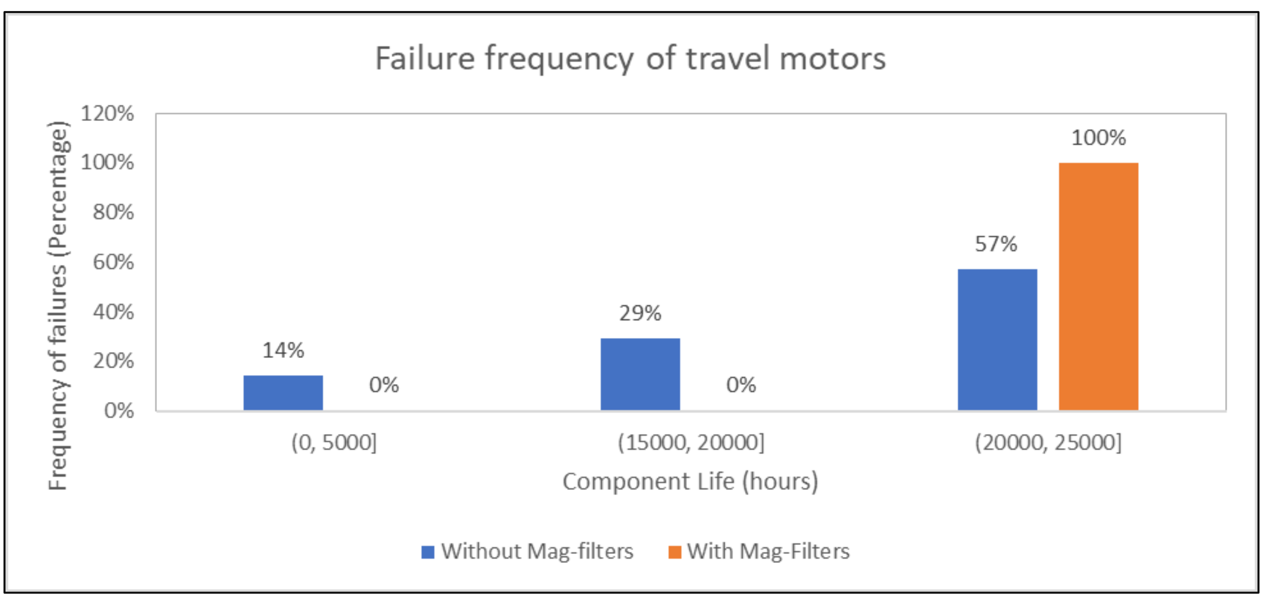


Figure 31. Failure Frequency of travel motors before and after Mag installation

3.9.4 Clam Cylinder Analysis

The average life of clam cylinders is 7172 hours, and the SD is 3308.12 hours. Data points above 2 standard scores, i.e., failures with TTF > 13,000 hours, were eliminated. The following plot represents the outliers detected in TTF data of clam cylinders. The X-axis represents the meter reading (SMU hours) of the shovel fleet, and the Y-axis represents the TTF hours of the clam cylinders. 7% of clam cylinder failures had errors in recordings with cylinder life greater than 13,000 hours. These data points were eliminated.

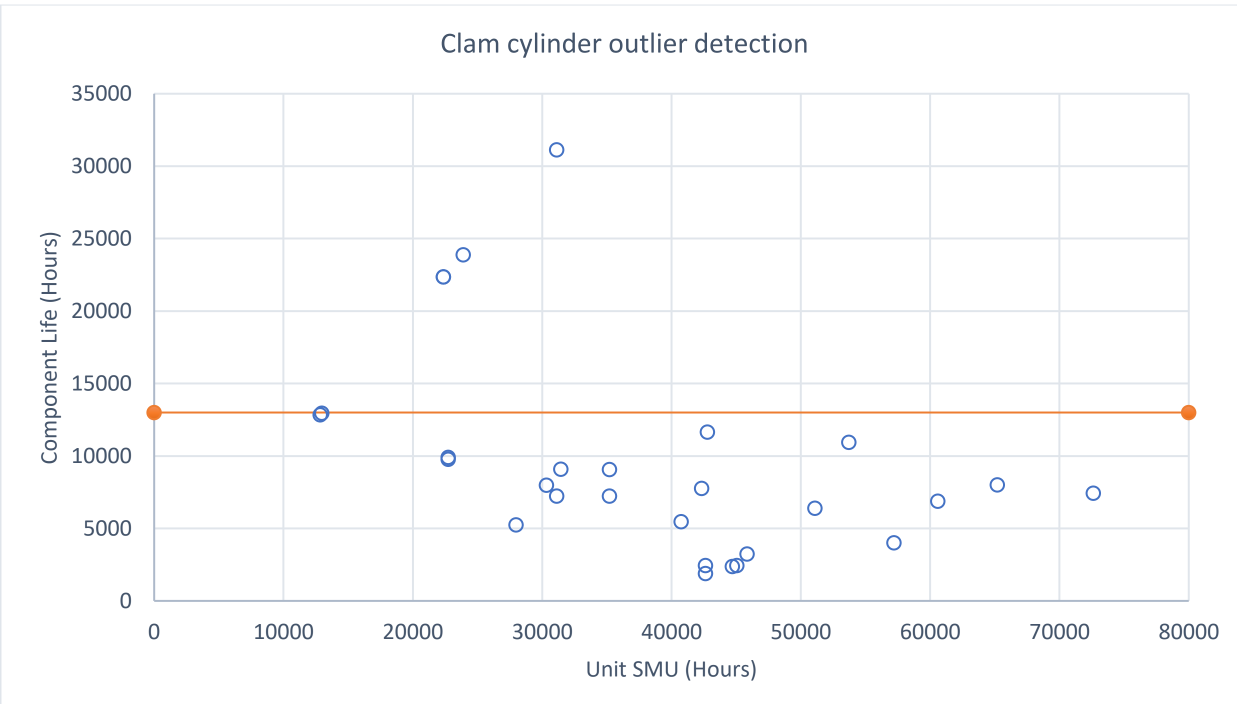


Figure 32. Outlier detection of clam cylinder failures

Clam cylinder average life was calculated after eliminating the outliers. The average life of cylinders before installation of Mag-filters is 8823 hours, and post-installation is 5659 hours. The average performance of clam cylinders is 7172 hours. 64% of the cylinders performed better than average life before the filter installation, and 42% of the cylinders performed better than average after installation. When compared to the average life of clam cylinders before Mag installation (i.e., 8824 hours), 17% of the magnetic filter installed cylinders have performed better than the average life before installation. On average, in the S1 unit, two clam cylinders were replaced every 10,000 hours before installation, and two cylinders were replaced every 10,000 hours post-Mag installations. On average, in S2, two cylinders were replaced every 10,000 hours before installation and three were replaced post-Mag installations. Table. 17 through Table. 19 present component life analysis of clam cylinders.

ANOVA test was used to check if the mean of the two groups, i.e., (the clam cylinder average life before installation of Mag and the average cylinder life after installation of Mag-filters) varied significantly. Normality of the data is checked by comparing mean and median values of the distribution, and Normal Probability Distribution (NPP plot). The details of the analysis are mentioned in Appendix A. With the above tests, the data is nearly normally distributed and hence ANOVA test was used to check the significance. Table. 20 displays the ANOVA results for clam cylinder average life before and after installation of Mag-filters.

The null hypothesis of the test states: Means of the population i.e., component life of clam cylinders before and after filter installation are equal, and both populations have a common variance. The alternate hypothesis states that the means of two populations vary significantly. For the clam cylinders, component analysis, $F > Fcrit$ and $p < 0.05$. Hence, the two means vary significantly. This means there is a significant difference in the component life before and after installation of Mags. The component life is significantly lesser than the average component life before the installation of filters.

MAG-FILTERS	AVERAGE LIFE	#CYLINDERS PERFORMING ABOVE AVERAGE LIFE
BEFORE INSTALLATION	8823	64%
AFTER INSTALLATION	5659	42%

Table 17. Average Life analysis of clam cylinder failures

X- Performance of Cylinders before Mag installation	*Avg Life – 8824*
% Mag installed Cylinders Performing above Average Life X	17%

Table 18. Performance of Mag installed clam cylinders compared to average life pre-installation

	#CYLINDERS REPLACED PER 10000 HOURS BEFORE MAGNETIC FILTER INSTALLATION	#CYLINDERS REPLACED PER 10000 HOURS AFTER MAGNETIC FILTER INSTALLATION
S1	2	2
S2	2	3

Table 19. Frequency of clam cylinder failures before and after installation of Mag-filter

ANOVA- CLAM CYLINDER

SOURCE OF VARIATION	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
BETWEEN GROUPS	57479153	1	57479153	6.212918	0.021114	4.324794
WITHIN GROUPS	1.94E+08	21	9251555			
TOTAL	2.52E+08	22				

Table 20. ANOVA test for clam cylinder average life analysis

Figure 33 and Figure 34 represent histograms for TTF data of the clam cylinders. TTF data is grouped into different bins with equal widths of 3000 component life hours, and the number of failures is plotted on the y axis corresponding to each bin. 17% of the total cylinders have failed in less than 3000 hours whereas only 9% failed before installation and 25% of the cylinders failed in less than 3000 hours after installation. 36% of the cylinders have achieved greater than 9000 hours and 18% of cylinders have achieved greater than 12,000 of component life before filter installation. Only 17% of the magnetic filter installed cylinders have greater than 9000 hours of component life, and no cylinders have achieved greater than 12,000 hours after installation of Mags.

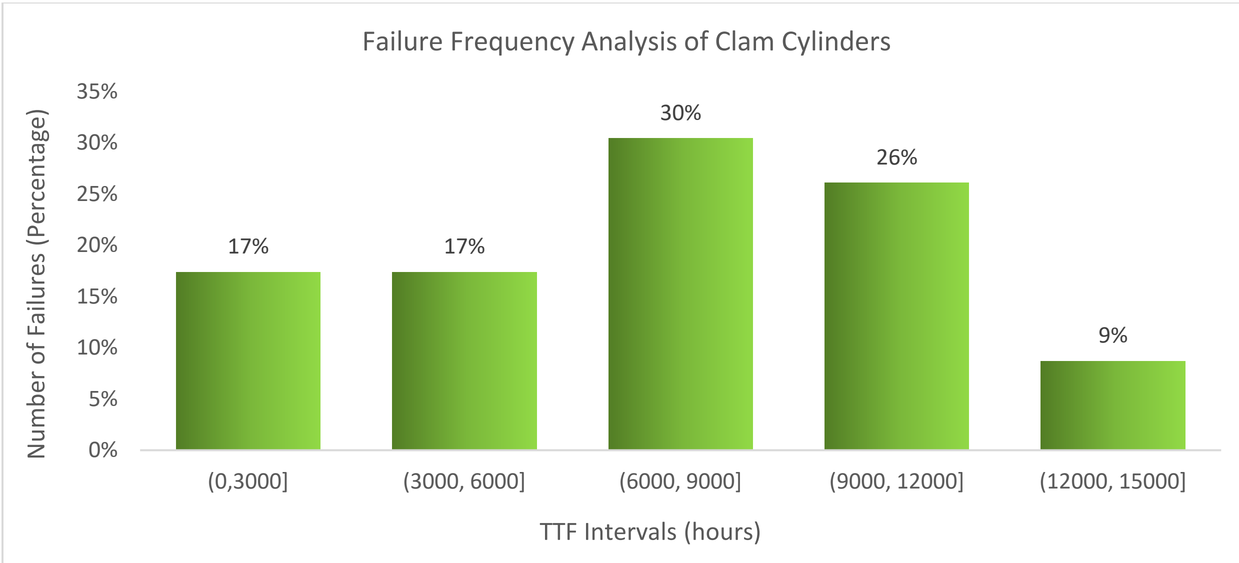


Figure 33. Frequency analysis of clam cylinder failures

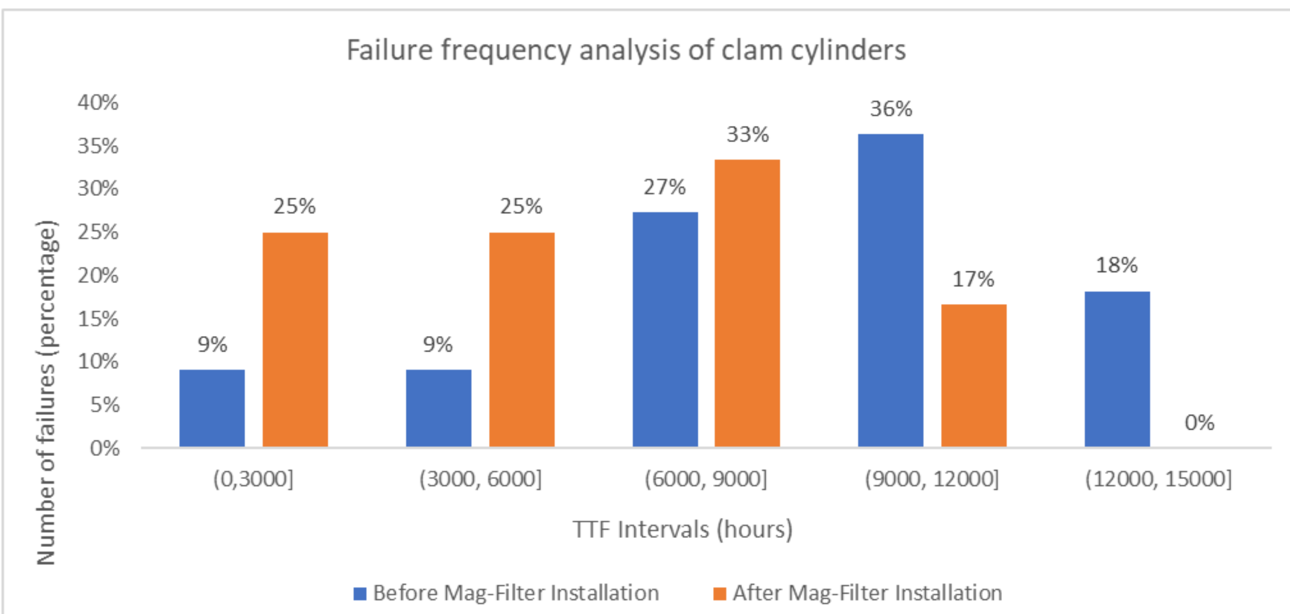


Figure 34. Failure frequency of clam cylinders before and after Mag installation

Figure 35 and Figure 36 represent the component life trend before and after the installation of Mags in both S1 and S2 units. The red line denotes Mag-filter installed SMU hour. The performance after the installation has somewhat increased and later decreased again. In S1, there is no trend in component life and failures are random. There is a random variation in cylinder performance before installation. The performance of the cylinders has increased after installing Mags. In S2, the first cylinder has achieved a good life of above 13,000 hours, and then the performance is constantly decreasing. The reasons might be the location of the equipment in the mine site or the previous failures and other components generating wear in the oil are influencing the average life.

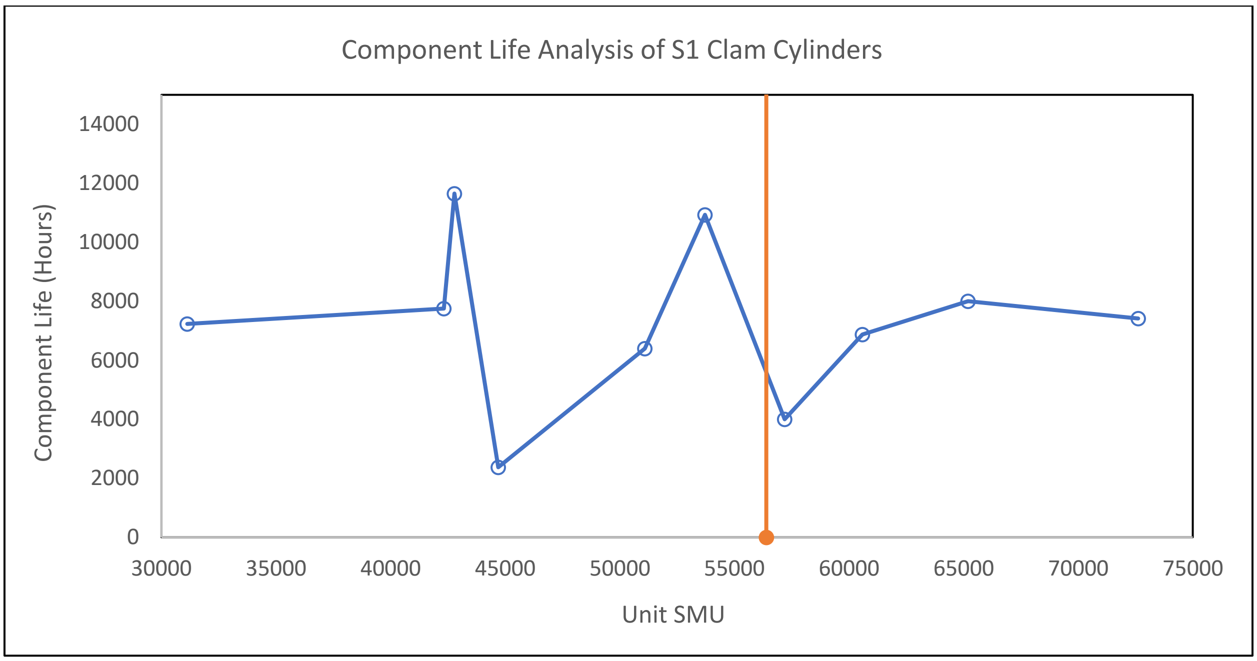


Figure 35. Component Life Analysis of S1 clam cylinders

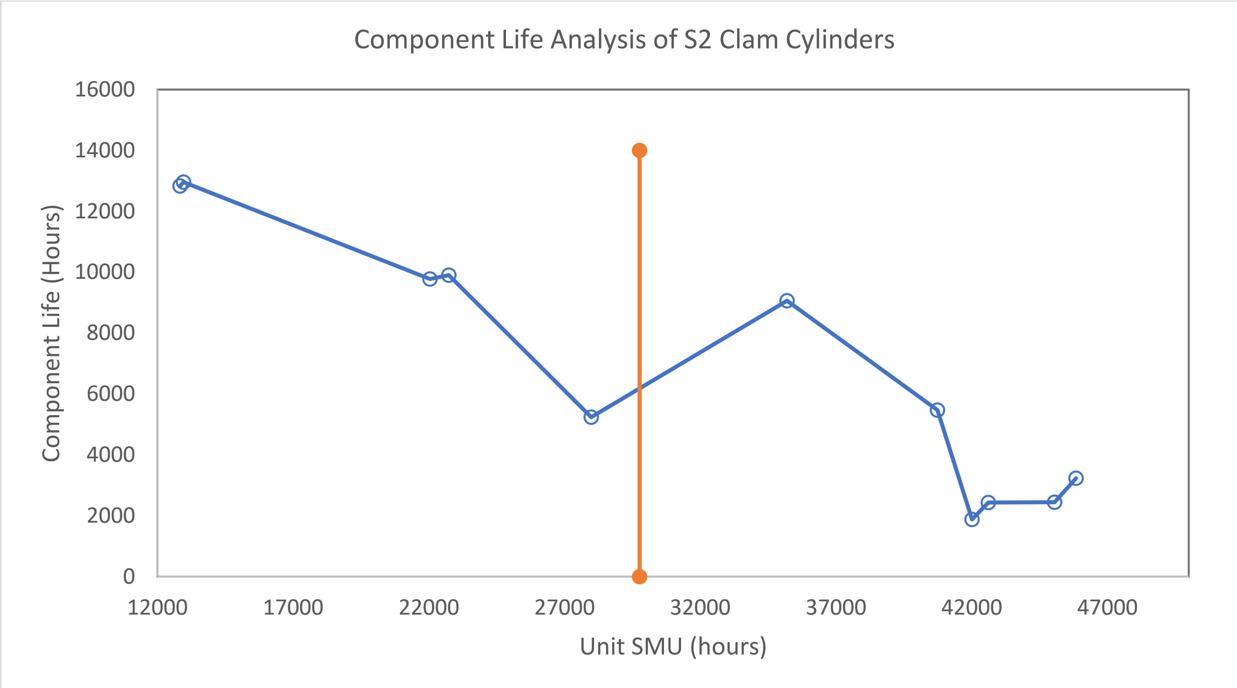


Figure 36. Component Life Analysis of S2 clam cylinders

3.9.5 Control Valves

The control valve is in the main circuit of the hydraulic system. A control valve in the shovel hydraulic system is tasked with the regulation of fluid speed, and by regulating the rate, it can control the speed of an actuator in the system. The control valve controls the oil flow in the actuators as directed by the DQR valves. The pump supplies hydraulic oil to the actuators through control valve. There are 6 control valves, three of which are located on the left side and three are located on the right side of the hydraulic system of the shovels considered in the analysis. Hydraulic oil delivered from a total 12 main pump units is routed to 6 control valve units via the high-pressure strainer. One unit of the control valve on each side has the same spool arrangement so that all hydraulic actuators such as cylinders and motors can be controlled by operating 3 control valve units located on one side.

The average life of control valves before outlier elimination is 27526 hours, and the SD is 4584.45 hours. Data points above 1.5 standard scores from the mean, i.e., failures with TTF > 36,694 hours were eliminated. The following plot, Figure 37 represents the outliers in TTF data of control valves. The X-axis represents the meter reading (SMU hours) of the shovel fleet, and the Y-axis represents the TTF hours of the control valve. These data points were eliminated. Around 20% of the data had a component life of above 35000 hours, and these data points were eliminated.

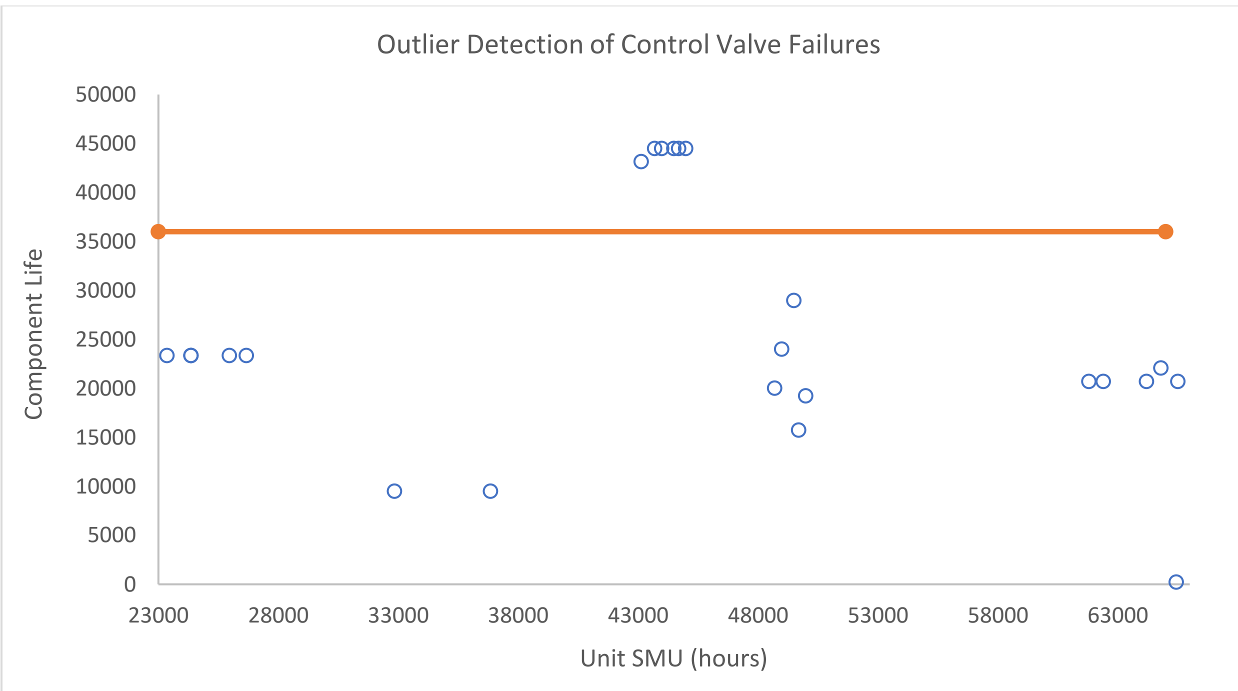


Figure 37. Outlier detection of control valve failures

Control valve average life was calculated after eliminating the outliers. The average life of valves before installation of Mag-filters is 23359 hours and post-installation is 18103 hours. The average performance of control valves is 20355 hours. 44% of the valves performed better than average before Mags installation and 46% of the valves performed better than average after installation. When compared to the average life of control valves before Mags's installation (i.e., 23359 hours), none of the Mag-installed control valves have performed better than the valve average life before installation. On average, in the S1 unit, one control valve was replaced every 10,000 hours before installation, and three valves were replaced every 10,000 hours post-Mag installations. On average, in S2, two valves were replaced every 10,000 hours before installation and one post-Mag installation. The component life analysis of control valves is represented in Table. 21 and Table. 22.

MAG-FILTERS	AVERAGE VALVE LIFE	#VALVES PERFORMING ABOVE AVERAGE LIFE
BEFORE INSTALLATION	23359	44%
AFTER INSTALLATION	18103	46%

Table 21. Average Life analysis of control valve failures

	#VALVES REPLACE PER 10000 HOURS BEFORE MAG-INSTALLATION	#VALVES REPLACE PER 10000 HOURS AFTER MAG- INSTALLATION
S1	1	3
S2	2	1

Table 22. Frequency of control valve failures before and after installation of Mag-filters

ANOVA test was used to check if the mean of the two groups, i.e., (the control valve average life before installation of Mags and the control valve average life after installation of Mags) varied significantly or not. Normality of the data is checked by comparing mean and median values of the distribution, and Normal Probability Distribution (NPP plot). The details of the analysis are mentioned in Appendix A. With the above tests, the data is nearly normally distributed, and hence ANOVA test can be used to check the significance. The following table displays the ANOVA results for control valve average life before and after installation of Mag-filters.

The null hypothesis of the test states: Means of the population of component life of control valves are equal, and all populations have a common variance. The alternate hypothesis states that the means of two populations vary significantly. From Table. 23, F< Fcrit and p >0.05. Hence, the two means do not vary significantly. This means there is no significant difference in the component life before and after installation of Mag-filters.

ANOVA - CONTROL VALVES

SOURCE OF VARIATION	*SS*	*Df*	*MS*	*F*	*P-value*	*F crit*
BETWEEN GROUPS	94711618	1	94711618	3.696116	0.34344	4.747225
WITHIN GROUPS	2E+08	12	16627403			
TOTAL	2.94E+08	13				

Table 23. ANOVA test for control valve average life analysis

Figure 38 represents histograms for TTF data of the main control valves. TTF data is grouped into different bins width of 5000 component life hours, and the number of failures is plotted on the y axis corresponding to each bin. All the failures before the installation of Mag-filters have achieved greater than 20000 hours of component life and post-installation of Mag-filters, 75% of valves has achieved component life of greater than 20000 hours. 25% of the failures have less than 10000 hours of component life.

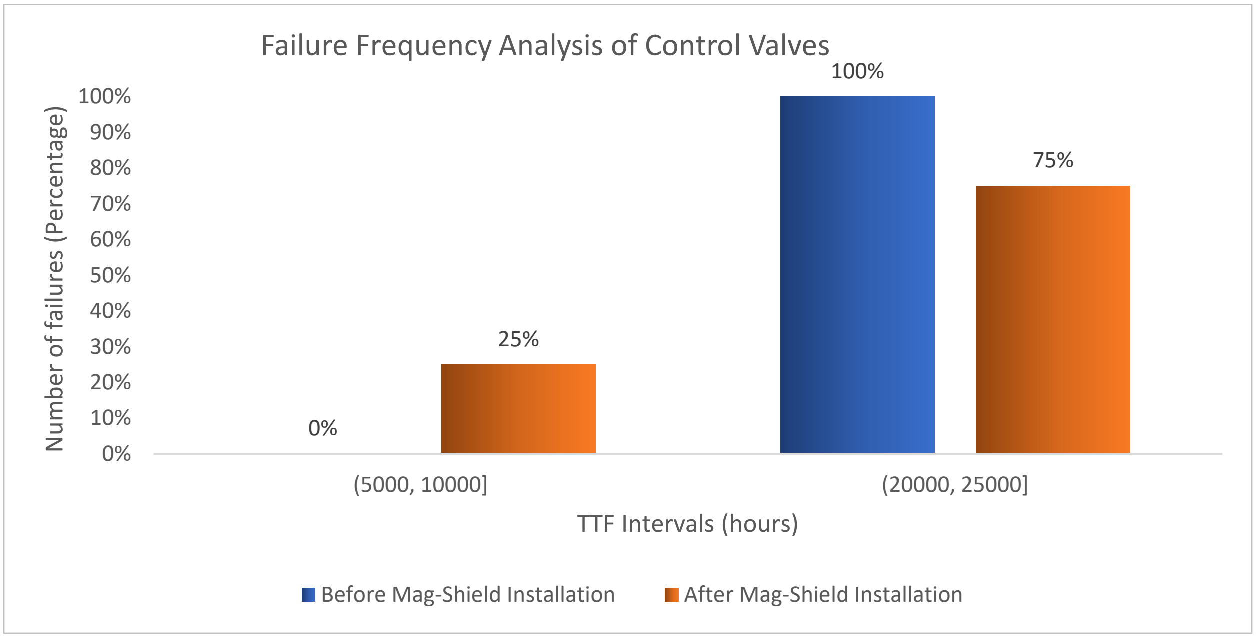


Figure 38. Failure frequency analysis of control valves

3.9.6 Bucket Cylinders

The bucket cylinder allows the movement of the bucket, or any other accessory mounted on the quick coupler. Bucket cylinders are in the delivery circuit of the hydraulic system. Bucket cylinders are supplied oil from all the control valves during tilting-in operations and from four of the six control valves during tilting-out operations.

The average life of bucket cylinders is 15,492 hours, and the SD is 6088.45 hours. Data points above 1.5 standard scores, i.e., failures with TTF > 25,000 hours, were identified as outliers. The following plot (Figure 39) represents the outliers in TTF data of control valves. The X-axis represents the meter reading (SMU hours) of the shovel fleet, and the Y-axis represents the TTF hours of the control valve. As indicated from the graph, no outliers are recorded in the data.

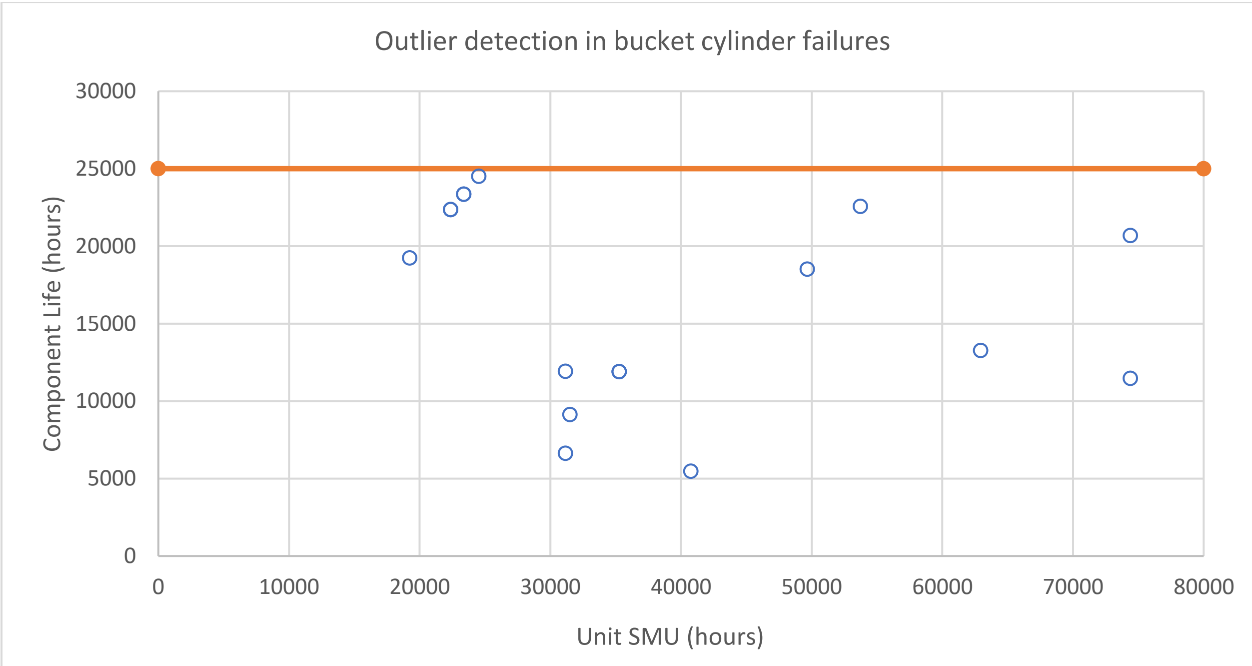


Figure 39. Outlier Detection of Bucket Cylinder

Bucket Cylinder average life was calculated before and after Mag installation. The average life of the cylinders before installation of Mags is 13347 hours, and post-installation is 12718 hours. The average performance of bucket cylinders is 13095 hours. 55% of the cylinders performed better than average before Mag installation, and 50% of the cylinders performed better than average after installation. When compared to the average life of bucket cylinders before Mag installation (i.e., 13347 hours), 50% of the Mag installed bucket cylinders have performed better than the cylinder average life before installation. On average, in the S1 unit, one bucket cylinder was replaced every 10,000 hours before installation, and one cylinder was replaced every 10,000 hours post-Mag-filter installations. On average, in S2, one cylinder was replaced every 10,000 hours before installation and one-cylinder post-Mag-filter installation. The component life analysis of bucket cylinders is represented in Table. 24 through Table. 26.

MAG-FILTERS	AVERAGE LIFE	#CYLINDERS PERFORMING ABOVE AVERAGE LIFE
BEFORE INSTALLATION	13347	55%
AFTER INSTALLATION	12718	50%

Table 24. Average Life analysis of bucket cylinder failures

X - Average Life of cylinders before the installation of Mags	*Avg Life - 13347*
#Cylinders Performing Above Average Life X	50%

Table 25. Performance of Mag installed bucket cylinders compared to average life pre-installation

	#CYLINDERS REPLACED PER 10000 HOURS BEFORE MAGNETIC FILTER INSTALLATION	#CYLINDERS REPLACED PER 10000 HOURS AFTER MAGNETIC FILTER INSTALLATION
S1	1	1
S2	1	1

Table 26. Frequency of bucket cylinders failures before and after installation of Mag-filters

ANOVA test was used to check if the mean of the two groups, i.e., (the bucket cylinder average life before installation of Mags and the bucket cylinder average life after installation of Mags) varied significantly or not. Normality of the data is checked by comparing mean and median values of the distribution, and Normal Probability Distribution (NPP plot). The details of the analysis are mentioned in Appendix A. With the above tests, the data represents a nearly normal distribution, and hence ANOVA test can be used to check the significance. The following table (Table. 27) displays the ANOVA results for bucket cylinder average life before and after installation of Mag-filters.

The null hypothesis of the test states: Means of the component life of bucket cylinder before and after installation of Mag-filters are equal, and all populations have a common variance. The alternate hypothesis states that the means of two populations vary significantly. For the bucket cylinders component analysis, F < Fcrit and p > 0.05. Hence, the two means do not vary significantly. This means there is no significant difference in the component life before and after installation of Mag-filters. Hence, the two means do not vary significantly.

ANOVA

SOURCE OF VARIATION	SS	Df	MS	F	P-value	F crit
BETWEEN GROUPS	44462904.92	1	44462905	1.118059	0.313005	4.844336
WITHIN GROUPS	437447534	11	39767958			
TOTAL	481910438.9	12				

Table 27. ANOVA test for bucket cylinder average life analysis

Figure 40 and Figure 41 represent histograms for TTF data of the bucket cylinders. TTF data is grouped into different bins width of 5000 component life hours and the number of failures is plotted on the y axis corresponding to each bin. Around 31% of the total cylinders have worked for greater than 20000 hours, of which 25% are Mag-filter installed cylinders. Around 16% of all cylinders failed have performed lesser than 10000 hours. 25% of Mag installed cylinders have less than 10000 hours of average life and around 11% of cylinder failures occurred in less than 10,000 hours before the installation of Mag-filters.

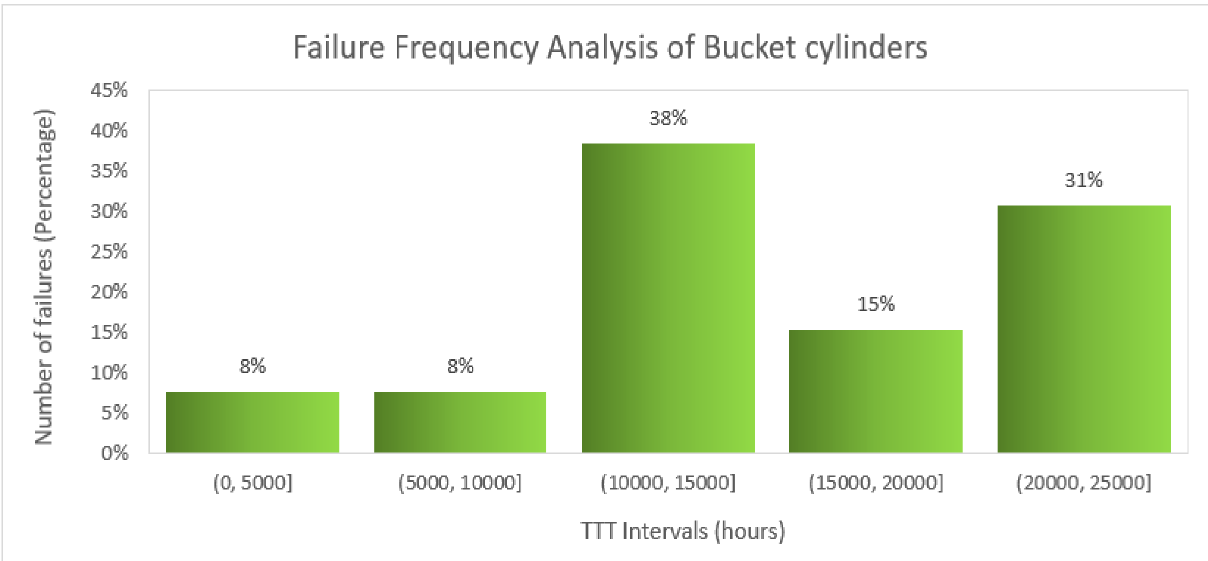


Figure 40. Failure frequency of bucket cylinder in different TTF Intervals

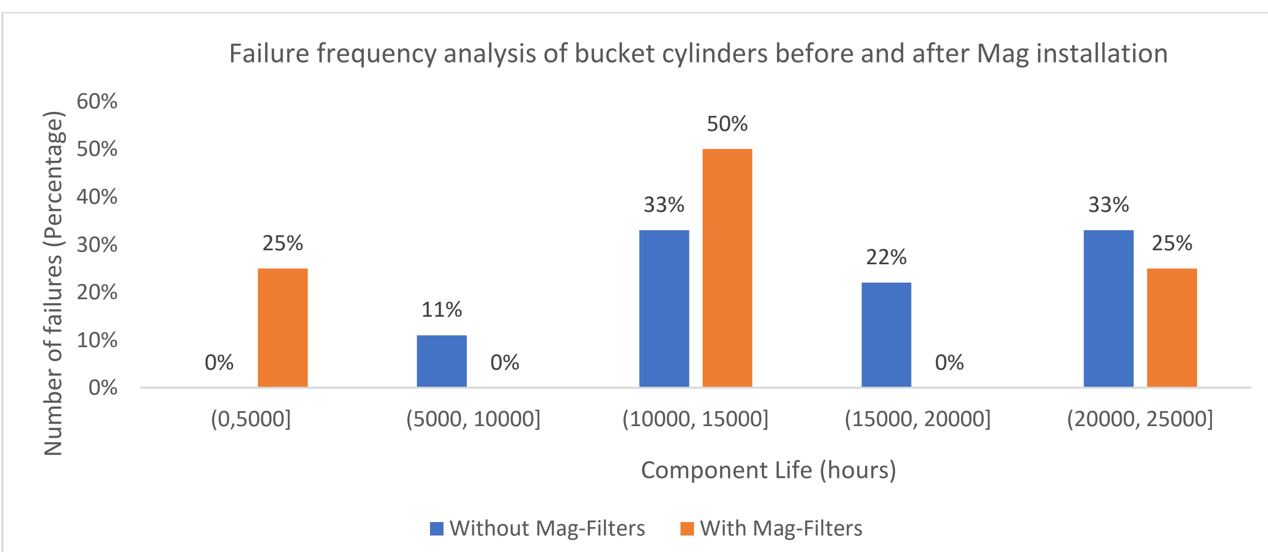


Figure 41. Failure Frequency of bucket cylinders before and after Mag installation

3.9.7 Propel Brake Valve

The propel flow control valve is used to reduce shock loads on the shovels. The shock loads are caused by the operation of shovel dig functions. The propel brake control valves are located in the brake release circuit of the hydraulic system. When a dig function is actuated while propelling the pilot pressure signal along with the control valve, oil flows through the orifice of the check valve assembly to gradually shift the propel flow control valve spool. The gradual shift of the spool slowly restricts the flow of supply oil to the control valve reducing any shock loads. The pilot pressure signal is released quickly through the check valve when all the dig functions are returned to neutral. The propel flow control valve spool is shifted quickly to the open position by the spring. There are four propel brakes in the hydraulic system of the considered shovels, two of which are located on the right side and two are located on the left side of the hydraulic system.

