Vaasan yliopisto
UNIVERSITY OF VAASA

Arif Hossain

A Data-Driven Downtime Reduction Framework for Traditional Plastic Extrusion Machines

School of Technology and Innovation
Master's thesis in Economics and Business Administration
Industrial Management

Vaasa 2026

UNIVERSITY OF VAASA
School of Technology and Innovation
Author: Arif Hossain
Title of the Thesis: A Data-Driven Downtime Reduction Framework for Traditional Plastic Extrusion Machines
Degree: Master of Science in Economics and Business Administration
Programme: Industrial Management
Supervisor: Jyri Naarmala
Year: 2026 **Pages:** 69

ABSTRACT:

Unplanned machine downtime is one of the major challenges in production environments. It reduces productivity, increase operational costs, and affects delivery deadline. Large manufacturing companies usually deal with this problem by implementing smart machine maintenance strategy, but small and medium-sized enterprises (SMEs) often lack the resources to implement such advanced maintenance strategy. Therefore, SMEs struggle a lot to keep the machine downtime under a tolerance limit. The opportunity here is that the SMEs already collect some operational data, such as interruption log and maintenance log, which are rarely analyzed together. This thesis aims to develop a practical decision-support framework that integrates these existing sources of data to reduce the downtime without any additional investment. The research is conducted at a local plastic profile manufacturing company. The study is based on three main theoretical perspectives, such as reliability engineering theory, root cause analysis theory, and decision theory. The methodology of this study follows a very structured approach. First, the downtime notes are categorized using text mining. Then, Pareto analysis and KPI analysis are performed to identify the dominant downtime category and evaluate machine performance. Next, a predictive model is developed to anticipate the failure risk of each machine. Finally, all the results from descriptive analysis and predictive analysis are combined into a framework to support maintenance decision-making. The Pareto analysis has found that only five categories, including equipment failure, tooling fault, component failure, calibration issue, and material issue were responsible for majority of the total downtime. It has been shown through KPI analyses that machine performance varies dramatically and the KPIs are highly interconnected. The accuracy rate of Random Forest model is high, but the recall rate is low, which stemmed from the class imbalance of the dataset. Nevertheless, the risk probabilities are useful in prioritizing maintenance task in machines. Feature importance analysis identifies that among the five features of the predictive model, rolling downtime is the most important feature. The main contribution of this study is to develop a decision support framework. The machines are ranked based on their failure probability achieved from the predictive model. Also, the underlying downtime reason of each machine is identified through machine-specific Pareto analysis. By combining these two information, actionable recommendations are generated for each machine. The thesis also provides a practical implementation roadmap for the company. The decision-support framework generated by this study is not just an analytical tool, it's a practical solution for SMEs or resource constrained manufacturing companies.

KEYWORDS: Downtime reduction, predictive maintenance, SMEs, Pareto analysis, KPI analysis, plastic extrusion

Contents

Figures

Tables

1 Introduction

Industry 4.0 is forcing the manufacturing sectors to make big changes in their production system by adding technology like cyber-physical systems, the Internet of Things (IoT), and data-driven analytics (Zheng et al., 2021). As technology reshape the sector, traditional manufacturing domains such as plastics production are facing increasing pressure to enhance efficiency, product quality, and operational reliability. Plastic extrusion is a manufacturing process that is used to make pipes, profiles, sheets, films, and many other polymer products (Hyvärinen et al., 2020). The extrusion lines are built to run for a long shift to reduce downtime, lower the cost, keep the material consistent, and increase throughput. So, even a short, single stoppage can make it very hard for the company to make high-quality and cost-effective goods. Downtime not only reduces daily production but also can obstruct delivery commitments and hurt customer relationships. Because of this, frequent stops of machine and long corrective repairs have become a major issue for many companies.

The modern manufacturing industry quickly adopts new technologies in order to create an interconnected, smart, and self-optimizing production systems (Jamwal et al., 2025). Advanced manufacturing facilities are increasingly deploying sophisticated sensor networks, real-time monitoring systems, and complex machine learning algorithms to detect machine failure before it occurs (Preethi et al., 2024). These technologies enable machine maintenance right on time, which decreases the unplanned downtime and increases the lifespan of machines. But the adoption of such advanced system is not common among manufacturing companies. Small and medium-sized enterprises (SMEs) often lack the critical infrastructure and resources, including sensors, data acquisition system, and advanced analytics platforms, to implement the Industry 4.0 technologies. Moreover, the shortage of dedicated research and development teams, insufficient IT expert, and limited awareness of Industry 4.0 benefits limits the ability of SMEs to reduce downtime and adopt predictive maintenance (Schröder, 2016, pp. 11-12).

Despite having these limitations, most traditional manufacturing companies collect two types of operational data: interruption logs and maintenance work orders. The interruption log records the sudden machine stoppage, and the maintenance work order documents the repairs and interventions of each machine. These datasets may not contain the details as high-frequency sensor data, but they provide important information about the equipment behavior, failure trends, and maintenance outcomes. However, they are typically stored in separate systems and rarely analyzed together in a systematic manner.

Therefore, there is a need for practical analytical approaches that will allow traditional manufacturing companies to extract insights from the operational data they already collect. Integrating interruption logs and maintenance records may help organizations better understand downtime causes and support data-driven maintenance decision-making.

1.1 Research background

Plastic extrusion is one of the most widely used polymer processing method. The applications of plastic extrusion ranges from packaging films to intricate profiles (Gogos & Tadmor, 2014). The traditional extrusion machines, mostly single-screw extruders, still dominate in many factories because of their robustness and relatively low cost. The main mechanical elements of extrusion machines are the motors, gearbox, screw, barrel, and cooling units. However, these components are very prone to wear and failure because they run under extreme condition of temperature, pressure, and mechanical stresses (Abeykoon et al., 2021).

There are many reasons for the machine failures, and each of them is different from one machine to another. These may involve issues with mechanical problems, electrical failures, operator errors, raw materials scarcity, and setup activities (Nwanya et al., 2017). Though it is not possible to get rid of all the system failures that cause the downtime, with identifying the root causes and proper maintenance planning, a large portion of these stops can be avoided. An intelligent maintenance policy is essential for effectively reducing machine downtime, as it minimizes the maintenance cost, prevents expensive production shutdowns, and increases product quality. Maintenance activities can be divided into three broad categories: corrective, preventive, and predictive (Duarte & Santiago Scarpin, 2023). Among these, predictive maintenance is the smartest and cost-effective one; it utilizes real-time data and condition monitoring to assess equipment's health. It also uses historical data of the machine to develop machine learning models and unfold the hidden patterns in machine failure. Maintenance is performed only when indicators or ML models suggest an impending failure. This method not only reduces downtime but also minimizes unnecessary servicing and extends machine life (He et al., 2017).

The advancement of Artificial Intelligence (AI) has made it easier for the company to reduce downtime. Companies can now keep an eye on the conditions of their machines in real time and make decision ahead of time, thanks to the invention of sensors, the Internet of Things (IoT), and data analytics. Also, new technology like edge computing and cloud integration have made it possible to quickly and accurately analyze a lot of machine data (Keshta et al., 2025). Machine Learning (ML), a core part of artificial intelligence, has shown its ability to successfully minimize machine downtime by predicting failures. For this reason, many big companies have adopted ML-enabled downtime management systems in their plants. For example, GE Aviation uses AI-backed preventive maintenance systems to keep an eye on its 44,000 in-service engines. This method cut the maintenance cost by 30% and boost aircraft availability (GE Aerospace, 2024). Meanwhile, at BMW's Regensburg plant, an AI-driven system track assembly-line conveyors and prevents around 500 minutes of disruption per year (BMW Group Corporate Communications, 2023).

However, the applicability of such an advanced framework to traditional manufacturing environments, especially traditional plastic extrusion, faces practical challenges. First, there are some financial hurdles that comes with this technology. For instance, the cost of setting modern equipment, upgrading existing systems, or hiring specialized personnel, can be very expensive for many companies. Second, to operate or maintain such advanced systems, companies need skilled labor, and many companies are short of that. Third, data collection is another big challenge for traditional manufacturing companies. The effectiveness of AI-backed advanced systems often rely on the historical data, but historical data are often noisy, incomplete, imbalanced, or simply missing. Many plants do not even keep a clearly labeled record of equipment failure logs (Phongmoo et al., 2025). Therefore, many manufacturing companies prefer lightweight, effective, and low-maintenance solutions that can easily be added to their existing PLC systems (Lei et al., 2018).

The academic research on machine downtime and maintenance has developed separately. Machine downtime studies mainly focus on identifying the reasons that causes the machine stoppage. This type of research uses tools like Pareto chart, fishbone diagram, and overall equipment efficiency (OEE) metrics (Muchiri & Pintelon, 2008). On the other hand, maintenance studies examine work order optimization, spare parts management, and scheduling without linking the operational interruptions that cause the interventions. This separation makes it harder to see how machine stoppages are connected to machine maintenance plan, which reduces the chances for focused improvements.

This research aims to fill these gaps by developing a practical framework that integrates both interruption logs and maintenance data. The goal is to find the main reasons for downtime, calculate important KPI and develop a predictive maintenance model for downtime reduction. The framework is designed specifically for traditional manufacturing companies, which has a lot of data but not much insight. It will help company to take data-driven decisions without requiring major infrastructure investment.

1.2 Research problem and research questions

The unplanned downtime imposes severe economic and operational burden on plastic extrusion manufacturing. A minor failure in the machine, like a broken screw or a burnt-out heater band, can cause the production line to stop for many days. This leads to missed delivery deadlines, wasted raw materials, and urgent, costly repairs. Although modern extrusion machines can avoid costly failures to some extent, the traditional extrusion machines are prone to frequent failures, hence, require a robust framework for downtime reduction.

In many factories, including the case company, the machine interruption data and maintenance log are collected routinely. But they are kept in separate systems and rarely analyzed together. Because of these, important relationships and patterns between machine failures and maintenance activities remain out of sight. Without the integrated analysis, it is difficult to determine which machine has the most downtime, which types of interruptions are more frequent, and which interruptions lead to maintenance. As a result, maintenance decisions are mainly dependent on experience rather than data-driven insights.