The average life of propel brake valves before outlier elimination is 19315 hours, and the SD is 8655 hours. Data points above 1.5 standard scores, i.e., failures with TTF > 33,000 hours, are identified as outliers. The following plot represents the outliers of TTF data of propel control valves. The X-axis represents the meter reading (SMU hours) of the shovel fleet, and the Y-axis represents the TTF hours of the control valve. As indicated from the graph (Figure 41), 4 failure points that are recorded in the data have component life of greater than 33,000 hours. These data points were eliminated.

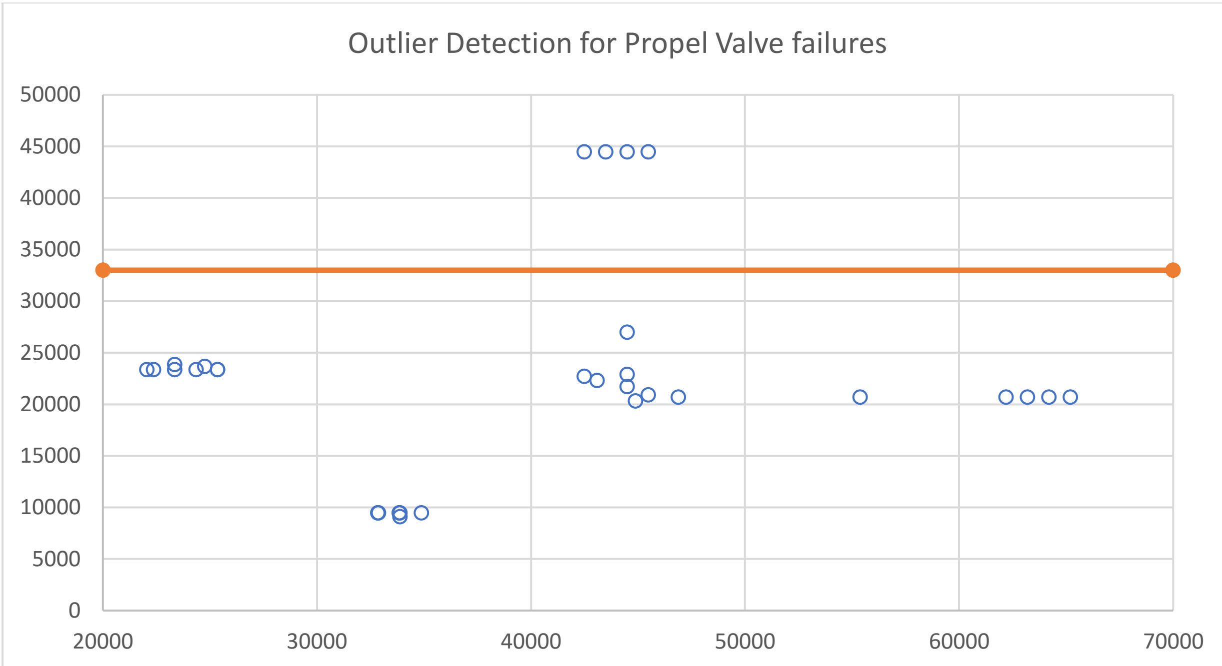


Figure 42. Outlier Detection of Propel Valve failures

Propel control valve average life was calculated after eliminating outliers. The average life of the valves before installation of Mag-filters is 15095 hours, and post-installation is 23359 hours. The average performance of propel valves is 17849 hours. 75% of the valves performed better than average before Mag installation, and 50% of the valves performed better than average after installation. When compared to the average life of propel valves before Mag installation (i.e., 13347 hours), 20% of the Mag installed propel valves have performed better than the average life before installation. On average, in the S1 unit, two valves were replaced every 10,000 hours before installation, and one valve was replaced every 10,000 hours post-Mag installations. On average, in S2, two valves were replaced every 10,000 hours before installation and one post-Mag installation. Table. 28 through Table. 30 present component life analysis of propel brake valves.

MAG-FILTERS	AVERAGE LIFE	#VALVES PERFORMING ABOVE AVERAGE LIFE
BEFORE INSTALLATION	23359	75%
AFTER INSTALLATION	15095	50%

Table 28. Average Life analysis of propel brake valve failures

X - AVERAGE LIFE OF PROPEL VALVES BEFORE THE INSTALLATION OF MAG-FILTERS	AVG LIFE - 23359
#VALVES PERFORMING ABOVE AVERAGE LIFE X	20%

Table 29. Performance of Mag installed propel valve compared to average life pre-installation

	#VALVES REPLACED PER 10000 HOURS BEFORE MAG INSTALLATION	#VALVES REPLACED PER 10000 HOURS AFTER MAG INSTALLATION
S1	2	1
S2	2	1

Table 30. Frequency of propel brake valves failures before and after installation of Mag-filters

ANOVA test was used to check if the mean of the two groups, i.e., (the propel valves' average life before installation of Mags and the average life after installation of Mags) varied significantly or not. Normality of the data is checked by comparing mean and median values of the distribution, and Normal Probability Distribution (NPP plot). The details of the analysis are mentioned in Appendix A. With the above tests, the data represents a nearly normal distribution, and hence ANOVA test can be used to check the significance. The following table (Table. 31) displays the ANOVA results for propel valve average life before and after installation of Mag-filters.

The null hypothesis of the test states: Means of the population are equal, and all populations have a common variance. The alternate hypothesis states that the means of two populations vary significantly. For the propel valves, component analysis, F >Fcrit and p< 0.02. Hence, the two means vary significantly. This means there is significant difference in the component life before and after installation of Mag-filters.

ANOVA- PROPEL VALVE

SOURCE OF VARIATION	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
BETWEEN GROUPS	1.82E+08	1	1.82E+08	7.217794	0.022833	4.964603
WITHIN GROUPS	2.52E+08	10	25231605			
TOTAL	4.34E+08	11				

Table 31. ANOVA test for propel valve average life analysis

3.9.8 Hydraulic Oil Cooler Fan Motor

The hydraulic oil cooler fan motor drives the pressure oil from the oil cooler fan motor pump to the hydraulic oil cooler. It is located in the oil cooler fan motor circuit. The oil cooler fan motor circuit prevents the oil from overheating which is a critical function in the hydraulic system. There are two hydraulic oil cooler fan motors one situated on the front and the other on the rear side of the hydraulic system.

The average life of the oil cooler fan motor before outlier elimination is 21307 hours, and the SD is 10453 hours. Data points above 1.5 standard scores, i.e., failures with TTF > 38,000 hours, are considered outliers. The following plot represents the outliers of TTF data of propel control valves. The X-axis represents the meter reading (SMU hours) of the shovel fleet, and the Y-axis represents the TTF hours of the control valve. As indicated from the graph (Figure 43), one failure point that is recorded in the data has a component life of greater than 38,000 hours. This data point was eliminated.

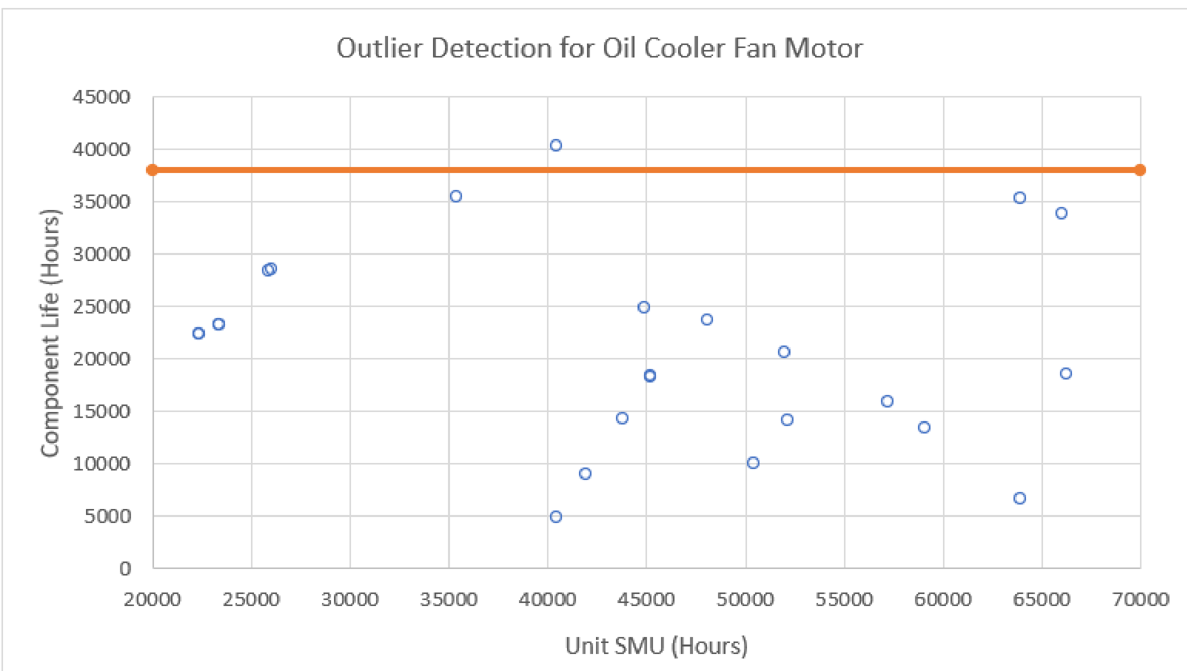


Figure 43. Outlier Detection of Oil Cooler Fan Motor failures

Hydraulic oil cooler fan motor average life was calculated after eliminating outliers. The average life of the motors before installation of Mags is 12348 hours, and post-installation is 14482 hours. The average performance of oil cooler fan motors is 13826 hours. 40% of the motors performed better than average before Mag installation, and 71% of the motors performed better than average after installation. When compared to the average life of motors before Mag- installation (i.e., 12348 hours), 71% of the Mag installed motors have performed better than the motor average life before installation. On average, in the S1 unit, two oil cooler fan motors were replaced every 10,000 hours before installation, and one motor was replaced every 10,000 hours post-Mag installations. On average, in S2, two motors were replaced every 10,000 hours before installation and one post-magnetic filter installation.

MAG-FILTERS	AVERAGE LIFE (HOURS)	#MOTORS PERFORMING ABOVE AVERAGE LIFE
BEFORE INSTALLATION	12348	40%
AFTER INSTALLATION	14882	71%

Table 32. Average Life analysis of oil cooler fan motor failures

X - AVERAGE LIFE OF OIL COOLER FAN MOTORS BEFORE THE INSTALLATION OF MAG-FILTERS	AVG LIFE - 19431
#FAN MOTORS PERFORMING ABOVE AVERAGE LIFE	71%

Table 33. Performance of Mag installed oil cooler fan motors compared to average life pre-installation

	#OIL COOLER FAN MOTOR REPLACED PER 10000 HOURS BEFORE INSTALLATION OF MAG-FILTERS	#OIL COOLER FAN MOTOR REPLACED PER 10000 HOURS AFTER INSTALLATION OF MAG-FILTERS
S1	2	1
S2	2	1

Table 34. Frequency of oil cooler fan motors failures before and after installation of Mag-filters

ANOVA test was used to check if the mean of the two groups, i.e., (the oil cooler fan motor average life before installation of Mags and the average life after installation of Mags) varied significantly or not. Normality of the data is checked by comparing mean and median values of the distribution, and Normal Probability Distribution (NPP plot). The details of the analysis are mentioned in Appendix A. With the above tests, the data represents a nearly normal distribution, and hence ANOVA test can be used to check the significance. The following table (Table. 35) displays the ANOVA results for oil cooler fan motor average life before and after installation of Mag-filters.

The null hypothesis of the test states: Means of the population are equal, and all populations have a common variance. The alternate hypothesis states that the means of two populations vary significantly. For the oil cooler fan motors, component analysis, $F < Fcrit$ and $p > 0.05$. Hence, the two means do not vary significantly. This means there is no significant difference in the component life before and after installation of Mag-filters. Hence, the two means do not vary significantly.

The null hypothesis of the test states: Means of the population are equal, and all populations have a common variance. The alternate hypothesis states that the means of two populations vary significantly. For the oil cooler fan motors, component analysis, $F > Fcrit$ and $p < 0.05$. Hence, the two means vary significantly. This means there is significant difference in the component life of oil cooler fan motors before and after installation of Mag-filters.

ANOVA -OIL COOLER FAN MOTORS

SOURCE OF VARIATION	SS	df	MS	F	P-value	F crit
BETWEEN GROUPS	703807.7357	1	703807.7	6.864	0.00935608	4.964603
WITHIN GROUPS	1025409913	10	1.03E+08			
TOTAL	1026113720	11				

Table 35. ANOVA test for oil cooler fan motors average life analysis

3.10 Reliability Analysis of Hydraulic Components before and after Installation of Mag-filters

3.10.1 Background Information

Reliability refers to the probability of the system meeting its desired performance standards in yielding output for a specific time duration when used under specific conditions (Dhillon, 2008). For instance, if a machine is designed to run continuously for 10,000 hours with no faults in between, the machine is said to be 100 % reliable for that period. However, if a failure occurs after 10,000 hours of operation, the machine's reliability after 10,000 hours is less than 100 % (Carlo, 2013). The component's reliability is a function of time and is always measured at a specific operating time. Reliable operation is interrupted or terminated by failures. A failure is an event that results in the inability to complete the required duties and meet the requirements. The theoretical definition of reliability is (Reliability = 1 – Probability of Failure), given by R(t). The probability of failure at a given time is the probability that a unit will be failed at a particular point in time.

Reliability assessment gives more details about the component and its operating conditions. The reliability analysis of components and equipment can reveal whether there is a link between new program implementation and cost savings. Keeping track of reliability can also help justify future investment in continued, improved, or new programs. In this study, understanding hydraulic component reliability can fairly give an idea on the chances that the hydraulic components are still working after specific time intervals and if the performance has improved after the installation of Mag-filters.

The objective of this study is to estimate the reliability of components at 5000 hours, 10000 hours, and 15000 hours of component operations for each hydraulic component before and after the installations of Mag-filters. This gives an idea about the changes in wear rate and particle contamination failures in the system after the installation of Mag-filters. Reliability analysis explains component behavior over time and if there is a difference in behavior before and after installation of Mag-filters. Different probability distributions are used in the analysis of reliability estimations. Relyence software is used to fit the given data in different distributions and calculate the best fit distribution that can describe the data. The different probability distributions that are used to fit the failure samples are Weibull 2-Parameter, Weibull 3-Parameter, Lognormal distribution, Normal distribution, Exponential 1-Parameter, Exponential – 2 Parameter, Gumbel-, Gumbel+, Rayleigh -1 Parameter, Rayleigh 2-Parameter. Prior to fitting distributions to the data, tests to validate the assumption of independent and identically distribution of data were performed. The software is used to plot reliability graphs, failure rate plots, PDF plots, and probability plots. The parameters of the best fit distribution are estimated using MLE or Rank regression method.

The reliability of each of the top 8 components is calculated and the system reliability is compared before and after Mag-filter installation.

3.10.2 Serial Correlation and Trend Tests

Trend analysis is used to check serial correlations and if the data is independently and identically distributed. Identically distributed data relates that one probability distribution can describe the characteristics of the entire dataset samples (Frost, 2020). For the reliability analysis of the systems, trend tests were used to check if there was any underlying trend in the failure samples. Pearson's correlation test was used to determine trend and correlation between i^{th} and $(i-1)^{th}$ failure. The failure data shows no trend or correlation for any of the top 8 components selected. Hence, probability distributions were used to estimate reliability.

3.10.3 Probability Distributions and Parameter Estimation for Component failure data

Probability distributions are the most widely used reliability data analysis tool. The basic idea of the probability distribution plots is to plot a nonparametric estimate of fraction failing as a function of time on a distribution-specific scale. Different probability distributions like the Weibull 2-Parameter, Weibull 3-Parameter, Lognormal distribution, Normal distribution, Exponential 1-Parameter, Exponential – 2 Parameter, Gumbel-, Gumbel+, Rayleigh -1 Parameter, Rayleigh 2-Parameter, and the others are used in plotting sample data. The idea is to assess whether the nonparametric estimate data is approximately linear to any of the distributions defined and if that distribution could adequately describe the data. The distributions can also describe different failure modes of the components like early failures, late failures, etc (Meeker, Hahn, & Escobar, 2017). The probability distributions help in estimating reliability at different given times of component life. This can help in understanding how reliability and failure rates have changed after the installation of Mag-filters.

Goodness-of-fit using residuals is used in finding the probability distribution that can best explain the data. The residuals from a fitted distribution are the differences between the response data and the fit to the response data at each value. It can be defined as the total error in describing the data by a probability distribution plot. Residuals and residual plots were used in the software to find the best fit distribution. The lesser the value of residuals, the more adequately the distribution describes the failure dataset.

MLE and Rank Regression methods were used for parameter estimation of the probability distributions. These parameters define the distribution. There are four parameters used in distribution fitting: location, scale, shape, and threshold. Not all parameters exist for each distribution. Distribution fitting involves estimating the parameters that define the various distributions. Maximum Likelihood estimate involves defining a likelihood function for calculating the conditional probability of observing the data sample given probability distribution. It involves maximizing the likelihood function in order to find the probability distribution that can best explain the observed data. This approach can be used to search a space of possible parameters for the distribution (Brownlee, 2019). The rank regression technique involves substituting data with their corresponding ranks. Following that, the data can be plotted such that it linearizes the failure data for a certain distribution. The failure times are represented by the x-axis coordinates, while the unreliability/failure rate estimations are represented by the y-axis coordinates. The process involves estimating unreliability using median rankings.

3.10.4 Reliability Comparison using different metrics

Failure rate and reliability vs time graphs are also used in the analysis to get more information on component behavior before and after the installation of Mag-filters. Failure rate vs time graphs gives information on the variation of the number of failures per unit of time at any given time in the hydraulic system. The failure rate can be defined as R/T where R is the number of failures and T is the total time. Reliability vs time graphs provides information on the probability that the components are successfully running at any given time during the component life. Component reliability can be calculated using failure rates in probability functions.

3.10.5 Results of Reliability Estimations

Correlation tests were used to check if the failure data of different components were independent and identically disturbed. Table. 36 shows the correlation analysis for different components of the hydraulic system. The table indicates that the failure data for each component follow i.i.d and mostly no correlation is seen between the failures.

CORRELATION BETWEEN ITH AND (I-1) TH FAILURE

	S1	S2
MAIN PUMPS	0.51	0.38
SWING MOTORS	0.32	0.13
TRAVEL MOTORS	0.72	0.31
CLAM CYLINDER	-0.39	0.68
BUCKET CYLINDER	-0.54	-0.03
HYDRAULIC OIL COOLER MOTOR	-0.03	0.49
HYDRAULIC OIL COOLER	0.34	0.52
MAIN CONTROL VALVE	-0.03	0.54
PROPEL BRAKE VALVE	0.39	0.65

Table 36. Correlation analysis for i.i.d assumption

The best fit distributions were plotted for failure data of components before and after installation of Mag-filters. Failure data of components including main pump, bucket cylinder before and after the installation of Mag-filters and travel motor, swing motor, clam cylinder, components post-installation of Mag-filters, and hydraulic oil cooler before installation of Mag-filters follow Weibull 3 – parameter distributions. The parameters of the distribution for different component failure data that follow 3-parameter Weibull distribution are represented in Table. 37. Travel motor failures, clam cylinders, control valves and oil cooler fan motors failure data follow Gumbel distributions. The parameters of the distribution for different component failures that follow Gumbel- distribution is represented in Table. 38 and Table. 39. Travel motor failure data after installation of filters follow Weibull 2- parameter distribution represented in Table. 40 and swing motor failure data before installation of Mag-filters follow log-normal distribution represented in Table. 41.

WEIBULL DISTRIBUTION- 3 PARAMETERS

	Β	Η	Ɣ
MAIN PUMPS - WITHOUT MAG-FILTERS	1.678	11131.88	4823.96
MAIN PUMPS - WITH MAG-FILTERS	1.867	12153.22	-838.25
SWING MOTORS WITH MAG-FILTERS	6.311	32423.5	-19275
BUCKET CYLINDER WITHOUT MAG-FILTERS	1.952	16673.02	2038.83
BUCKET CYLINDER WITH MAG-FILTERS	2.512	19091.84	-3983.68
CLAM CYLINDERS WITH MAG-FILTERS	1.106	4819.76	1481.24
PUMP DRIVE WITHOUT MAG-FILTERS	0.556	8361.69	18429.94

Table 37. Parameter Estimates of TTF data 3-Parameter Weibull Distribution

GUMBEL – DISTRIBUTION

	Δ	Ξ
TRAVEL MOTORS WITHOUT MAG-FILTERS	7.93.06	19616.75
CLAM CYLINDER WITHOUT MAG-FILTERS	2995.28	10400.75
OIL COOLER MOTORS WITH MAG-FILTERS	8059.31	18931.9
MAIN CONTROL VALVES WITHOUT MAG-FILTERS	7779.49	20129.37

Table 38. Parameter Estimates of TTF data Gumbel- Distribution

GUMBEL+ DISTRIBUTION

	Δ	Ξ
OIL COOLER MOTOR WITHOUT MAG-FILTERS	8692.39	6379.51

Table 39. Parameter Estimates of TTF data Gumbel+ Distribution

WEIBULL-2 PARAMETER

	β	η
TRAVEL MOTORS WITH MAG-FILTERS	20.528	22759.74

Table 40. Parameter Estimates of TTF data Weibull 2-parameter Distribution

LOGNORMAL DISTRIBUTION

	μ	σ
SWING MOTORS WITHOUT MAG-FILTERS	12540.22	7618.07

Table 41. Parameter Estimates of TTF data Lognormal Distribution

Reliability of the hydraulic components was calculated using the above mentioned best fit distribution and their estimated parameters at different times including 5000 hours, 10000 hours, and 15000 hours before and after the installation of Mag-filters. Table. 42 through Table. 45 show reliability estimated at different times for the hydraulic components considered in this analysis.

AT 5000 OPERATION HOURS

	Reliability before Magnetic filter Installation	Reliability after Nag-Shield Installation
MAIN PUMPS	0.99	0.775
SWING MOTORS	0.84	0.85
TRAVEL MOTORS	0.87	1
BUCKET CYLINDER	0.96	0.86
CLAM CYLINDER	0.84	0.49
OIL COOLER MOTOR	0.69	0.84
MAIN CONTROL VALVE	1	0.87

Table 42. Reliability of Components at 5000 hours of operation life before and after Magnetic filter Installation

AT 10000 HOURS OF OPERATION LIFE

	Reliability before Magnetic filter Installation	Reliability after Nag-Shield Installation
MAIN PUMPS	0.76	0.45
SWING MOTORS	0.63	0.59
TRAVEL MOTORS	0.769	0.97
BUCKET CYLINDER	0.79	0.64
CLAM CYLINDER	0.41	0.16
OIL COOLER MOTOR	0.48	0.71
MAIN CONTROL VALVE	0.78	0.76

Table 43. Reliability of Components at 10000 hours of operation life before and after Magnetic filter Installation

AT 15000 HOURS OF OPERATING LIFE

	Reliability before Magnetic filter Installation	Reliability after Nag-Shield Installation
MAIN PUMPS	0.42	0.21
SWING MOTORS	0.37	0.24
TRAVEL MOTORS	0.59	0.95
BUCKET CYLINDER	0.54	0.37
CLAM CYLINDER	0.18	0.09
OIL COOLER MOTOR	0.27	0.53
MAIN CONTROL VALVE	0.56	0.6

Table 44. Reliability of Components at 15000 hours of operation life before and after Magnetic filter Installation

In order to understand in detail regarding the component behaviour, failure-rate vs. time and reliability vs. time graphs were plotted for the top 4 highest failing components of the hydraulic system to understand the overall change in behaviour of the component. Figure 42 through Figure 48 represent the reliability over time graphs for failure data before and after installation of Mag-filters. The x-axis represents time in hours and y-axis represents reliability of the component at the given time. The dotted line in the 3-parameter Weibull graph represents 2-parameter Weibull distribution of the function. Figure 49 through Figure 54 represent the failure-rate vs. time graphs for failure data before and after installation of Mag-filters. The x-axis represents time in hours and y-axis represents failure-rate of the component at the given time. The graphs indicate early failures, late failures and variation in reliability and failure rate before and after installation of Mag-filters.

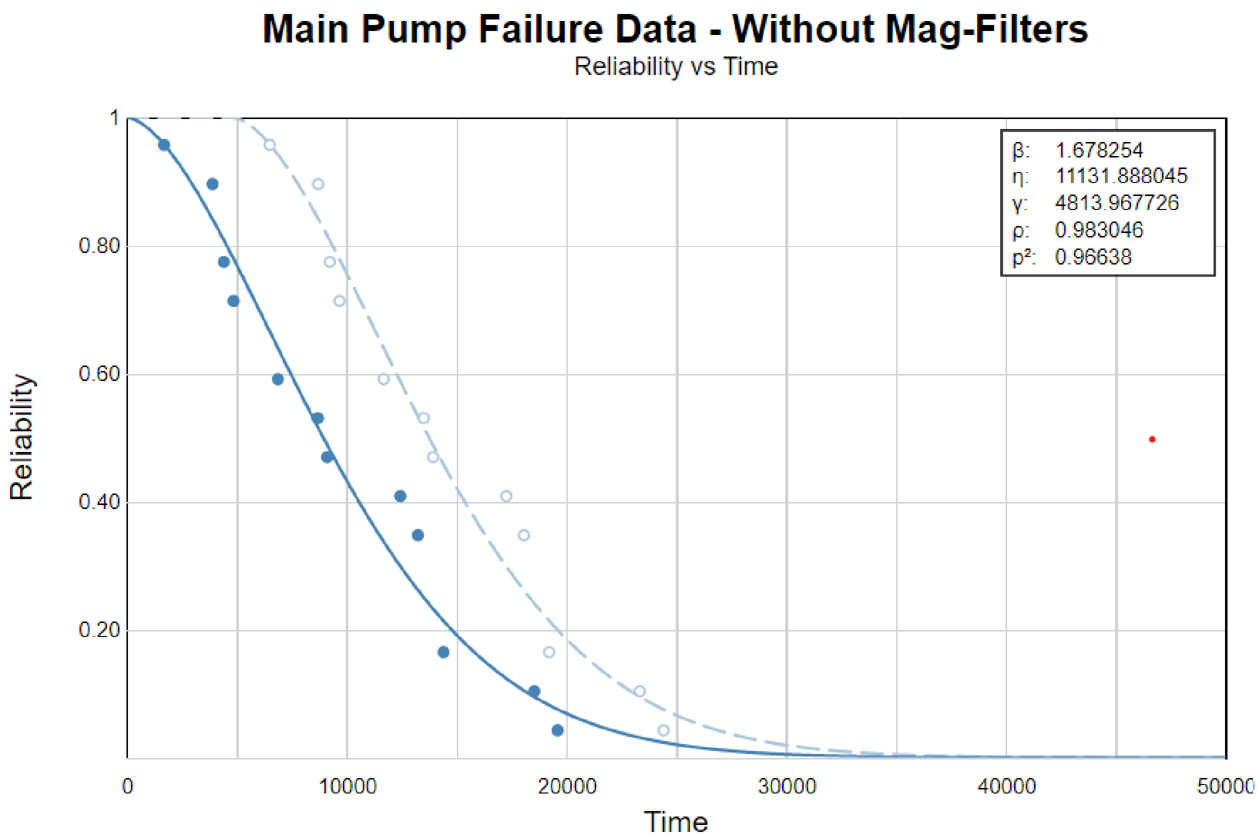


Figure 44. Reliability Vs time (hours) plot for Main Pumps before installation of Mag-filters

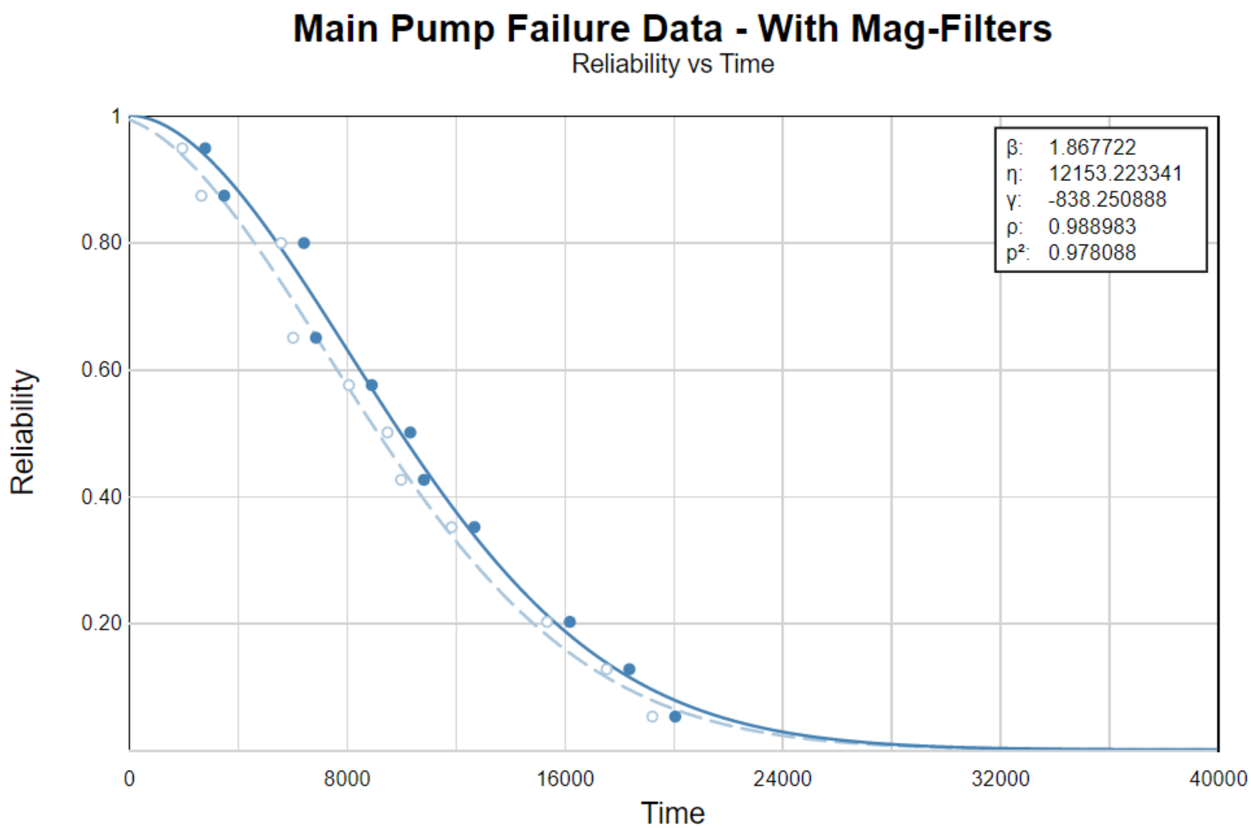


Figure 45. Reliability Vs time (hours) plot for Main Pumps before installation of Mag-filters

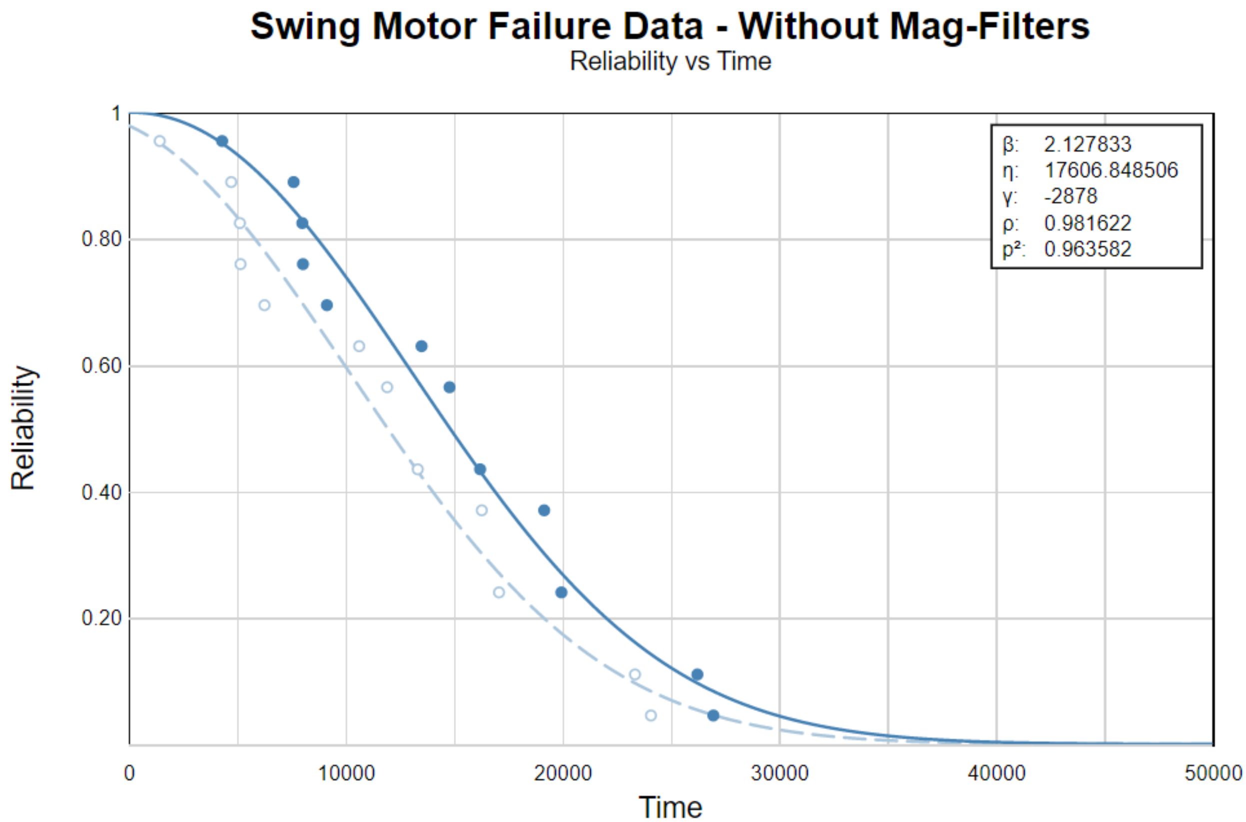


Figure 46. Reliability Vs time (hours) plot for Swing Motors before installation of Mag-filters

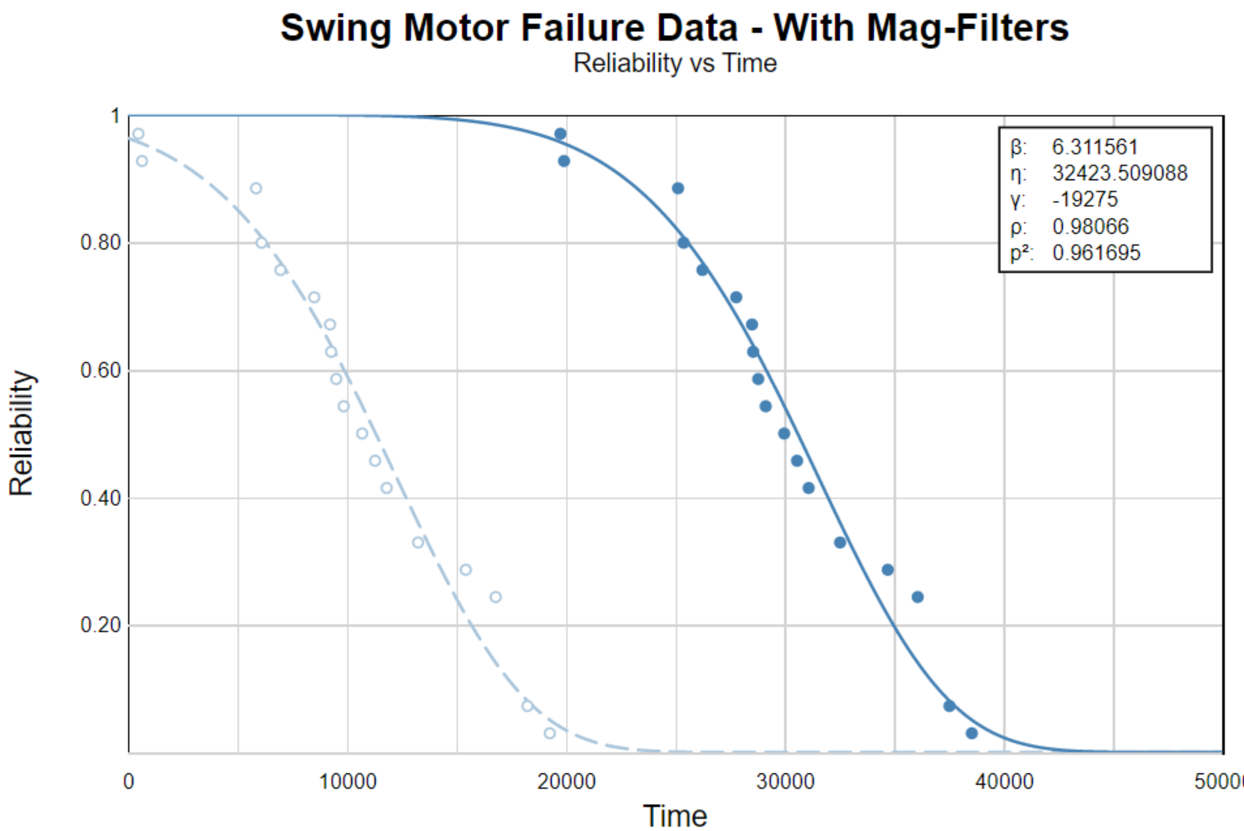


Figure 47. Reliability Vs time (hours) plot for Swing Motors after installation of Mag-filters