Therefore, this thesis aim to solve the problem of how machine interruptions data and maintenance logs can be integrated and analyzed together. This will help to understand existing downtime patterns, identify root causes, and develop strategies for unplanned stops. The goal is to design a framework that can effectively reduce the downtime and optimize maintenance work orders of traditional plastic extrusion machines.

To address this problem, the research will try to answer the following questions:

1. What are the most frequent and severe causes of machine interruption in the case company?
2. How can key performance indicators (KPI) and interruption data be used to assess machine health and identify high-risk machines?
3. How can predictive modelling and descriptive analysis be integrated together to develop a decision-making framework for maintenance activities?

1.3 Research objectives

The goal of this thesis is to develop a data-driven framework for the case company to reduce its overall machine downtime and develop a futuristic maintenance plan. To reach the goal, the main objectives are:

I. To review literature on downtime reduction and predictive maintenance, identifying gaps specific to traditional plastic extrusion.
II. To clean, standardize, and integrate the historical data of machine interruptions and maintenance work orders.
III. To perform a descriptive analysis, like Pareto analysis, to identify the main reasons for the machine downtime.
IV. To develop a predictive maintenance model for critical machines by exploring the historical interruption patterns and maintenance logs.
V. To provide practical recommendations and decision-making tools to reduce the machine downtime in manufacturing operations.

1.4 Scope and limitations

This research mainly focuses on the plastic extrusion lines at the case company. The company keeps records of machine interruptions and maintenance work orders. This study used the historical data of machine interruptions, production data, and maintenance logs from April 2024 to September 2025. There are many reasons for machine stoppage, such as equipment issues, quality issues, and raw material issues. But for this study, only the equipment issue (tool/calibrator, extrusion line, or other equipment) is considered.

However, there are a few limitations of this study. First, the historical data that are used in this research were primarily recorded for operational purposes, not for research purposes. For that reason, the dataset may contain missing information, inconsistencies, and limited details. Second, the study does not use any advanced sensor monitoring data, such as machine vibration or temperature. Therefore, the analysis is based on operational records and the operator's perspective. Third, the predictive analysis is limited to a few machines because the historical data is not enough to develop a predictive maintenance model for all machines. Finally, the research is done for a single case company, so the findings may not be generalizable to all manufacturing companies. However, the results provide valuable insight into the production system of the case company and the similar industries that collect the interruption data and maintenance log can adopt the framework.

1.5 Case company description

The study was conducted at a local company in Finland that operates in the field of plastic extrusion. The company produces plastic profiles with intricate shapes that is used in various industries, such as construction industry, door and window industry, greenhouse industry, transportation industry and energy industry. The production system is consist of multiple extrusion lines and each functions as an independent work center. The long extrusion process is used to produce the plastic profiles in the company and to maintain the quality during the whole process the company needs reliable machine. However, the sudden machine stoppage or unplanned downtime often hinders the company’s effort to produce quality products. Therefore, there is a need of implementing smart maintenance strategy.

The interruption logs and maintenance work order are recorded by the production operators and maintenance team. The interruption log contains information, such as downtime events, interruption notes, duration, start date, end date, etc. Whereas, the maintenance report contains information, including failure events, reason of failure, part details, etc. However, the data is recorded for operational purposes, not for research purposes, which leads to variations in data. The company provides a suitable case study as it maintains historical data on interruption log and maintenance report. This allows the research to explore how can operational data be analyzed to make data driven maintenance decision.

1.6 Structure of this thesis

The thesis consists of six chapters. Each chapter covers specific part of the research, and together they make a clear and logical flow of the study. Chapter 1, Introduction, provide the information on the research background, research problem, objectives, scope, and limitation of the study. Chapter 2 represents the literature review, where previous research on similar topic has been studied, and a theoretical framework has been developed. The three main areas of this chapter are the downtime analysis methods, maintenance strategies, and operational performance measurement. The methodology of this research is discussed in Chapter 3. How the data is collected, what steps are taken for processing the data set, and what methods are used to analyze the data are mentioned in this chapter. Chapter 4 represents the results and outcomes of the analysis. It shows dominant causes of downtime using Pareto analysis, maintenance log analysis, key performance indicator calculations, and predictive analysis of critical machines. In chapter 5, the research questions are answered, the theoretical contributions are discussed, and a clear implementation roadmap is provided. Finally, chapter 6 concludes the summary of the research by highlighting the key findings and what needs to be done to mitigate the unplanned downtime.

2 Literature Review

This chapter reviews the existing academic literature related to downtime analysis, maintenance strategies, and decision-making in manufacturing environments. The main focus was given to the areas such as evaluation of maintenance strategies in small and medium-sized enterprises (SMEs), methods of downtime analysis and machine performance measurements in traditional manufacturing companies, and growing role of data-driven decisions making in industrial settings. Finally, the chapter presents the theoretical frameworks that have been used in this study and the research gaps that the thesis aims to address.

2.1 Maintenance Strategies

The machine maintenance strategies in manufacturing companies have changed drastically over the recent decades. Previously companies used to follow the corrective maintenance, where failures were addressed only after they occurred. But now, companies have moved to predictive maintenance, where the machine failures are anticipated prior to their occurrence by using historical machine data (Molęda et al., 2023). This shift has occurred due to the increasing demand for operational reliability, cost efficiencies and asset longevity in the industrial sector. Generally, maintenance activities can be divided into three broad categories: corrective, preventive and predictive maintenance.

2.1.1 Corrective Maintenance

Corrective maintenance also known as reactive maintenance is the oldest maintenance strategy that is used in the industry. In this strategy, machines are only fixed or get maintenance treatment when a breakdown occurs. Depending on the severity of the failure, the maintenance personnel repair or replace the faulty components. The main advantage of corrective maintenance is its simplicity. It requires minimal planning, little monitoring and low initial investment, which makes it very convenient for the industrial use (Cao et al., 2025). However, this benefit comes with some hidden costs, which sometimes outweigh the benefits.

According to Gholipour (2025), although reactive maintenance is easy to apply, it often times leads to unplanned machine downtime and expensive machine repair due to emergency. This problem is more serious in continuous production lines like plastic extrusion, glass manufacturing, or food processing. In continuous production lines, the machines are expected to running for no-stop, but when failure occurs, it not only stops the machine but also disrupts the whole process. The unexpected stoppage of machines leads to delayed shipments, missed production schedules and other financial losses (Jardine et al., 2006). Moreover, emergency machine repair is far more expensive than the planned machine maintenance. This is because they often involve overtime labor, urgent purchase of spare parts, and additional damage caused by sudden component failures (Weidner, 2023).

Despite the limitations, corrective maintenance is appropriate for non-critical assets, where failure does not impact the safety, operations and costs, or where preventive maintenance cost exceeds the immediate repairing cost (Gholipour, 2025). But in continuous manufacturing process, where equipment failure is very crucial, corrective maintenance is not that effective or appropriate.

2.1.2 Preventive Maintenance

Preventive maintenance is the upgraded version of the reactive maintenance. It was basically introduced to overcome the shortcomings of the corrective maintenance. In this approach, the maintenance work is done on a regular interval, such as after a specific period, or a specific operating hours. The main goal here is to reduce the probability of machine failure. By maintaining the sudden failure, the organization can minimize downtime, avoid emergency repair and ensure smooth production operations (Basri et al., 2017).

Time-based preventive maintenance has some limitations as well. One of them is that maintenance activities are done on a fixed time interval, regardless of the equipment conditions (Nunes et al., 2024). So, maintenance activities can be performed on a perfectly good equipment, leading to unnecessary costs of labour, materials and spare parts. On the other hand, even though machines are serviced regularly, failures can still happen between the two consecutive maintenance periods, which means failures are not one hundred percent preventable. Because of these drawbacks, this approach may lead to over-maintenance or under-maintenance of the machine. Over-maintenance causes extra cost to the company without clear benefits, while under-maintenance results in unexpected failures with huge losses. For this reason, many companies are shifting from time-based preventive maintenance strategy to condition-based predictive maintenance strategy, which allows the company to prioritize their maintenance decision based on equipment health, not just fixed time intervals (Ciocoiu et al., 2017).

Even with the limitations, preventive maintenance is widely used in various industrial settings. This is because it is easy to apply and needs less resources than the condition-based predictive maintenance. In predictive maintenance, companies need extra sensors, continuous data collection process and analytical tools to monitor the condition of machines. Therefore, traditional manufacturing companies that do not have advanced data processing tools mostly rely on preventive maintenance strategies.

2.1.3 Predictive Maintenance

Predictive maintenance (PDM) tries to overcome the weakness of preventive maintenance. By using historical machine data, predictive maintenance estimates when a failure is likely to happen. Besides, unlike corrective maintenance, instead of acting when a problem is detected, predictive maintenance analyzes the trends of machine behavior over time and plans maintenance strategies accordingly. This approach enables the company to reduce unplanned downtime and avoid unnecessary early intervention of the machines (Zonta et al., 2020).

Predictive maintenance builds on three key elements: data, analytical models and performance evaluation. These three elements make the approach more proactive and pragmatic. As the complexity of machines has evolved in modern manufacturing environments, predictive maintenance has also changed its way of analysis based on three key elements. Previously, the method mainly relied on physical and mathematical models to analyze how machines degraded over time. Today, predictive maintenance uses data-driven solutions based on machine learning and artificial intelligence. This shift highlights the robustness of the predictive maintenance strategy. With high-quality data and well-maintained models, predictive maintenance can accurately anticipate the machine failures (Zhu et al., 2026).

Recent review studies have shown that research on predictive maintenance is growing quickly and advanced manufacturing companies are adopting this strategy rapidly. A systematic review by Resanovic & Balc (2024) examined the importance of predictive maintenance in the era of Industry 4.0. The review discuss the use of artificial intelligence, machine learning, and deep learning techniques in PdM applications. It also identifies the new methods, such as transformer model and self-supervised learning have great potential in improving the performance of predictive maintenance. Finally, it suggested that companies should adopt this maintenance strategy to increase the production efficiency and survive in the global competition.