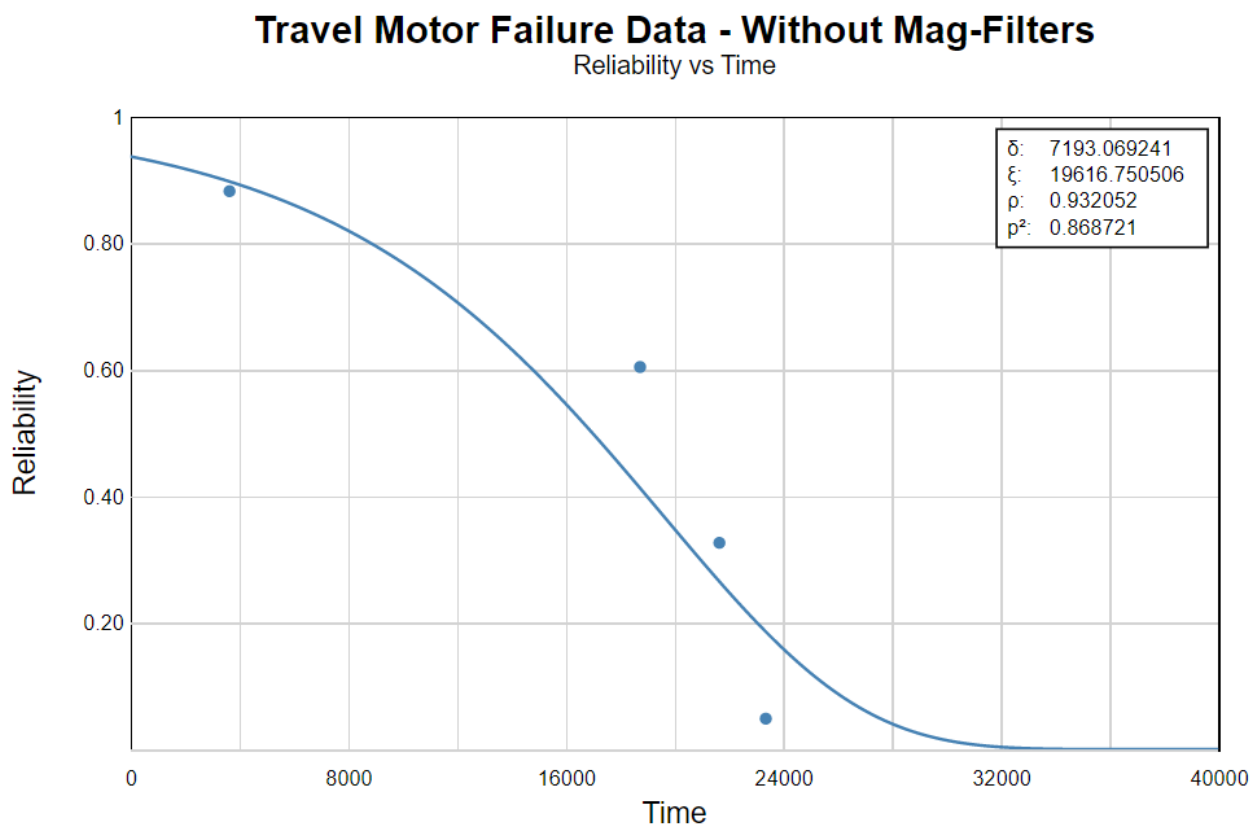


Figure 48. Reliability Vs time (hours) plot for Travel Motors before installation of Mag-filters

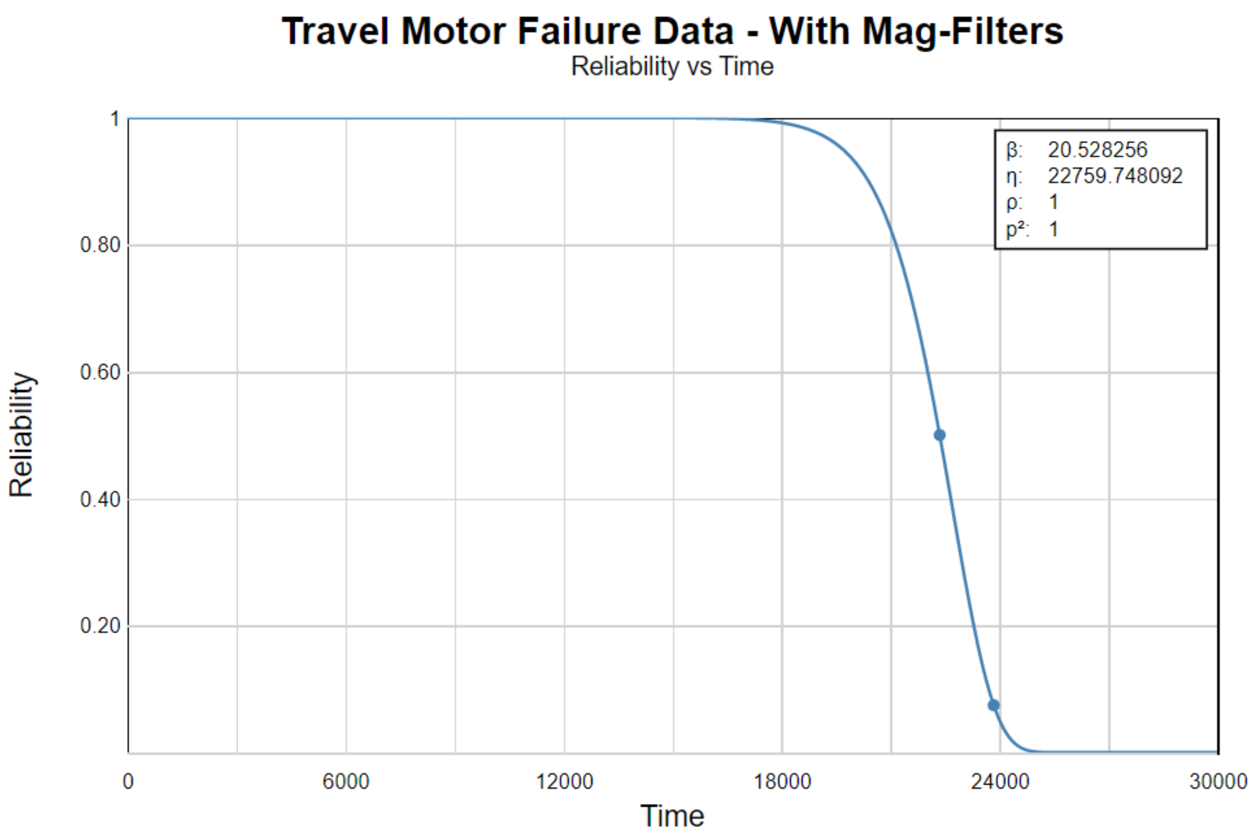


Figure 49. Reliability Vs time (hours) plot for Travel Motors after installation of Mag-filters

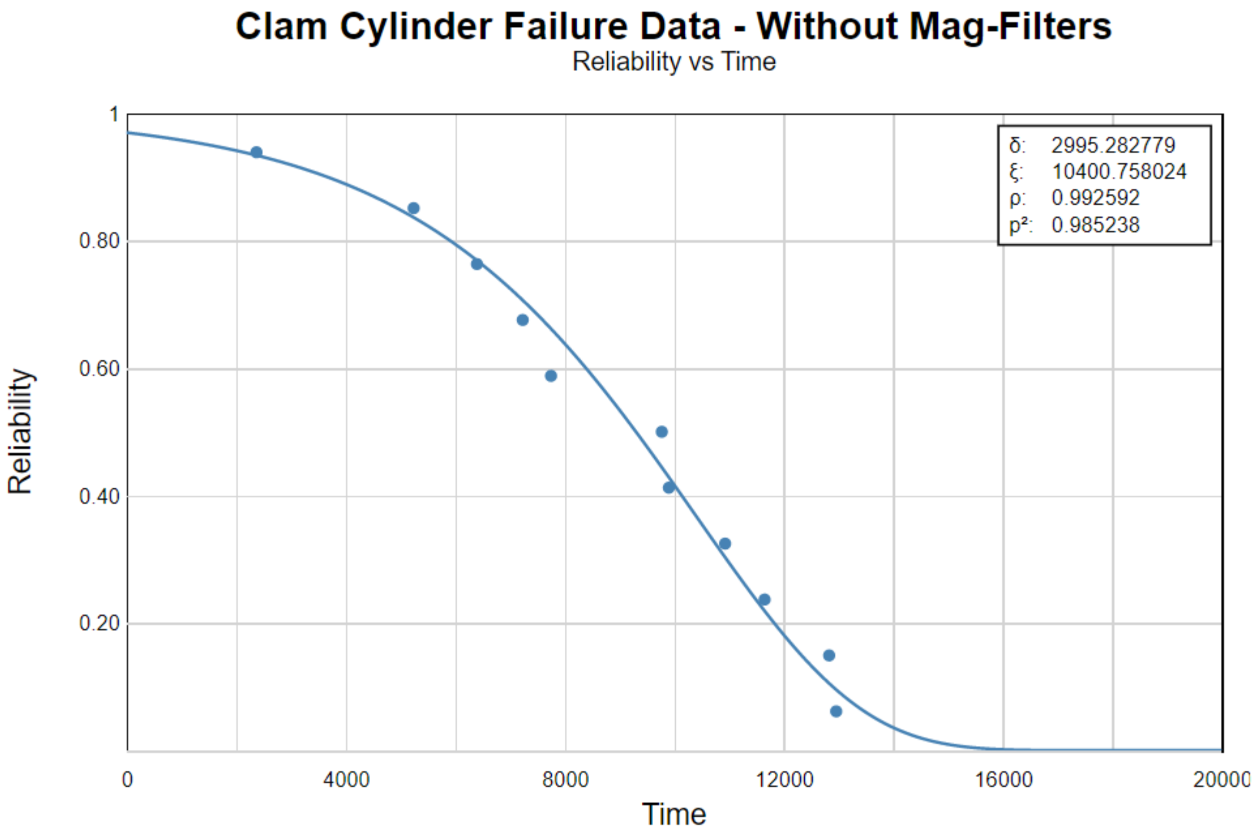


Figure 50. Reliability Vs time (hours) plot for Clam Cylinders before installation of Mag-filters

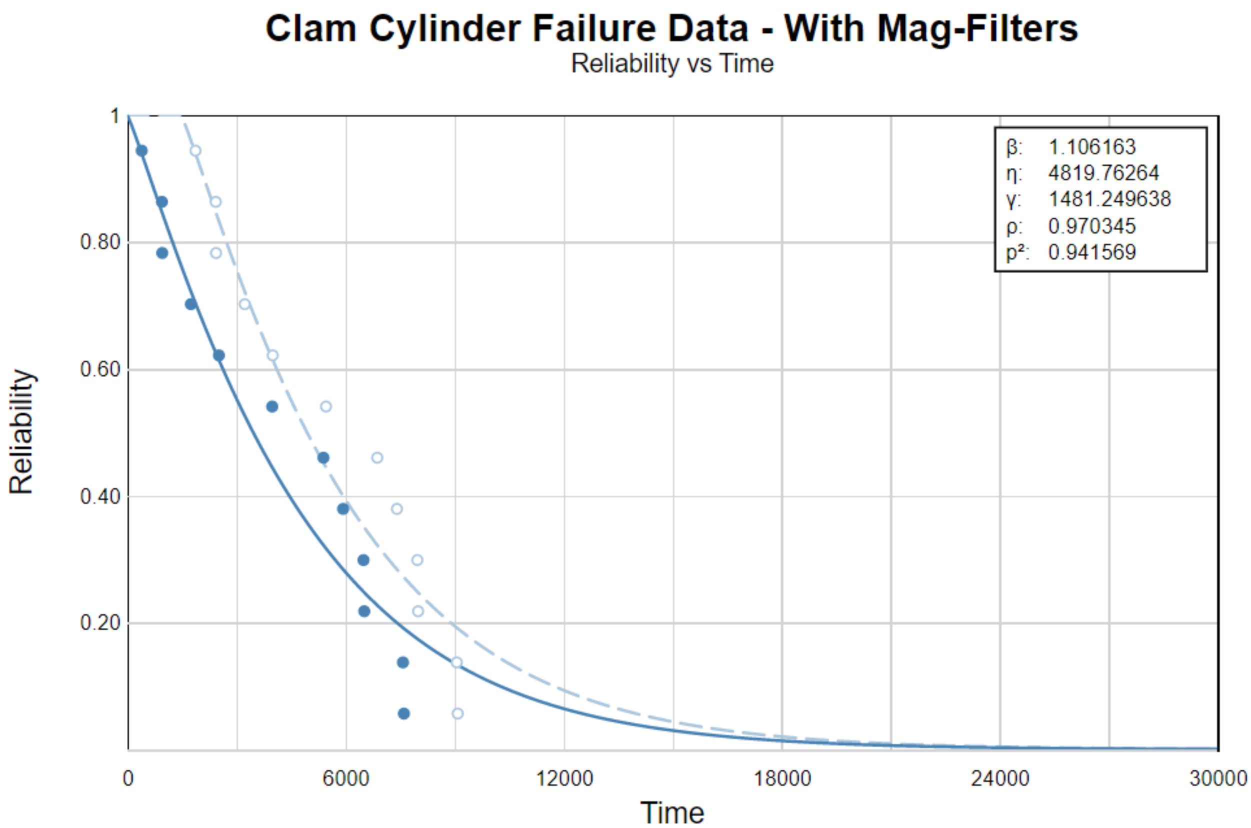


Figure 51. Reliability Vs time (hours) plot for Clam Cylinders after installation of Mag-filters

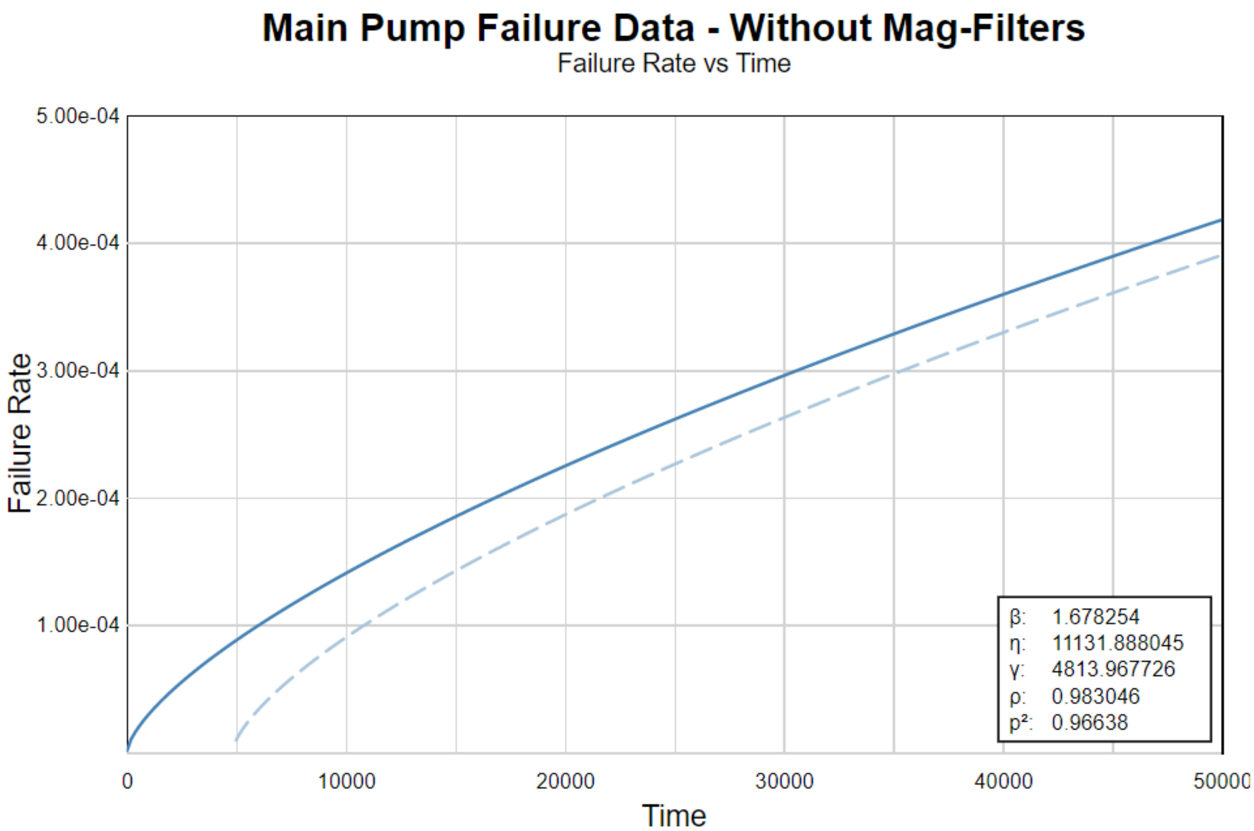


Figure 52.Failure-Rate Vs time (hours) plot for Main Pump before installation of Mag-filters

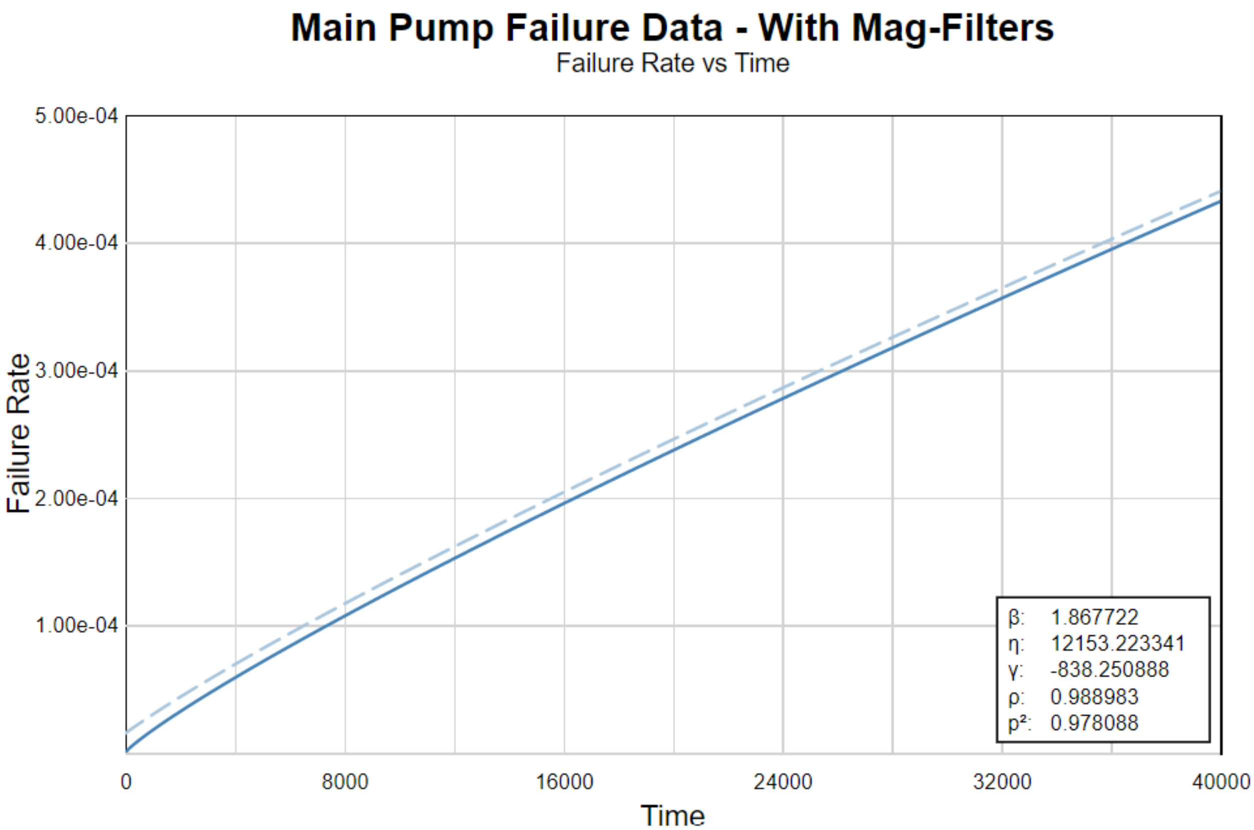


Figure 53. Failure-Rate Vs time (hours) plot for Main Pump after installation of Mag-filters

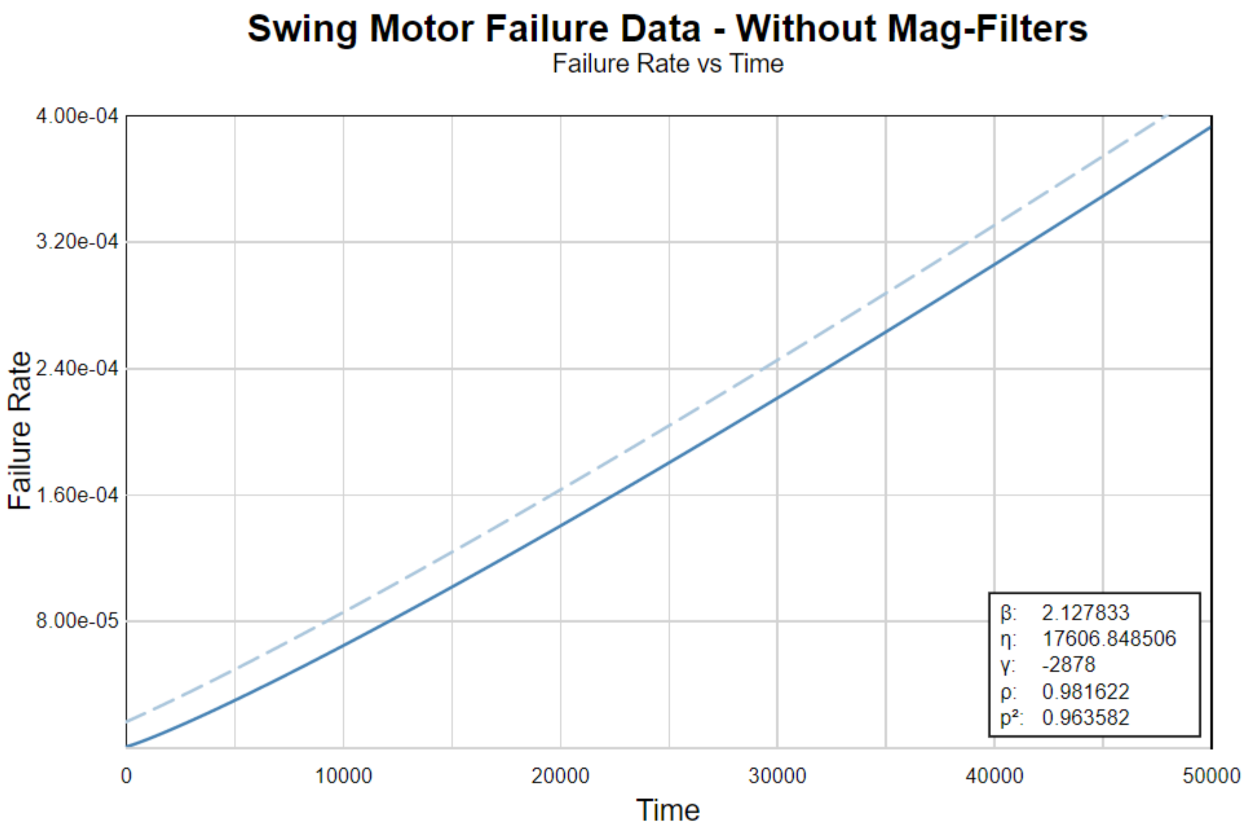


Figure 54. Failure-Rate Vs time (hours) plot for Swing Motors before installation of Mag-filters

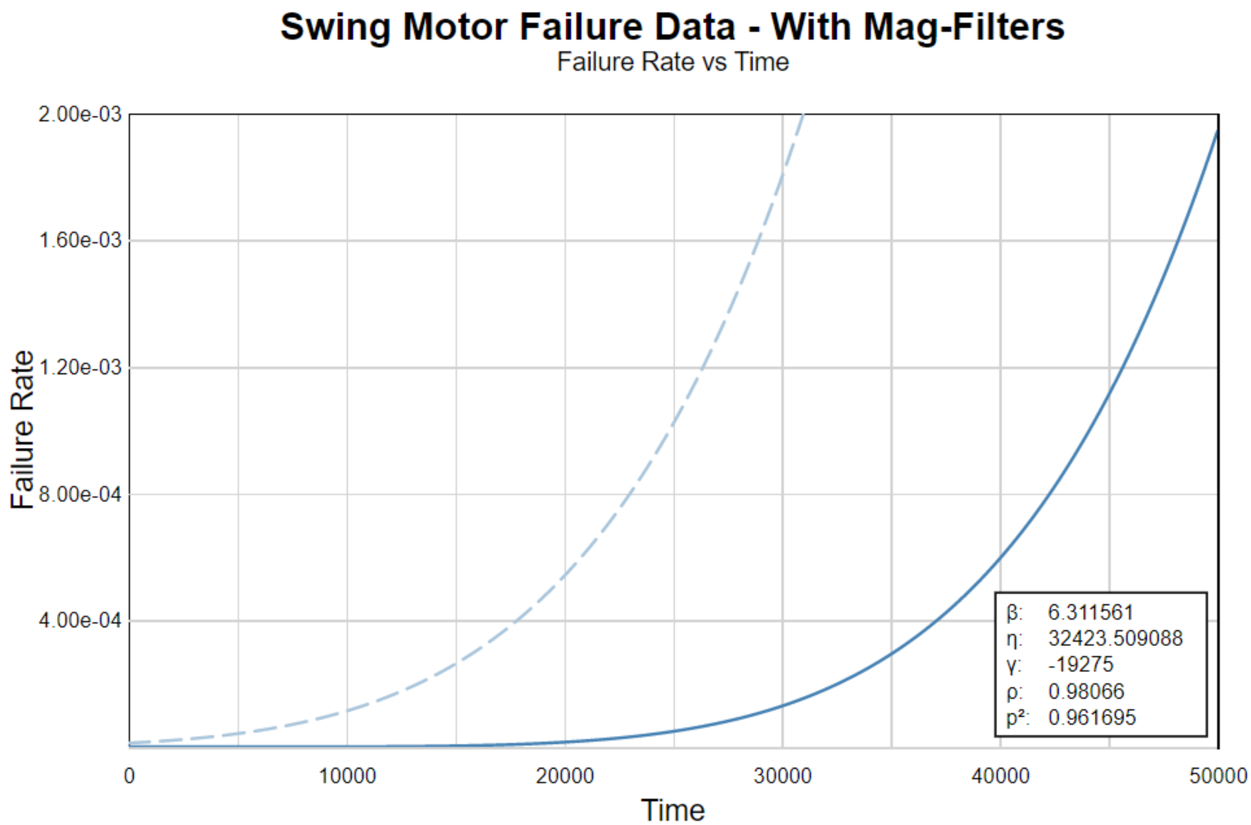


Figure 55. Failure-Rate Vs time (hours) plot for Swing Motors after installation of Mag-filters

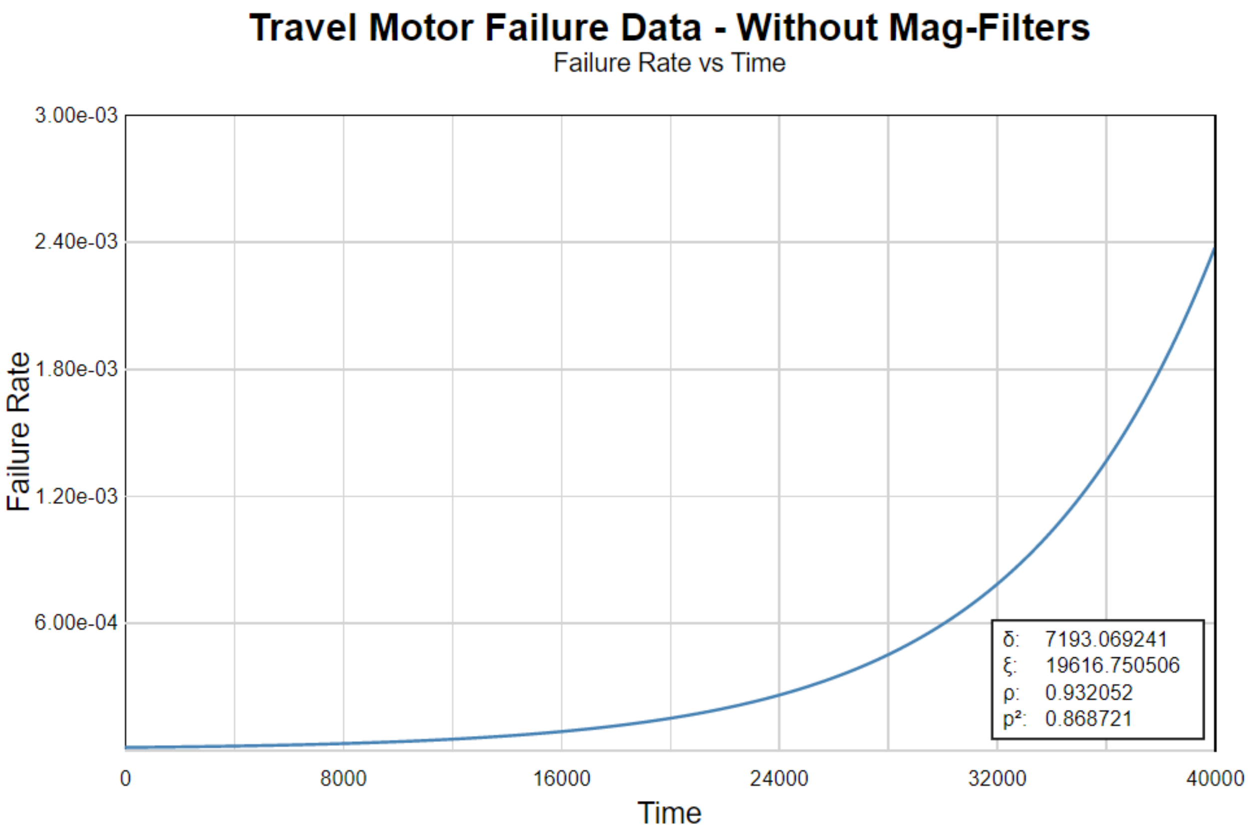


Figure 56. Failure-Rate Vs time (hours) plot for Travel Motors before installation of Mag-filters

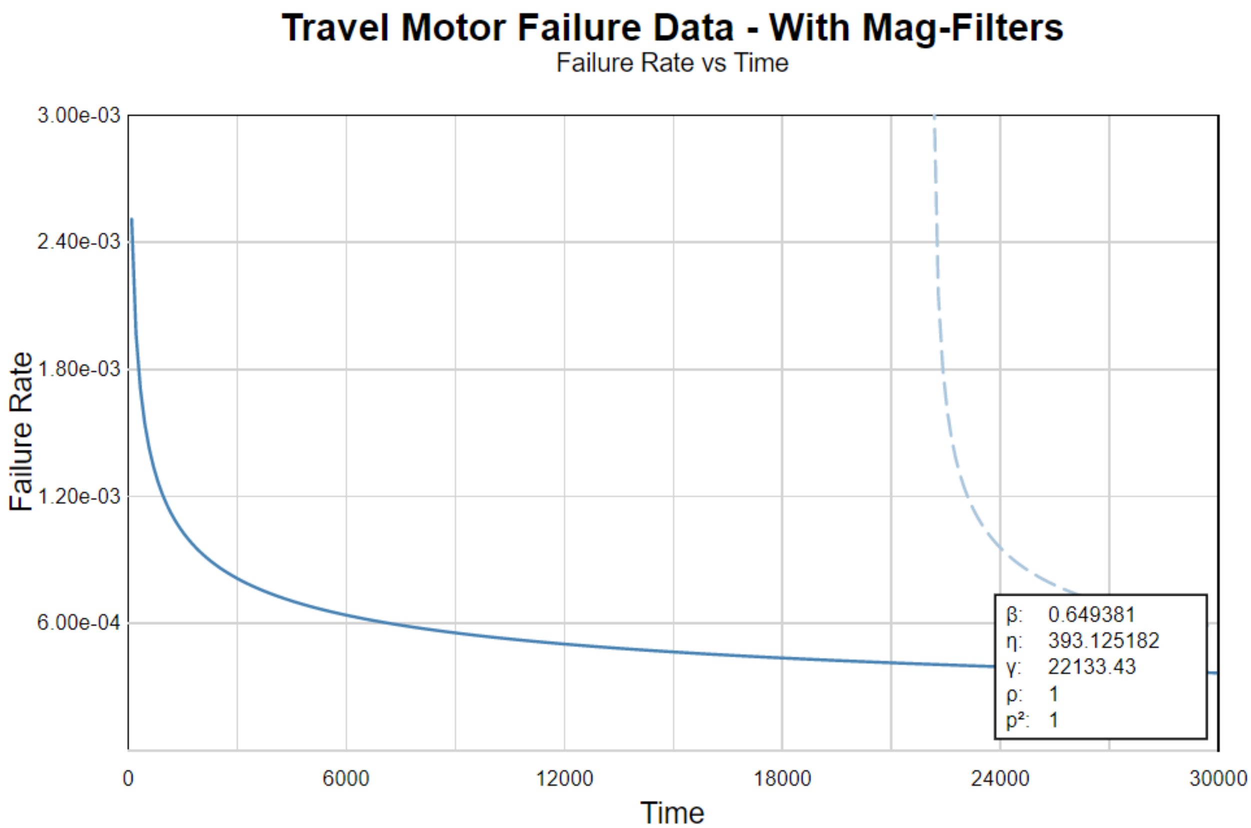


Figure 57. Failure-Rate Vs time (hours) plot for Travel Motors after installation of Mag-filters

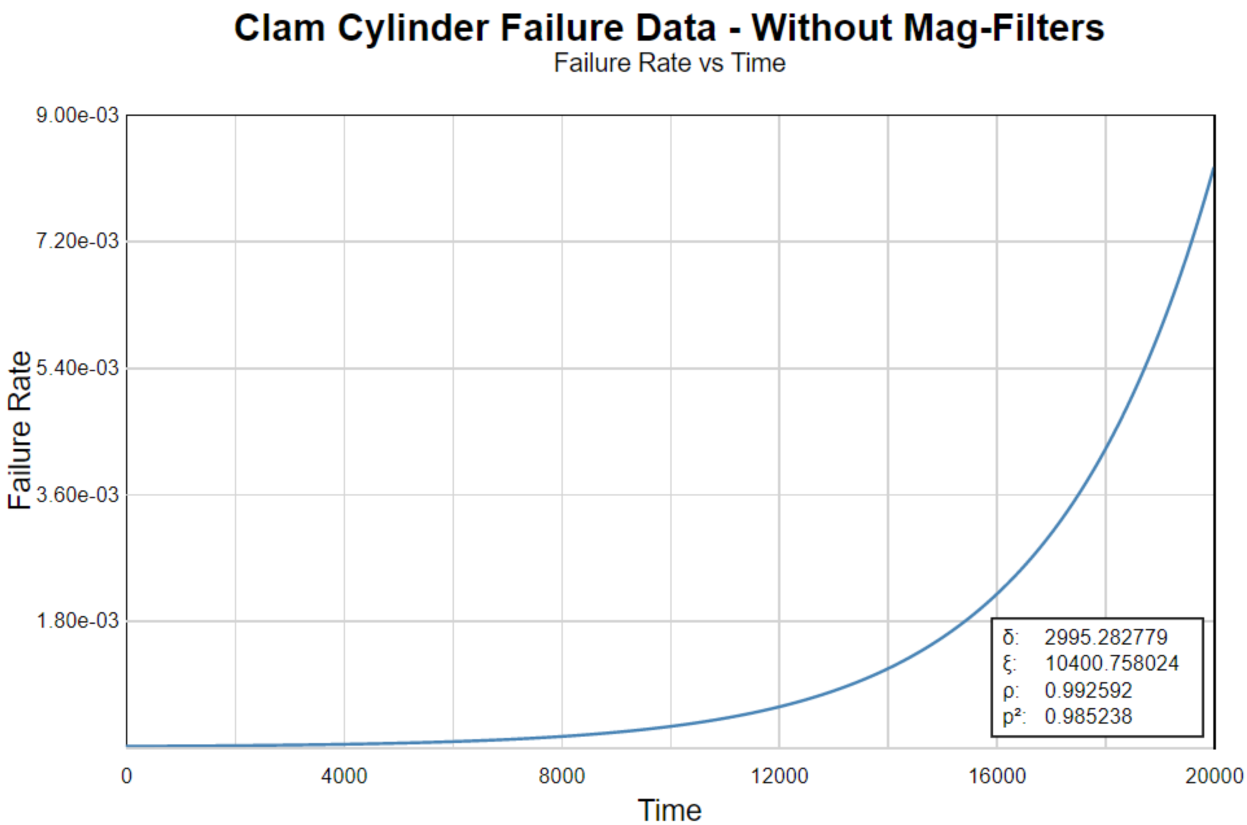


Figure 58. Failure-Rate Vs time (hours) plot for Clam Cylinders before installation of Mag-filters

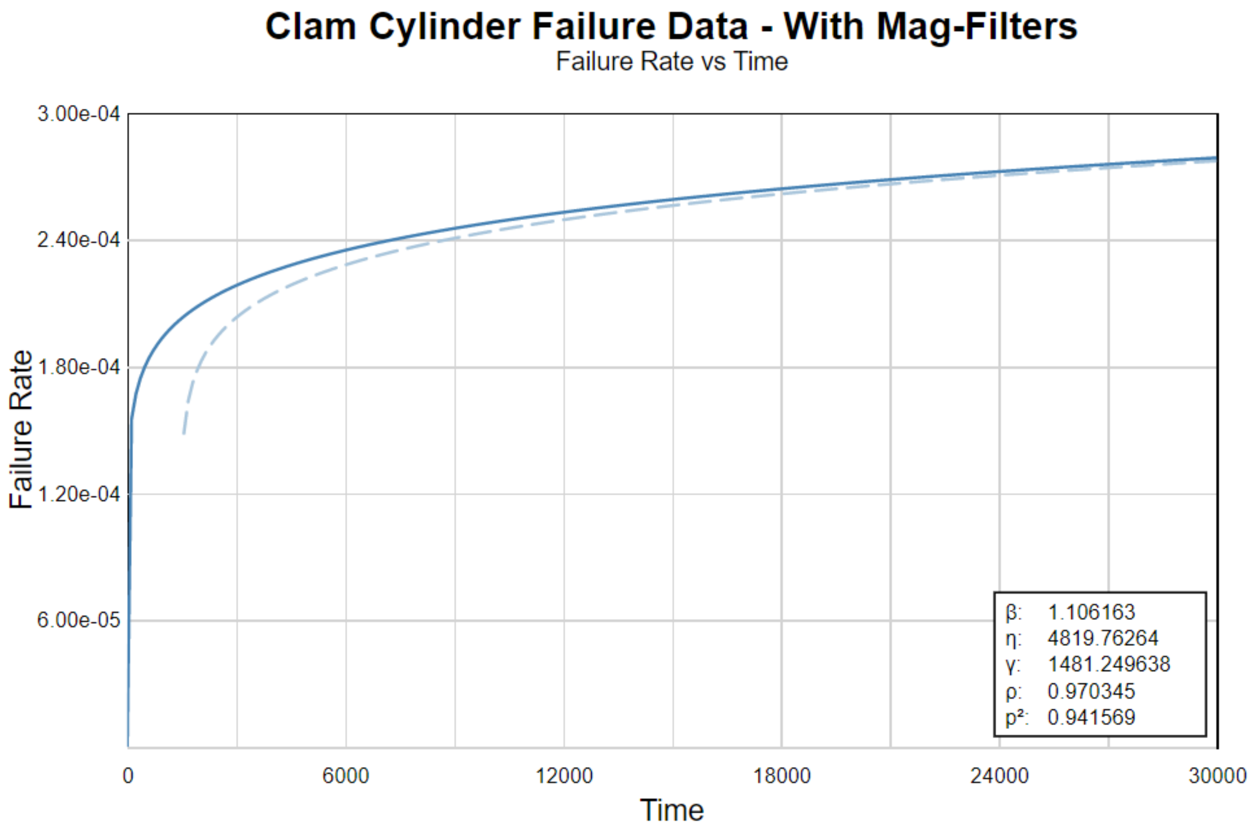


Figure 59. Failure-Rate Vs time (hours) plot for Clam Cylinders after installation of Mag-filters

3.11 Particle Count Analysis in Oil before and after Installation of Mag-filters

3.11.1 Background Information

The main objective of this study is to analyze failures based on the particle count in the oil prior to the failure before and after installation of Mag-filters. Oil samples are collected from return circuit of the hydraulic system every 600 hours and these samples are used for analysis of particles present in oil and other important factors like viscosity, H_2O levels and dirt contamination present in oil using ICP elemental analysis methods. The concentration of wear metals present in the oil i.e., iron, chromium, nickel, aluminum, copper, lead, and tin, and contaminant metals like sodium, potassium, and silicon are measured using ICP methods. The impending failures of the components can be predicted using particle count present in oil. The higher number of particles indicate highly likely chances of contamination indicating early failures. Particle count in oil are reported in terms of ISO cleanliness code. ISO code of less than 20 indicate acceptable level of particles in oil, from 20-23 indicate reportable level of particles in oil and greater than 23 indicate critical levels that alarm the system to be immediately cleaned and fix the root cause of contamination. In this study, percentage of failures in acceptable levels, reportable levels and critical levels were studied before and after installation of Mag-filters. This gives an indication on the role of Mag-filters in preventing wear and debris related failures.