Similarly, Tsallis et al. (2025) conducted a systematic review study where they mentioned how machine learning-based predictive maintenance is applied in different industries. This study reviewed 60 peer-reviewed articles between 2020 and 2024. The results showed that if machine learning technique is combined with the data from different sensors in manufacturing plants, the prediction accuracy of PdM increases significantly. The study also mentioned that the commonly used open dataset for predictive maintenance are CMAPSS (commercial modular aero-propulsion system simulation), MIMII (malfunctioning industrial machine investigation and inspection), and SECOM (semi-conductor manufacturing). However, challenges like data imbalance and differences in data structures still hinders the advancement of predictive maintenance.

U. Khan et al. (2026) conducted another comprehensive review study on predictive maintenance method based on deep learning. In this study, they provided updated information on the application of preventive maintenance in different industries and its strengths and limitations. The survey showed that deep learning can significantly improve the maintenance tasks, such as remaining useful life prediction, anomaly detection, and fault classification. They also showed that in some cases, deep learning is better than the machine learning approach especially when dealing with complex data.

2.2 Advanced Maintenance Challenges for SMEs

The adaptation rate of the advanced maintenance strategy, such as predictive maintenance, is not the same across all industries. While large companies easily adopt the method with the help of IoT, machine learning and data analytics, small and medium-seized enterprise (SMEs) often face unique challenges in implementing it (M. Khan et al., 2022). In reality, the number of financial resources, technical infrastructure, and skilled personnel in SMEs is less than that of big organizations, which makes it more difficult for them to adopt Industry 4.0-based maintenance strategies (Nasirinejad et al., 2025).

Although there are a lot of research studies have been conducted on predictive maintenance strategies for large industries, SMEs domain has been neglected in these studies. A recent study on predictive maintenance in SMEs found that the existing literature on PdM doesn't focus on small and medium-sized enterprises, particularly the financial part of the SMEs (M. Khan et al., 2022). The study also identifies several challenges in this field, such as limited access to quality data, lack of monitoring of PdM systems, and shortage of skilled personnel. Furthermore, in another comprehensive study on the adoption of smart maintenance in SMEs, the authors mentioned that although smart maintenance has been studied since 2011, very little focus has been given to the attributes of SMEs (Nasirinejad et al., 2024).

Another challenge of implementing predictive maintenance in small and medium enterprises is digitalization (Grooss, 2024). Unlike large corporations the SMEs have different levels of digitalization which makes it difficult to connect one system to another or bring them all together in a common ground. This leads to some technical hurdles such as, data becomes locked in multiple storages, different software systems cannot communicate, and the available information is incomplete or inconsistent (Fathi & Wernher van de Venn, 2024). The accuracy of failure prediction in predictive maintenance largely depends on the data quality and software compatibility, because the AI model is trained on data. Therefore, if these issues are not addressed, the PdM strategy will remain ineffective in the SMEs domain. Addressing these challenges require robust data management strategies, including data integration, cleansing, and preprocessing (Surendran Pillai et al., 2026).

Despite all the challenges that SMEs have faced, they play a vital role in developing economic growth, creating jobs and promoting innovation. Smart maintenance adaptation can be a breakthrough for SMEs, because it can enhance the competitive performance of SMEs by minimizing cost, increasing reliability and improving efficiency. Therefore, there is an urgent need to take a practical approach that allows the traditional manufacturing companies to analyze their existing production data and implement smart maintenance.

2.3 Downtime Analysis in Manufacturing

Machine downtime has become a big issue for modern manufacturing. Breakdown means machine stoppage, and as long as a machine remains idle, there is a loss of production, delayed delivery and poor cost performance. Therefore, companies are always in search of the best solutions to minimize machine downtime. Previous studies rigorously investigated the cause and characteristics of downtime using descriptive analytical tools. The most common tools used in downtime analysis are Pareto analysis, root cause analysis, failure mood and effect analysis, statistical process control, and overall equipment effectiveness (Muchiri & Pintelon, 2008).

Pareto analysis is one of the most extensively used downtime analysis methods. This approach is based on the 80/20 rule, which means 80 percent of the problems come from 20 percent of the causes. In different sectors like manufacturing, energy, transportation and healthcare, Pareto analysis is used in identifying the severe reasons of a problem. For example, ALMashaqbeh & Hernandez (2024) had conducted a study to reduce downtime in a plastic production line. They applied Pareto analysis on six the big losses (downtime categories) of the line and found out that only a few failures are responsible for the most machine stoppages. This study allowed the case company to focus its resources on only the crucial problems. Similarly, in another study conducted by Muchiri & Pintelon (2008) on maintenance performance measurement, they mentioned that the Pareto tool is very easy to apply because it does not need any software or advanced statistics.

Root cause analysis (RCA) is another technique that is widely used to evaluate the reason for downtime. Unlike Pareto analysis, which only identifies the most frequent reasons for downtime, root cause analysis identifies why failure happened in the first place. The main purpose of using root cause analysis is to prevent the occurrence of failure by addressing its underlying causes. Many research studies have been conducted on the effectiveness of using RCA technique to minimize downtime. For example, Wolniak (2019) studied a car factory to identify the machine stoppage. He used the 5-why technique and the Ishikawa diagram to identify the main causes of downtime. At last, the study concluded that RCA techniques can significantly reduce the downtime of a production line. 5 whys is the most common tool of RCA techniques, it helps structure the analysis. In this approach, an operator is asked ‘why’ a failure occurred until the root cause of the problem is identified. RCA is a well-established and systematic downtime analysis tool; because of its ease of use and effectiveness in failure detection, it is widely applied in large companies and SMEs.

Overall equipment effectiveness (OEE) is generally used to measure the performance of machines. OEE considers three things in a machine: first, how often the machine is available, second, how fast the machine is running and third, what is the quality it produces (Muchiri & Pintelon, 2008). Among these three pillars of OEE, availability is directly linked to the machine downtime. Availability is calculated by dividing the actual machine runtime by the planned production time. If the machine availability is low, that means time has been lost due to some unexpected machine breakdown. Therefore, by tracking the availability, companies can analyze the downtime. For example, Setiawan et al. (2022) conducted a case study to analyze the downtime of a PVC compound industry. To identify the downtime, they first calculated the availability rate (76.72%) of the production line, which was well below the standard (85%). By deeper analysis, they found that 'setup and adjustment' was the main reason for the downtime and suggested standardizing the cleaning process and implementing TPM.

However, OEE has some limitations when used alone. OEE is only a score which depends on availability, machine speed and quality. A machine can have a high OEE score, but still have some unexpected stops, which can only be known from machine breakdown logs. Many researchers have suggested that overall equipment effectiveness analysis should be paired with other downtime analysis tools. Therefore, this study focuses on individual KPI (Availability, MTBF, MTTR, downtime, frequency) rather than relying on a single OEE score. The study also incorporates Pareto analysis to evaluate the downtime, and together these two approaches provide a comprehensive and transparent insight into downtime minimization.

2.4 Decision Framework for Downtime Reduction

One of the most important steps in downtime analysis is taking decision. A good decision-making framework can help the maintenance team to turn the raw data into some actionable recommendations. There are some frameworks that researchers have been using in their studies to analyze downtime and generate practical recommendations for machine maintenance. Among them, the layered framework is very famous, it consists of three levels: data collection, analysis and actions. In this approach, first, various data is gathered from machines, logs and other systems. Second, the raw data is analyzed using different statistical tools or machine learning. Finally, based on the analysis, a decision is taken about when a repair should be scheduled or when a parts should be ordered (Fathi & Wernher van de Venn, 2024). This layered framework is very helpful because it separates the information that is needed for the analysis from the actions that need to be taken. Besides, it allows the companies with low digitalization to start at the first level and gradually move up.

Another well-established framework is risk-based decision making. In this approach, maintenance decision is taken based on machine risk instead of machine failure prediction. Risk is calculated by multiplying the probability of failure with consequences of failure. In the study conducted by Gholipour (2025), it is found that the risk-based decision-making framework is very convenient for SMEs to reduce machine downtime. SMEs are often lack of digital tools to predict machine failure accurately, hence, by focusing on the high-risk failure mode, they still can reduce the downtime significantly. For example, a machine that fails frequently (high probability of failure) can be fixed within five minutes. On the other hand, a machine that fails rarely but takes days to be fixed (high consequence) may receive much attention compare to the machine with high probability of failure.

Condition-based trigger is an advanced decision-making framework; in this method, maintenance activities are started when machine health such as temperature or vibration, reach to a predefined threshold. The machine health is determined automatically with the help of sensors or manually with the help of operators. In a study on condition-based maintenance, the authors Zhao et al. (2025) mentioned that it is critical to determine the appropriate threshold level. If the threshold level is set too low, the maintenance activities will be triggered frequently which is a waste of resources. If the threshold level is set too high, the failure will occur before the trigger is activated. Hence, it is essential to find out the right balance. Generally, historical data is used to determine the threshold level, the more the data, the accurate the level. SMEs can also use this framework without setting expensive sensors. In this case specific operation hours or number of production can be used to set threshold level.

Besides these three widely used frameworks, researchers use other conceptual models as well for decision making, such as the economic decisions model and the hybrid framework. In the economic model, the maintenance decision is taken based on cost rather than failure probability. The basic rule for this model is that if the cost of failure exceeds the cost of preventive actions, the maintenance will be performed, otherwise not (Singh et al., 2021). Hence, the framework requires an accurate cost estimation of the downtime to make better decisions, which the SMEs often do not track. Alongside this economic model, many researchers have used hybrid frameworks for maintenance decision-making. A hybrid framework is a combination of different models. For example, a hybrid framework may use a risk-ranking framework to determine which machine to work on, a condition-based trigger to decide when to act and an economic model to choose between repair and replacement.

Choosing the right framework is crucial for the company to make the right decision. But it is not easy, it depends on many things, such as company's digitalization level, data availability, and budget. For the case company, a hybrid framework has been selected, which combines the risk ranking model with the Pareto-based root cause analysis. Risk-ranking model tells which machine carries the highest risk of failure and the Pareto analysis tells what could be the underlying cause of the failure. By using these information the framework generates practical recommendations.