3.11.2 Initial Data Exploration of Oil Sample Data

The 4µ particle contamination in oil for the top eight components is measured before and after installation of Mag-filters. Before the installation of Mag-filters, 49% of the failures were at acceptable levels, 44% of failures were at reportable levels and 7% of the failures were at critical contaminant levels. After the installation of Mag-filters, 61% of failures were at acceptable levels, 33% of the failures were at reportable levels and 6% of the failures were at critical levels. In the S1 unit, before the installation of Mag-filters, 81% of the failures were at acceptable levels and 19% of failures were at reportable contaminant levels. After the installation of Mag-filters, 41% of failures were at acceptable levels, 41% of the failures were at reportable levels and 18% of the failures were at critical levels. In S2, before the installation of Mag-filters, 59% of the failures were at acceptable levels, 24% of failures were at reportable levels and 18% of them were critical failures. After the installation of Mag-filters, 54% of failures were at acceptable levels, 46% of the failures were at reportable levels. Figure 60 through Figure 66 shows percentage of failures in acceptable, reportable, and critical range of 4µ particles. The red segment indicates critical failures, the orange segment indicates percentage of failures that occurred at reportable levels, and the blue segment indicates failures when the 4µ particles in the oil were in acceptable range. From the graphs it can be inferred that on an overall level, the number of failures with particles in critical range has increased post installation of Mag-filters. The percentage of failures with reportable level of particles has increased in S2 post Mag-installation and the number of critical failures has increased in S1 post installation of Mag-filters. This can be indicative that the magnetic-filters are very effective in containing normal wear and can enhance component life of wear- related failures by not increasing wear-rate but cannot contain huge number of particles released specially during debris related failures.

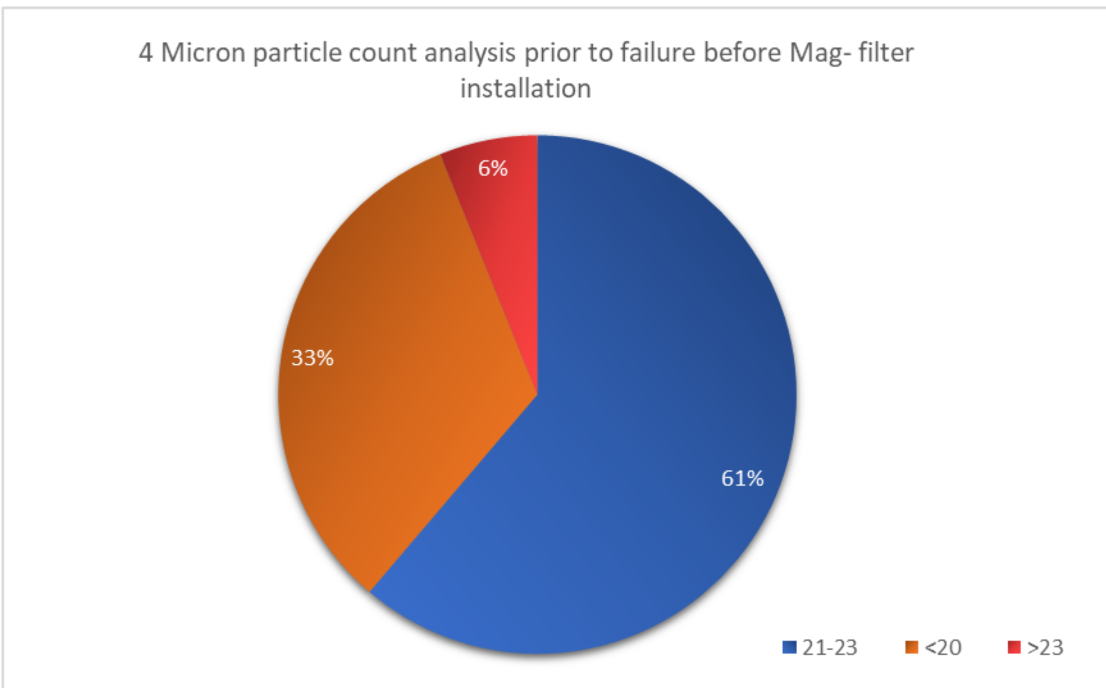


Figure 60. Percentage of failures in different oil conditions before Magnetic filter Installation

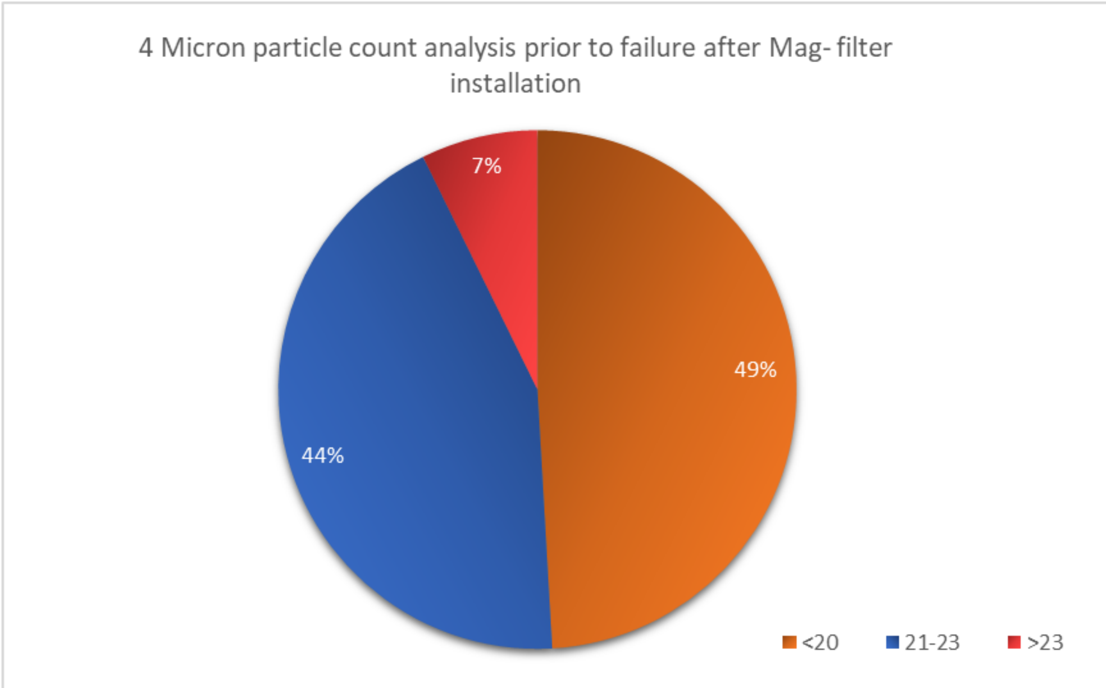


Figure 61. Percentage of failures in different oil conditions after Magnetic filter Installation

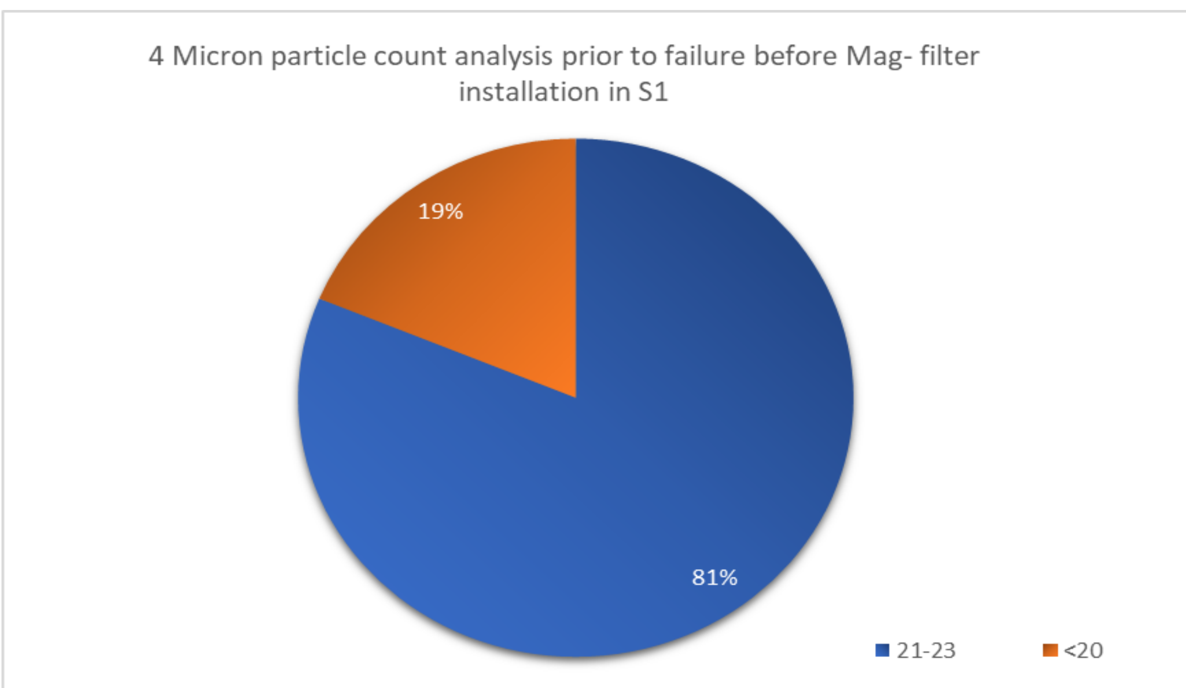


Figure 62. Percentage of failures in different oil conditions before Magnetic filter Installation in S1

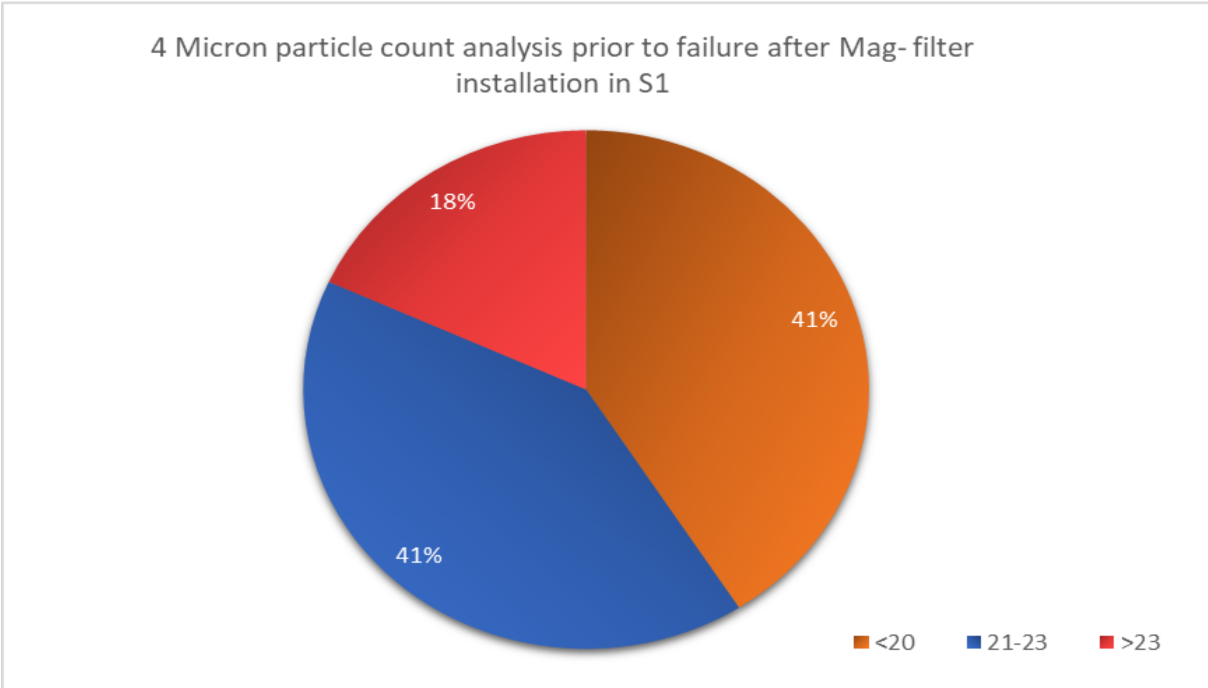


Figure 63. Percentage of failures in different oil conditions after Magnetic filter Installation in S1

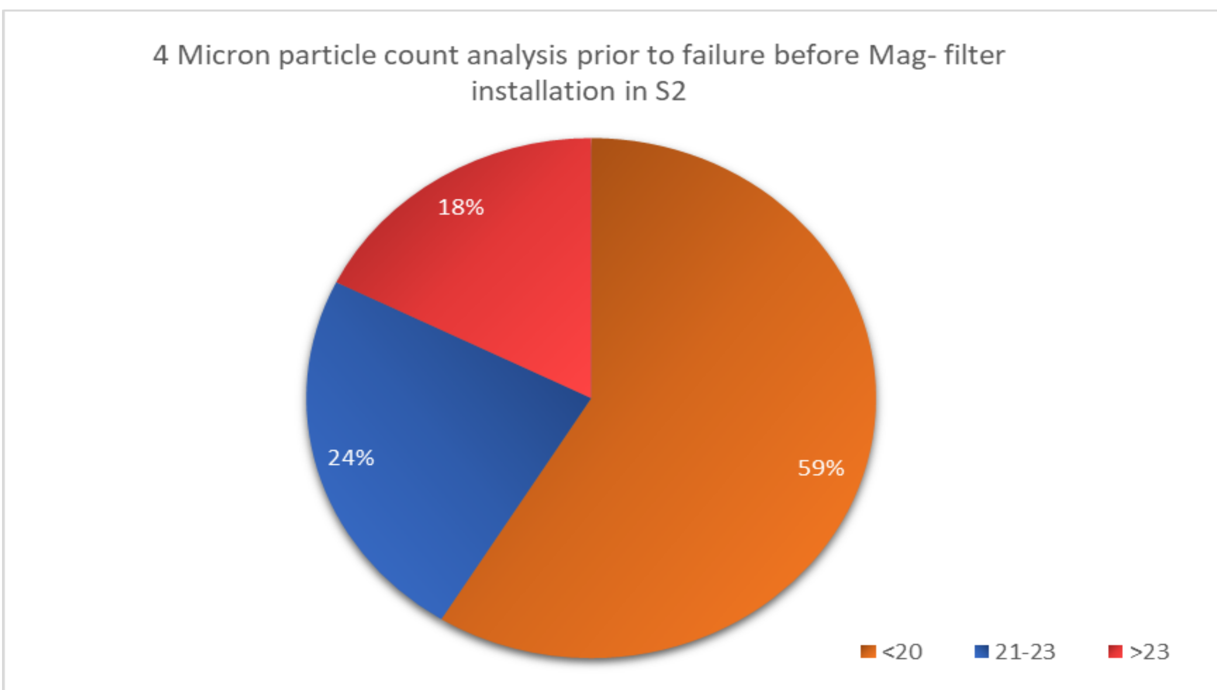


Figure 64. Percentage of failures in different oil conditions before Magnetic filter Installation in S2

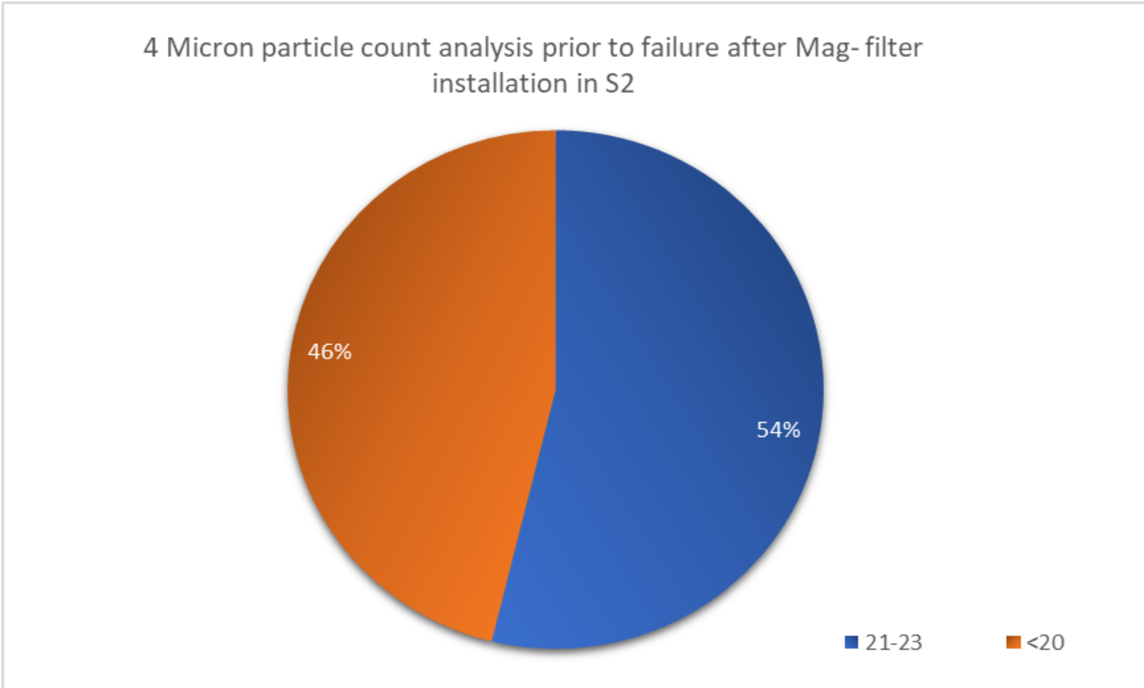


Figure 65. Percentage of failures in different oil conditions after Magnetic filter Installation in S2

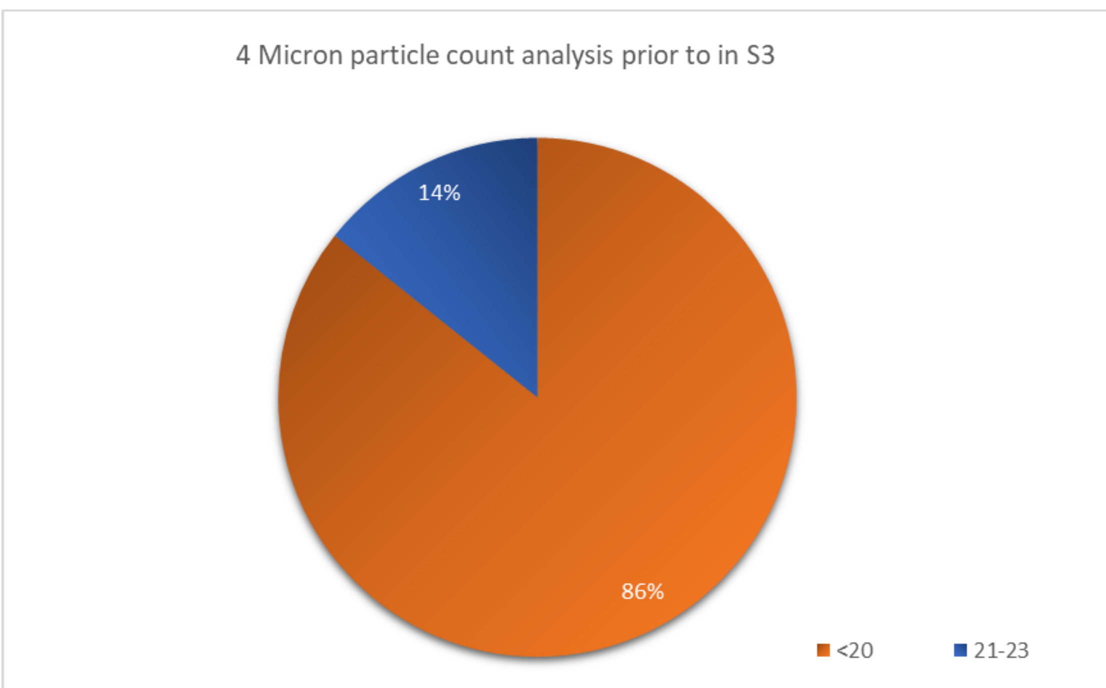


Figure 66. Percentage of failures in different oil conditions after Magnetic filter Installation in S3

The particle count variation in oil was plotted against meter reading (Unit SMU hours) for S1, S2 and S3 units and variations in oil count before and after the installation of Mag-filters was studied. In S1, the oil particle count shows a few spikes after 50,000 SMU units that might have occurred due to the working location in the mine. The frequency of failures has also increased in S1 in the interval of 50,000 to 60,000 SMU hours. The particle count has consistently reduced after the installation of filters. In S2, the particle counts in oil shows deviation from the acceptable range even before Mag-filter installation. The particle count spikes have significantly increased after 25,000 SMU hours. The particle count of 4μ continues to fluctuate and elevate even after the installation of Mag-filters. The particle count in S3 is consistently low from the start. The Mag-filters for this unit were installed at 22,000 hours before the operation started. Figure 67 to Figure 69 shows variation of 4 microns particles in oil in different shovel units. The orange line indicates magnetic filter installed SMU hour.

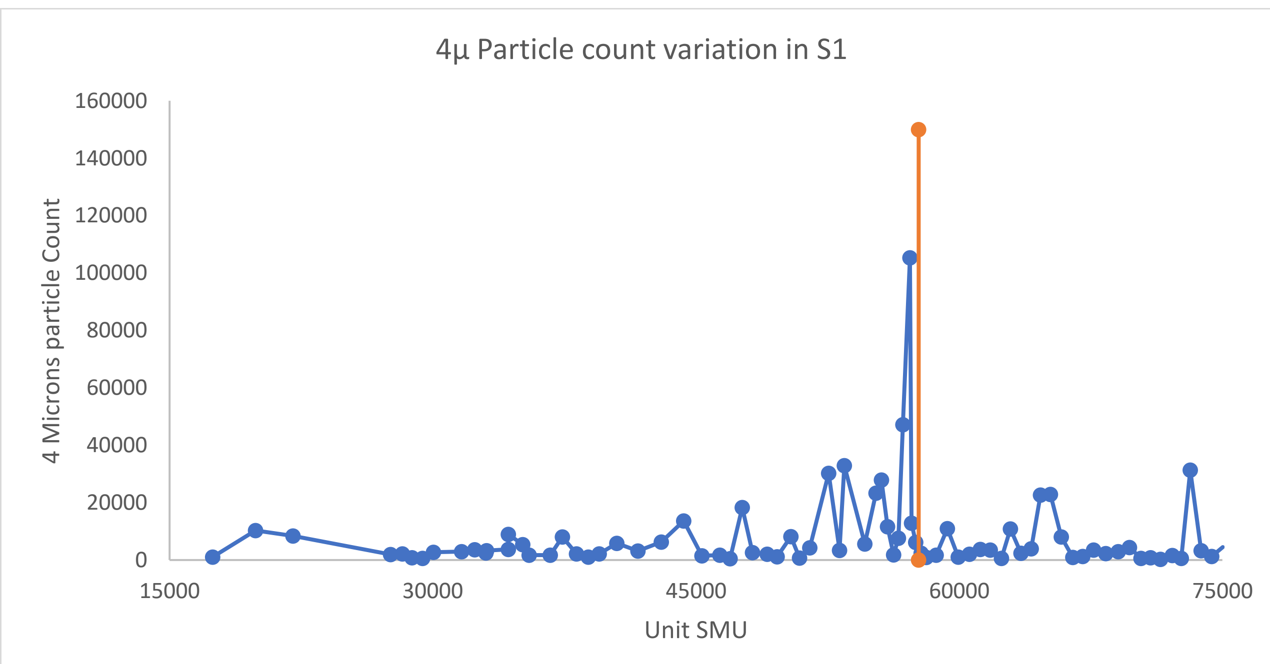


Figure 67 Particle count variation in S1 unit

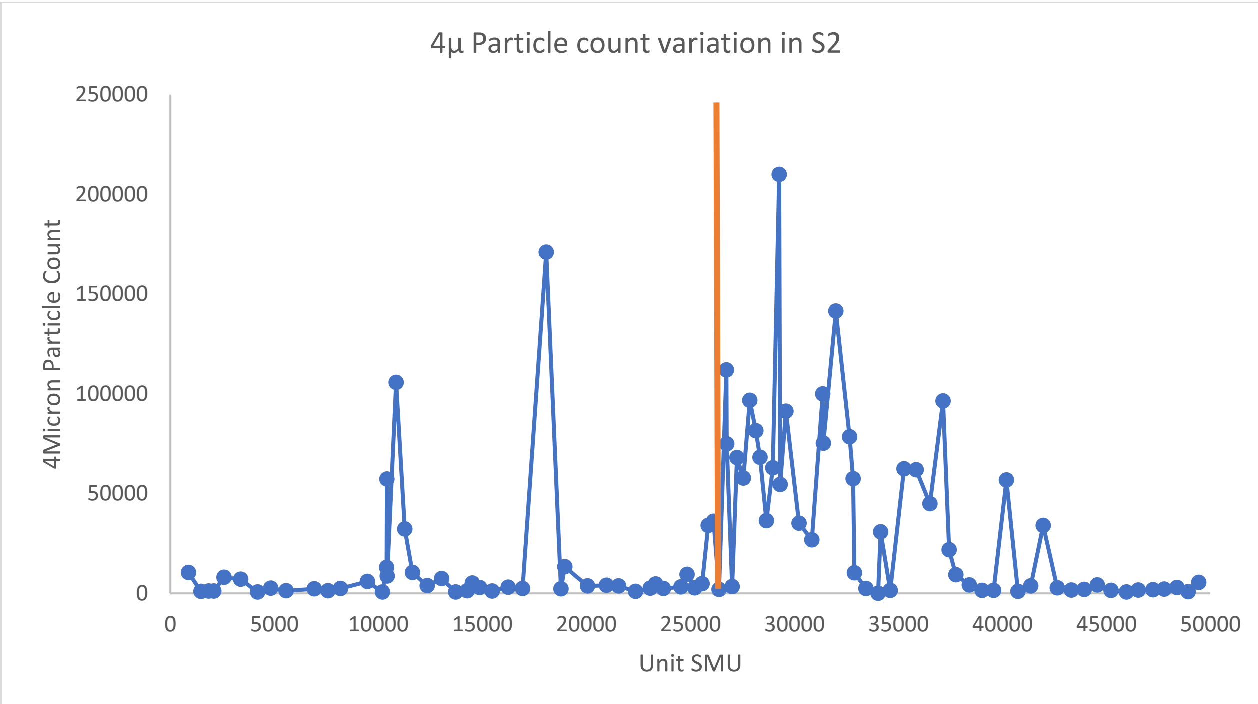


Figure 68. Particle count variation of 4 Microns particles in S2 unit

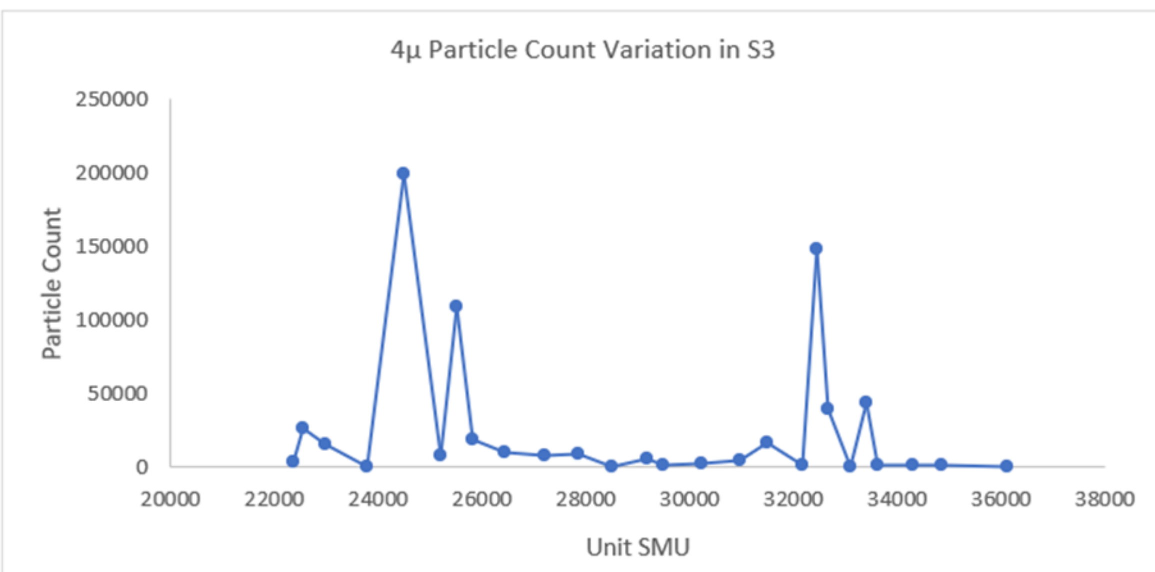


Figure 69. Particle count variation of 4 Microns particles in S3 unit

Figure 70 and Figure 71 were plotted to study the correlation of 4µ, 6µ and 14µ particle count in the oil. From the three graphs, it can be noted that 4µ and 6µ particle counts are highly correlated. If the 4µ particle count in oil is increasing, the 6µ particle count is also increasing. The 14µ particles are always at lower and at consistent levels. This means that the 4µ and 6µ particles mainly influence the failures of hydraulic components.

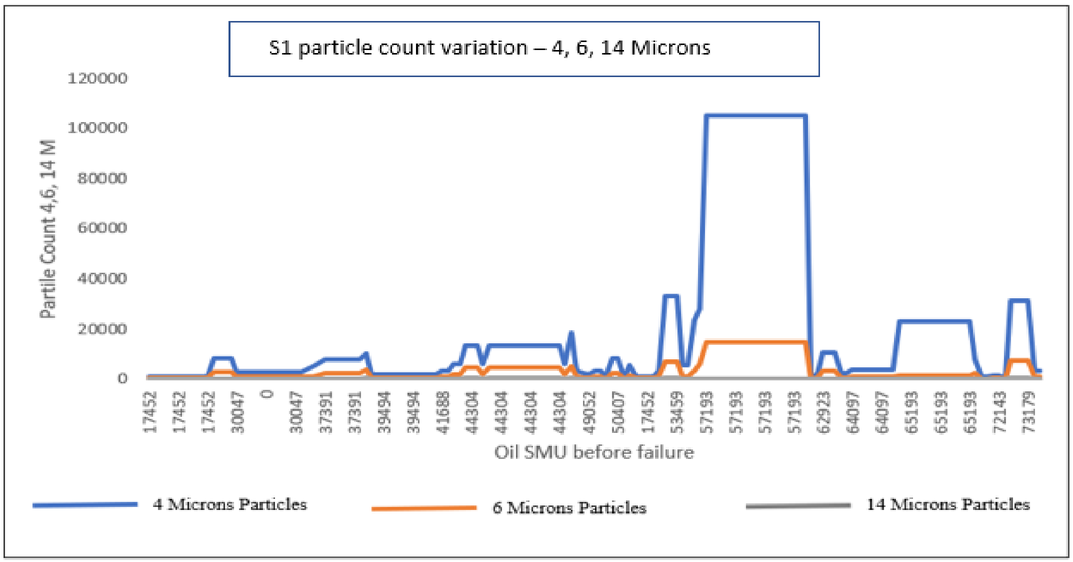


Figure 70. Particle Count Variation in S1

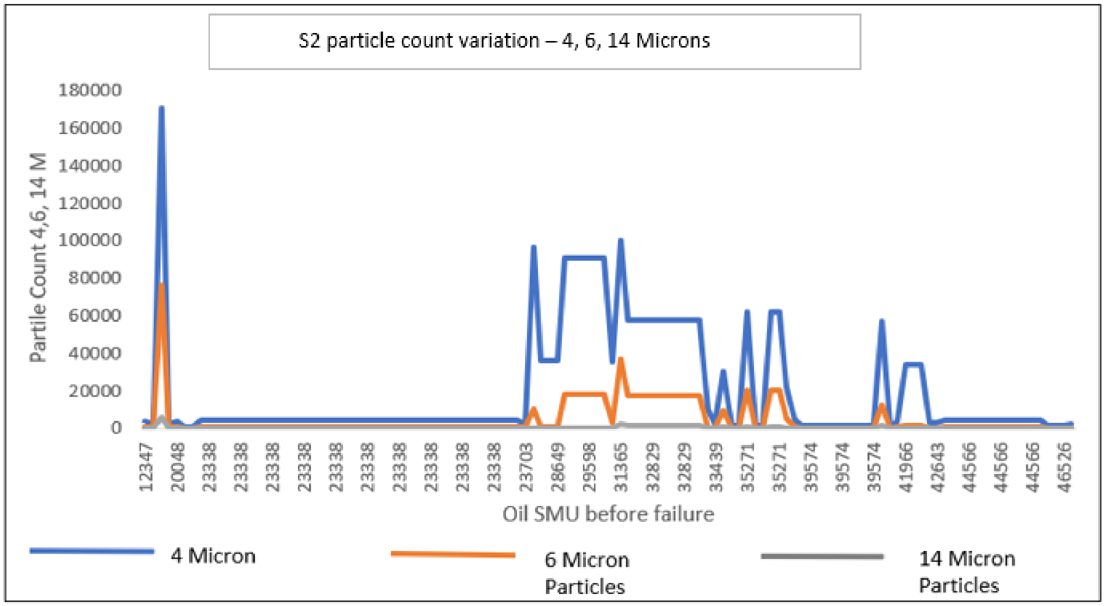


Figure 71. Particle Count Variation in S2

3.12 Comeback Failure Analysis

3.11.1 Background Information

Hydraulic wear and contamination failures often influence the failures of other components in the hydraulic system. A large number of particles are released in the hydraulic oil when a hydraulic wear/contamination failure occurs. These particles circulate in the oil and increase the wear rate of other components. When a debris failure occurs in the system, debris starts traveling downstream of the system and eventually causes clogs and may lead to catastrophic failure of the system. Thus, comparing the system's comeback failure rate before and after the installation of Mag-filters can aid in determining the effectiveness of Mag-filters. When a wear or system contamination failure occurs, a large number of particles are released into the oil. These particles can enhance the wear rates of other hydraulic components or can result in oil debris leading to catastrophic failures as the debris circulates in the oil. Mag-filters can effectively capture small particles and prevent oil contamination which would ultimately prevent successive failures. Hence, the comeback failure rate method is used to identify successive failures in the next 1000 hours of operation after a given component failure and compare how it has changed after the installation of Mag-filters. This helps in quantifying and analyzing the effects of Mag-filters on successive hydraulic failures. The comeback failure rate of a hydraulic component in the shovel fleet is calculated as follows:

$$CR = \frac{\sum_{t=i}^{i+1000} F_i}{\sum_{t=i}^{n} F_i} \tag{39}$$

Where CR is the comeback failure rate, F_i *is* the ith Failure, n is the total number of component failures, t is the meter reading (SMU hours) at the time of failure. Table. 45 indicates comeback rate for different components of hydraulic system.

	COMEBACK FAILURE BEFORE INSTALLATION OF MAG-FILTERS	COMEBACK FAILURE AFTER INSTALLATION OF MAG-FILTERS
MAIN PUMPS	0.92	0.87
SWING MOTOR	1.21	1
TRAVEL MOTOR	3	0.5
BUCKET CYLINDER	0.17	1.5
CLAM CYLINDER	1.36	1.78
HYDRAULIC OIL COOLER	0.33	0
MAIN CONTROL VALVE	0.33	2
PROPEL BRAKE VALVE	0.5	2.33
HYDRAULIC OIL COOLER FAN MOTOR	2.5	0.5
PUMP DRIVE	2	0.67

Table 45. Comeback rates of different hydraulic components

3.12 Cost Analysis

3.12.1 Background Information

Cost analysis involves the comparison of costs of component replacement failures before and after the installation of magnetic filters. This can be beneficial in evaluating the effects of Mag-filters in terms of the cost saved per failure after installation of the magnetic filters. The goal of this study is to compare the failure costs of the identified hydraulic system components for every 10,000 hours before and after the installation of Mag-filters. The analysis includes component part costs, labor costs, and ancillary costs (miscellaneous costs including seal hoses, bolts, etc that are required to install new components).

3.12.2 Cost Analysis Results

The cost per 10,000 working hours for each component before installation given by C_B of Mag-filters and cost per 10,000 hours after installation of Mag-filters, C_A is as follows

$$C_B = \frac{\sum_{i=1}^{n} C_k + \sum_{i=1}^{n} LC + \sum_{i=1}^{n} Misc}{Total\ SMU\ hours\ before\ installation\ of\ Mag\_filters} * 10{,}000 \qquad (40)$$

$$C_A = \frac{\sum_{i=1}^{n} C_k + \sum_{i=1}^{n} LC + \sum_{i=1}^{n} Misc}{Total\ SMU\ hours\ after\ installation\ of\ Mag\_filters} * 10{,}000 \qquad (41)$$

Where C_k is the cost of replaced component, LC is the labor cost for replacing the component, Misc. is the approximate miscellaneous cost spent on the component replacement. The total SMU hours before the installation of Mag-filters for the data collected for three units of hydraulic shovel is approximately 86,158 hours and the total SMU hours after the installation of Mag-filters for the data collected for three units of shovel is approximately around 64,137 hours. Table. 46 indicates cost analysis for different components of hydraulic system before and after installation of Mag-Filters.

COMPONENT	COST PER 10,000 HOURS BEFORE MAG-FILTERS, C_B	COST PER 10,000 HOURS AFTER MAG-FILTERS, C_A
HYDRAULIC OIL COOLER	$ 9,285	$ 2,495
BUCKET CYLINDER	$ 24,691	$ 28,430
CLAM CYLINDER	$ 19,917	$ 26,755
HYDRAULIC OIL COOLER FAN MOTOR	$ 5,777	$ 5,160
SWING MOTOR	$ 31,976	$ 47,477
TRAVEL MOTOR	$ 31,091	$ 11,933
MAIN PUMP	$ 69,277	$ 89,340
MAIN CONTROL VALVE	$ 36,115	$ 32,343
PROPEL/BRAKE CONTROL VALVE	$ 13,525	$ 18,169

Table 46. Cost of component replacement per 10,000 hours before and after installation of Mag-filters

3.13 Summary and Conclusions

Different metrics were used in the analysis of effectiveness of Mag-filters. Pareto analysis was used to identify top eight hydraulic components with highest failure frequency. From the pareto analysis it was noted that main pumps were responsible for almost 25% of hydraulic wear and contamination related failures. In the component life analysis method, average life of components before and after the installation of Mags was compared and ANOVA test was used to check if the two means varied significantly. Different components had different results with regards to average life before and after Mag installation and ANOVA test results were also different. The summary of component life analysis for all eight components are presented below in Table. 47. Of the identified eight components, the components where ANOVA test shows a significant difference in mean between the two groups (i.e., component average life before and after installation of Mag-filters) are considered in quantifying the effects of Mag-filters by comparing different key performance indicators used in this study. Table. 47 summarizes ANOVA test results for different components.

Component	Average life before Mag installation	Average life after Mag installation	ANOVA Test Variation
Main Pumps	14915	12197	Significant
Swing Motors	12885	11577	Not Significant
Travel Motors	18734	23017	Significant
Clam Cylinder	8823	5659	Significant
Bucket Cylinder	13347	12718	Not Significant
Hydraulic Oil Cooler Fan Motor	12348	14842	Significant
Control Valve	23359	18103	Not Significant
Propel Valve	23359	15509	Significant

Table 47. Summary of component life analysis results

From Table. 47, it can be noted that main pumps, travel motors, clam cylinders and hydraulic oil cooler fan motor have significant difference in the component life before and after installation of Mag-filters. Hence these components were further considered in quantifying the effects of Mag-filters. The percentage difference in component life, reliability analysis, comeback failure rate and cost impacts for these failures are calculated. Table. 48 through Table. 51 present the percentage difference in different metrics post installation of Mag-filters.