2.5 Theoretical Framework

The theoretical framework is a set of theories and ideas based on which the research is developed. This study is grounded in three theoretical perspectives that guides the analysis of machine downtime at case company. The reliability engineering theory provides the mathematical foundation for quantifying the key performance of the machines using mean time between failure (MTBF), mean time to failure (MTTR), and availability metrics. The root cause analysis theory provides the analytical support to identify the dominant downtime reason of each machine through Pareto analysis. Finally, the decision theory helps translate the analytical results from the reliability engineering theory into actionable recommendations. Together, these two theories create a robust tool through which historical downtime data is analyzed and a practical decision-making framework is developed.

2.5.1 Reliability Engineering Theory

Reliability engineering theory provides the mathematical foundation to calculate the KPIs used in this study. The theory focuses on measuring the likelihood of a machine to work for a certain period of time without breaking down. From this framework, three key metrics are used: MTBF, MTTR, and availability. These metrics are widely accepted for academic uses and industrial practices. They help the company to identify how reliable their maintenance policy is. MTBF is the mean time between two consecutive failures. It is calculated by dividing the total operating time by the total number of failures in a period. If a machine shows high MTBF scores, it means the machine is reliable (Smith, 2011). MTTR refers to the time a machine needs to get fixed after a breakdown. It covers all the time from breakdown to machine restart, such as problem diagnosis, getting spare parts and repair time. A low MTTR score means the maintenance team is efficient in solving problems. Availability means what fraction of the total time the machine was actually running. It is calculated by dividing the machine operating time by the planned production time (Birolini, 2017). These three metrics are mathematically connected, affects in one can lead to changes in the others. In this study, the metrics are calculated for each machine on a weekly basis. The combination of these three KPIs provide a complete picture of the machine health.

2.5.2 Root Cause Analysis Theory

Root cause analysis (RCA) is a systematic problem-solving method where instead of simply addressing the symptoms, the underlying causes of the problem are identified. In the manufacturing and maintenance context, RCA aims to determine the main reasons for a failure to occur and help organizations to take the necessary steps to prevent the failure from recurring. The premise of RCA is that a problem has multiple reasons, but not all reasons have equal contributions, and only a few are the root cause (Pietsch et al., 2024). Pareto analysis and fishbone diagram are two widely used root cause analysis tools. Pareto principle, also known as 80/20 rule, assumes that approximately 80% consequences of a problem comes from the 20% reasons. It is calculated by ranking the reasons in descending order based on their impact (e.g. total downtime, frequency of occurrence). In the maintenance context, Pareto analysis is done to find out the "vital few" reasons of machine failure and take action accordingly, this help the maintenance team to allocate the limited resources properly (Fadeyi et al., 2016). Another tool, fishbone diagram, also known as Ishikawa or cause and effect diagram, is used to systematically explore the potential causes of a specific problem. In this method, the problem is placed on the head of the fish and its causes are written on the bones of the fish. This study only used the Pareto analysis from RCA theory. By ranking the categories by total downtime, the "vital few" reasons are identified and maintenance strategies are planned accordingly.

2.5.3 Decision Theory

Decision theory provides a framework where results from different analyses are combined together to generate practical recommendations for the problem. The main purpose of the decision theory is to support the decision maker to take decision based on available information rather than intuition or impulse. In maintenance context, the decision theory is widely used because it helps prioritize the maintenance tasks based on risk, cost and available resources. In this study, the decision theory is applied in a practical way by combining risk assessment, root cause analysis, and action planning. First, the machines are categorized into high, medium and low risk using the probability from a predictive model. Then, the dominant cause of downtime of each machine is determined by Pareto analysis. Finally, based on the machine risk and the underlying cause, maintenance priority has been set. This approach ensures that machines problems are addressed according to their needs and at the same time, maintenance resources are used efficiently.

2.6 Research Gaps and Study Focus

The literature review reveals three main research gaps. First, most of the studies conducted on predictive maintenance assume that factories have different sensors, data collection processes and analytical tools. But this is not the case in SMEs, most of the small and medium-sized enterprises are lack of such resources. Second, downtime analysis usually focuses on the identification of the problem that causes the machine breakdown. On the other hand, the maintenance strategy-related study focuses on scheduling the maintenance work order in an efficient way. These two domains are rarely combined together, hence, the organizations know what causes the failure but don't know how to integrate the knowledge into the maintenance strategy. Third, predictive maintenance strategy has been extensively used in high-tech manufacturing companies, but the traditional plastic extrusion lines have always been understudied.

The study develops a simple, data-driven framework for traditional plastic extrusion machines. No sensor data or advanced data collection tools are needed for this study. Only the machine breakdown data and maintenance logs are used here. These two datasets are very basic and most of the SMEs collect these. The framework combines the descriptive analysis, such as Pareto charts and KPI tracking, with predictive modelling (Random Forest) to rank the machine risk and generate practical recommendations. This way, the study integrates two distinct domains- downtime analysis and maintenance strategy- to develop a unique solution for the company.

3 Methodology

This chapter outlines the research methods used in this study. It describes each step from data collection to model development. The tools that are used in the data analysis process are also discussed in this section.

3.1 Research Methods

The study adopts a quantitative, data-driven case study approach to develop a framework for reducing downtime in plastic extrusion lines. The research is conducted at a single manufacturing facility, which allows to analyze the real-time data of machine interruption and maintenance. The case study design is used because the study explores how existing operational data can be turned into actionable insights. The research is also aligned with the elements of design science research (DSR), as it develops an artifact in the form of a downtime reduction framework. The design science research aims to build and evaluate artifacts that can be used to solve real-world problems (Hevner et al., 2004). In this study, the artifact is the framework that integrates descriptive analysis and predictive analysis to develop a model that supports maintenance decisions. The methodological approach used in this study is depicted in Figure 1.

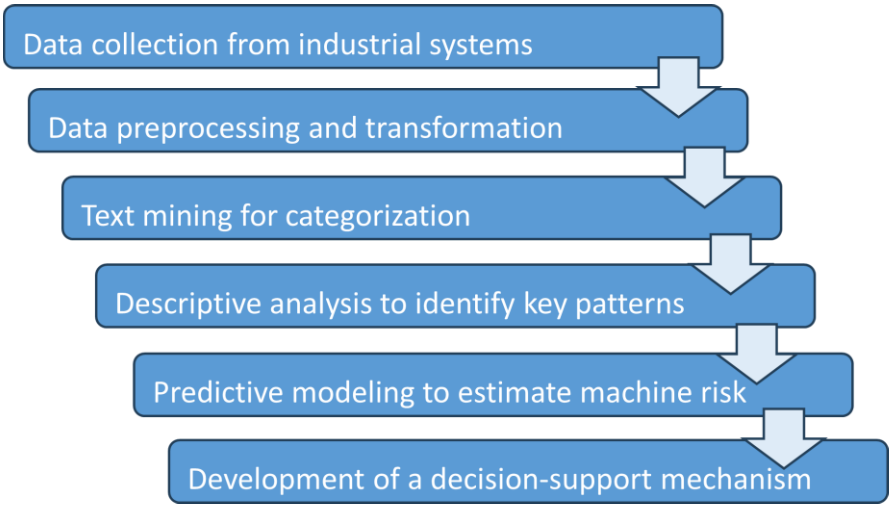

Figure 1. Research Method Steps.

3.2 Data Collection

Three primary data sources are used in this study. Data were collected from the company's ERP system. The description of the data is as follows:

- **Interruption clocking logs** – It contains the records of every machine stop from 1st April, 2024 to 30th September, 2025. It also includes timestamp (start and end), downtime duration, work center, parts no., and free-text interruption notes written by operators. It provides a comprehensive view of machine behavior.
- **Production data –** The dataset contains valuable information about actual production timestamp, planned production timestamp, machine runtime, machine setup time, machine productivity, etc. These records are used to calculate availability and other time-based KPIs.
- **Maintenance logs/Fault report –** The maintenance dataset is recorded by the maintenance department. It includes valuable information like work center (machine), actual start and finish times, maintenance duration, priority description (critical, high, medium), fault descriptions, etc. Not all interruptions require maintenance intervention, some issues are resolved by the operators. This dataset only contains those cases where maintenance personnel were involved.

All three data were extracted from the company's internal database covering a historical period of almost 17 months. No new sensor were installed for data collection to ensure that the framework can be applied in existing setting without needing any extra resources.

3.3 Data Preparation and Processing

In this study, each production line or work center is considered as a single machine and used as a unit of analysis. It helps to maintain the consistency across multiple datasets. The interruption and maintenance events happened in different parts of a machine, but the datasets are not fully aligned with the part-level information. Although the interruption logs maintain the part-level details, the maintenance report does not. As a result, the integration of these two datasets on a common term has become very difficult. Additionally, the same extrusion dies, calibrators and other equipment can be used in different production lines, and there are a large number of these parts involved in the production systems. Hence, it increases the complexity of the study if part-level analysis is considered. Therefore, to ensure a common basis of analysis, the work center (machine) level was selected.

The raw data consisted of individual events which were both noisy and difficult to analyze. To make the data easy to analyze and consistent across all the datasets, a 7-day time window was created. After the creation of the window, all the valuable information across multiple datasets was recorded within the time period. The weekly aggregation helped capture the important machine behavior trend, reduce data fragmentation and support predictive modeling.

All three datasets are structured according to the 7-day windows. From the interruption logs, total hours of machine stoppage and number of events within the time period were calculated. In the production data, the machine runtime is provided based on production orders, but for the ease of analysis, the runtime was assumed to be uniformly distributed across each 7-day window and calculated accordingly. From the maintenance log, only the frequency of the event was calculated. The final dataset was organized in a machine–week format, where each row represents one machine in one time window. This type of format is suitable for machine learning and predictive analysis.

3.4 Text Mining for Downtime Classification

Text mining technique was applied to the downtime notes to identify and organize the downtime reasons. First, the notes were translated and cleaned using a Python script, then important keywords were extracted from the notes using the TF-IDF technique, and lastly, by implementing rule-based classification, the similar keywords were grouped together and assigned to different categories. The categories were then used to simplify the downtime analysis, support Pareto analysis and provide input to the final framework.