Component	Before Installation	After Installation	Percentage Change
Travel Motors	18734	23017	23%
Clam Cylinder	8823	5659	-36%
Hydraulic Oil Cooler Fan Motor	12348	15842	28%
Main Pumps	14915	12197	-18%

Table 48. Percentage change in component life post-installation of Mag-filters

Component	Before Installation	After Installation	Percentage Change
Travel Motors	0.87	1	15%
Clam Cylinder	0.84	0.49	-42%
Hydraulic Oil Cooler Fan Motor	0.69	0.84	22%
Main Pumps	0.99	0.775	-22%

Table 49. Percentage change in reliability at 5000 hours of operation

Component	Before Installation	After Installation	Percentage Change
Travel Motors	3	0.5	83%
Clam Cylinder	1.36	1.78	-31%
Hydraulic Oil Cooler Fan Motor	0.33	0.01	97%
Main Pumps	0.92	0.87	5%

Table 50. Percentage change in comeback failure rate post Mag installation

Component	Before Installation	After Installation	Percentage Change
Travel Motors	$31,091	$11,933	62%
Clam Cylinder	$19,917	$26,755	-34%
Hydraulic Oil Cooler Fan Motor	$5,777	$5,160	11%
Main Pumps	$69,277	$89,340	-29%

Table 51. Percentage change in component replacement cost post-Mag installation

From the above analysis it can be inferred that different metrics are differently affected by the installation of Mag-filters. In the component life analysis, certain components have an increased life, and the other components have lower average life after the installation of Mag-filters. The lower average life post installation of Mag-filters can be attributed to various reasons like aging of the machine, location of the equipment and rock hardness variability where the machine is working. The reliability metrics also infer that certain component indicate increase in reliability whereas a other components have lesser reliability at 5000 hours of operation when compared to before filter installation. Comeback failures have significantly decreased post installation of filters for mostly all components. Cost analysis also indicates a few components have lesser component cost per 10,000 hours post installation of filters whereas few other components indicate an increase in component cost per 10,000 hours post installation of Mag-filters. Hence weights were assigned to each metric based on the variability the variable can explain. For example, if all the components have a positive Mag impact or all components have a decreased performance impact then the metric is assigned a higher weight rather than a metric where the component shows both positive and no impact. The weight of each parameter was calculated based on the variability explained by the parameter. The weight of each metric is calculated as follows:

$$W_{M_i} = \frac{\left|C_{P_i} - C_{N_i}\right|}{\sum_{i=1}^{4}\left|C_{P_i} - C_{N_i}\right|} \tag{42}$$

Where W_{M_i} is the weight assigned to the metric, $\left|C_{P_i} - C_{N_i}\right|$ is the difference in count of variables explaining positive increament in percentage change post installation of magnetic filters and count of variables explaining no change (decreament in terms of negative percentage change post installation of magnetic filters).

Accordingly following weights are assigned to each metric:

Metric	Weight
Component Life Analysis	0.22
Reliability Analysis	0.22
Comeback Rate	0.33
Cost Analysis	0.22

Using the following weight, percentage impact of each of the metric is calculated for each component. The weighted impact of a variable on each component is the product of weight assigned to that component and the percentage change in the variable of the component post installation of magnetic filter.

$$\mathit{Weighted\ impact,}\ M_i = \text{Percentage change in the variable after Mag installation} * \text{Weight of the metric } W_{M_i} \tag{43}$$

A weighted score is then assigned to each component by summing the weighted impact of all the variables on the component. The sum of weighted scores is used to calculate the total score. The total score explains the overall impact of magnetic filters based on the parameters/ variables considered in this analysis. The total score explains the overall weighted difference in performance change post Mag installation of all components across each parameter. If the value is positive, then it indicates that the overall Mag effects are positive and if the overall sum is negative then it indicates the Mag-filters are not effective in containing failures. Table. 52 shows the overall impact of Mag-filters on the four components with a significant variation in component life post-Mag installation.

Component	Component Life	Reliability	Comeback	Cost	Weighted Score
Travel Motors	0.05	0.03	0.28	0.14	**0.49**
Clam Cylinder	-0.08	-0.09	-0.10	-0.08	**-0.35**
Hydraulic Oil Cooler Fan Motor	0.06	0.05	0.32	0.02	**0.45**
Main Pumps	-0.04	-0.05	0.02	-0.06	**-0.13**
				Total Score	0.46

Table 52. Overall Impact of Mag-filters

As seen from Table. 52, the total score calculated by summing the weighted scores of each component is +0.46 indicating the overall change post installation of Mag-filters is positive, and hence it is beneficial installing Mag-filters in the hydraulic system. It can also be inferred that magnetic filters have significantly decreased comeback failure rate. From the analysis, it is not indicative that component life is improving post installation of Mag-filters. The reliability of the all the components also do not improve significantly after the installation of Mag-filters. However, cost per 10,000 hours of different components has reduced to a good extent with the installation of Mag-filters. Hence, it can be concluded that magnetic filters have a positive impact on the hydraulic system however they are not the only influencing factor of wear and contamination failures. The other external factors like location of working of equipment, oil cleanliness, type of oil and maintenance activities after a contamination failure also influence the wear and contamination failure of hydraulic system.

4: SUCCESSIVE FAILURE PROBABILITY PREDICTION OF HYDRAULIC COMPONENTS USING MACHINE LEARNING TECHNIQUES

4.1 Background Information

Failures of hydraulic system components have a significant impact on the failure of other hydraulic components. When a hydraulic component wears out, a huge volume of wear particles are released into the system. During typical shovel operations, particles are carried in the hydraulic oil, and if they are not cleaned by the filters, they increase the wear rate of other components, reducing their performance and average life. A significant number of particles are also released into the system during a debris failure. The debris particles will eventually lead to catastrophic failure of the downstream components of the system.

The objective of this chapter is to establish a method for determining the probability of hydraulic component failures in the next 1000 hours of operation after a given failure. Previous component failures, as well as the condition of the oil at the time of failure and 600 hours prior to failure, are used to predict the probability of successive component failures. Figure 72 presents a flowchart with the steps involved in predicting successive failures using machine learning algorithms.

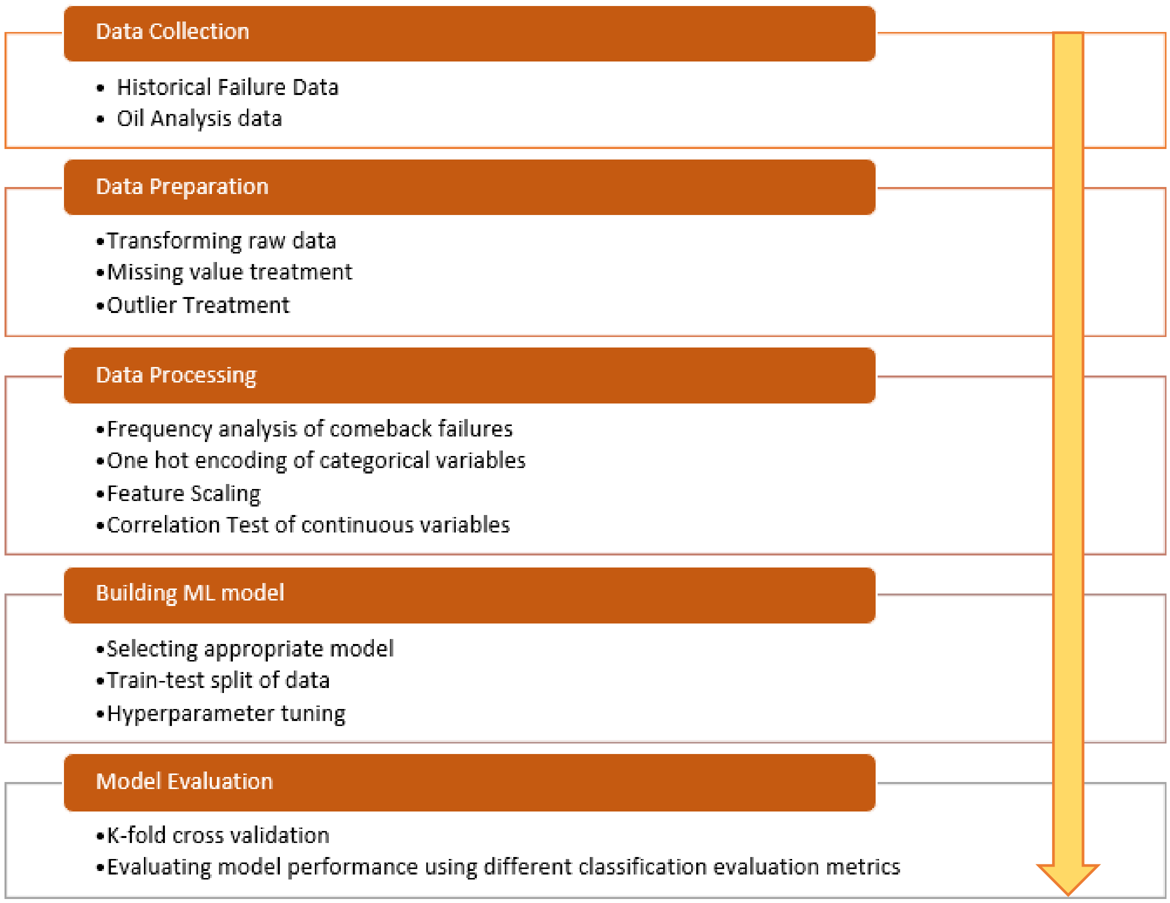


Figure 72. Flow chart detailing steps involved in ML process to identify successive failures

The following sections of this chapter presents a detailed overview of the various steps involved in diagnosing successive failures in next 1000 hours after a hydraulic failure such as data collection, data preparations, data processing, building machine learning models, hyperparameter tuning and evaluating the performance of models. The model performance was also validated by a K-fold validation method using different train-test samples from the dataset.

4.2 Data Collection

In the analysis and prediction of equipment failures, the machine learning models use a variety of input data types. Vibration data, historical failure data, oil analysis and contamination censor data, imagery data, and time series data are the most prevalent types. This analysis uses historical failure data as well as oil analysis particle count data at the time of failure and 600 hours prior to every failure to predict future failures. Failure description data and failure condition data are gathered for hydraulic component replacement failures at the SMU* Component level, as specified in Chapter 3. Only leakage wear and debris-related failures were used in the analysis to assess the probability of successive failures in the next 1000 hours. The objective of using historical failure data was to find failure patterns in 1000 hours of operation prior to a component failure. If a certain failure pattern of hydraulic component failures in the system is established, the impact of a component failure on other components can be investigated, and a relationship between these failures can be derived to predict future failures. As a result, the model can assist in identifying future failures after a hydraulic component failure. Downtime and maintenance costs for these components can be decreased by implementing a proper maintenance plan.

Oil samples data corresponding to each failure is also used for the analysis. Oil samples from the shovels are collected at every 600 hours and particle count of (4\6\10\14\25\35\70)μ particles in the oil are measured. The results also indicate if there are any elevations in particle levels and information about viscosity and water levels in the oil. As the wear and debris related failures are mainly influenced by oil contamination, 4μ, 6μ and 10μ particles at the time of failure and variation in particle count 600 hours prior to the failure are considered in the ML model. The hydraulic system of shovel consists of 45 critical components. For ease of analysis and improved model performance, similar components with same function and capacity were grouped together as one component. (For example, Swing Motor Front Left, Swing Motor Front Right, Swing Motor Rear Left, and Swing Motor Rear Right were grouped as Swing Motors). Hence, there were 15 key components included in the analysis. Historical failure data of these 15 components were used in the analysis. Other minor failures (like rotary joint failures etc.) were excluded from the analysis. A total of 490 failure records for the three units of hydraulic shovel were used to predict failures. This study uses a total of 9 features, 6 of which are continuous variables and 3 of which are categorical. Appendix B contains a sample of the data collected. The list of 45 major components constituting the hydraulic system is mentioned in Appendix B. The list of 15 majorly grouped components is presented in the later sections.

4.3 Data Preparation

4.3.1 Data Transformation

Two machine learning algorithms are used in this study to identify the probability of successive failures. K-NN algorithm is used to predict the probability of a successive failure in the next 1000 hours after a hydraulic component has failed. The model uses historical component data of failures to check if there are any failures in the next 1000 hours of a failure given the oil particle conditions. Based on the training data, the KNN model learns to predict the probability of a future failure in the next 1000 hours after a hydraulic component failure given the oil particle conditions at the time of component failure. The Naïve Bayes algorithm is used to predict the probability of failures of all the hydraulic components after a component has failed. Based on the combined probability of both the models, decisions can be taken on preventive maintenance actions and component replacements.

The unique key identifier used in data collection for K-NN model is WO (work order). A unique work order is created for each hydraulic component replacement. SMU (meter reading) at the time of failure and shovel unit number are also recorded. Information on whether the magnetic filter was installed at the time of failure, flagged (1/0) is also tabulated for each failure. Oil particles at the time of failure and 600 hours prior to failure are recorded using oil samples data. A component-failure column is created that is flagged (1/0) to indicate if there is a failure in the next 1000 hours of that component failure. The dataset used in the model analysis includes information on WO, SMU hours at the time of failure, component replaced, Mag-filter information and 4μ, 6μ and 10μ particle count at the time of failure and 600 hours prior to failure. A sample of data collected for K-NN algorithm is presented in Appendix B.

The data used for naïve bayes algorithm records failure status of all 15 components included in the analysis at every SMU failure hour. A "failure-status" column is created to indicate the failure state (1/0) of components at each SMU failure hour. A new column called "Failure status in the last 1000 hours" was added to identify components that have failed in the 1000 hours previously to the SMU failure hour. For each failure, information on whether the magnetic filter was installed at the time of failure, flagged (1/0), is tabulated with 1 indicating Mag filter was installed at the time of failure. Using oil samples data, oil particles at the time of failure and 600 hours prior to failure are collected. The dataset includes information on work order, SMU hours at the time of failure, component replaced, failure status of components at the time of failure, failure status of components in the last 1000 hours of failure, Mag-filter information and 4μ, 6μ and 10μ particle count at the time of failure and 600 hours prior to failure. A sample of data collected for naïve bayes algorithm is presented in Appendix B.

The methodologies used in this study to identify subsequent failures within 1000 hours of a component failure using K-NN and Nave-Bayes machine learning approaches are detailed in the sections below.

4.3.2 Identifying and Handling Null Values

ML and DL-based algorithms cannot handle missing values, hence missing values in the input features must be addressed. The most popular techniques for dealing with missing values in a dataset are to replace them with a value of 0; replace them with the mean or median or mode value of the feature; use linear interpolation to fill in the missing values, or to remove the missing data points from the dataset. In this study, the mean value of a continuous feature was used to replace the null values of that feature. There were missing values in features tabulated from oil sample data. Around 13% of data from features including 4μ particle count at the time of failure, 6μ particle count at the time of failure, and 14μ particle count at the time of failure was missing. Around 19% of data was missing from the 4μ particle count 600 hours prior to failure, the 6μ particle count 600 hours prior to failure, and the 14μ particle count 600 hours prior to failure.

4.3.3 Encoding of Categorical Variables

Categorical data is information that is divided into distinct categories within a dataset. Encoding categorical data is the process of transforming categorical data into integer format so that data with converted categorical values can be fed into various models. Label encoding, one hot encoding, dummy encoding, binary, and target encoding are the methods for encoding categorical data.

One hot encoding is used when there is no order in the categorical variable. A new variable is created for each level of a category feature in one hot encoding. Each category is represented by a binary variable with a value of 0 or 1. The absence of that category is represented by 0 while the presence of that category is represented by 1. The number of variables depends on the levels present in the categorical variable. One hot encoding was employed in the Naive Bayes dataset to encode component failures that failed over the past 1000 hours for each failure. Failures for different components in the past 1000 hours prior to each hydraulic failure were encoded in 15 newly created categories with binary values 0 and 1. An illustration of the dataset is represented in the later section.

Label encoding technique is used when the categorical feature is ordinal. In label encoding, each label is converted into an integer value. Label encoding was used in K-NN and Naïve Bayes dataset to encode different components of the hydraulic system.

4.3.4 Features Selection by Correlation Analysis

Model training with highly correlated variables may lead to multicollinearity that reduces the precision of estimation. Training model with a large number of features is also time-consuming and computationally expensive. The Naïve Bayes algorithm also assumes the training features to be independent of each other. Hence in this study, if the variables exhibit a strong correlation correlation of greater than 0.75 then one of the features is omitted from the model training dataset. Different tests are used to estimate the correlation between two variables. Pearson's correlation, Spearman's correlation, and Kendall Tau's correlation tests are the most commonly used correlation tests. The correlation of continuous variables in this analysis is determined using Spearman's correlation test. The Spearman’s correlation test is defined as:

$$\rho = 1 - \frac{6 \sum {d_i}^2}{n(n^2 - 1)} \tag{44}$$

Where n is the total number of observations, d_i is the difference between the ranks of corresponding variables. The feature selection of continuous variables in this study are presented in the later section.

4.3.5 Feature Scaling

Feature scaling marks the end of the data preprocessing in machine learning. It is a method to standardize the independent variables of a dataset within a specific range. In other words, feature scaling limits the range of variables so that the variables can be compared on common grounds. Feature transformation is an essential step of the DM process that rescales the input features to a smaller range and is crucial where the values of input features do not have the same order of magnitude. The most popular choices for rescaling the input features in ML are min-max scaler and standard scaler. A standard scaler is used in K-NN analysis to standardize continuous variables. Naive Bayes doesn't require and is not affected by feature scaling. In fact, any algorithm which is not distance-based is not affected by Feature Scaling. Standard scaler rescaling is defined as:

$$Standard\ Scaler\ \bar{x} = \frac{x - \mu}{\sigma} \tag{45}$$

where μ is the mean and σ is the standard deviation of an input feature.

4.4 Building ML Models

4.4.1 k-NN Model

The k-nearest neighbors (k-NN) method is a supervised machine learning algorithm that can be used to address classification problems (Harrison, 2018). k-NN is a kind of instance-based learning in which the function is only estimated locally, and all computation is deferred until classification. The data points are categorized based on how their neighbors are classified. The algorithm's idea is that all data points with similar characteristics will be found near together. To classify an unknown instance represented by some feature vectors as a point in the feature space, the *k*-NN classifier calculates the distances between the point and points in the training data set. The algorithm chooses k nearest points to the new point and classifies the new point to that category for which the number of neighbors is maximum from the chosen k points. Usually, the Euclidean distance is used as the distance metric. Then, it assigns the point to the class among its *k* nearest neighbors (where *k* is an integer). If A and B are two points represented by feature vectors A = $\{x_1, x_2, x_3, \ldots x_n\}$ and B = $\{y_1, y_2, y_3, \ldots y_n\}$ where n is the dimensionality of the feature space then the normalized Euclidean metric is generally used as

$$dist(A,B) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \tag{46}$$

The choice of K, as well as the distance measure used to pick the nearest K points, determine the performance of a k-NN classifier. The sensitivity of K selection can significantly decrease K-NN classification performance. (Imandoust & Bolandraftar, 2013) (Jabbar, Deekshatulu, & Chandra, 2013) (Zhang Z. , 2016). The minimal error method and grid search method are used to find the best k value using the training and test dataset. KNN algorithm is used to predict the probability of a successive failure in the next 1000 hours after a component has failed. The details of using this algorithm are discussed in the later section.

4.4.2 Naïve Bayes Algorithm

Naïve bayes is a supervised machine learning algorithm that assumes an underlying probability distribution and captures uncertainty about the model in a logical manner by calculating the probabilities of occurrence. The Naive Bayes algorithm is a straightforward probability classifier that derives a set of probabilities by counting the frequency and combinations of values in a data set. When assessing the value of the class variable, the method applies Bayes' theorem and assumes that all variables are independent. In a range of controlled categorization challenges, the algorithm learns quickly (Saritas & Yasar, 2019). Bayes' Theorem is a simple mathematical formula used for calculating conditional probabilities. The basic mathematical formula of Bayes theorem is given by

$$P(\mathrm{y}|X) = \frac{P(y) * P(X|y)}{P(X)} \tag{47}$$

Where P (y |X) is the posterior probability, P (X |y) is the likelihood probability, P(y) is the prior probability and P(X) is the marginal probability.

Naïve Bayes classifier is a machine learning classifier that is an extended version of Bayes theorem. The variable y is the output class variable with binary or multiclass categories.

X = ($x_1, x_2, x_3, x_4, x_5, x_6, \ldots, x_n$) represent n features of the model input data. Then the probability of y given the X conditions is denoted by

$$P(y|x_1, x_2, \ldots., x_n) = \frac{P(x_1|y) * P(x_2|y) \ldots \ldots \ldots * P((x_n|y) * P(y)}{P(x_1)\, P(x_2) \ldots \ldots . P(x_n)} \tag{48}$$

Since the denominator in the dataset is constant and does not change,

$$P(y|x_1, x_2, \ldots., x_n) = P(x_1|y) * P(x_2|y) \ldots \ldots \ldots * P((x_n|y) * P(y)$$

Therefore,

$$y = argmax_y P(y) \prod_{i=1}^{n} P(x_i|y) \tag{50}$$

From the above equation, the class of the sample can be obtained, given the predictors. There are different types of Naive Bayes classifiers. When characteristic values are continuous, it is assumed that the values associated with each class are spread according to the Gaussian distribution, which is the normal distribution. On multinomial distributed data, Multinomial Naive Bayes is preferred. Bernoulli Naive Bayes is employed when data is distributed according to multivariate Bernoulli distributions (Gandhi, 2018) (Prabhakaran, 2018). Since the dataset used in the model has both continuous and categorical data, a gaussian distribution classifier was used for the analysis. When the predictors take up a continuous value and are not discrete, these values are assumed to be sampled from a gaussian distribution (Gandhi, 2018). A probability density function of the variable can be assumed by estimating the parameters of the distribution. The formula for conditional probability for continuous variables changes to

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} * e^{-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}} \tag{51}$$

Where μ_y is the sample mean, σ_y is the sample standard deviation and σ_y^2 is the sample variance.

The Naïve Bayes model is used to predict the probability of failures of all the hydraulic components after a component has failed. Based on the combined probability of both the models, decisions can be taken on preventive maintenance actions and component replacements. The details on using this algorithm are represented in later sections.

4.4.3 Train-test Split Method

The train-test split procedure is used to estimate the performance of machine learning algorithms when they are used to make predictions on data not used to train the model. It is a fast and easy procedure to perform, the results of which are used to compare the performance of model algorithms (Brownlee.J, 2020). A train-test split of 80:20 was used in the KNN algorithm, where 80% of the data was used in the training set and 20% of the data was used as the test set. In case of naïve bayes algorithm, the outcome(dependent) variable is not balanced. Since the data also consists of information of components that have not failed along with the failure component at each failure SMU hour, the data is biased more towards 0s than 1s. Hence, a stratified sampling method was used to balance the binary categories of dependent variable in the train and test groups. A stratified train-test split divides the data into train and test sets in such a way that it preserves the same proportion of examples in each class of dependent variable observed in the original dataset (Brownlee.J, 2020).

4.5 Model Evaluation and Tuning

4.5.1 K-fold Cross Validation

Cross-validation is a resampling technique for evaluating machine learning models that have smaller data size. The process includes a single parameter, k, which specifies the number of groups into which a given data sample should be divided. Cross-validation is most commonly used in applied machine learning to validate model efficiency on unknown data. That is, to use a limited sample to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model (Jason, 2020). Since the dataset for naïve bayes algorithm was smaller and biased, a k-fold cross validation with the value of k =3 was used to evaluate model performance and efficiency. The value of K =3 generated 3 different models with different samplings of train-test groups using stratified sampling method. The model efficiency was evaluated by resampling training data into 3 different sets.

4.5.2 Hyperparameter Tuning

A hyperparameter is a control parameter for the learning process, and hyperparameter tuning is the process of identifying a set of optimal hyperparameters for the chosen learning algorithm in machine learning and deep learning. Tuning hyperparameters can be done in a variety of ways, the most basic of which is hand tuning, which is labor intensive. Automated hyperparameter tuning frameworks such as grid search and random search have been proposed to solve the limitations of manual hyperparameter tuning. Grid search includes searching at various hyperparameter value combinations in the space of a grid that the user has selected. The Grid search does an exhaustive search by calculating the error on a validation set using multiple combinations of hyperparameter values and selecting the combination of parameters that provides the least error as the optimal hyperparameters (Mishra, 2019). A grid search hyperparameter tuning method was used in the KNN model to find the best value of K with minimum error. The search was also validated by graphing the minimum error rate for all values of K on training and validation set.

4.5.3 Performance Evaluation Metrics Used in Failure Prediction Models

Performance evaluation metrics help in understanding how well the model describes the dataset. Model evaluation is a process through which the quality of the system's prediction can be quantified. The trained model measurement is tested on validation and test datasets and the labeled data (dependent variable) in the validation or test dataset is compared with its own predictions (Mishra, 2019). The model performance metrics give a rough estimate of the accuracy of the algorithm on the dataset being used and the chances of over-fitting and under-fitting problems.

For classification models, accuracy, log loss, precision, recall, confusion matrix, AUC score, F1 score, sensitivity and specificity are the mainly used metrics to evaluate model performance. This study mainly uses accuracy, confusion matrix and AUC score to estimate the performance of test dataset using KNN and Naïve Bayes algorithm

- **Accuracy** measures the proportion of true results to total cases. Accuracy is presented as

$$accuracy = \frac{True\ Positive + True\ Negative}{True\ Postive + True\ Negative + False\ Positive + False\ Negative} \quad (52)$$

- Confusion Matrix assesses the relationship between the label and the classification of the model. A confusion matrix has two axes: one for the predicted label and the other for the actual label and represents the number of classes. Confusion Matrix is represented as

True Positive	False Positive
False Negative	True Negative

- The area under the curve (AUC) is calculated by plotting true positives on the y axis and false positives on the x axis. This metric is useful since it gives a single value to compare different sorts of models.

4.6 Results and Discussions

4.6.1 Data Preparation

Table. 53 shows the components of the hydraulic system that were considered for successive failure analysis of system components in the next 1000 hours. The dataset prepared for k-NN, and Naïve Bayes algorithm are presented in Appendix B. Around 72% of the times, a component failure was associated with a comeback failure in the next 1000 hours.

HYDRAULIC OIL COOLER FAN PUMP	HYDRAULIC OIL COOLER	BUCKET CYLINDER
PILOT PUMP	MAIN CONTROL VALVE	CLAM CYLINDER
A/C COMPRESSOR PUMP	TRAVEL MOTOR	MAIN PUMP, 5 & 6
BOOM CYINDER	HYDRAULIC OIL COOLER FAN MOTOR	MAIN PUMP, 1 & 2
PROPEL/BRAKE CONTROL VALVE	MAIN PUMP, 9 & 10	SWING MOTOR

Table 53. Components of hydraulic system

4.6.2 Data Processing and Preliminary Analysis

An initial analysis was performed to identify components with the highest rates of comeback failures in the 1000 hours following the component failure. Figure 73 describes the probability of comeback failure of different hydraulic components considered in this study. The swing motor failures have the highest probability of having a component failure in the next 1000 hours of operation. After a swing motor failure, there are 48% chances that a component fails within 1000 hours. Main Pumps, 1 &2 have 33% chances of a comeback failure, followed by Main Pumps 5 &6 and clam cylinders, which have a 25% risk of a successive failure in the next 1000 hours. For the hydraulic components evaluated in the study, the following graph depicts the chance of a subsequent failure within 1000 hours of the component failure.

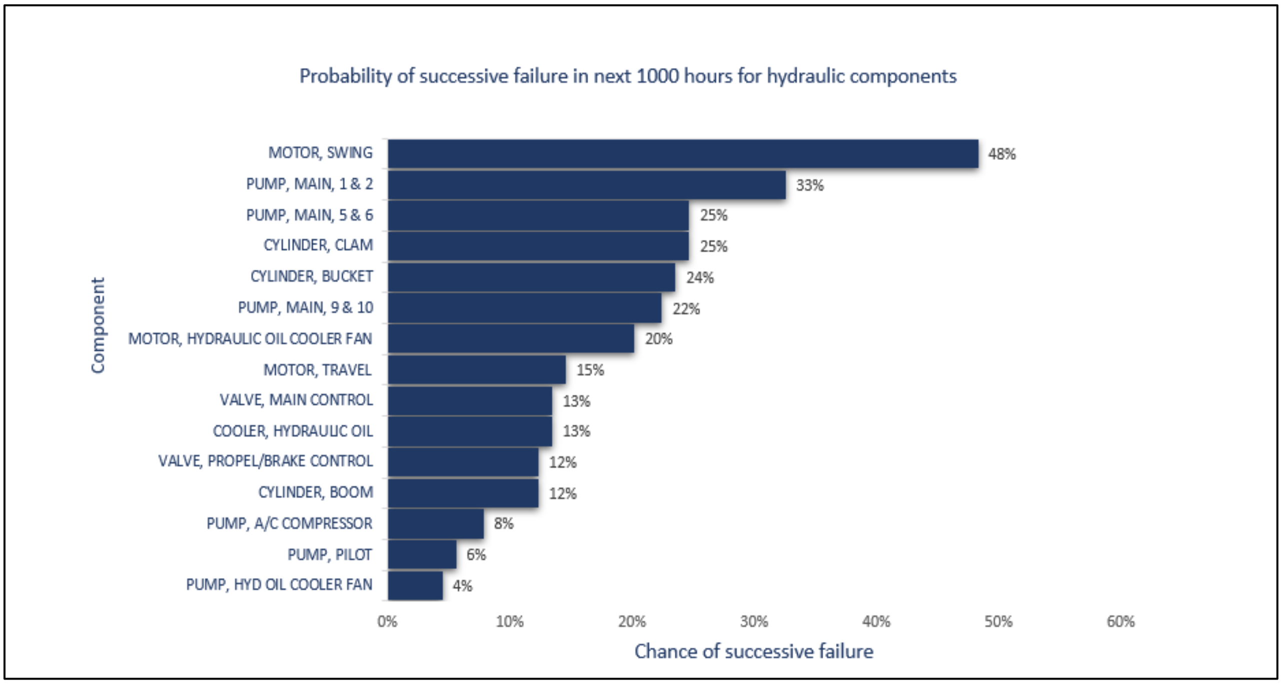


Figure 73. Probability of successive failures of different hydraulic components

The next phase in the data exploration involves investigating components that have higher chances of failure after a swing motor and main pump, 1&2 failure. The objective of this phase is to determine the likelihood of hydraulic component failures in 1000 hours of operation following the loss of swing motors and main pumps 1&2. Figure 74 depicts the likelihood of various component failures following a swing motor failure. After a swing motor fails, there is a 44% probability of another swing motor failing in the next 1000 hours of operation, followed by a 12% possibility of a bucket cylinder failure, a travel motor failure, and main pump, 9&10 failures. Figure 75 depicts the likelihood of distinct component failures following a main pump failure. Following the failure of main pump 1&2, there is a 24% chance of main pump 3&4 (grouped as main pump 1&2) failure, followed by a 21% possibility of a swing motor failure and a 17% chance of main pump 9&10. A bar graph represented in Figure 76 is used to identify if the failures in the next 1000 hours of a swing motor failure and a main pump failure are occurring randomly due to chance (i.e., they have achieved their expected life) or if they are influenced by the failure of swing motor and main pumps. The graph shows difference in failure frequency of components after a swing motor and a main pump failure, and the average standard deviation of the components (as indicated in section 3.4) is greater than at least 7000 hours for all hydraulic components indicating the failures are not random (i.e., a component is failing because it has achieved its full life) rather there is an influence of previous failures on the successive failures. The probability of successive failures of components in the next 1000 hours after a swing motor failure and main pumps is calculated using conditional probability theorem as

$$P(A_i \backslash B) = \frac{P\ (A_i\ \cap B)}{P(B)} \tag{53}$$

where component A_i is any one of the 15 hydraulic components considered and component B is either swing motors or main pumps, 1 &2.

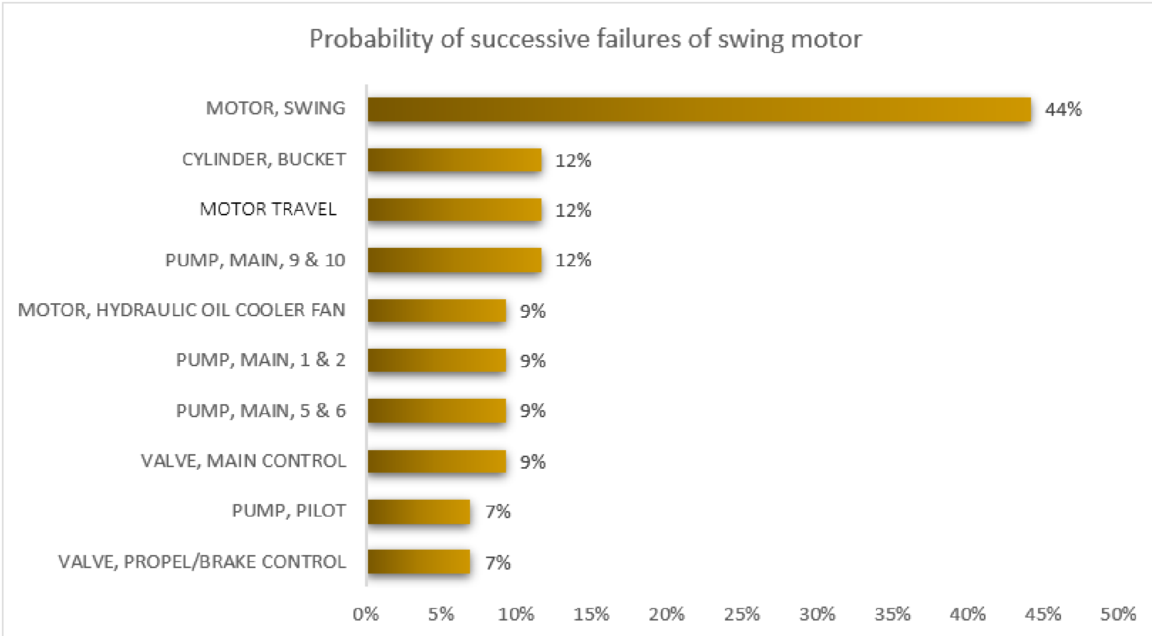


Figure 74. Probability of successive failures of swing motors

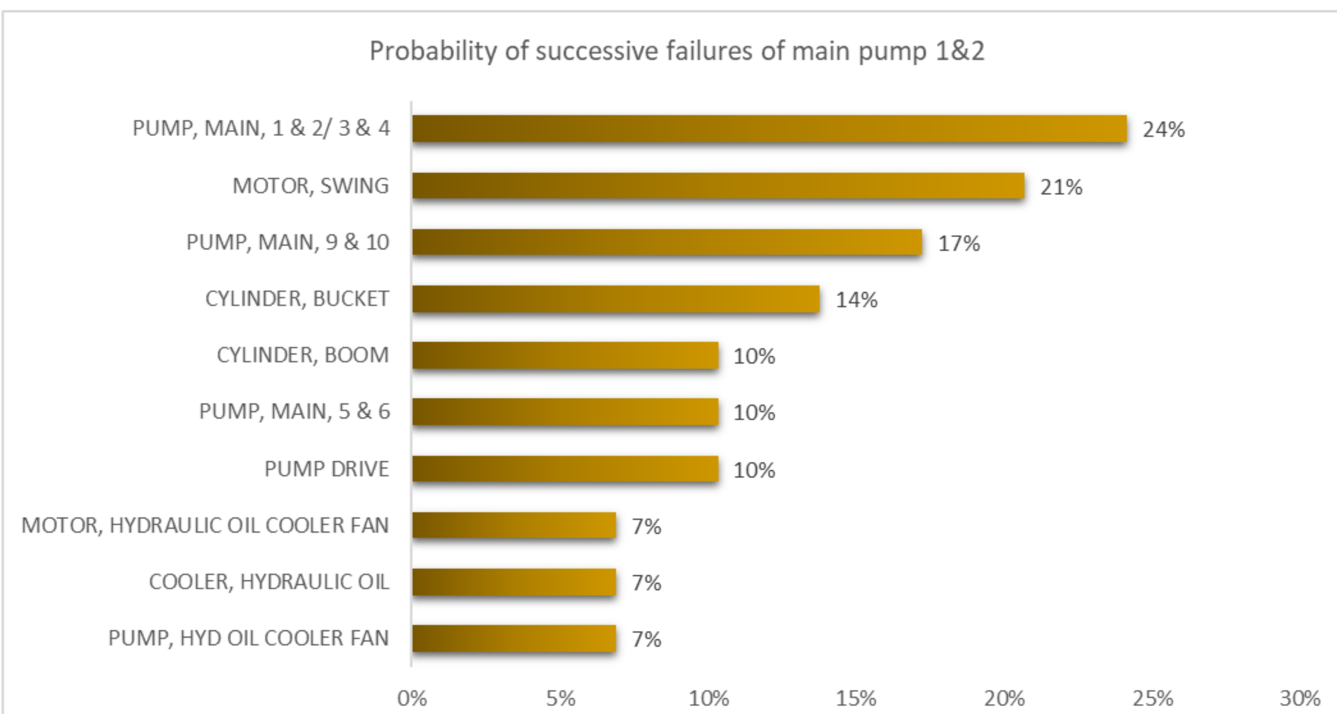


Figure 75. Probability of Successive Failures of Main pump 1&2

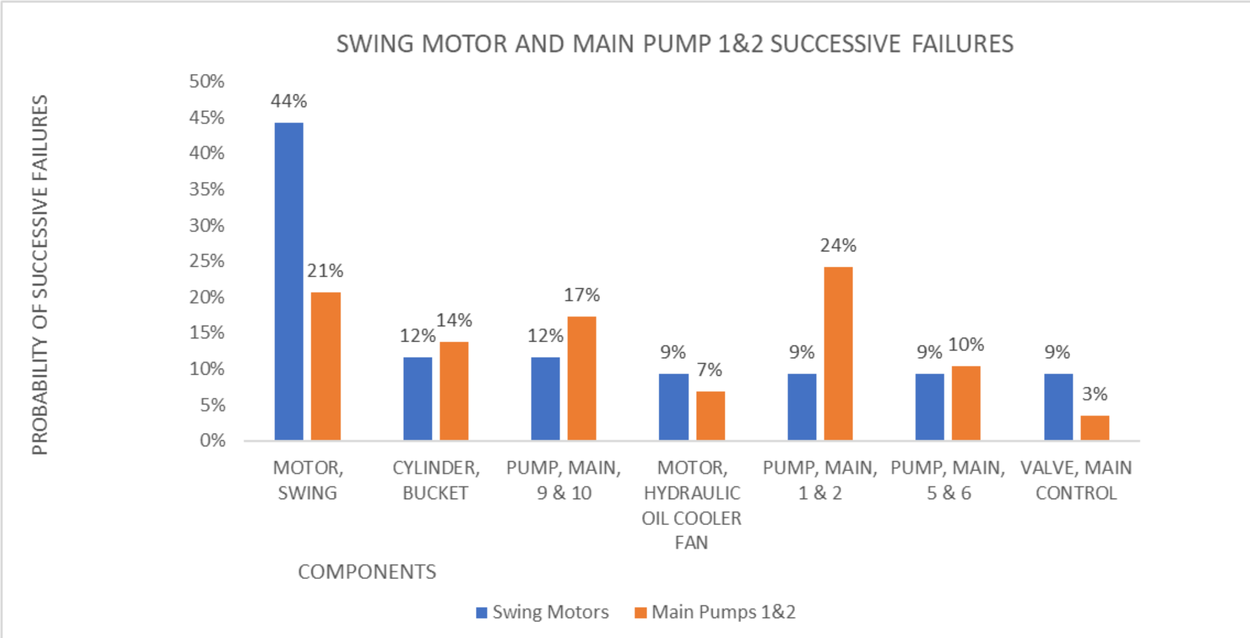


Figure 76. Comparison of Swing Motor and Main Pump 1&2 Successive failures

Figure 77 to Figure 82 indicate the probability density of failures in the next 1000 hours of a component failure for the given particle count in the system. The x- axis represents particle count in oil at the time of failure and y-axis represents the density of distribution. The preliminary analysis indicates that there is a change in probability of successive failures with the change in particle count in the oil. The chances of a successive failure in the next 1000 hours of operation after a component failure are higher when the 4µ particle count are in the ISO code range of 18 – 21 at reportable levels. The observations from the graphs infer that when there is a debris related failure or critical failures (with particle count in oil greater than 22) the hydraulic system might be cleaned, and the particles might be flushed out system which would not initiate any further failures. But when the particle counts are at reportable levels (18-22), there are higher chances of successive failure indicating the particles remain in the system even after the components are replaced and these particles flow in the oil eventually causing more failures. Blue bars represent that there are no successive failures in the next 1000 hours after a component failure and orange bars represent successive failures in the next 1000 hours of the component failure. The particle count at the time of failure is represented on the x axis and the probability density of the particle count is represented on y axis.