3.5 Descriptive Data Analysis

Two main approaches were used in this stage: Pareto analysis and Key Performance Indicator (KPI) analysis. Pareto analysis was applied to identify the most critical interruption categories of machine stoppage. For this analysis, total downtime and total number of frequencies of each category were calculated and organized in ascending order. The cumulative percentage of downtime of each category was also determined to show relative importance.

Five KPIs were used in this study. First, total downtime, which represents the total time a machine was not operational during the 7-day window. It shows the overall production loss caused by the machine interruption. Second, downtime frequency, it represents the total number of interruption events and also indicates whether the interruption is small and frequent or large and rare. Third, mean time to repair (MTTR), it is the average time a machine require to be functional again after a failure. It represents the maintenance efficiency. Fourth, mean time between failure (MTBF), it is the average runtime between two consecutive failures. It represents the machine reliability. Lastly, availability, which is referred to as the ratio of runtime to total operational time.

3.6 Predictive Modeling

The predictive modeling in this study is defined as the forecasting of machine failure in the upcoming time window based on its current interruption behavior. The input features of this model were derived from the interruption dataset. These features were aggregated within the 7-day time window. The selected features include total downtime, downtime frequency, mean time to repair (MTTR), mean time between failures (MTBF), availability and rolling downtime.

The target variable for this modeling was created from the maintenance dataset. It indicates whether the maintenance activity for a machine will happen in next week. If a maintenance event happens in the next week, the target variable is set to 1 and if not, the variable is set to 0. While making the predictions for the next week, the model always considers the information from the previous weeks. That means the model looks at how the machine has been performing for a certain period of time and predicts the future outcome accordingly. This approach closely reflect the real-world decision making, where future action is made based on past observations rather than assumptions.

The model was developed using Random Forest classifier. Random Forest is an ensemble learning method which constructs multiple decision trees and take the output of the trees to improve the performance of the prediction. The reasons behind using this random forest model are its ability to handle inconsistent datasets and nonlinear relationships between variables. The model performance was evaluated by classification metrics, including accuracy, precision and recall. In addition to these outputs, the model also provides the probability of maintenance events for each machine. These probabilities are later on used to develop a risk-based ranking in the decision support stage.

3.7 Decision-Support Framework

The machine failure probability generated from the predictive model was ranked as high, medium, and low risk. From the machine-specific Pareto analysis, dominant categories of downtime were extracted. By combining the machine risk and root cause of downtime, a decision-making framework was developed. The aim of the framework was to generate targeted recommendations for each machine considering these two factors.

3.8 Validity and Reliability

The use of real industrial data and setting logical connection between interruption logs and maintenance reports strongly supports the validity of this study. Also, the application of key performance indicators (KPI) and other analytical tools enhances the credibility of this research. But the incomplete maintenance reports and small dataset limit the validity of this thesis to some extent.

The study is reliable because it follows a standard 7-day time window to perform all its calculations. Also, the data processing procedures remain consistent across all analyses. In addition, the results of this study can be replicated in the same manufacturing setting. Therefore, the structured data, standardized processing tools, and reproducibility of the results ensure the overall reliability of the findings.

4 Results

The findings of this study from the text mining, descriptive analysis, and predictive modeling are presented in this chapter. The results are organized in three main parts. First, the results from the text mining and descriptive analysis are presented to understand the behavior of interruptions. In the second part, the output from the predictive modeling is evaluated. Finally, the decision-making framework has been developed using the results from the descriptive analysis and predictive modeling.

4.1 Text Mining Results

The notes for machine stoppage in the interruption logs were originally written in Finnish. To better understand the notes and enable effective text mining, the notes were translated into English using a automated Python-based translation process. The translated text was noisy and full of redundant entries. So, to ensure consistency, the text was cleaned by converting all words in lower case, removing special characters and filter out irrelevant terms. Figure 2 shows the Finnish to English translation by Python.

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

	Clocking Note (Finnish)	clean_text
54	Työkalusta tulee naarmu pintaan purku ja tutki...	the tool will scratch the surface disassemble ...
55	terä hajosi	the blade broke
56	apukuskin puoli hirtti calibrointii kiinni	auxiliar half hinged calibration closed
57	työkalun putsaus apukuskin puoleinen kupariura...	the copper groove on the side of the auxiliary...
58	pora alkaa olla liian kulunut	the drill is getting too worn
59	kierteet menny	the threads are gone
60	Robotti ei toimi	the robot is not working
61	jyrsintä jälki huono	milling trace bad
62	megapakki ei ota kontaktia virrat otin pois/pä...	megapakki doesnt make contact i took the power...
63	linja poikki maquire temppuilee	across the line maquire tricks
64	Ajo poikki extruuderi sammui liian korkean sis...	across the drive the extruder shut down due to...
65	Vesijärjestelmän lauhduttimet piti resetoida v...	the water system condensers had to be reset if...
66	valmis. Ajon lopetus ja työkalun purku	ready stopping the drive and disassembling the...
67	kuskin puolen neula oli katkennut myös koitan ...	the drivers side pin was broken im also trying...
68	porayksikkö täynnä purua ja profiili hypännyt ...	the drill unit is full of chips and the profil...
69	Extruuderin kanssa ongelma. Rasitus korkea eik...	problem with the extruder the stress is high a...
70	ROSKAA SUUTTIMES	garbage nozzle
71	sahan imuri ei ole päällä ja saha takkuaa	the saws vacuum cleaner is not on and the saw ...

Figure 2. Finnish to English translation by Python.

Using Term Frequency-Inverse Document Frequency (TF-IDF) in Python, keywords were extracted from the notes. This step helped transform the unstructured text into structured form. The keywords were then analyzed from industrial perspective and similar words are grouped together to make meaningful downtime categories. 22 categories were formed from the keywords, which means 22 categories of downtime had been selected. The downtime categories and keywords are shown in Table 1.

Table 1 Downtime categories and keywords.

Category	Keywords
Equipment Failure	broken, stopped, failed, shut, died, asleep, working, malfunctioned, repeatedly, cutter, pull, machine, dryer, line, stopped
Tooling Fault	drill, saw, punch, blade, tool, presser, milling, jamming, stuck, down, rising, return, home, slipping, lifting, actuation, roller
Component Failure	bolt, nut, screw, thread, belt, spring, connector, needle, bearing, coupling, fastener, lock, pin, washer, gasket, fitting, clamp
Calibration Issue	calibration, table, calibration piece, water jets, side plates, crooked, bent, scratches, fiberglass, tape, objects
Material Issue	granulate, material, silo, fresh, moist, plastic, pieces, onbone, bag, moisture, contamination, foreign, clump, lump
Robot/PLC Fault	robot, portal, frequency, converter, fault, jammed, hot, stayed, milling, servo, controller, connection, screen, alarm
Subsystem Issue	vacuum, mass, suction, cyclone, iceva, err350, dropping, zero, filter, sucking
Quality Defect	poor, surface, wavy, gloss, rubber, tearing, round, edges, burn, streak, mark, short, wrong, label, barcode, finish, scratch, flash
Electrical/Heating Failure	burned, heating, element, zone, temperature, nozzle, relay, adapter, dryer, tool, rise, error, cable, thermostat, surface
Process Anomalies	fluctuating, pumping, cork, screw, width, length, speed, unstable, vibrating, drift, changed, poor dimensions,
Missing Component	missing, cannot find, no cages, no spare, ran out, hidden, lost, not found, shortage, absent, copper
Setup Issue	setup, hatch closed, belt not pressed, loosened too much, not tightened, hurried, centered, power taken, wrong installation
Extruder Fault	extruder, rpm, strain, high, inlet pressure, melt pump, differential pressure, accessible, whining, barrel, feed throat
Cooling/Water System Failure	water, cooling, flooded, pump, condensers, tank, level, calibration water, weak, no water, hose, leak, overflow
Welding Fault	welding, welder, end plate, gaps, crooked, melted, back wall, leaves gaps, pulls, plates

Category	Keywords
Nozzle Failure	nozzle, seal, leaking, bolt hole, trouser piece, paste, blocked, debris, loose, worn, tip
Sensor/Detection Fault	eye, sensor, blocked, alarm, detected, passed, no alarm, claims, guard, thickness, tolerance, light curtain, recognition
System Fault	alarm, baston, servo, controller, connection failure, error, fault code, diversion, control, screen dark
Material Flow Problem	flow, maguire, hopper, feeding, weighing cup, pelletizer, solidifies, colors, not flowing, bridging
Mechanical Failure	mechanical, structural, leg cracking, frame twisted, roll stand, bent, cracked, deformed, fatigue
Misalignment	collided, askew, hit, between rollers, pushed, out of line, off position, crooked, misaligned, shifted
Pneumatic Issue	air, hose, pneumatic, compressed, leaking, cylinder, valve, blow pipe, cup air, pressure

Once the categories have been defined, a rule-based classification was implemented to the data. By this approach a keyword dictionary in Python was used to categorize downtime notes. The notes were assigned to specific categories automatically based on keyword matching. The categories were reviewed personally to make sure that it reflects the actual meaning of the downtime notes. Figure 3 depicts the downtime notes categorization.

	clean_text	keywords	Category
54	the tool will scratch the surface disassemble ...	tool, nozzle, surface	Quality Defect
55	the blade broke	blade, broke	Tooling Fault
56	auxiliar half hinged calibration closed	calibration, closed	Calibration Issue
57	the copper groove on the side of the auxiliary...	copper	Missing Component
58	the drill is getting too worn	drill, getting	Tooling Fault
59	the threads are gone	gone	Other
60	the robot is not working	working, robot	Equipment Failure
61	milling trace bad	bad, milling	Tooling Fault
62	megapakki doesnt make contact i took the power...	work, power, doesnt	Electrical/Heating Failure
63	across the line maquire tricks	line, maquire, tricks	Equipment Failure
64	across the drive the extruder shut down due to...	drive, extruder, pressure, high	Extruder Fault
65	the water system condensers had to be reset if...	warm, water	Cooling/Water System Failure
66	ready stopping the drive and disassembling the...	tool, drive	Tooling Fault
67	the drivers side pin was broken im also trying...	broken, new	Equipment Failure
68	the drill unit is full of chips and the profil...	profile, drill, blade, cut, unit	Tooling Fault
69	problem with the extruder the stress is high a...	extruder, problem, high	Extruder Fault
70	garbage nozzle	nozzle	Electrical/Heating Failure
71	the saws vacuum cleaner is not on and the saw ...	vacuum, saw, cleaner, stuttering	Equipment Failure

Figure 3. Downtime notes assigned to the category.