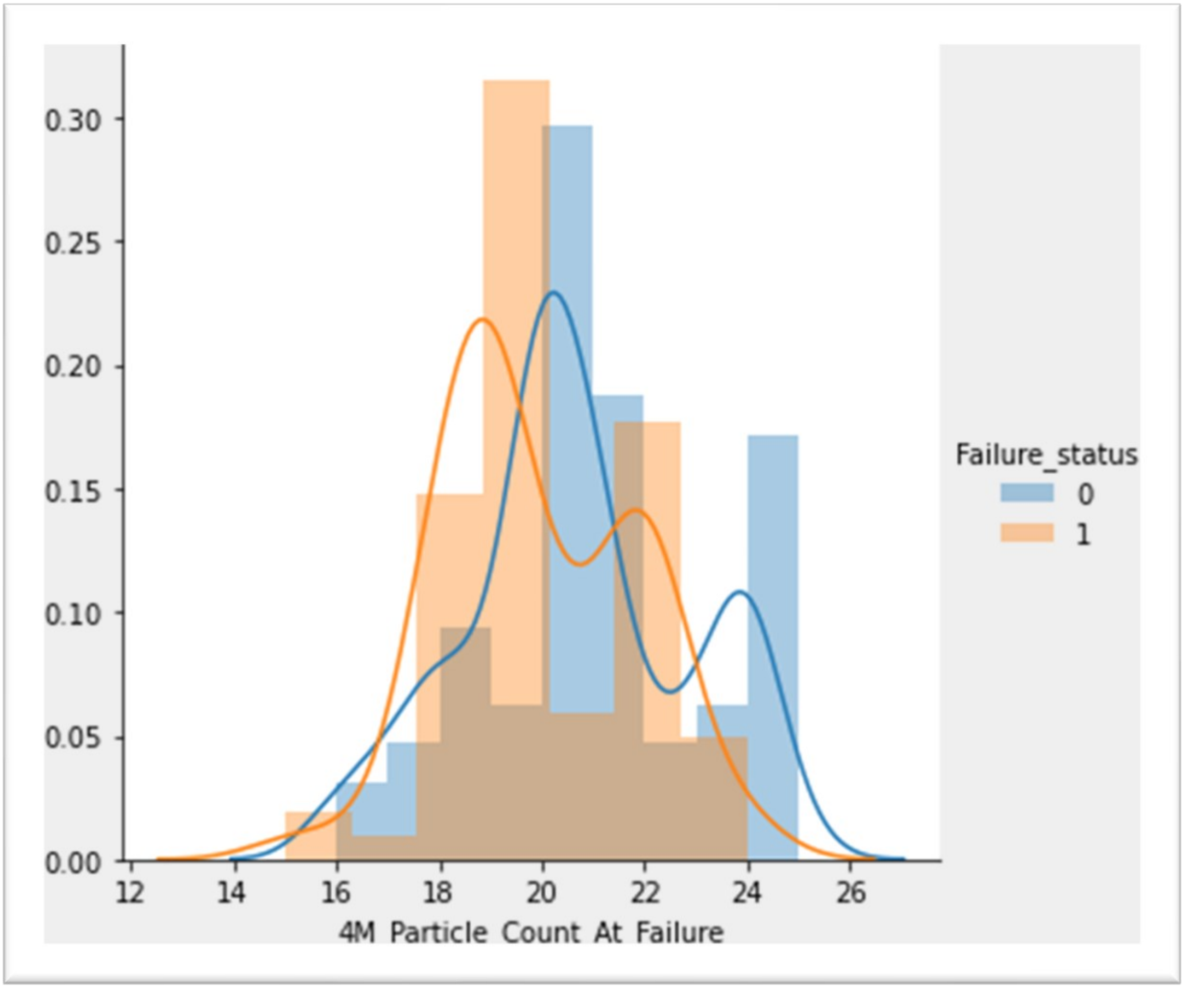


Figure 77. PDF Plot of Successive failures for variation in 4μ particle

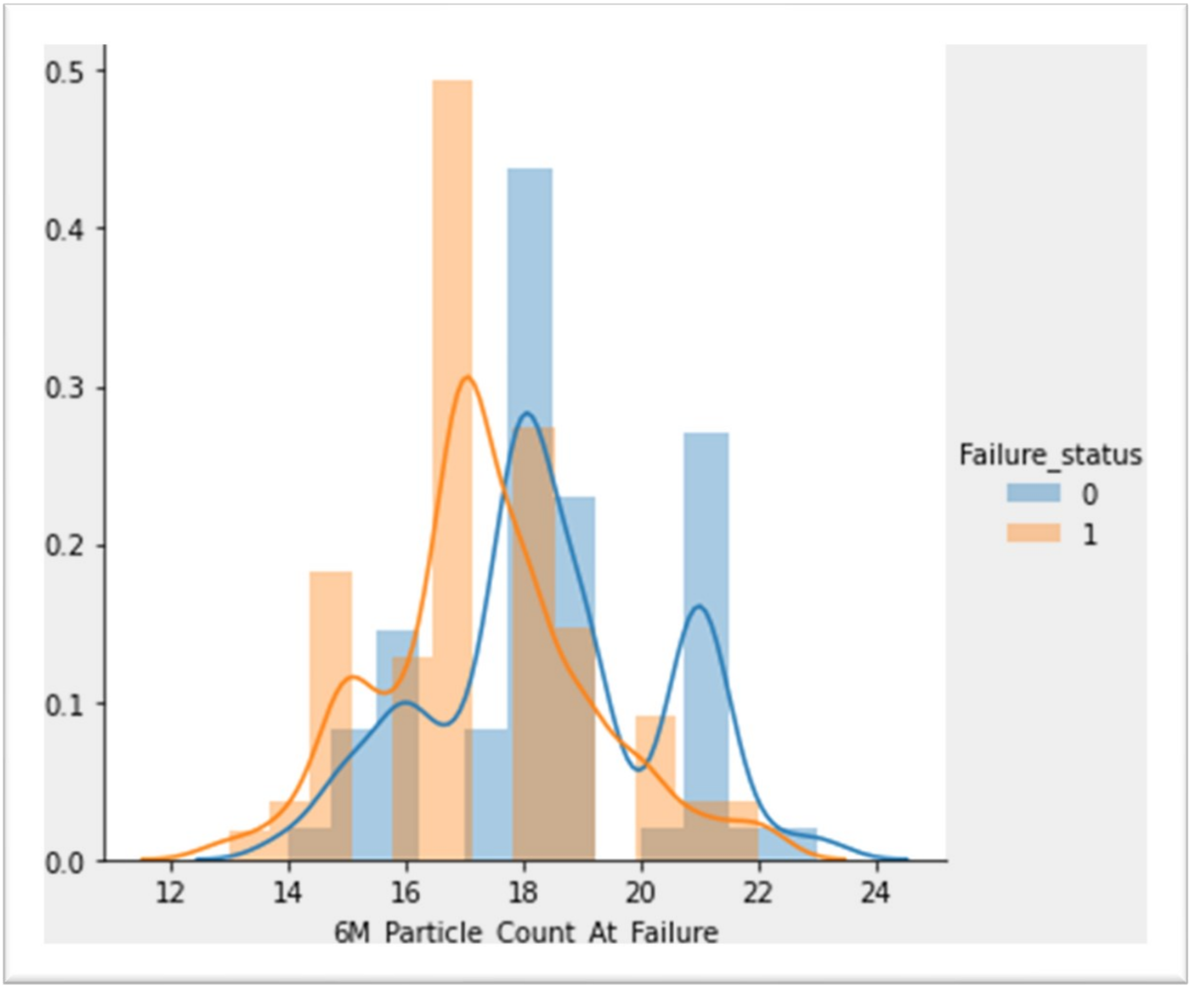


Figure 78. PDF Plot of Successive failures for variation in 6μ particle

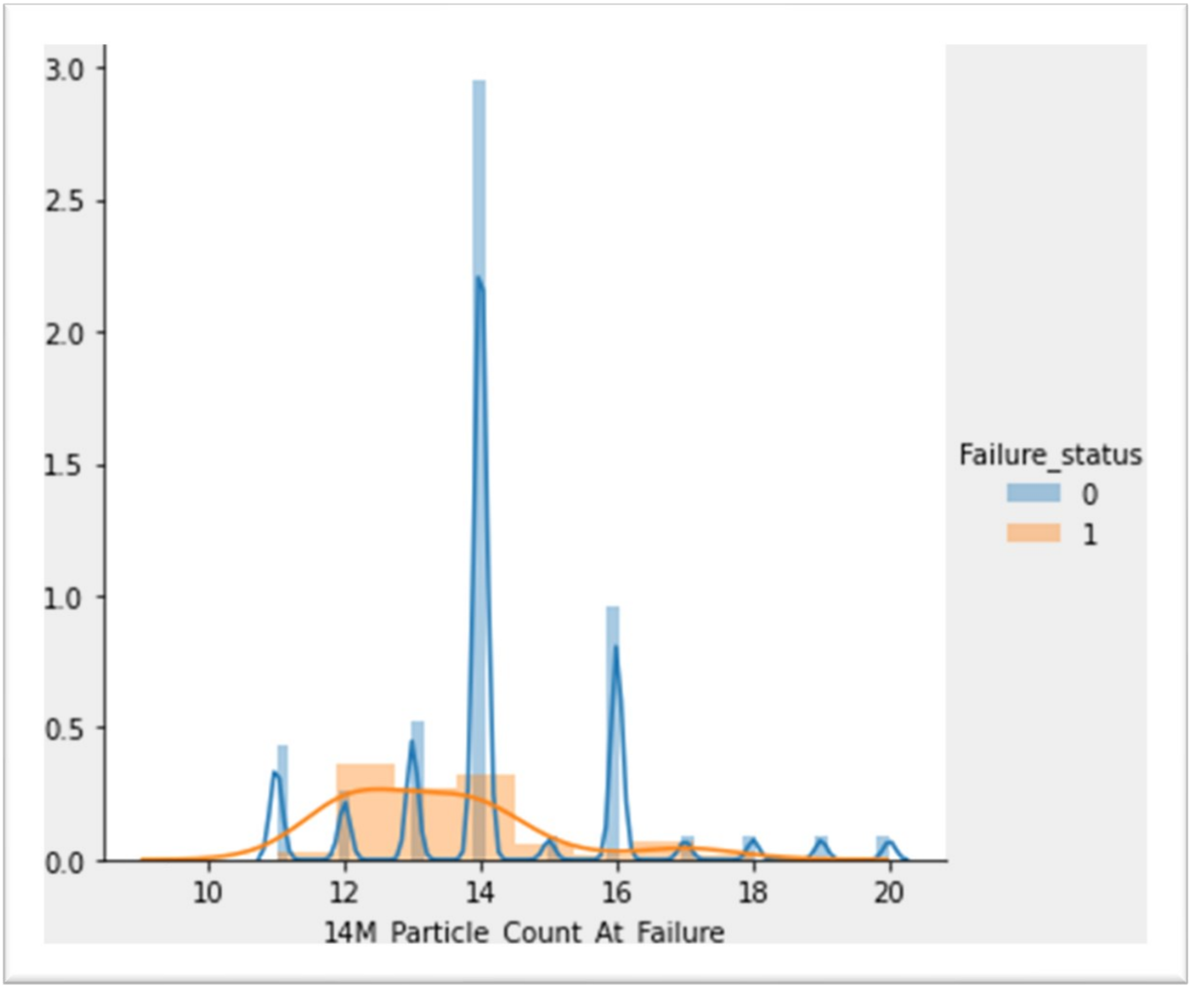


Figure 79.PDF Plot of Successive failures for variation in 14µ particle

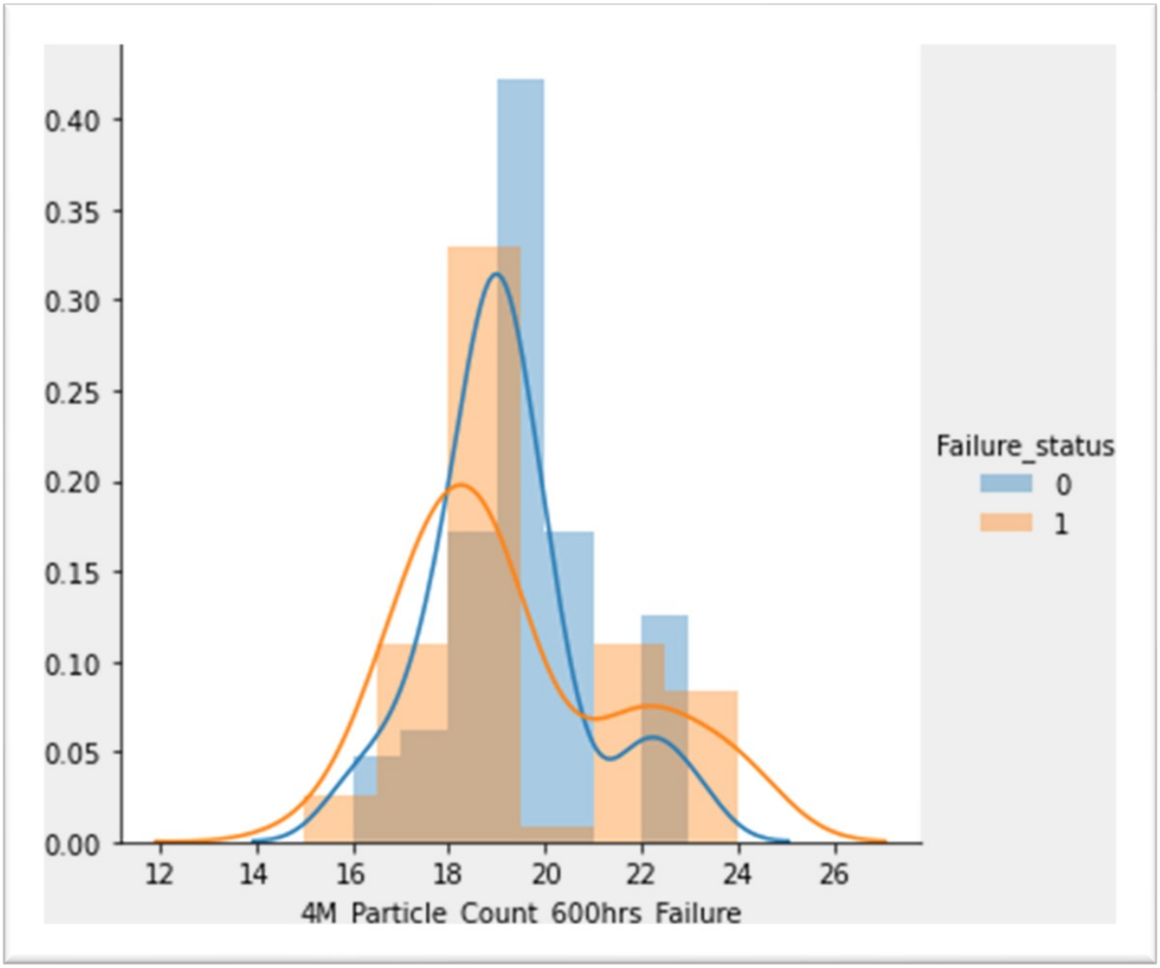


Figure 80.PDF Plot of Successive failures for variation in 4µ particle 600 hours prior to failure

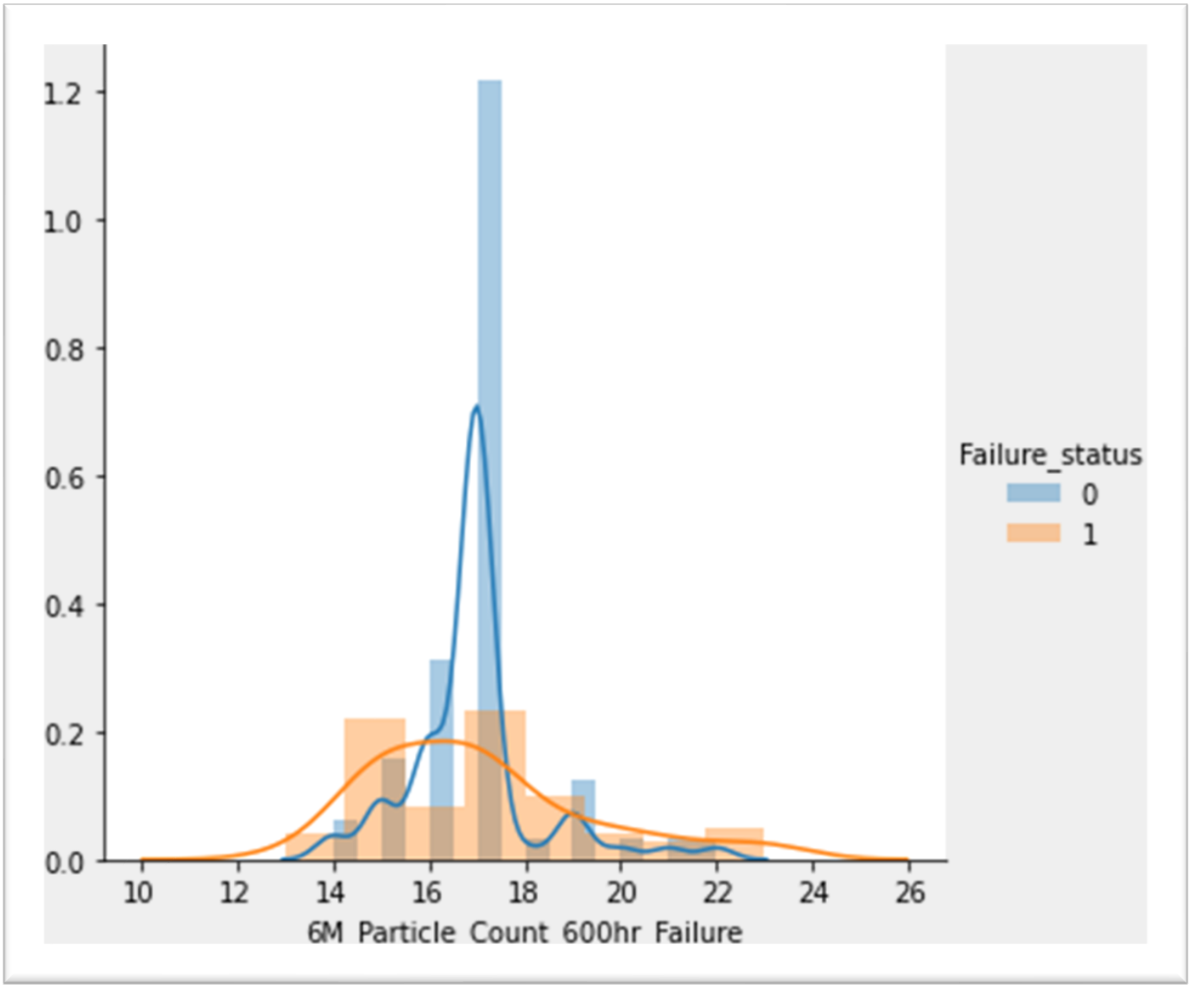


Figure 81.PDF Plot of Successive failures for variation in 6µ particle 600 hours prior to failure

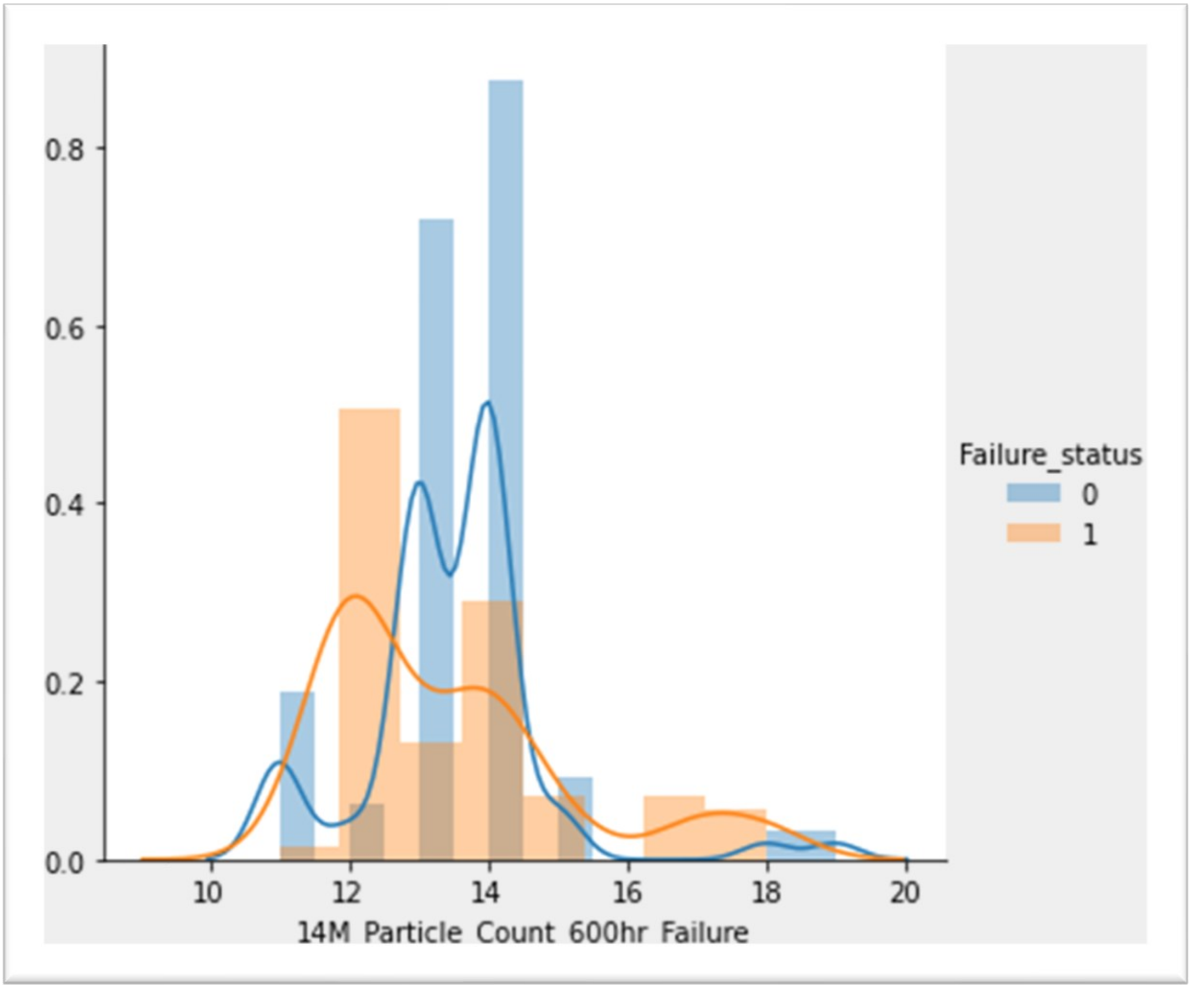


Figure 82. PDF Plot of Successive failures for variation in 14µ particle 600 hours prior to failure

There are different Naïve bayes algorithms used for classification analysis. This study uses gaussian Naïve Bayes classifier to analyze failure dataset. When working with continuous data, an assumption often considered is that the continuous values associated with each class are distributed according to a normal (or Gaussian) distribution. The likelihood of the features is assumed to be probability density function of normal distribution. Although, the Naïve bayes framework does not assume a distribution on the feature themselves if the data does not fit a near normal distribution, this may have a performance hit. Hence Q-Q plots were used to check if the data was normally or nearly normally distributed. All the features used in this study show a near normal distribution of data and hence gaussian Naïve Bayes model is used without affecting the model performance. Figure 83 to Figure 86 show Q-Q plots of features including 4μ, 14μ particle count at the time of failure and 600 hours prior to the failure.

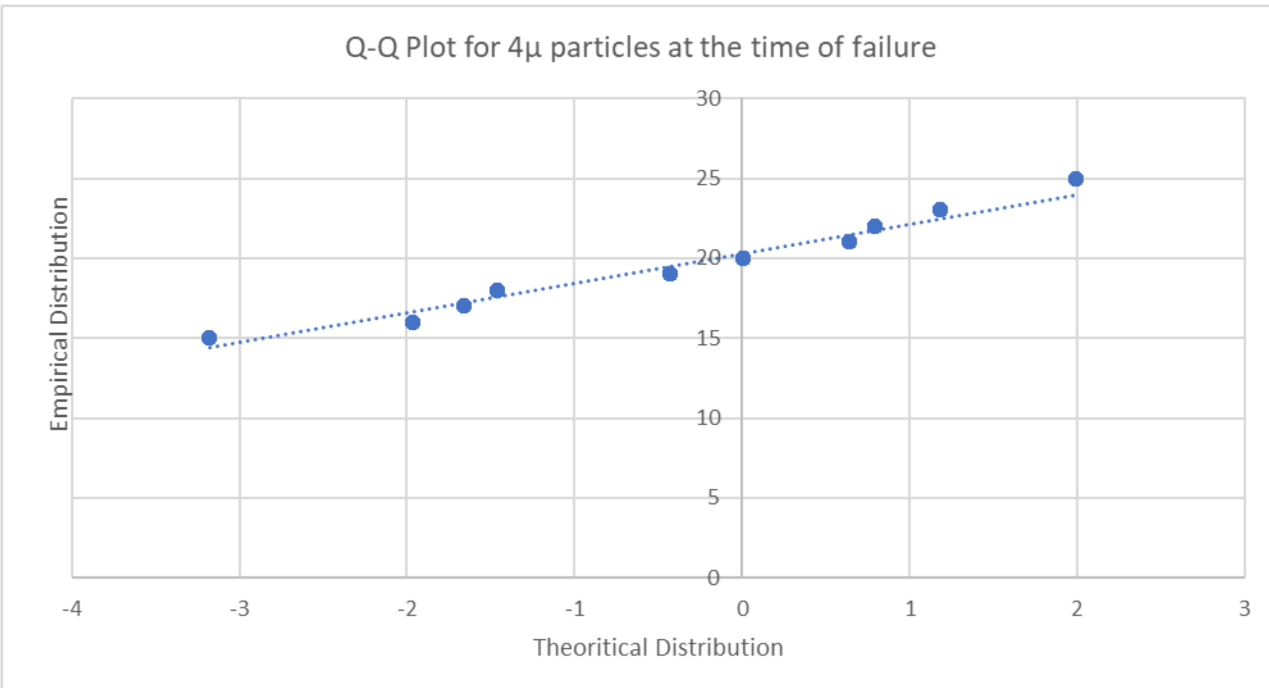


Figure 83.Q-Q plot of 4μ particle count at failure feature

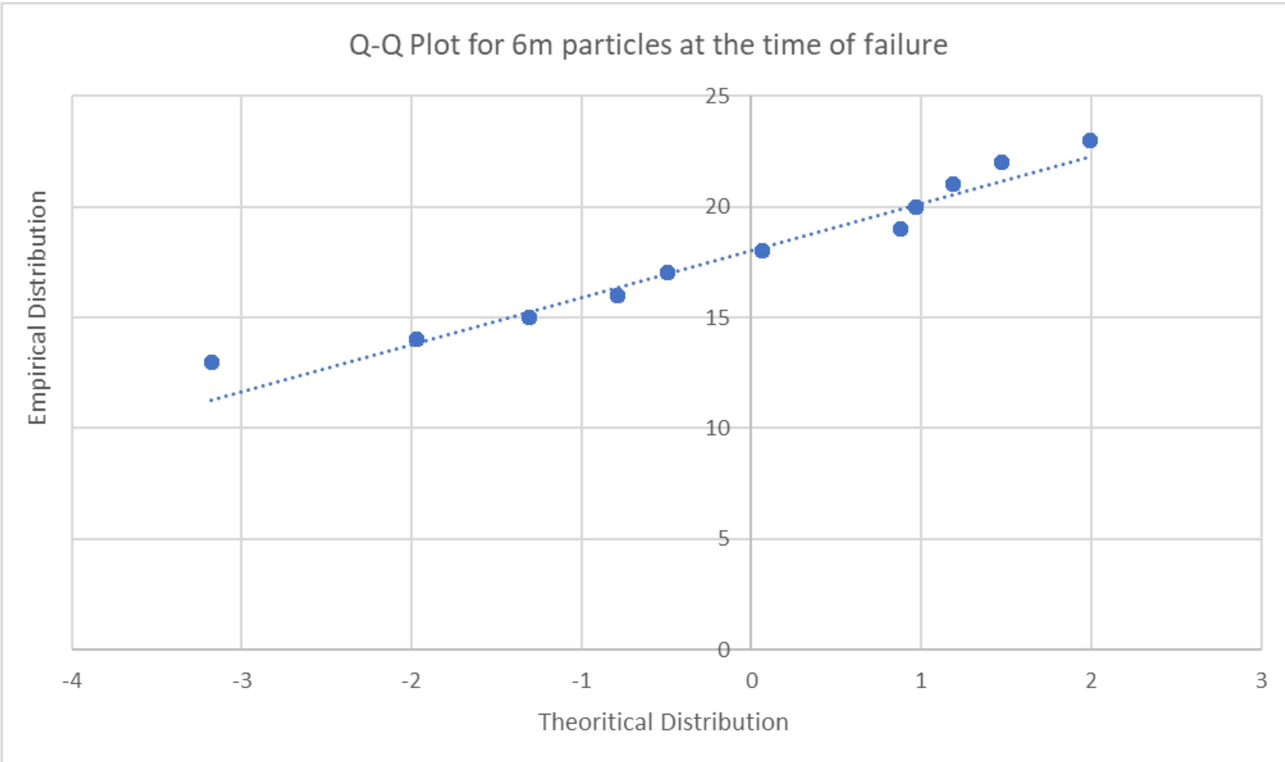


Figure 84.Q-Q plot of 6μ particle count at failure feature

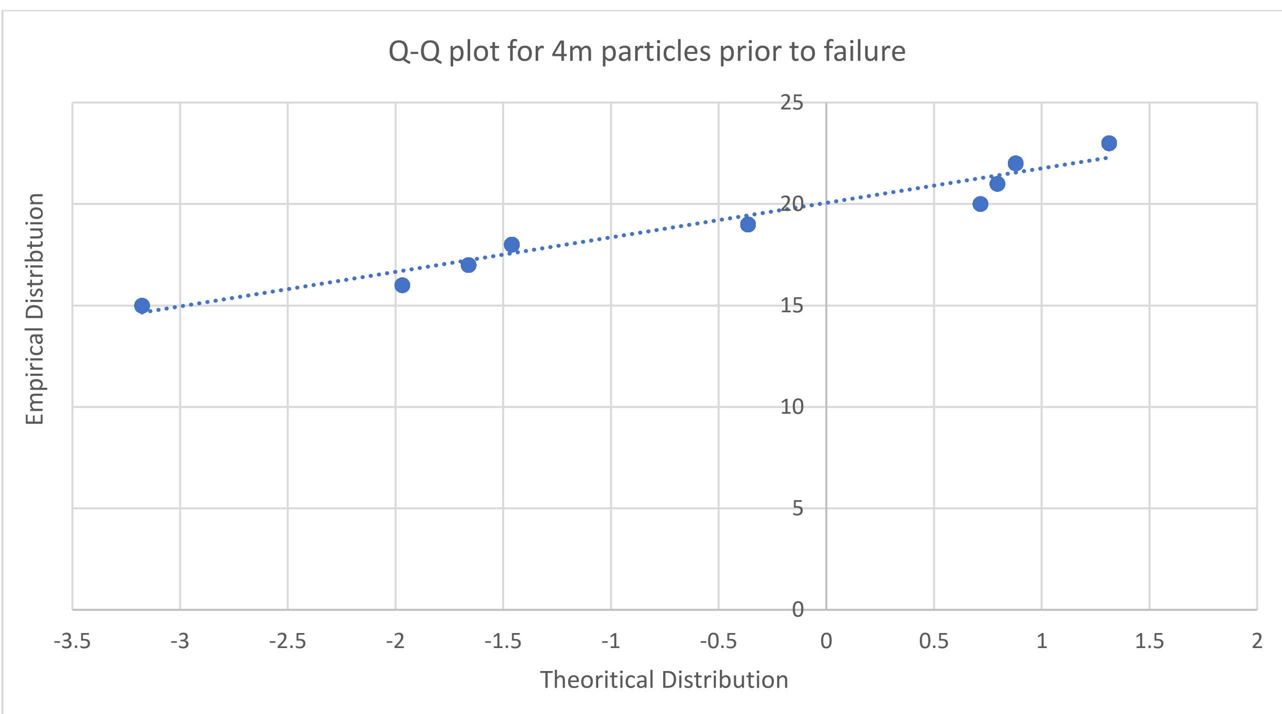


Figure 85.Q-Q plot of 4μ particle count 600 hours prior failure feature

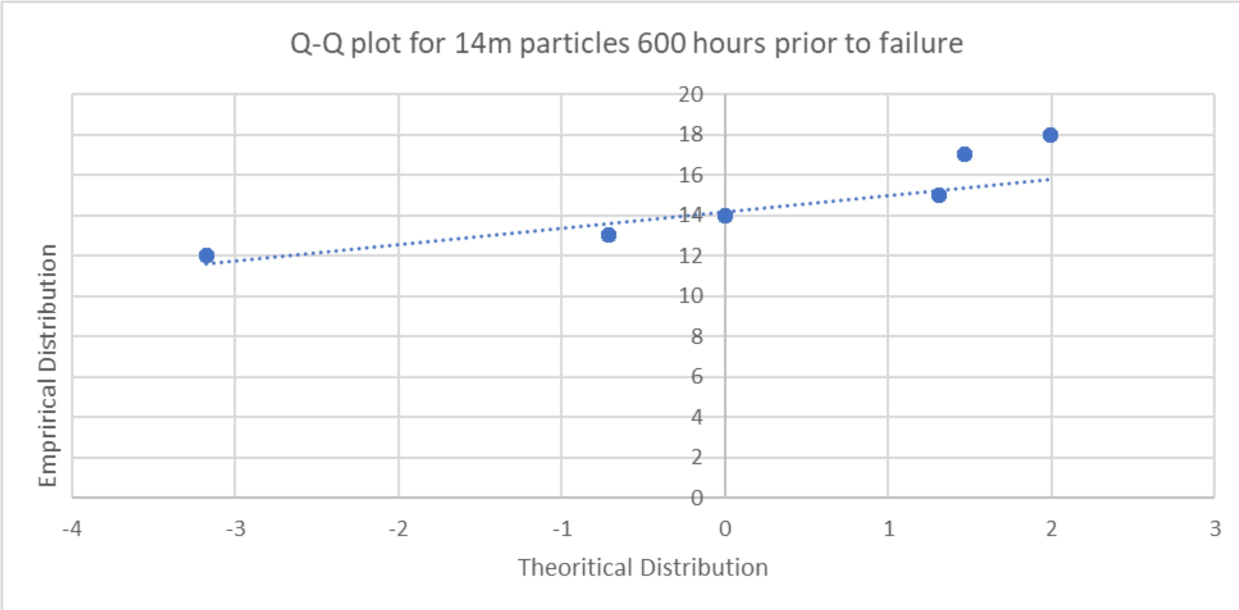


Figure 86.Q-Q plot of 14µ particle count 600 hours prior failure feature

For the Naïve Bayes model, the component condition for all 15 components is described at each failure. For every SMU hour recording of a hydraulic component failure, the information about the failure status of all the 15 components are tabulated. A component-failure column was created (flagging 1/0) to indicate if the component had failed at that SMU hour. One hot encoding was used to identify failures in the 1000 hours prior to the failure SMU hour. A sample of data tabulated for Naïve Bayes model is presented in Appendix B.

For a better understanding of the data preparation for the model input, an example is shown below on the tabulation of failure data in Table. 54.

SMU hours	Component-failed	Component-failed	A	B	C	Particle Count at the time of failure (4\6\10) μ	Particle Count 600 hours prior to the failure (4\6\10) μ
2000	A	1	0	0	0	14	13
2000	B	0	0	0	0	14	13
2000	C	0	0	0	0	14	13
2500	A	0	1	0	0	22	15
2500	B	1	1	0	0	22	15
2500	C	1	1	0	0	22	15
5000	A	1	0	0	0	19	16
5000	B	1	0	0	0	19	16
5000	C	1	0	0	0	19	16

Table 54. Illustration of data input for Naive Bayes Model

Suppose if there were three hydraulic components and the components failed at 2000, 2500 and 5000 SMU hours. The columns are tabulated in the following way,

- Each component status (whether it has failed or not) is recorded against each SMU hour.
- A, B, C columns are created to record if component A, component B, component C failed in the previous 1000 hours of each failure. As shown in the table, the first SMU failure hour is recorded at 2000 meter-hours.
  - ➢ Component A has failed at 2000 hours which is flagged 1 in the component-failed column. Since B and C have not failed at 2000 hours, they are flagged 0. Since the first hydraulic failure is at 2000 meter-hours, no components have failed previously. Hence A, B, C are marked 0.

- The next instance of failure is at 2500 meter-hours. Component B and C have failed at 2500 meter-hours which are flagged 1 in the component-failure column. Since A has not failed it is marked 0. Since A had failed in less than 1000 hours, Component A is marked 1 in column A and the others are flagged 0.
- In scenario 3, all the 3 components have failed at 5000 SMU hours and are flagged 1 in the component-failure column and since there was no failure in previous 1000 meter-hours, all columns A, B, C are flagged 0.

- Particle count (4\6\10) μ at the time of failure and 600 hours prior to the failure are recorded for each SMU failure hour.

Table. 55 shows the correlation coefficients determined by performing a Pearson's correlation test because of its ability to measure the degree of linear relationships, and the following features have a high correlation coefficient greater than 0.75

- 6μ particle count at the time of failure and 4μ particle count at the time of failure are highly correlated with correlation coefficient of 0.898. Hence 6μ particle count at the time of failure was eliminated.
- 6μ particle count 600 hours prior to the time of failure and 4μ particle count 600 hours prior to the time of failure are highly correlated with correlation coefficient of 0.919. Hence 6μ particle count 600 hours prior to failure was eliminated.

correlation Matrix	4M Particle Count at Failure	6M Particle Count at Failure	10M Particle Count at Failure	4M Particle Count 600hrs Failure	6M Particle Count 600hr Failure	10M Particle Count 600hr Failure
4M Particle Count at Failure	1					
6M Particle Count at Failure	0.898	1				
10M Particle Count at Failure	0.511	0.713	1			
4M Particle Count 600hrs Failure	0.670	0.519	0.198	1		
6M Particle Count 600hr Failure	0.586	0.512	0.293	0.919	1.000	
10M Particle Count 600hr Failure	0.315	0.358	0.526	0.592	0.784	1

Table 55.Correlation metrics of Continuous Variables

KNN algorithm was modeled using train dataset as described in the initial sections. Standard libraries in python were used for the analysis of KNN model. The data split of dependent variable (1 & 0) in the dataset of KNN model is shown in Figure 86. Since the dataset is nearly balanced, a 70:30 train-test split was used to divide the data into training and test sets respectively. The continuous variables were standardized using StandardScaler library after eliminating variables discussed in the previous sections. The classification accuracy with K=1 was 0.7. A grid search method was used to identify best value of K from a value range between 1 to 45 with a 10-fold cross validation for each K. According to the grid search method, the best value of K was estimated as 13.