The classification results shows that almost every downtime notes in the interruption dataset can be explained by a limited number of categories. Downtime notes are valuable information in the dataset, it represents the true reasons behind the machine failure, but it is difficult to analyze the data in its raw form. Therefore, the classification offers a great opportunity to analyze the data at a granular level. For example, as depicted in Figure 4, now the types of downtime that occurred frequently and their respective durations can be calculated.

Although this downtime classification through the text mining process is mostly automated, it still requires a significant amount of human intervention. Same keywords can appear in different categories, because the same word can have different meanings in different contexts. So, while doing the category, not only the keywords but also the context and meaning of the notes were taken into consideration.

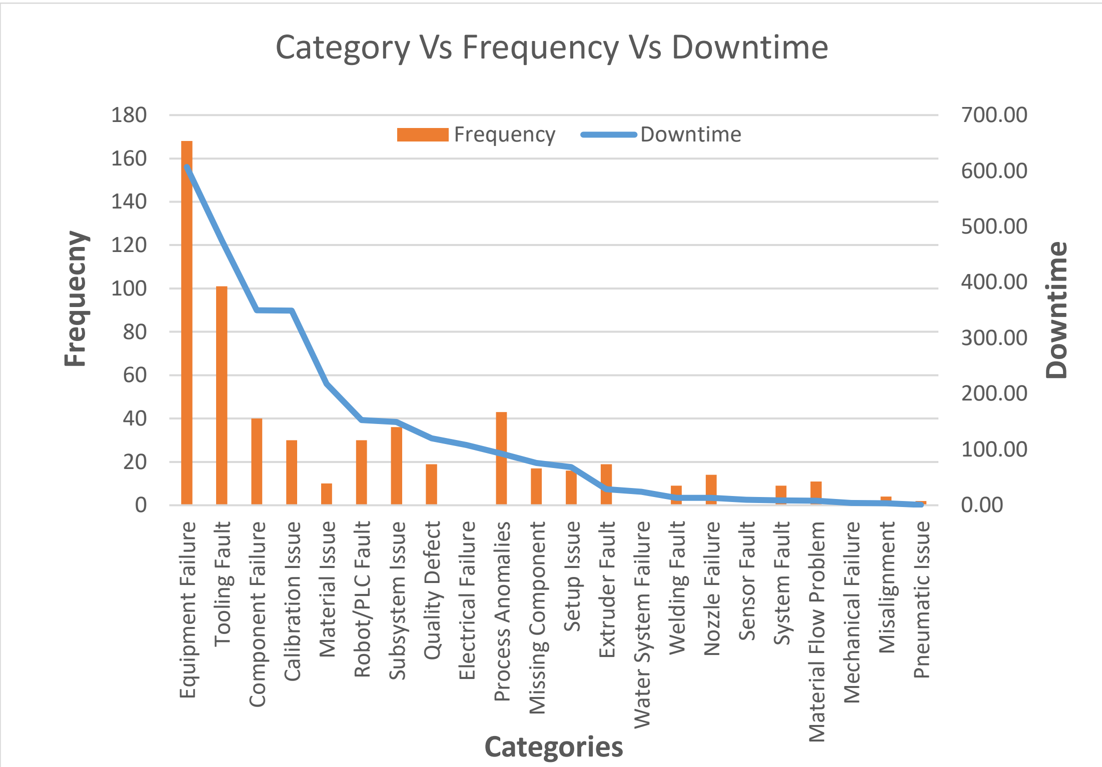

Figure 4. Relationship among interruption categories, frequency, and time.

4.2 Descriptive Analysis Results

The descriptive analysis was performed to understand the characteristics of machine interruptions and their hidden patterns. Using historical interruption data, this analysis tried to find out how often the interruptions occur, how long they stay, and which machine is affected the most. The objective here was to identify the most dominant reasons that caused the downtime and evaluate machine performance using key indicators. Two types of analysis were performed in this section: 1) Pareto analysis of downtime and 2) KPI analysis of machine performance.

4.2.1 Pareto Analysis of Downtime

The dataset has 22 categories of downtime and not all categories are equally important. According to the Pareto principle, also known as the 80/20 rules, 80 percent of problems are caused by 20 percent of the reasons. Therefore, by focusing on top two or three main reasons, a large portion of problems can be solved, which will save both time and resources. In this study, the Pareto analysis was conducted to identify the most significant categories that caused the interruptions.

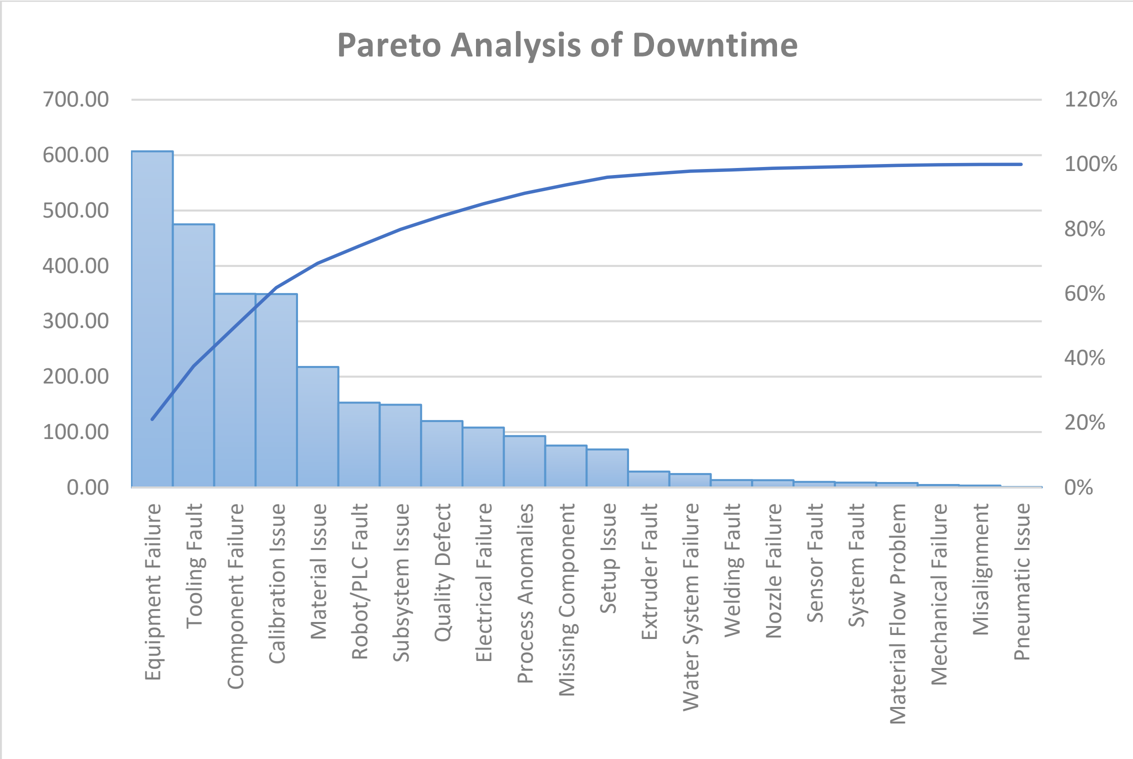

Figure 5. Pareto chart of machine interruptions.

From the results illustrated in Figure 5, it is shown that 5 downtime categories, including Equipment Failure, Tooling Fault, Component Failure, Calibration issue, and material issue are responsible for 70% of the total downtime. These 5 categories are the "vital few" that should be prioritized for improvement.

4.2.2 Machine Performance Based on KPIs

Machine health can provide valuable information about the downtime generated in the machine. Therefore, in addition to Pareto analysis, several KPIs were also calculated to evaluate the performance of the machine. The data structure followed a 7-day time window, which means each KPI of each machine was initially assessed on weekly basis.

Due to the large number of machine-week data, KPIs results are summarized at the machine level in this section. It makes the interpretation easy and presentable as shown in Table 2. For each machine, average values of downtime, frequency, machine runtime, MTTR, MTBF, and availability were calculated across the 7-day time windows.

Table 2 KPI summary per machine.

Machine	Avg. Downtime	Avg. Frequency	Avg. Runtime	Avg. MTBF	Avg. MTTR	Avg. Availability
A	1.90	0.69	6.06	8.76	2.74	0.76
B	1.21	0.59	4.91	8.33	2.05	0.80
C	0.44	0.10	2.42	23.63	4.28	0.85
D	2.90	0.72	9.92	13.82	4.04	0.77
E	10.68	1.03	13.03	12.70	10.41	0.55
F	0.27	0.01	0.17	16.49	16.13	0.51
G	5.39	0.33	4.78	14.33	16.16	0.47
H	0.93	0.38	7.50	19.51	2.41	0.89
I	2.21	0.31	9.06	29.44	7.17	0.80
J	0.90	0.12	2.29	19.83	7.76	0.72
K	0.01	0.03	0.21	8.13	0.26	0.97
L	0.33	0.13	7.38	57.55	2.57	0.96
M	1.49	0.12	2.37	20.53	12.95	0.61
N	0.38	0.40	6.64	16.71	0.95	0.95
O	1.06	0.08	1.20	15.64	13.74	0.53
P	1.05	0.55	12.83	23.26	1.90	0.92
Q	2.94	1.04	18.14	17.47	2.83	0.86
R	5.70	0.88	28.23	31.92	6.45	0.83
S	2.91	0.51	16.02	31.25	5.68	0.85
T	3.37	0.18	2.94	16.36	18.80	0.47
U	1.55	0.22	4.93	22.63	7.12	0.76
V	1.44	0.27	6.87	25.50	5.34	0.83
W	0.21	0.03	0.73	28.37	8.27	0.77
X	0.97	0.56	6.15	10.90	1.72	0.86

From the results, it is shown that the KPIs of the machines are connected to one another, hence, they should be studied together not separately. When one KPI changes, it immediately affects the others because they are linked both mathematically and operationally. For example, if the downtime of the production line increases, the availability of the machine on that line decreases because the machine spends less time on running. This relationship is clearly illustrated in Figure 6.

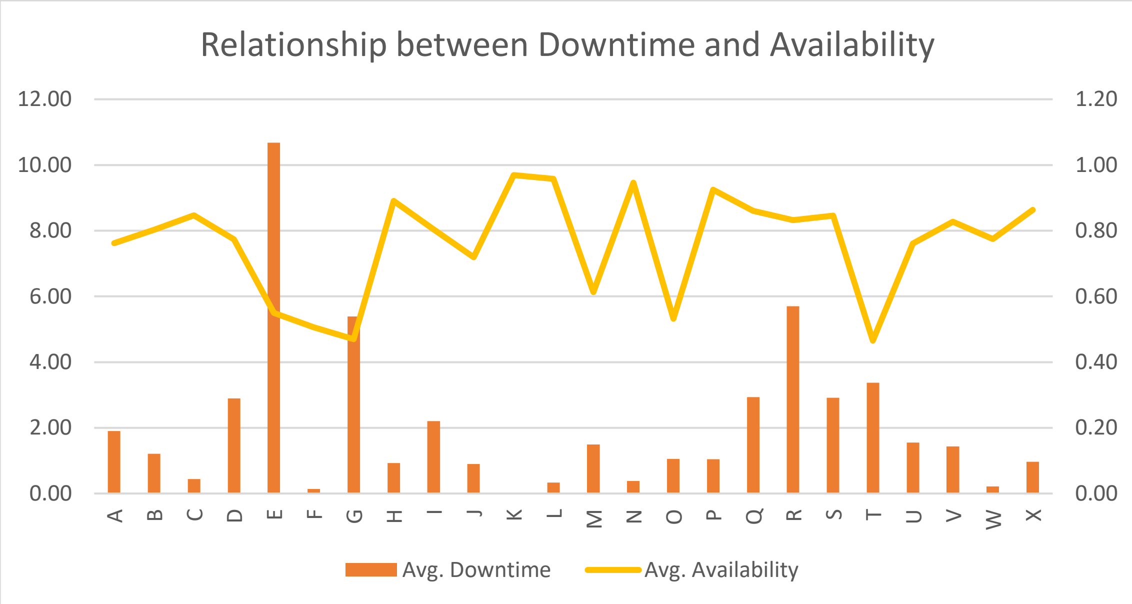

Figure 6. Relationship between downtime and machine availability.

Similarly, higher downtime and failure frequency lead to lower Mean Time Between Failure (MTBF) and higher Mean Time to Repair (MTTR). Which means machines will have low reliability and low maintenance efficiency, respectively. These relationships show that machine performance is a combination of multiple factors, not just a single indicator.

4.3 Predictive Modeling Results

The predictive modeling in this study analyzes the relationship among machine interruption, machine performance and maintenance history in a given week and predict whether a maintenance intervention is needed for the machine in the following week. The prediction was performed at the machine-week level, the input was machine's operational condition, and the output was the probability of machine failure in the next week. The formulation of this predictive modeling allows the company to move from reactive maintenance to proactive maintenance, where the need for machine repair can be known in advance.

4.3.1 Model Performance

The Random Forest model was used to predict whether a machine failure will occur in the following week. The model was trained on 80% dataset and tested on the remaining 20% dataset. The model performance was evaluated based on accuracy, precision, recall, and F1-score. The summary of these metrics is presented in Table 3.

Table 3 Overall model performance.

Metric	Value
Accuracy	0.80
Precision	0.33
Recall	0.14
F1-score	0.20

The model achieved an overall accuracy of 80%, which means a large number of predictions of the model are correct. However, the accuracy alone is not sufficient for evaluating the model performance, because the classes are highly imbalanced. The majority of the observations of the model detected no maintenance events, which can cause the model to be biased toward predicting majority class.

To better understand the model, a detailed classification report was analyzed. Table 4 represents the classification report.

Table 4 Classification report.

Class	Precision	Recall	F1-score	Support
0 (No maintenance)	0.84	0.94	0.89	302
1 (Maintenance)	0.33	0.14	0.20	63

The results have shown that the model performs significantly better in predicting the no-maintenance events (class 0). The precision, recall and F1-score are very high for this class. On the other hand, the model's prediction for the maintenance events (class 1) is very low, with precision and recall score 33% and 14%, respectively. This indicates that a large portion of maintenance events may not be correctly identified by the model.

To further illustrate the performance, the confusion matrix is presented in Figure 7.

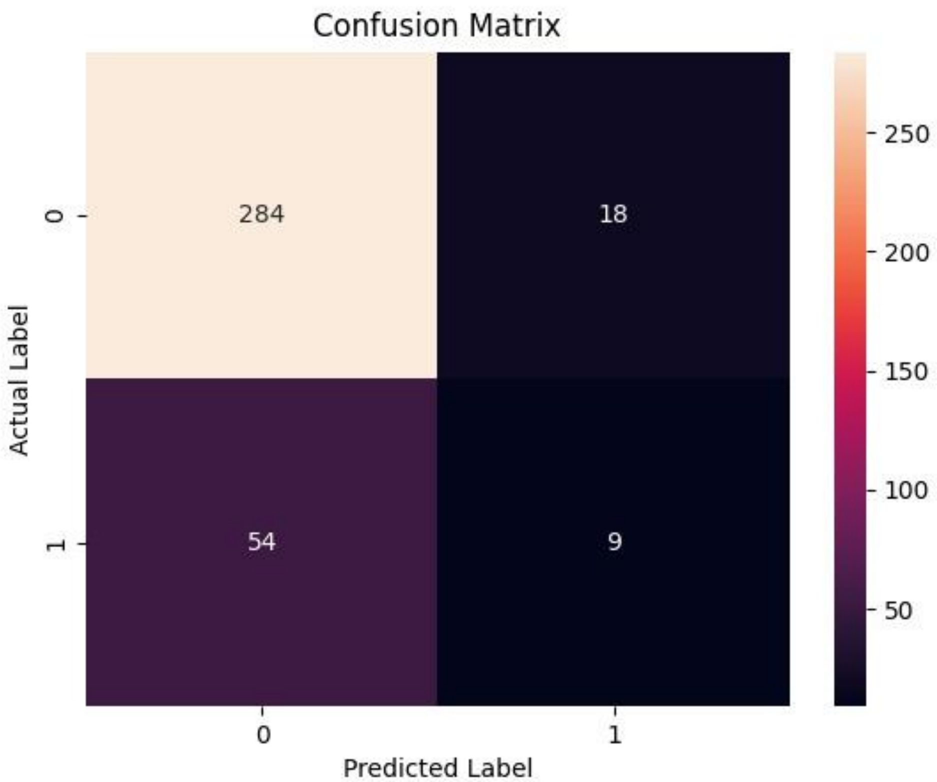

Figure 7. Confusion matrix of model predictions.

The matrix shows that a large number of non-maintenance events (284) are correctly classified. The number of maintenance events (54) that are misclassified (false-negative) is also large. But the number of maintenance predictions (9) that are correct is very small. This imbalance highlights the fact that the model is highly inclined to predict the major class (non-maintenance events).

4.3.2 Impact of Class Imbalance

From the model performance, it is observed that the model is highly influenced by dataset structure. In this study, the machine behavior was summarized in 7-day time windows. But in reality, maintenance activities in the case company didn't occur on weekly basis and were relatively infrequent compared to other machine behaviors.

As a result, most of the time windows do not contain any maintenance events which leads to a highly imbalanced dataset. When the model is trained on such a dataset, it learns to predict the majority class accurately, in this case, the non-maintenance events. But when it comes to learning the minority class, in this case, maintenance activities, the model shows less sensitivity to that, meaning many actual events are missed. This behavior explains why the model has a low recall rate but high accuracy.

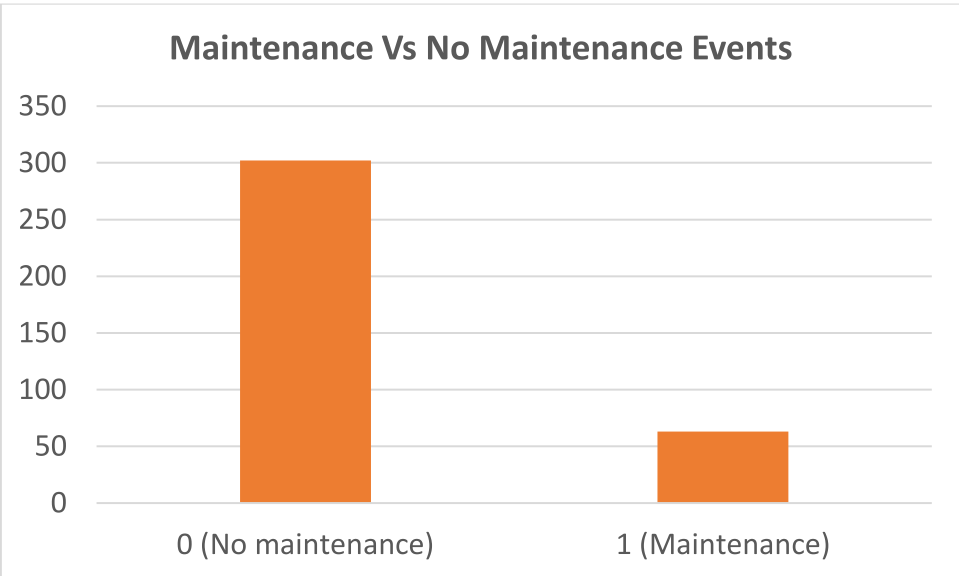

Figure 8. Distribution of Target Variable.

Figure 8 clearly illustrates that the non-maintenance events dominate the tested dataset of the model. This makes it very difficult for the model to learn patterns associated with the maintenance occurrence.