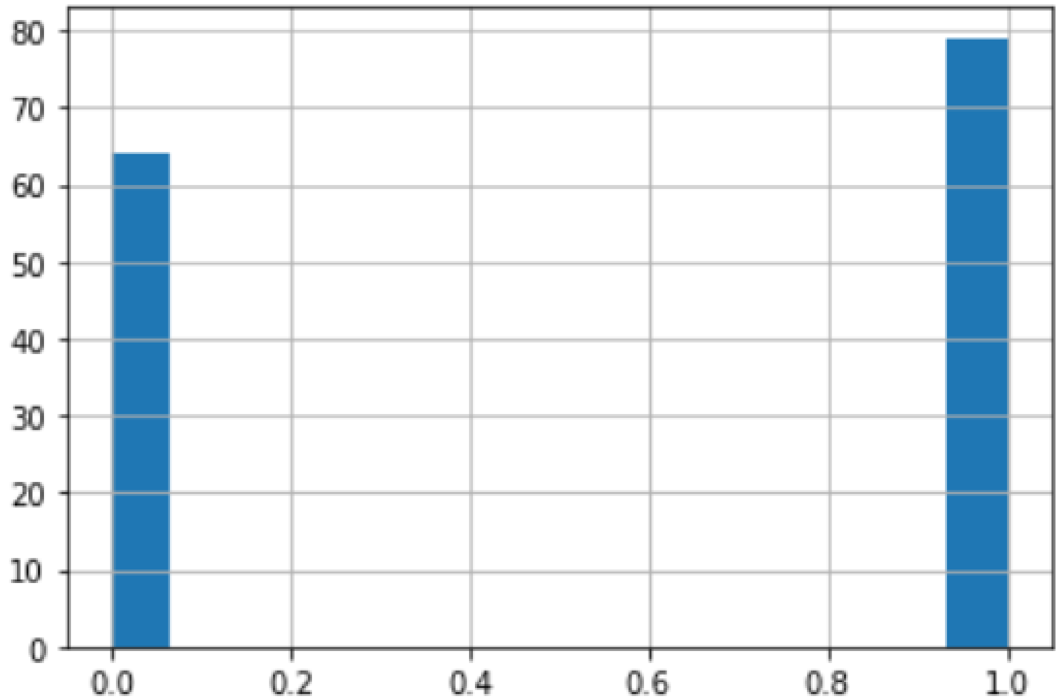


Figure 87.Category split of dependent variable

Hence the KNN model with the value of K =13 was used on the training set. The accuracy of the model improved to 0.8 with good recall and precision values. Figure 88 shows the classification report of KNN analysis with K=13.

	precision	recall	f1-score	support
0	0.79	0.65	0.71	17
1	0.79	0.88	0.84	26
accuracy			0.79	43
macro avg	0.79	0.77	0.77	43
weighted avg	0.79	0.79	0.79	43

Figure 88. Classification results of KNN model with k= 13

Naïve Bayes algorithm is used to determine the probability of failure for each of the 15 components in the next 1000 hours given a component failure. Dataset for Naïve Bayes was prepared as described in section 4.2 with component-failure flagged 1/0 representing if that component failed at that SMU failure hour and using one-hot encoding, columns were created for each component to indicate if there was a failure in the previous 1000 hours for the given SMU failure hour. The particle count at the time of failure and 600 hours prior to failure are also used in the model. Analysis results of probability of successive failure for 15 components using Naïve Bayes resulted in very low model accuracy of 0.4. Hence, only comeback failures of swing motor failures were considered for failure analysis. Since the probability of successive failure in next 1000 hours is highest for swing motors, a Naïve Bayes model was built considering only the swing motor failures. The model is used to predict probability of component failures after a swing motor failure. The failure status of top 5 component failures (from Figure .73) that had highest comeback rate were also considered along with swing failure. The data split of dependent variable (0/1) for the Naïve Bayes algorithm is shown in Figure 89. Since the dataset is imbalanced, a stratified train-test split is used in the analysis to divide the data into train and test group. A 3-fold cross validation method was used to analyze model results.

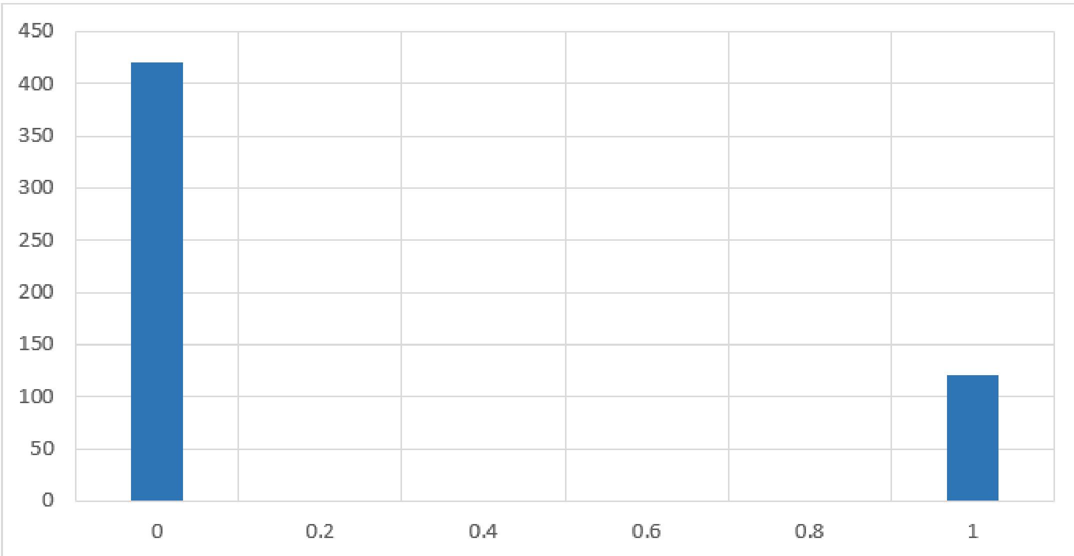


Figure 89. Binary split of dependent variable

Three different models were run with different training dataset and model performance was similar in the three models. Below is the accuracy and confusion matrix for the three Naïve bayes models that were used for predicting probability of component failure given a swing motor failure. Table 56 through Table. 59 indicate performance metrics of the three Naïve Bayes models used in predicting successive failures in the next 1000 hours of operation after a component failure.

	MODEL 1	MODEL 2	MODEL 3
ACCURACY	0.88	0.89	0.9

Table 56. Accuracy scores of Naive Bayes Models

	0	1
0	92%	8%
1	55%	45%

Table 57.Confusion Matrix for Naive Bayes Model1

	0	1
0	96%	4%
1	60%	40%

Table 58. Confusion Matrix for Naive bayes Model2

	0	1
0	93%	7%
1	60%	40%

Table 59. Confusion Matrix of Naive Bayes Model3

4.7 Summary and Conclusions

The main objective of this study was to identify wear/contamination-related failures that were influenced by the previous failures. Using the historical data, the aim was to predict the probability of successive failures given a component has failed. This can be a crucial step in the maintenance activities. If the failures can be identified in advance, then the components that are highly likely to fail after a particular component failure can be replaced along with the initially failed component to prevent on additional labor costs, downtime hours and other maintenance and repair costs. Machine learning algorithms help in finding failure patterns using historical failure data to predict future failures. Two machine learning models were used in this study for successive failure analysis.

k-NN model was used to predict if the possibility of a successive failure after a component has failed. The model accuracy is around 0.8 which means that the model can accurately predict 80% of the times if there is a possibility of future failure. The model has more than half of the times correctly predicted the possibility of a successive failure within 1000 hours of operation and has 65% of the times has correctly predicted that there is not going to be a successive failure when there was no failure in the next 1000 hours.

Naïve Bayes algorithm was to predict the probability of failure of all 15 components after a component failure. 3 folds of training samples were used to test the algorithm efficiency. The average model accuracy is around 0.89 which means the algorithm has correctly predicted the outcomes 89% of the times. The confusion matrix helps in deciding to use the algorithm in maintenance practices. The model has correctly predicted more than 92% of the time that a component is not going to fail in the next 1000 hours. The model has correctly predicted the chances of successive failure 40% of the time. As the false positive rate of the model is less than 3%, this is suitable to be adopted in equipment replacement plans. The false-positive rate is the number of times the model has predicted a failure when there was no failure. As the hydraulic component parts are expensive, a higher false-positive rate would add to a loss on purchasing component parts. Thus, a lower false-positive rate ensures that unnecessary costs are not spent on buying new components when the component still has not achieved the expected life. The model has identified successive failures 40% of the time. This means, without a prior knowledge on successive failures, out of every 10 components failures after a swing motor failure, 4 of them can be mitigated using prediction model.

In order to be more accurate about a successive component failure after a given failure, the results of the two algorithms together can be used. If the output of both the KNN and Naïve Bayes models are [1,1] there is a highly likely chance of a failure, if the output of the models are [1,0] or [0,1], a combined decision can be taken based on the component life and the component type and oil conditions. If the output of the models are [0,0], it means there is a highly likely chance that there would not be a failure in the next 1000 hours of operations.

5. CONCLUSIONS

This chapter presents the summary and conclusions drawn from the research work. The chapter also discusses the significance and contribution of this research and suggests recommendations for future work.

5.1 Summary of Research

Large hydraulic shovel – truck system forms the backbone of surface mining industry. Hydraulic systems are often subjected to wear and contamination related failures. The component failures generated by wear particles also lead to debris accumulation in the oil leading to the failure of downstream components. Eventually the whole system breaks down. Hydraulic system failures are very costly and are associated with longer downtimes. Hence companies are trying to mitigate hydraulic failures by introducing new filters to contain contaminant particles, practice CBM and use data driven techniques to prevent these failures. With the addition of industry 4.0 to the mining industry, the aim is always to extend component lives of equipment, reduce repair bills, get back to work faster and reduce follow on failures. This research mainly focused on using data driven techniques to identify the effectiveness of hydraulic Mag-filters and predict successive failures to improve reliability and availability of the shovels.

Four distinct sections of a thorough literature review were reviewed and presented. The overview of the hydraulic system, the different types of hydraulic filters, hydraulic failures, and their criticality levels are discussed in the first section of the literature review. The cleanliness of hydraulic oil, its effects on the system, and the oil analysis method for condition-based monitoring are reviewed in the second section of the literature review. In the third section of the literature study, various statistical and data analytics methods for failure analysis are presented. In the fourth section of the literature review, various machine learning approaches used in failure analysis and predictive maintenance, primarily for mining equipment are presented. The literature review helped in understanding in detail about the hydraulic system of giant shovels, the design of the hydraulic system of shovels, the importance of hydraulic system filters, different failure modes, and failure causes in the hydraulic system, condition monitoring of the hydraulic system based on oil samples, and crucial factors affecting oil contamination and hydraulic system wear. The literature review section also summarized various statistical techniques that can be applied in different analyses and the major contributions of previous researchers which helped in better understanding and implementation of various machine learning techniques for failure analysis and predictions in this study. Researchers have implemented numerous machine learning models and achieved satisfactory results for fault diagnostics and predictive maintenance analysis in mining and other industries.

Although numerous statistical techniques are applied in the analysis of failures of mining equipment, there is no such framework described to quantify the pre and post effects of implementing a solution to prevent equipment failures. The first part of this study aims to apply different statistical techniques to quantify the effects of magnetic filters installed in the shovel hydraulic system.

Several machine learning techniques are already used in the analysis of mine equipment failures. Despite the popularity and application of machine learning techniques, no previous work is found related to the prediction of the probability of successive failure of the mine equipment components that are influenced by the previous failure. The second part of this study is an attempt in implementing machine learning models to find the probability of successive failures of different hydraulic components in the next 1000 hours of operation after a component has failed.

6.2 Research Conclusions

Through the first part of this research, an integrated framework has been developed to analyze the pre and post effects of implementing a solution to prevent mine equipment failures using different statistical techniques. The study demonstrates the use of different statistical techniques like descriptive and inferential statistical analyses, ANOVA test, probability distributions, IID tests, Central Limit Theorem, and goodness-of-fit tests. Different key performance indicators are used in the pre-post analysis of implementing a solution to mitigate mine equipment failure. The study throws light on how a pre-post analysis can be studied from different perspectives using different KPIs to assess before and after performance of the mine equipment. To conclude on the overall impacts of magnetic filters on the hydraulic system, a weighted variable method was used to identify the effects of Mag-filters on different components across different KPIs used in this study. Hence, the research work highlights the significance of filter impacts on different KPIs and to what extent the performance of hydraulic components has changed across these KPIs post Mag-filter installation. The study uses different parameters like component life, reliability, oil particle count, comeback failure rate and failure costs to study the impacts of Mag-filters and how these parameters are affected post installation of Mag-filters. Most hydraulic component failures have seen a large reduction in comeback failures, while component life, reliability, and failure costs have improved significantly for a few system components. Different weights are assigned to the parameters based on the variability explained by each parameter in order to assess overall Mag impacts. A total weighted score is calculated based on the factor change in performance post Mag installation for different hydraulic components considered across different parameters. The total score indicates a positive change concluding that the Mag-filters are beneficial on the hydraulic system. However, as noted in the analysis, not all the hydraulic components across the different KPIs considered show an improvement in performance post Mag installation. As a result, it can be stated that mag filters are advantageous in extending the life of hydraulic systems, although they are not the only element affecting the system performance. Other factors like the location of the equipment in the mine, working conditions, and rock hardness also influence hydraulic system performance.

Through the second part of the analysis, an integrated methodology has been established using machine leaning techniques to predict the probability of successive failures in the next 1000 hours of operation after a component failure. Supervised classification models were explored and a framework to assess failure probability using KNN and Naïve Bayes algorithms was presented using historical failure data. This study demonstrates the use of different machine learning models and probability theory to assess likelihood of successive failures. KNN model was used to predict the chance of occurrence of a successive failure in the next 1000 hours of operation of a component failure. The accuracy of prediction model was 0.8 indicating that 80% of the times the model can correctly predict if there is a successive failure occurrence in the next 1000 hours of operation after a component failure. Naïve Bayes models were used to predict the chance of a component failure in the next 1000 hours of operation after a swing motor failure that had the highest comeback failure rate. The average accuracy of prediction was around 0.89. An ensemble of the two models can be used to make a decision on the output. This can be used in predictive maintenance of shovels to schedule component replacements that would reduce the run to failures and increase production.

In summary, the first part of research developed an integrated framework for pre-post analysis of implementing a solution to mitigate mine equipment failures. The second part of research presented machine learning classification algorithms to predict successive failures in the next 1000 hours of operation after a component failure. The results presented in Chapter 3 and Chapter 4 show that several statistical and machine learning techniques can be employed for solving hydraulic system problems of giant shovels.

6.3 Research Contribution

The main contribution of this research is the development of a framework for pre-post analysis of the effects of implementation of a solution to prevent mine equipment failures and develop and integrate a methodology to assess the likelihood of component failures of the hydraulic system that are impacted from previous failures. This provides a better understanding of the applicability of various statistical and machine learning techniques and facilitate the diagnosis of hydraulic failures using historical failure data and oil analysis samples. The integrated framework and the KPIs developed and used in the first part of the thesis that presents the analysis of pre-post effects of magnetic filters can be generalized and used as a hypothesis framework solution to test the pre-post effects of implementing a solution to prevent equipment failures. The methodology developed in the second part of the thesis to predict successive failure probability prediction in the hydraulic system using machine learning algorithms can be used to diagnose and predict equipment failures. The second part of the research demonstrates several important pre-processing and data preparation steps. The research also demonstrates the use and application of different performance evaluation metrics and their importance in solving different machine learning problems.

6.3 Challenges and Limitations

Various statistical and machine learning algorithms are used in this study to analyze hydraulic system failures. The algorithms generate effective results on the mag-filter analysis and probability of failure prediction. However, no initial framework was available to analyze the effects of implementation of solution to mitigate failures. Hence, one of the challenges of the project was to understand and design a framework and cautiously select KPIs that would best describe the effects of Mag-filters. The key aspect of all statistical learning and machine learning algorithms is the amount and quality of data. Missing values and wrongly recorded failure data affect the results and need initial treatments. As the volume of data increases, the complexities increase. One of the main challenges of this research was the difficulty in collecting correct failure data from the data base files and dealing with missing values and outliers. The first section of the project draws the conclusion regarding the efficiency of Mag-filters, however, the cause of effects of change in performance could not be further investigated due to the lack of data. Additionally, the database's missing IFRs and CRs led to the information loss, making it difficult to conduct a thorough investigation of the cause and nature of hydraulic failures. Understanding and extracting information about the cause and type of the failures for which IFR and CR data were available was a time-consuming task since the information had to be manually extracted. Based on the results presented in this research and the key challenges listed in this section, the choice of data along with suitable algorithm will have a significant impact on the outcome of future work based on the framework suggested in this research.

6.4 Recommendations for future work

The most important and integral part of any statistical and data analysis is the quality and quantity of information obtained for the analysis. The higher the quality of data, the closer the analysis results are to the actual observances. One of the biggest challenges in the mining industry is the collection and collaboration of quality data. In this research, one of the most time-consuming tasks was to access and collect data from different data sources and manually inspect work order reports to study the hydraulic failures. With the advancement of big data analytics tools, the foremost recommendation is to automate the process of data collection and integrate and store all data sources of a company’s data using big data tools so that the data used for analysis is more reliable and error free. The first part of the study is to analyze the effectiveness of Mag-filters. A few questions regarding the cause of the hydraulic system's performance change after the installation of Mag-filters were challenging to answer due to a lack of knowledge of other external factors influencing hydraulic system performance. The extent of each factor's influence on performance can be determined if the information on external factors, such as the location and working environment of the shovels, can be collected. The second part of this research aims to predict successive failure probability of components in the next 1000 hours of operation given a hydraulic component has failed. Only oil conditions and component failures in the previous 1000 hours of operation before the hydraulic failure are considered in the machine learning models employed in this study. The model can be more robust and generalized to hydraulic wear failures of large mining shovels if more information about external conditions, the condition of the mines, and the cause and type of failure for all shovels can be gathered. Machine learning models are used in this analysis to determine the likelihood of subsequent failures. It is not possible to train more powerful models, such as deep learning and neural network techniques, using the amount of failure data that has been gathered. More complex algorithms can be applied to the training dataset for better predictions and enhanced accuracy of the prediction model, and the models can be generalized for hydraulic failure analysis if sufficient data on hydraulic failure is gathered.

BIBLIOGRAPHY

(2013, March). Retrieved from Weibull: https://www.weibull.com/hotwire/issue145/tooltips145.htm

A, J. C., Correa, J., & Guzman, A. A. (2020). *Mechanical Vibrations and Condition Monitoring.*

Abdelhadi, A. (2017). Heuristic Approach to schedule preventive maintenance operations using K-Means methodology. *International Journal of Mechanical Engineering and Technology, Volume 8, Issue 10*, 300-307.

Abdi, A. (2016). *Three types of Machine Learning Algorithms.*

Ahmadi, S., Hajihassani, M., Moosazadeh, S., & Moomivand, H. (2020). An Overview of the Reliability Analysis Methods of Tunneling Equipment. *CrossMark, Volume 14*, 218-219.

Akersten. (1987). The double TTT-plot, a tool for the study of non-constant failure intensities. *Proceedings of Reliability '87*, (pp. 2B/3/1-2B/3/8.). Birmingham, UK.

Akersten, P. A. (1986). The bivariate TTT-plot--a tool for the study of non-constant failure intensities. *Proceedings of Society of Reliability Engineers.* Otaniemi, Finland.

Alguindigue, I. E., Loskiewicz-Buczak, A., & Uhrig, R. (1993). Monitoring and diagnosis of rolling element bearings using artificial neural networks. *IEEE Transactions on Industrial Electronics*, 209 - 217.

Alla, H. R., Hall, R., & Apel, D. (2020). Performance evaluation of near real-time condition monitoring in haul trucks. *International Journal of Mining Science and Technology 30*.

Almasi, A. (2014, April 21). *Oil analysis methods and lubrication monitoring.* Retrieved from https://www.plantservices.com/: https://www.plantservices.com/articles/2014/oil-analysis-methods-lubrication-monitoring/

Amy, H. (2020, October 15). *WHAT IS EQUIPMENT RELIABILITY AND HOW DO YOU IMPROVE IT?* Retrieved from https://nonstopreliability.com: https://nonstopreliability.com/equipment-reliability/

An, L., & Sepehri, N. (2005). HYDRAULIC ACTUATOR LEAKAGE FAULT DETECTION USING EXTENDED KALMAN FILTER. *International Journal of Fluid Power*.

Andreev, R. (2015). *Evaluation of Hydraulic Excavator and Rope Shovel Major Maintenance Costs in Operation.*

Ascher, H., & Feingold, H. (1984). Repairable Systems Reliability: Modeling, Inference, Misconceptions and Their Causes.

Ayodele, T. (2010). *New Advances in Machine Learning.* InTech.

Azadeh, A., & Zadeh, S. A. (2015). An integrated fuzzy analytic hierarchy process and fuzzy multiple-criteria decision-making simulation approach for maintenance policy selection. *Transactions of The Society for Modeling and Simulation International 92*.

Babcock, W., & Battat, B. (2006). *Reducing the Effects of Contamination on Hydraulic Fluids and Systems.* Retrieved from https://www.machinerylubrication.com/.

Baker, M. (2020). *Most Common Causes of Hydraulic Systems Failure*. Retrieved from https://yorkpmh.com.

Battifarano, M., DeSmet, D., Madabhushi, A., & Nabar, P. (2018, April). *Predicting Future Machine Failure from Machine State Using Logistic Regression.* Retrieved from https://www.researchgate.net/publication: https://www.researchgate.net/publication/324584543_Predicting_Future_Machine_Failure_from_Machine_State_Using_Logistic_Regression

Bengtsson, M., & Lundstrom, G. (2018). On the importance of combining "the new" with "the old" – One important prerequisite for maintenance in Industry 4.0. *Swedish Production Symposium* (pp. 118-125). Stockholm, Sweden: Procedia Manufacturing 25 .

Bennett, F. (2013, December). *Oil Analysis Explained.* Retrieved from https://www.machinerylubrication.com: https://www.machinerylubrication.com/Read/29598/oil-analysis-report

Bhattacharjee, P., Dey, V., & Mandal, U. K. (2020). Risk assessment by failure mode and effects analysis (FMEA) using an interval number based logistic regression model. *Elsevier Safety Science, Volume 132*.

Bin, G., Bin, G., Gao, J., Li, X. J., & Dhillon, B. (2012). Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network. *Mechanical Systems and Signal Processing, Vol 27*, 696 -711.

Brain, M. (2021). *How Hydraulic Machines Work*. Retrieved from https://science.howstuffworks.com: <https://science.howstuffworks.com/transport/engines-equipment/hydraulic.htm>

Brownlee, J. (2019, November 5). *A Gentle Introduction to Maximum Likelihood Estimation for Machine Learning*. Retrieved from https://machinelearningmastery.com/: https://machinelearningmastery.com/what-is-maximum-likelihood-estimation-in-machine-learning/

Brownlee, J. (2020, August 26). *Train-Test Split for Evaluating Machine Learning Algorithms.* Retrieved from https://machinelearningmastery: https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/

Bui, X.-N., & Drebenstedt, C. (2009). Use of hydraulic backhoe excavator in Surface Mining.

Burt, C. N., & Caccetta, L. (2018). *Equipment Selection for Mining: with Case Studies.*

Carlo, F. D. (2013). *Reliability and Maintainability in Operations Management.* Schiraldi, Massimiliano: InTech.

Celikmih, K., Inan, O., & Uguz, H. (2020). Failure Prediction of Aircraft Equipment Using Machine Learning with a Hybrid Data Preparation Method. *Intelligent Decision Support Systems Based on Machine Learning and Multicriteria Decision-Making*, 10.

Chang, K.-H. (2015). Reliability Analysis. In K.-H. Chang, *e-Design* (pp. 523-595). Elsevier Inc.

Chatterjee, S., & Bandopadhyay, S. (2012). Reliability estimation using a genetic algorithm-based artificial neural network: An application to a load-haul-dump machine. *Expert Systems with Applications, Volume 39, Issue 12*, 10943-10951.

Chaulya, S., & Prasad, G. (2016). *Sensing and Monitoring Technologies for Mines and Hazardous Areas.*

Coetzee, J. L. (2004). *Maintenance.* Trafford Publishing.

Deighton, M. G. (2016). *Facility Integrity Management.*

Dhillon, B. S. (2002). *Engineering Maintenance: A Modern Approach.* CRCISBN: 9780429132209.

Dhillon, B. S. (2008). *Mining Equipment Reliability, Maintainability, and Safety.* London: 2008 Springer-Verlag London Limited.

Dhillon, B. S. (2008). *Mining Equipment Reliability, Maintainability, and Safety.* London: Springer-Verlag London Limited.

Din, I. U., Guizani, M., Rodrigues, J. J., Hassan, S., & Korotaev, V. V. (2019). Machine learning in the Internet of Things: Designed techniques for smart cities. *Elsevier - Future Generation Computer System - Volume 100*, 826-843.

Doostan, M., & Chowdhury, B. H. (2017). Power distribution system equipment failure identification using machine learning algorithms. *2017 IEEE Power & Energy Society General Meeting.* Chicago, IL, USA: IEEE.

Drummond, C., & Yang, C. (2008). *Reverse-Engineering Costs : How much will a Prognostic Algorithm save ?* Retrieved from semanticscholar.org: https://www.semanticscholar.org/paper/Reverse-Engineering-Costs-%3A-How-much-will-a-save-Drummond-Yang/

Du, C.-J., & Sun, D.-W. (2008). Object Classification Methods. In C.-J. Du, & D.-W. Sun, *Computer Vision Technology for Food Quality Evaluation* (pp. 57-80). Elsevier Inc., Academic Press.

Du, J. (2008). *Evaluation of Equipment Reliability, Availability and Maintainability in an oil sand processing plant.*

Duc, H. N., Kamwa, I., Dessaint, L.-A., & Cao-Duc, H. (2017). A Novel Approach for Early Detection of Impending Voltage Collapse Events Based on the Support Vector Machine. *International Transactions on Electrical Energy Systems*.

Fahad, S. A., & Mahbub, M. A. (2016). Clustering, A Modified K-Means Algorithm for Big Data. *| IJCSET(www.ijcset.net) |April 2016 | Vol 6, Issue 4*, 129-132.

Fan, Q., & Fan, H. (2015). Reliability Analysis and Failure Prediction of Construction Equipment with Time Series. *Journal of Advanced Management Science Vol. 3, No. 3, September 2015*.

Fitch, B. (2013, October). *Anatomy of Wear Debris.* Retrieved from https://www.machinerylubrication.com: https://www.machinerylubrication.com/Read/29537/wear-debris-anatomy

Forthofer, R. N., Lee, E. S., & Hernandez, M. (2007). Logistic and Proportional Hazards Regression. In R. N. Forthofer, E. S. Lee, & M. Hernandez, *Biostatistics - second edition* (pp. 387-419). Elsevier Inc, Academic Press.

Frimpong, S., Hu, Y., & Inyang, H. (2008). Dynamic Modeling of Hydraulic Shovel Excavators for Geomaterials. *International Journal of Geomechanics*.

Frost, J. (2020, September 14). *Independently and Identically Distributed Data (IID)*. Retrieved from https://statisticsbyjim.com/: https://statisticsbyjim.com/basics/independent-identically-distributed-data/

Gandhi, R. (2018, May 5). *Naive Bayes Classifier.* Retrieved from https://towardsdatascience.com/: https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c

Gannon, M. (2018). *What types of hydraulic filters are available?* Retrieved from https://www.sealingandcontaminationtips.com/.

Garvey, R. (2003). *Is Your Particle Counter Giving You PPM and Size Distribution?* Retrieved from https://www.machinerylubrication.com/.

Gordon, G. (2004, June 15). *Support Vector Machines and.* Retrieved from https://www.cs.cmu.edu: https://www.cs.cmu.edu/~ggordon/SVMs/new-svms-and-kernels.pdf

Hall, R. (1997). *Analysis of Mobile Equipment Maintenance Data In an Underground Mine.*

Hall, R. A., Knights, P. F., & Daneshmend, L. K. (2000). Pareto analysis and condition-based maintenance. *Mining Technology > Transactions of the Institutions of Mining and Metallurgy*.

Harrison, O. (2018, September 10). *Machine Learning Basics with the K-Nearest Neighbors Algorithm.* Retrieved from https://towardsdatascience.com/: https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761

Heney, P. (2014). *Hydraulic filters: Size, location, and quality.* Retrieved from https://www.fluidpowerworld.com/.

Hildreth, J., & Dewitt, S. (2016). Logistic Regression for Early Warning of Economic Failure. *52nd ASC Annual International Conference Proceedings.* Associated Schools of Construction.

Holmberg, K. a. (2017). Global energy consumption due to friction and wear in the miningindustry. *Tribology International,*.

Hsu, H.-H., & Hsieh, C.-W. (2010). Feature Selection via Correlation Coefficient. *JOURNAL OF SOFTWARE, VOL. 5, NO. 12, DECEMBER 2010* .

Hu, W., & Men, Z. (2020). Study on Design of Working Device and Hydraulic System of HydraulicExcavator Based on Computer Aided Technology.

Huang, Y. Q., Nie, S., & Ji, H. (2014). Identification of Contamination Control Strategy for Fluid Power System Using an Inexact Chance-Constrained Integer Program. *Journal of Applied Mathematics 2014*.

Hwang, S., & Jeong, J. (2018). SVM-RBM based Predictive Maintenance Scheme for IoT-enabled Smart Factory . *Thirteenth International Conference on Digital Information Management.* ICDIM.

Imandoust, S. B., & Bolandraftar, M. (2013). Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background. *S B Imandoust et al. Int. Journal of Engineering Research and Applications Vol. 3, Issue 5*, 605 - 610.

Jabbar, M. A., Deekshatulu, *. B., & Chandra, P. (2013). Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm. *International Conference on Computational Intelligence: Modeling Techniques and Applications.* Procedia Technology.

Jakkula, V. (2011). Tutorial on Support Vector Machine (SVM).

Jan, B., Farman, H., Khan, M., Talha, M., & Din, I. U. (2019). Designing a Smart Transportation System: An Internet of Things and Big Data Approach. *IEEE Wireless Communications, vol. 26, no. 4*, 73-79.

Jason. (2020, August 3). *A Gentle Introduction to k-fold Cross-Validation.* Retrieved from https://machinelearningmastery.com/: https://machinelearningmastery.com/k-fold-cross-validation/

Jia, F., Lei, Y., Lin, J., Zhou, X., & Lu, N. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing, Volumes 72–73*, 303-315.

Jiarula, Y., Gao, J., Gao, Z., Jiang, H., & Wang, R. (2016). Fault mode prediction based on decision tree. *IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC).* Xi'an, China: IEEE.

Jin-Hee, K. (2018). A Study on Prediction Model of Equipment Failure Through Analysis of Big Data Based on RHadoop. *Wireless Personal Communications*.

Jonathan P Peck, J. B. (1994). On-line condition monitoring of rotating equipment using neural networks. *ISA Transactions, Volume 33, Issue 2,*.

Jordan, J. (2017, November 2). *Hyperparameter tuning for machine learning models.* Retrieved from https://www.jeremyjordan.me: https://www.jeremyjordan.me/hyperparameter-tuning/

Junior, M. F., Leite, J. C., Carvajal, T. L., Nascimento, M. H., Júnior, J. d., & Freitas, C. A. (2020). Decision Making Applications in Modern Power Systems. In S. H. Aleem, A. Y. Abdelaziz, & R. Bansal, *Decision Making Applications in Modern Power Systems* (pp. 337-364). Academic Press.

K. R. M. Rao, P. V. (2001). Graphical methods for reliability of repairable equipment and maintenance planning. *International Symposium on Product Quality and Integrity*, (pp. 123-128).

Kamepalli, S., & Mothukuri, R. (2014). Implementation of Clustering-Based Feature. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3*.

Kelsh, H. (2008). Parameters of hydraulic mining excavators for successful rope shovels replacement in low ambient temperatures.

Kendon, P. (2019, September 24). *How Well Do You Understand the Key Metrics for Reliability and Maintenance?* Retrieved from https://www.prometheusgroup.com/: https://www.prometheusgroup.com/posts/how-well-do-you-understand-the-key-metrics-for-reliability-and-maintenance

Khayal, O. (2014). *Study of Failure in Hydraulic Systems (Case study of machinery used in local).*

Kissell, R., & Poserina, J. (2017). Advanced Math and Statistics. In R. Kissell, & J. Poserina, *Optimal Sports Math, Statistics, and Fantasy.* Elsevier Ltd.

Kleinbaum, D. G., & Klein, M. (2002). *Logistic Regression - Second Edition.* New York: Springer-Verlag.

Knights, P. F. (2001). Rethinking Pareto Analysis: maintenance applications of logarithmic scatterplots. *Emerald Insight*, 252-263.

Kohli, M. (2021). *Predicting Equipment Failure On SAP ERP Application Using Machine Learning Algorithms.*

Ku, J.-H. (2018). A Study on Prediction Model of Equipment Failure Through Analysis of Big Data Based on RHadoop. *Wireless Personal Communications 98(12)*.

Kumar, D., & Kumar, D. (2018). Maintenance Management. In D. Kumar, & D. Kumar, *Sustainable Management of Coal Preparation* (pp. 369-380). Woodhead Publishing.

Kumar, U. (1990). *Reliability analysis of Load-Haul-Dump machines.*

Kumar, U., Klefsjo, B., & Granholm, S. (1989). Reliability investigation for a fleet of load haul dump machines in a Swedish mine. *Elsevier*, 341-361.

Lai, C.-D., Murthy, D. N., & Xie, M. (2006). Weibull Distributions and Their Applications.

Lazzari, L. (2017). Engineering Tools for Corrosion. In L. Lazzari, *Statistical Analysis of Corrosion Data* (pp. 131-148). Elsevier Ltd.

Li, L., Mechefske, C. K., & Li, W. (2004). Electric motor faults diagnosis using artificial neural networks. *Insight - Non-Destructive Testing and Condition Monitoring, Volume 46*, 616 -621.

Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing, Vol 108*, 33 -47.

Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry 27(2)*.

Macuzic, I. (2010). Development of Mobile Device for Oil Analysis.

Madenas, N., Tiwari, A., Turner, C. J., & Woodward, J. (2014). Information flow in supply chain management: A review across the product lifecycle. *CIRP Journal of Manufacturing Science and Technology*, 335-346.

Marcus, B. (2004). Condition Based Maintenance System Technology – Where is Development Heading? *International Conference of Euromaintenance.* Barcelona, Spain.

Mayer, A. (2009). *Nine Reasons Why Oil Analysis Programs Fail.* Retrieved from https://www.machinerylubrication.com.

Maymon, G. (2018). Some Important Statistical Distributions. In G. Maymon, *Stochastic Crack Propagation* (pp. 9-18). Elsevier Inc.

McCloy, D., & Martin, H. (1980). *Control of fluid power : analysis and design.* Chichester [Eng.] : E. Horwood ; New York : Halsted Press, 1980.

Meeker, W. Q., Hahn, G. J., & Escobar, L. A. (2017). *Statistical Intervals: A Guide for Practitioners and Researchers, 2nd Edition.* Wiley.

Mencik, J. (2016). Reliability of Systems. In J. Mencik, *Concise Reliability for Engineers* (p. 214). InTech.

Mishra, A. (2019, December 2). Evaluating Machine Learning Models. In *Machine Learning in the AWS Cloud.* Indianapolis: Wiley. Retrieved from https://www.analyticsvidhya.com/.

Mitrev, R., Gruychev, R., & Pobegailo, P. (2011). CAD/CAE Investigation of a large mining excavator.

Moghaddass, R., & Zuo, M. J. (2012). Fault Diagnosis for Multi-State Equipment with Multiple Failure. *IEEE*.

Monroe, W. (2017, August 14). Logistic Regression.

Moon, M. (2008). *Taking Lubricant Cleanliness to the Next Level.* Retrieved from https://www.machinerylubrication.com.

Moosavian, A., Ahmadi, H., & Khazaee, A. T. (2013). Comparison of two classifiers; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing. *Shock and Vibration 20 (2013) 263–272*.

Murthy, D., Atrens, A., & Eccleston, J. (2002). Strategic maintenance management. *Journal of Quality in Maintenance Engineering 8*.

Nakamura, J. (2007). *Predicting Time-to-Failure of Industrial Machines with Temporal Data.*

Narkhede, S. (2018, June 26). *Understanding AUC - ROC Curve*. Retrieved from https://towardsdatascience.com/: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

Ng, F., Harding, J. A., & Glass, J. (2016). Improvinghydraulicexcavatorperformancethrough. *Mechanical Systems and Signal Processing*.

Nirmal, K. S., Agarwal, P., & Reddy, K. N. (2019). Early Detection of Equipment Failure Using One-class Support Vector Machine. *Proceedings of the International Conference on Industrial Engineering and Operations Management.* Bangkok: IEOM Society International.

Ozdogan, M. (2003). Electric Rope Shovels versus Hydraulic Excavators.

Park, C.-G., Yoo, S., Ahn, H., Kim, J., & Shin, D. (2020). A coupled hydraulic and mechanicalsystem simulation for hydraulicexcavators.

Park, Y. S., & Lek, S. (2016). Ecological Model Types. In S. E. Jørgensen, *Developments in Environmental Modelling, Vol 28* (pp. 1-26). Elsevier Ltd.

Patil, D. D., Wadhai, V. M., & Gokhale, J. A. (2010). Evaluation of Decision Tree Pruning Algorithms for Complexity and Classification Accuracy. *International Journal of Computer Applications*.

Pintelon, L., & Parodi-Herz, A. (2008). *Maintenance: An Evolutionary Perspective.* Springer, London.

Prabhakaran, S. (2018, November 4). *How Naive Bayes Algorithm Works? (with example and full code).* Retrieved from https://www.machinelearningplus.com/: https://www.machinelearningplus.com/predictive-modeling/how-naive-bayes-algorithm-works-with-example-and-full-code/

Ray, S. (2017, September 13). *Understanding Support Vector Machine(SVM) algorithm from examples.* Retrieved from https://www.analyticsvidhya.com/: https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

Reik, M., & Oberli, F. (2014). *Selection of Filters for Lubricating Oil and Hydraulic Systems.* Springer, Berlin, Heidelberg.

Riantama, R. N., Prasanto, A. D., Kurniati, N., & Anggrahini, D. (2020). Examining Equipment Condition Monitoring for Predictive Maintenance, A case of typical Process Industry. *Proceedings of the 5th NA International Conference on Industrial Engineering and Operations Management.* Detroit, Michigan, USA.

Robert, K., & Jim, P. (2017). Advanced Math and Statistics. In R. Kissell, & J. Poserina, *Optimal Sports Math, Statistics, and Fantasy* (pp. 103-135). Elsevier Ltd.

Rokach, L., & Maimon, O. (2005). Decision Trees. In L. Rokach, & O. Maimon, *The Data Mining and Knowledge Discovery Handbook* (pp. 165-192). Boston, MA: Springer.

Rokach, L., & Maimon, O. (2015). *Data Mining With Decision Trees - Second Edition.* Singapore: World Scientific Publishing Co. Pte. Ltd.

Ron LeBlanc, S. (2019). The real world benefits of used oil analysis. *Canadian Mining journal*.

Rosario, S. F., & Thangadurai, K. (2015). RELIEF: Feature Selection Approach. *International Journal of Innovative Research and Development, volume 4, No 11*.

S.Cundiff, J. (2001). *Temperature and Contamination Control.* Retrieved from abe.ufl.edu.

Sahu, A. R., & Palei, S. K. (2020). Fault prediction of drag system using artificial neural network for prevention of dragline failure. *Engineering Failure Analysis, Volume 133*.

Sanz, J., Perera, R., & Huerta, C. (2012). Gear dynamics monitoring using discrete wavelet transformation and multi-layer perceptron neural networks,. *Applied Soft Computing, Volume 12, Issue 9,*, 2867-2878.

Saritas, M. M., & Yasar, A. (2019). Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification. *International Journal ofIntelligent Systems and Applications in Engineering*.

Sellathamby, C., Moore, B., & Slupsky, S. (2010). Increased Productivity by Condition-Based Maintenance Using Wireless Strain Measurement System. *Canadian Institute of Mining (CIM) MEMO Conference.* Sudbury, ON, Canada.

Sharma, A., Jigyasu, R., Mathew, L., & Chatterji, S. (2018). Bearing Fault Diagnosis Using Weighted K-Nearest Neighbor. *Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018).* IEEE.

Sharma, R. (2020, December 10). *Guide to Decision Tree Algorithm: Applications, Pros & Cons & Example.* Retrieved from https://www.upgrad.com/: https://www.upgrad.com/blog/guide-to-decision-tree-algorithm/

Short, M., & Twiddle, J. (2019). An Industrial Digitalization Platform for Condition Monitoring and Predictive Maintenance of Pumping Equipment. *Sensors 19*.

Singh, M., Lathkar, G. S., & Basu, S. K. (2012). Failure Prevention of Hydraulic System Based on Oil Contamination. *Journal of The Institution of Engineers (India)*.

Singh, M., Lathkar, G. S., & Basu, S. K. (2012). Failure Prevention of Hydraulic System Based on Oil Contamination. *Journal of The Institution of Engineers (India)*.

Smiley, A. (2021). *Troubleshooting Hydraulic Pumps*. Retrieved from https://www.machinerylubrication.com.

Smith, D. J. (2005). Basic reliability prediction theory. In D. J. Smith, *Reliability, Maintainability and Risk* (pp. 75-88). Elsevier Ltd.

Snetkov, D. S., & Kosolapov, A. I. (2019). Substantiation of equipment complexes for the development of browncoal deposits in the mode of coal quality management.

Strelnikov, A., Markov, S., Rattmann, L., & Weber, D. (2018). Theoretical Features of Rope Shovels and Hydraulic Backhoes Using at Open Pit Mines. *IIrd International Innovative Mining Symposium.*

Suryo, S. H., & Bayuseno, A. (2018). Study on Reliability Analysis of Hydraulic Components and Excavator Engine in Maintenance of Mine Heavy Equipment. *International Journal of Mechanical Engineering and Technology, Volume 9, Issue 8*, 1244-1254.

Thibault, R. (2013). *Mining Success With World-Class Oil Analysis.* Retrieved from https://www.efficientplantmag.com/.

Thomas, D. S. (2018). *The Costs and Benefits of Advanced Maintenance in Manufacturing*. Retrieved from National Institute of Standards and Technology : https://nvlpubs.nist.gov/nistpubs/ams/NIST.AMS.100-18.pdf

Troy, D. (2018, February 12). *The Importance Of Efficient Mining Equipment*. Retrieved from https://industrytoday.com/: https://industrytoday.com/importance-efficient-mining-equipment/

Unger, R., & Conway, K. (1994). *Mining Publication: Impact of Maintainability Design on Injury Rates and Maintenance Costs for Underground Mining Equipment*. Retrieved from https://www.cdc.gov: https://www.cdc.gov/niosh/mining/works/coversheet812.html

V.Ramani, R. (2012). Surface Mining Technology: Progress and Prospects.

Vahed, A. T., Ghodrati, B., & Hosseinie, S. H. (2018). Enhanced K-Nearest Neighbors Method Application in Case of Draglines Reliability Analysis. *27th International Symposium on Mine Planning and Equipment Selection.* Santiago, Chile.

Walczak, S., & Cerpa, N. (2001). Artificial Neural Networks. In R. A. Meyers, *Encyclopedia of Physical Science and Technology (Third Edition).* 609-629: Academic Press.

Weber, A. (2005, November). *Key Performance Indicators - Measuring and Managing Maintenance Function*. Retrieved from https://www.plant-maintenance.com: https://www.plant-maintenance.com/articles/KPIs.pdf

Wei, G., Luo, Z., & Yu, a. X. (2019). Clustering Research on Ship Fault Phenomena Based on K-means Algorithm. *Chinese Control And Decision Conference (CCDC)* (pp. 4412-4415). IEEE.

Yaghini, A., Hall, R. A., & Apel, D. (2020). Modeling the influence of electric shovel operator performance on mine.

Yi, X. J., Chen, Y. F., & Hou, P. (2017). Fault diagnosis of rolling element bearing using Naïve Bayes classifier. *Vibroengineering PROCEDIA, Vol. 14, 2017*, 64-69.

Zhang, N., Wu, L., Yang, J., & Guan, Y. (2018). Naive Bayes Bearing Fault Diagnosis Based on Enhanced Independence of Data. *MDPI*.

Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine 4(11)*.

Zheng, H., Dai, Y., & Zhou, Y. (2019). Fault Prediction of Fan Gearbox Based on K-Means Clustering and LSTM. *IOP Conf. Series: Materials Science and Engineering.* IOP Publishing.

Zhu, X., Hou, D., Zhou, P., Han, Z., Yuan, Y., Zhou, W., & Yin, Q. (2019). Rotor fault diagnosis using a convolutional neural network with symmetrized dot pattern images. *Measurement, Volume 138*, 526 -535.

Zisserman, A. (2015). *The SVM Classifier.* Retrieved from https://www.robots.ox.ac.uk/: https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf

APPENDIX A: OIL ANALYSIS TOOL, WO REPORTS AND NORMALITY TESTS

A1: An example of working of oil analysis tool:

Hydraulic Volume	2200 L						
	Wear Metal						
Element Added	**Iron**	**Chromium**	**Nickel**	**Aluminium**	**Copper**	**Lead**	**Tin**
Volume of patricles added l	0.05	0.01	0.01	0.06	0.01	0.80	0.01
Density of Metal kg/l	7.87	7.14	8.91	2.70	8.94	11.34	7.27
Mass Added kg	0.39	0.09	0.12	0.16	0.12	9.07	0.10
Hydraulic Metal Limits Moderate ppm	**22.00**	**10.00**	**5.00**	**10.00**	**30.00**	**30.00**	**15.00**
Hydraulic Metal Limits Critical ppm	30.00	30.00	10.00	30.00	60.00	60.00	30.00
Particle Size Added	> 4	> 4	> 4	> 4	> 4	> 4	> 4
Resultant ISO Particle Count ppm	23	6	6	27	6	364	6
Resultant ICP Particle Count ppm	23	6	6	27	6	364	6
ISO Code	**12**	**10**	**10**	**12**	**10**	**16**	**10**

A2: An example of Initial Failure Report (IFR)

COMPONENT FAILURE – INITIAL REPORT

SITE FAILURE OCCURED: -

UNIT#: S 5577

MACHINE S/N: FF018NQ001016

COMPONENT FAILED NAME: Final drive (RH Travel Gearcase Assembly)

LOCATION ON MACHINE: R/H

FIELD REPORTS ATTACHED ☐

PICTURES ATTACHED ☐

TRACK REPORT ATTACHED ☒ -N/A -

FORM COMPLETION DATE: March 5/17

COMPONENT FAILURE DATE: MARCH 04, 2017.

UNIT SMU: 32, 803.

COMPONENT SMU: 32, 803

COMPONENT BENCHMARK SMU: 36,000

TROUBLESHOOTING WO: 354923

CONTRACTOR TROUBLESHOOTING WO: - N/A -

WARRANTY CONSIDERATION: Yes☐ No☒

FAILURE INFORMATION:

Was this failure the result of an incident (damage, abuse or neglect) and requires an incident report? If YES include Incident Report number. Yes ☐ No ☒

Incident report #:

Why was the initial troubleshooting or repair work order created? What was the complaint? Explain:

- Machine stuck during removal. R/H Travel motors failed.

What were the results of the initial troubleshooting work order to justify change out? Explain:

found excessive plug in upper high speed final drive pinons

What was the machine/operator doing at the time of failure or when he/she first noticed something was wrong?

Trying to remove machine, Travelling and pushing with front structure putting extra force through final drive.

Was the component operated after initial indication of failure? If so, for how long?

no

Was there any recent work or lubrication issues that may have affected this failure: YES ☒ NO ☐

If YES please explain and include relevant work order numbers:

Final drive Duo Cone failed 2 years ago. filled with grease. water, dirt and sand has gone through final drive causing failure. * see WO # 284153 opened 10/06/15

Has the VHMS or monitoring system been downloaded and analyzed if applicable? YES ☐ NO ☐ N/A ☒

Is there a record of abuse or events recorded which may have contributed to or indicated impending failure? Explain:

Did the oil analysis indicate an impending failure (if applicable): YES ☐ NO ☒

If "YES" were Reliability Alert(s) issued: YES ☐ NO ☐

Was an appropriate action taken and documented in response to the Reliability Alert(s)? Explain:

Are there any known design / product issues possibly affecting this failure: YES ☑ NO ☐

If YES please explain:

MFT is causing Duo cones on fixed drives
To get mud packed and fail.

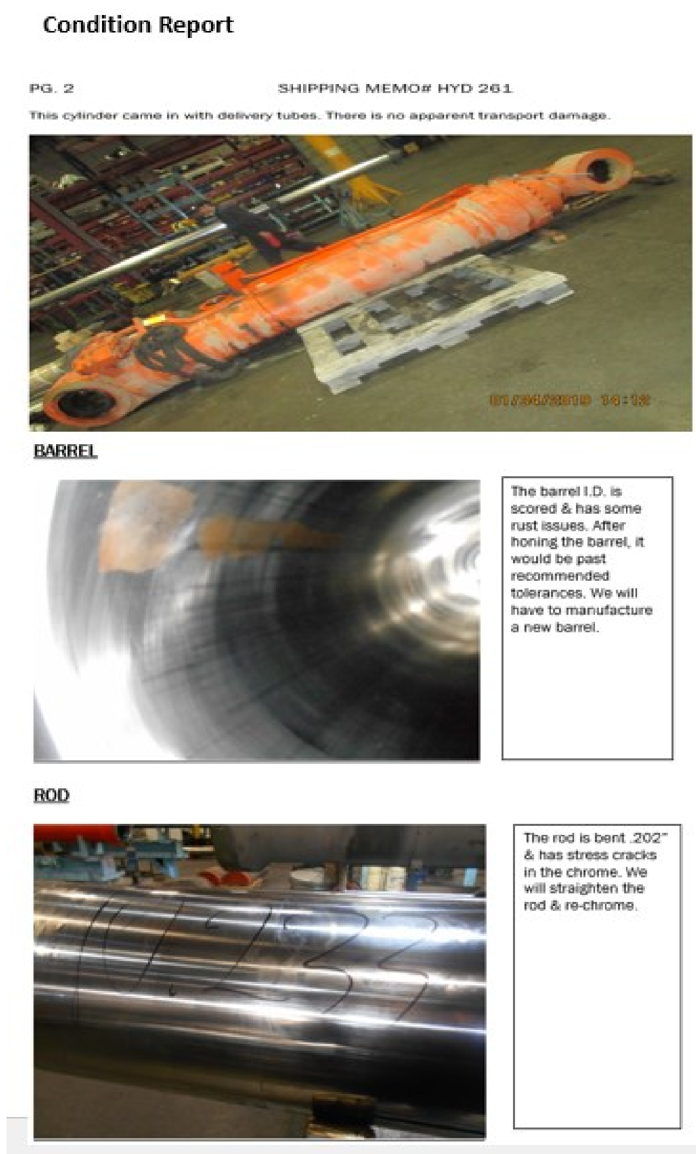
Is there any immediate action recommended to prevent future premature failures similar to this one? YES ☐ NO ☒

If YES please explain:

Are there any additional details that affect the failure / need for a new component? Explain:

- Recomend trying to repair in house. Depending
on cost and parts Avaiblity

A3: Example of Condition Report (CR)


Condition Report

PG. 2 SHIPPING MEMO# HYD 261

This cylinder came in with delivery tubes. There is no apparent transport damage.

BARREL

The barrel I.D. is scored & has some rust issues. After honing the barrel, it would be past recommended tolerances. We will have to manufacture a new barrel.

ROD

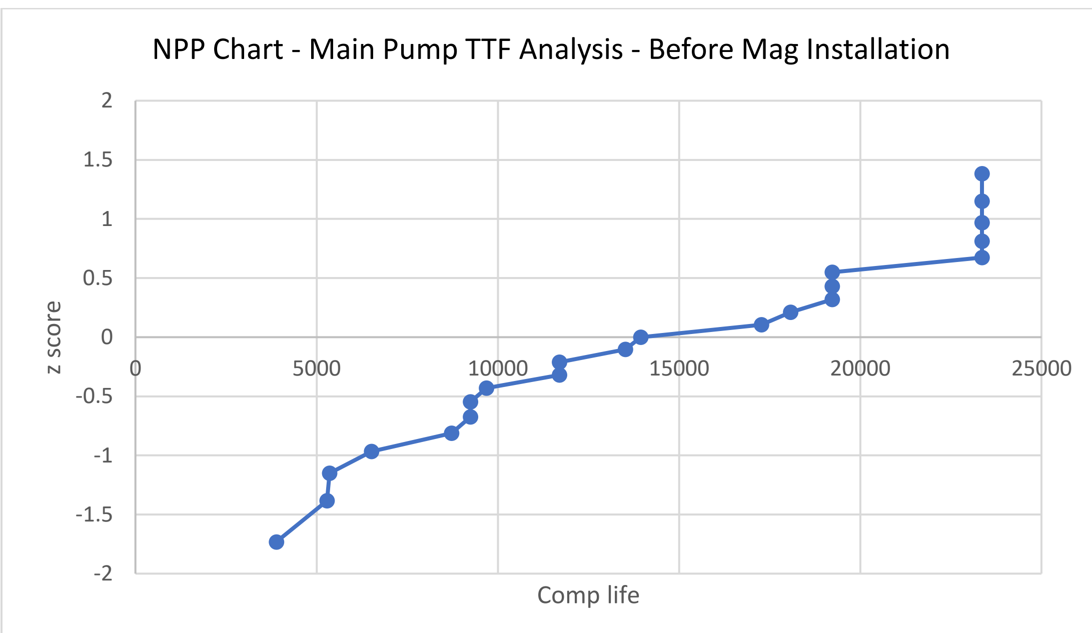
The rod is bent .202" & has stress cracks in the chrome. We will straighten the rod & re-chrome.

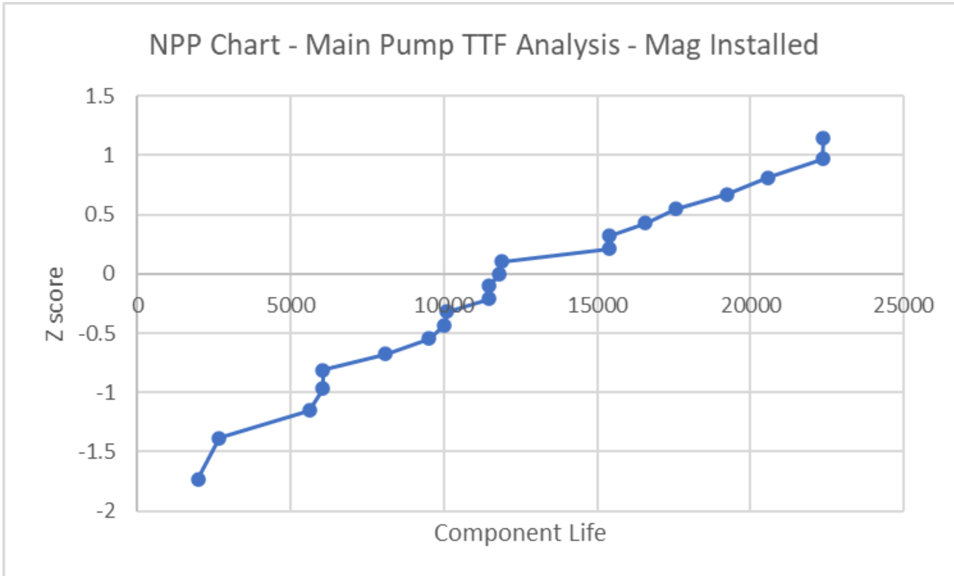
Normality Test: The normality test for TTF data of each component is used to check if the data is near normally distributed to use ANOVA test to see the significance in variation in component life after Mag-filters were installed. Descriptive statistics and NPP charts are used to check normality of the data. The mean and median values are compared, and skewness of the data is \checked. If the mean and median values are closer, and skewness is <0.5 then the data is indicating normality and NPP charts are used to see if the data follows near normal distribution. Normal Test Plots (also called Normal Probability Plots or Normal Quartile Plots) are used to investigate whether process data exhibit the standard normal "bell curve" or Gaussian distribution. Component Life vs their Z-score values are plotted and if the line is linear or somewhat linear then the data is assumed to follow normal distribution or a near normal distribution and ANOVA is used for the test of significance. The details of descriptive statistics and NPP charts for all the 8 hydraulic components considered are represented below.

A4: Normality test: Main Pumps

Main Pump - TTF Summary (Before Mag Installation)	
Mean	14915.34783
Standard Error	1439.413429
Median	13943
Mode	23359
Standard Deviation	6903.184296
Sample Variance	47653953.42
Kurtosis	-1.457291689
Skewness	-0.054401758
Range	20545
Minimum	3887
Maximum	24432
Sum	343053
Count	23

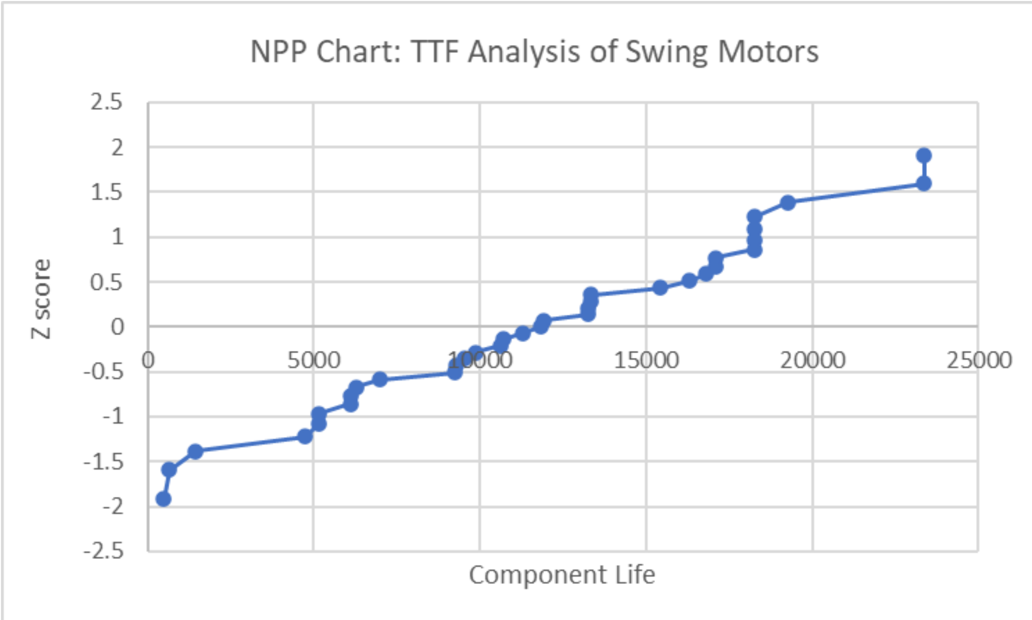
Main Pump - TTF Summary (After Mag Installation)	
Mean	12197
Standard Error	1330.000473
Median	11476
Mode	6048
Standard Deviation	6094.82784
Sample Variance	37146926.4
Kurtosis	-0.8490595
Skewness	0.134092936
Range	20378
Minimum	1979
Maximum	22357
Sum	256137
Count	21





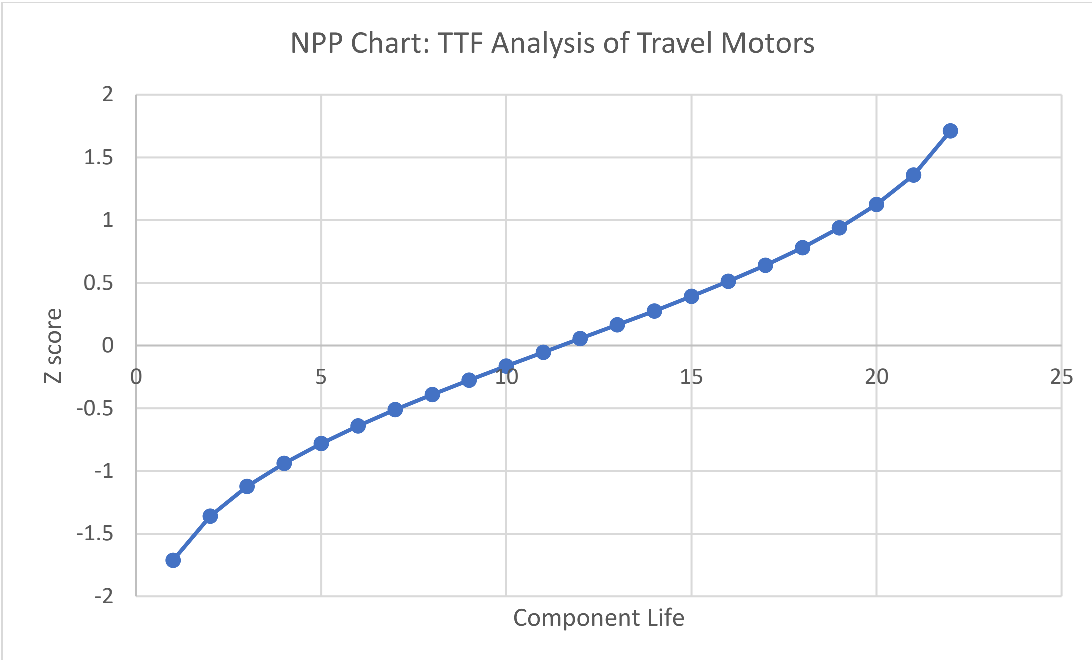
A5: Normality test: Swing Motors

TTF Summary: Swing Motors	
Mean	12121.78
Standard Error	1054.809
Median	11875.5
Mode	18255
Standard Deviation	6328.852
Sample Variance	40054363
Kurtosis	-0.655
Skewness	0.008067
Range	23612
Minimum	484
Maximum	24096
Sum	436384
Count	36



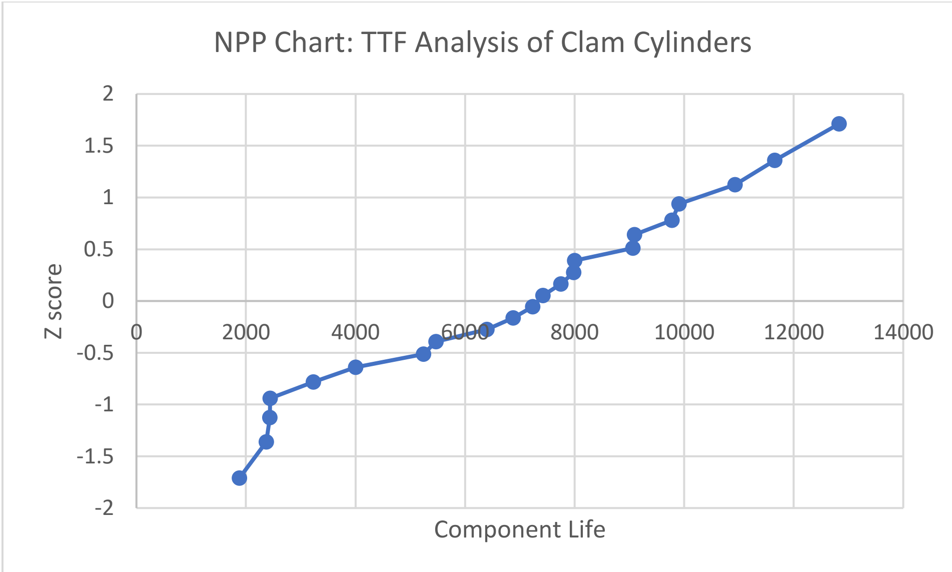
A6: Normality Test: Travel Motors

TTF Summary: Travel Motors	
Mean	20409.95652
Standard Error	1159.864188
Median	22357
Mode	22357
Standard Deviation	5562.513236
Sample Variance	30941553.5
Kurtosis	6.473972529
Skewness	-2.644377991
Range	20203
Minimum	3640
Maximum	23843
Sum	469429



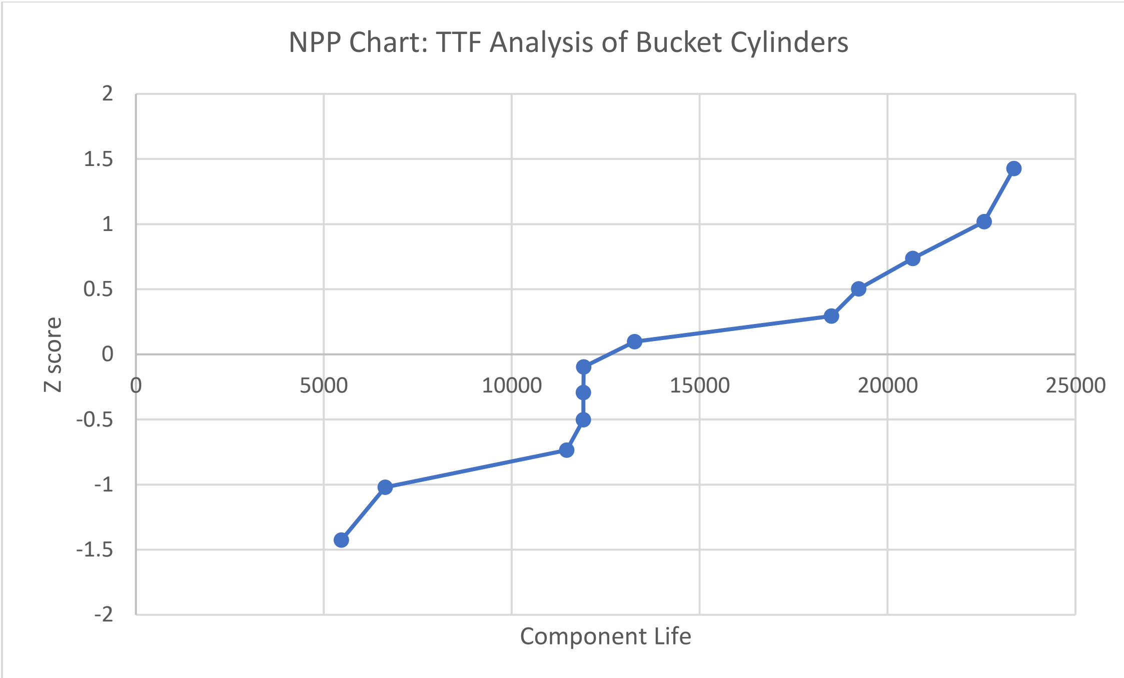
A7: Normality Test: Clam Cylinders

TTF Analysis: Clam Cylinders	
Mean	7172.217
Standard Error	705.3744
Median	7421
Mode	#N/A
Standard Deviation	3382.857
Sample Variance	11443719
Kurtosis	-0.91869
Skewness	-0.0068
Range	11082
Minimum	1878
Maximum	12960
Sum	164961



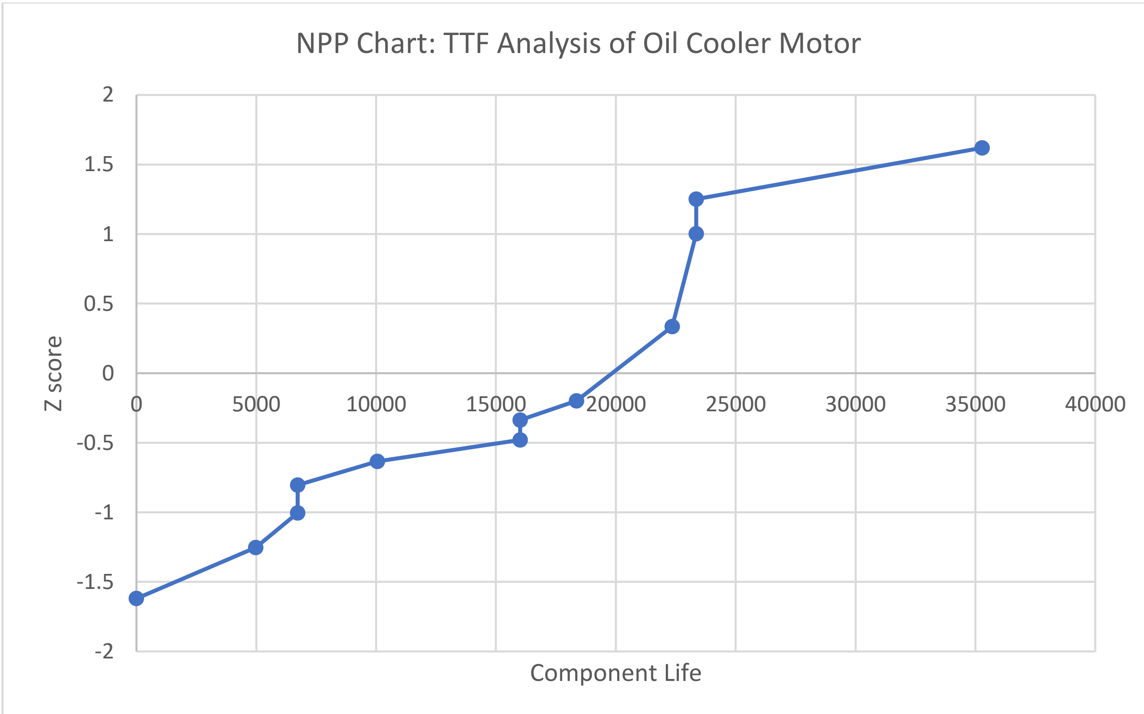
A8: Normality Test – Bucket Cylinder

TTF Analysis: Bucket Cylinders	
Mean	15492.08
Standard Error	1757.603
Median	13272
Mode	11901
Standard Deviation	6337.129
Sample Variance	40159203
Kurtosis	-1.27746
Skewness	-0.04792
Range	19045
Minimum	5468
Maximum	24513
Sum	201397



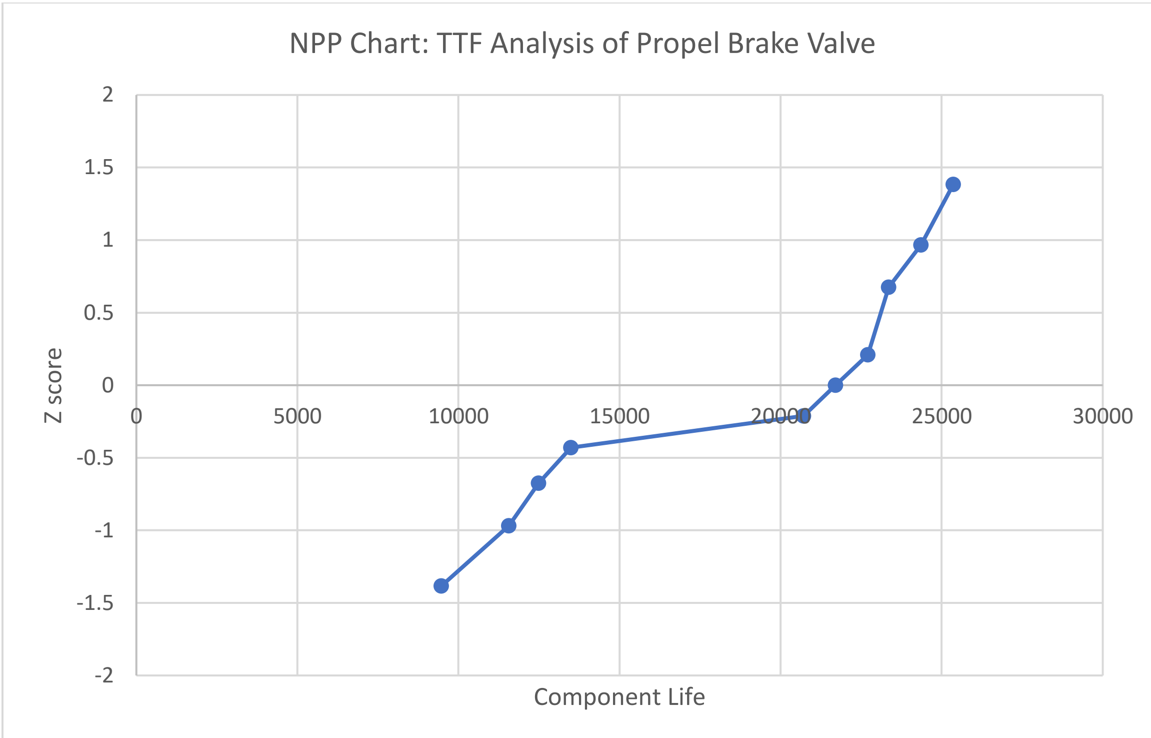
A10: Normality Test – Oil Cooler Motor

TTF Analysis: Oil Cooler Motors	
Mean	17936.32
Standard Error	2132.149
Median	18367
Mode	22357
Standard Deviation	9293.824
Sample Variance	86375160
Kurtosis	0.112277
Skewness	-0.00386
Range	35415
Minimum	0
Maximum	35415
Sum	340790



A11: Normality Test – Propel Brake Valve

TTF Analysis: Propel Brake Valve	
Mean	17849.67
Standard Error	1814.154
Median	20711
Mode	20711
Standard Deviation	6284.415
Sample Variance	39493870
Kurtosis	-1.64999
Skewness	-0.69599
Range	13890
Minimum	9469
Maximum	23359
Sum	214196


NPP Chart: TTF Analysis of Propel Brake Valve
Z score
Component Life
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5000
10000
15000
20000
25000
30000

APPENDIX B: SAMPLE DATA FOR ML MODELS

B1: Data Collected for ML Models

WO	Unit	Component	SMU (hours)	Mag-filter	Component Life Achieved	Warranty	Damage	Particle Count at time of failure	Particle Count 1200 hours prior to failure
8392	1	PUMP, MAIN, 1 & 2	18616	No	23892	No	No	21/19/16	20/19/15
9467	1	MOTOR, TRAVEL, FRONT LH	18736	No	24096	No	Yes	21/19/16	20/19/15
9438	1	MOTOR, TRAVEL, FRONT RH	18736	No	24432	No	No	21/19/16	18/16/14
8332	1	MOTOR, TRAVEL, REAR LH	18736	No	24513	Yes	No	21/19/16	16/15/10
8332	1	MOTOR, TRAVEL, REAR RH	18736	No	28438	No	No	21/19/16	16/15/10
12068	2	CYLINDER, ARM	19229	No	11917	No	No	21/19/16	16/15/10
15074	2	CYLINDER, BUCKET, RH	19229	No	11917	No	No	21/19/16	16/15/10
15872	2	PUMP, MAIN, 5 & 6	23452	No	31042	Yes	No	21/19/16	16/15/10
22134	2	PUMP, A/C COMPRESSOR, LH	17648	No	31042	No	No	21/19/16	19/17/13
35317	2	PUMP, HYD OIL COOLER FAN, FRONT	15342	No	11898	No	No	17/15/12	19/17/13
37196	3	PUMP, MAIN, 1 & 2	17614	No	31127	No	Yes	19/17/13	16/15/10

B2: List of 45 hydraulic components considered for analysis

HYDRAULIC OIL COOLER, FRONT	TRAVEL MOTOR, FRONT RH	MAIN CONTROL VALVE, BOTTOM LH
HYDRAULIC OIL COOLER, REAR	TRAVEL MOTOR, REAR LH	MAIN CONTROL VALVE, BOTTOM RH
BOOM CYLINDER, LH	TRAVEL MOTOR, REAR RH	MAIN CONTROL VALVE, MIDDLE RH
BOOM CYLINDER, RH	A/C COMPRESSOR PUMP, LH	MAIN CONTROL VALVE, MIDDLE LH
BUCKET CYLINDER, LH	A/C COMPRESSOR PUMP, RH	MAIN CONTROL VALVE, TOP LH
BUCKET CYLINDER, RH	HYDRAULIC OIL COOLER FAN PUMP, FRONT	MAIN CONTROL VALVE, TOP RH
CLAM CYLINDER, LH	HYDRAULIC OIL COOLER FAN PUMP, REAR	PILOT CONTROL, VALVE
CLAM CYLINDER, RH	MAIN PUMP, 1& 2	PROPEL/BRAKE CONTROL VALVE, LH1
HYDRAULIC OIL COOLER FAN MOTOR, FRONT	MAIN PUMP, 11 & 12	PROPEL/BRAKE CONTROL VALVE, LH2
HYDRAULIC OIL COOLER FAN MOTOR, REAR	PUMP MAIN, 3 & 4	PROPEL/BRAKE CONTROL VALVE, RH1
SWING MOTOR, FRONT LH	PUMP MAIN, 5 & 6	PROPEL/BRAKE CONTROL VALVE, RH2
SWING MOTOR, FRONT RH	PUMP MAIN, 7 & 8	HYDRAULIC OIL COOLER FAN, FRONT
SWING MOTOR, REAR LH	PUMP MAIN, 9 & 10	HYDRAULIC OIL COOLER FAN, REAR
SWING MOTOR, REAR RH	PILOT PUMP, LH	PUMP DRIVE OIL COOLER PUMP, RH
TRAVEL MOTOR, FRONT LH	PILOT PUMP, RH	PUMP DRIVE OIL COOLER PUMP, LH

B3: Sample data used in K-NN algorithm

SMU	Components	Component Code	Unit	4M Particle Count at Failure	6M Particle Count at Failure	10M Particle Count at Failure	4M Particle Count 600hrs Failure	6M Particle Count 600hr Failure	10M Particle Count 600hr Failure	Failure status
30927	PUMP, MAIN, 1 & 2	9	1	20	18	14	19	17	15	1
31042	CYLINDER, BOOM	2	1	19	17	13	19	17	15	1
31146	COOLER, HYDRAULIC OIL	1	1	20	18	14	19	17	15	0
31146	CYLINDER, BUCKET	3	1	20	18	14	19	17	15	0
31146	MOTOR, SWING	5	2	20	18	14	19	17	15	0
35415	MOTOR, HYDRAULIC OIL COOLER FAN	4	2	18	16	13	20	17	13	0
36502	PUMP, MAIN, 5 & 6	10	2	18	16	12	19	17	14	0
37949	PUMP DRIVE	6	2	20	18	14	19	17	14	0
37949	PUMP, HYD OIL COOLER FAN	8	2	20	18	14	19	17	14	0
37949	PUMP, MAIN, 5 & 6	10	3	20	18	14	19	17	14	0
37949	PUMP, MAIN, 9 & 10	11	3	20	18	14	19	17	14	0
37949	PUMP, PILOT	12	3	20	18	14	19	17	14	0
40389	MOTOR, SWING	5	1	20	18	14	18	16	13	1

B4: Sample data for Naïve Bayes algorithm

SMU	Components	Failure status	4M Particle Count at Failure	6M Particle Count at Failure	10M Particle Count at Failure	4M Particle Count 600hrs Failure	6M Particle Count 600hr Failure	10M Particle Count 600hr Failure	Swing Motor failure in last 1000 hours	Main Pump, 1 & 2 failure in last 1000 hours	Bucket Cylinder in last 1000hours
12831	COOLER, HYDRAULIC OIL	0	20	18	15	19	17	13	0	0	0
12831	CYLINDER, BOOM	0	20	18	15	19	17	13	0	0	0
12831	CYLINDER, BUCKET	0	20	18	15	19	17	13	0	0	0
12831	CYLINDER, CLAM, LH	0	20	18	15	19	17	13	0	0	0
12831	CYLINDER, CLAM, RH	1	20	18	15	19	17	13	0	0	0
12831	MOTOR, HYDRAULIC OIL COOLER FAN	0	20	18	15	19	17	13	0	0	0
12831	MOTOR, SWING	0	20	18	15	19	17	13	0	0	0
12831	PUMP DRIVE	0	20	18	15	19	17	13	0	0	0
12831	PUMP, A/C COMPRESSOR	0	20	18	15	19	17	13	0	0	0
12831	PUMP, HYD OIL COOLER FAN	0	20	18	15	19	17	13	0	0	0
12831	PUMP, MAIN, 1 & 2	0	20	18	15	19	17	13	0	0	0
12831	PUMP, MAIN, 5 & 6	0	20	18	15	19	17	13	0	0	0