4.3.3 Interpretation of Model Performance

Despite the highly imbalanced dataset and the model's tendency to predict non-maintenance events, the model provides important insights into the machine behavior. The Random Forest model not only tells whether a maintenance event will occur in next windows, but also generates the probability values that indicate how likely a machine will fail. These probabilities as shown in Figure 9, are used to assess the machine risk level. The machine with the higher risk is considered more likely to require maintenance in the next window and prioritized accordingly.

The model is not a standalone tool for failure prediction, instead, it's a component for a broader decision-making framework. The output of the model is combined with results from the descriptive analysis, particularly the dominant downtime categories of each machine. This integration allows the maintenance personnel to take decision not only based on failure likelihood but also on the underlying cause of machine interruptions.

	Machine	Risk_Probability
1840	Q	0.770000
1828	E	0.720000
1844	U	0.680000
1827	D	0.370000
1839	P	0.230000
1826	C	0.135632
1825	B	0.135632
1824	A	0.135632
1846	W	0.135632
1830	G	0.135632
1829	F	0.135632
1833	J	0.135632
1838	O	0.135632
1836	M	0.135632
1835	L	0.135632
1834	K	0.135632
1847	X	0.135632

Figure 9. Probability of machine failure generated by the model.

4.3.4 Feature Importance

Figure 10 shows the feature importance scores generated from the trained model. The value of each feature indicates how much its contribution is in predicting whether a maintenance event will occur in the next week. According to the analysis, Rolling down-time is the most important feature, followed by MTBF and Availability.

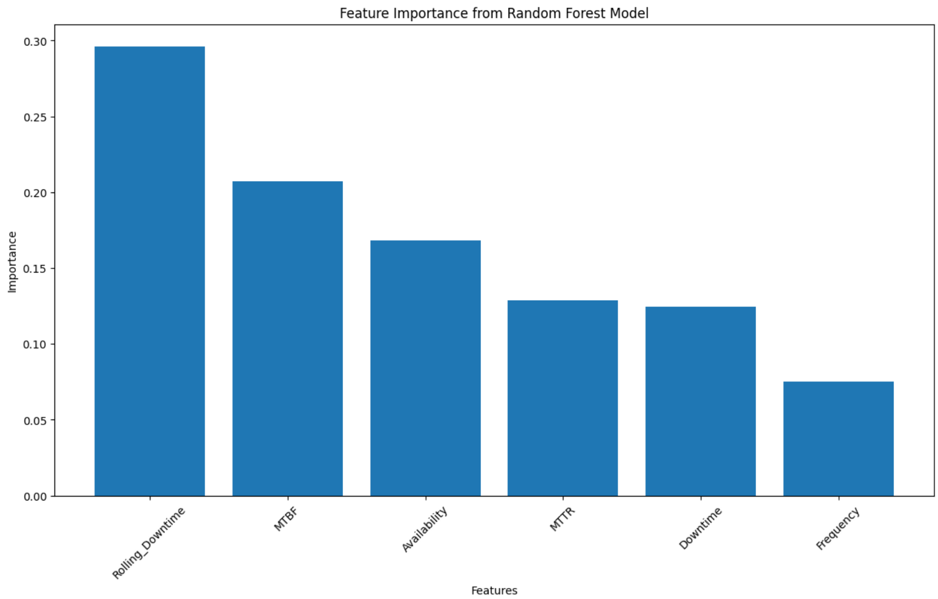

Figure 10. Feature Importance from Random Forest Model.

Rolling downtime is the moving average of total machine interruptions in two weeks. From the figure, it is shown that the rolling downtime is the most influential feature in this model. It has an importance score of approximately 0.30 which indicates that the downtime trends, especially the past two weeks average of interruptions, are the strongest indicators of machine failure.

The second most important feature is mean time between failures (MTBF). It has an importance score of approximately 0.20, which indicates that machine reliability also plays an important role in predicting machine failure. Availability with an importance score of approximately 0.15, further contributes to the model's effort in predicting failures.

Other features, including MTTR (Mean Time To Repair), Downtime and Frequency, have a moderate to least influence on the model. The reason rolling downtime is more important than single-week downtime is because of its increasing interruption patterns. A machine that shows gradually worsening downtime over two weeks requires more urgent maintenance than the machine that shows high downtime in a single week.

4.4 Decision-Support Framework Results

The decision-support framework integrates results from both predictive analysis and descriptive analysis. The output generated in the predictive modelling was mainly used for making prioritization decisions of maintenance activities. Whereas the dominant downtime causes from the descriptive analysis were used to direct the maintenance task according to the underlying machine problem.

4.4.1 Framework Logic

The Random Forest model generates probability scores of machine maintenance in the next time windows. In other words, these probabilities indicate the risk of machine failure in upcoming windows. Using these risk scores, machines were ranked or classified into three categories: high, medium, and low. Risk levels were assigned using the following thresholds:

- High risk: Probability > 0.70
- Medium risk: 0.40 < Probability ≤ 0.70
- Low risk: Probability ≤ 0.40

For each machine, the major downtime categories were identified using machine-specific Pareto analysis. This step helps the decision maker to take action not only based on machine failure risk but also on the underlying interruption cause. Different machines are affected by different major issues, such as tooling faults, component failures, equipment failures, etcetera. Therefore, the maintenance strategy should be machine-specific, not just a uniform action plan for all machines.

The suggested framework combines these two factors, machine risk and downtime causes, and generates targeted recommendations accordingly. High-risk machines are assigned immediate intervention strategies, whereas medium-risk machines are recommended for close monitoring, and those machines that are at low-level risk are suggested to receive minimal attention.

4.4.2 Framework Output

Table 5 presents the complete output of the framework for 10 machines.

Table 5 Decision-Support Framework Output.

Machine	Risk Scores	Risk Level	Dominant Downtime Reason	Recommendation
Q	0.77	High	Tooling Fault	Immediate inspection and maintenance
E	0.72	High	Component Failure	Component replacement and preventive maintenance
U	0.68	Medium	Equipment Failure	Condition monitoring and ensuring the availability of critical parts
D	0.37	Low	Electrical Failure	Temperature monitoring and control
P	0.23	Low	Equipment Failure	Routine monitoring
C	0.14	Low	Extruder Fault	Routine monitoring
B	0.14	Low	Tooling Fault	Routine monitoring
A	0.14	Low	Process Anomalies	Routine monitoring
W	0.14	Low	Calibration Issue	Routine monitoring
G	0.14	Low	Component Failure	Routine monitoring

4.4.3 Interpretation of Results

The results of the framework indicate that a very few machines fall into the high-risk level. Which means that the maintenance effort can be effectively prioritized rather than distributed evenly across all machines.

For instance, machines such as Q and E show a high probability of failure in upcoming week. Also, these two machines are associated with the critical issues such as tooling faults and component failures. Therefore, they require immediate intervention and preventive maintenance to avoid potential production disruptions in future. A moderate-risk machine like U doesn’t need as much attention as high-risk machines. Condition monitoring (e.g., weekly vibration or temperature checks) and ensuring the stock of critical parts, is sufficient maintenance measures for that machine. Lastly, the machines with low level of risk show stable operational behavior, hence only need routine monitoring.

Overall, the suggested approach integrates both predictive insights and descriptive findings to develop a decision-making framework and ensure efficient allocation of maintenance resources.

4.5 Summary of Results

The study aims to develop a decision-making framework for traditional plastic extrusion machines by analyzing the historical data of interruptions and maintenance logs. The machine interruption records were processed first, using text mining (TF-IDF) and manual grouping. Around 800 downtime notes were analyzed, and 22 categories were derived from the notes.

A Pareto analysis was performed on the categories and found that only 5 categories were responsible for 70% of the downtime. Equipment failure, tooling fault, component failure, calibration issue, and material issue are the “vital few” causes of machine interruption. Then, the KPI evaluation was conducted, which showed that machine performance varies significantly across the work center. KPIs are highly interconnected, meaning that if one factor changes, the other factor will change automatically. For example, if downtime increases, the machine's availability decreases.

A Random Forest model was used to predict the machine failure in next week. All the data were structured in a 7-day time window. The model achieved an accuracy rate of 80.3% and a recall rate of 14.3%. The main reasons for the low recall rate are class imbalance and data quality constraints. The model’s output was used to assign the risk level of the machine. If the failure probability is > 0.7, it is assigned a high-risk machine. If the probability is 0.40 -0.70, it is assigned a medium-risk machine and if the probability is <0.40, it is recorded as a low-risk machine.

The decision-making framework takes into account the machine risk generated by the predictive modelling and the dominant root cause of machine downtime generated by Pareto analysis, while making targeted recommendations for each machine. The framework successfully demonstrates how existing interruption logs and maintenance data can be used to make decisions for maintenance actions. It also didn’t require any advanced sensors or costly infrastructure. The combination of predictive risk ranking and descriptive cause identification provides a low-cost decision support tools for traditional manufacturing companies.

5 Discussion

This chapter explains the results presented in Chapter 4 and discusses the significance of them in terms of the research questions. It first answers the research questions and compares the findings of this thesis with prior studies to highlight similarities and differences. The chapter also addresses the theoretical and managerial implications of this study. After that, a detailed explanation of how the case company can implement the framework in practice is discussed. Finally, the limitations of this study are mentioned, and the scope of future work is suggested.

5.1 Addressing the Research Questions

RQ1: What are the most frequent and severe causes of machine interruption in the case company?

Pareto analysis results showed that machine interruption is not evenly distributed across all the categories. From the Pareto chart in Figure 5, the five most frequent categories were identified - Equipment Failure, Tooling Fault, Component Failure, Calibration Issue, and Material Issue – together they account for 70% of the total downtime. Among these top 5, Tooling fault and Component Failure are the dominant categories in high-risk machines (Table 5). The study generates a keyword table (Table 1) for each category using a text-mining process. This table can help the production operators and maintenance staff to classify future interruptions in the right categories.

RQ2: How can key performance indicators and interruption data be used to assess machine health and identify high-risk machines?

The KPIs, such as downtime, frequency, MTBF, MTTR, and availability, were calculated for each machine over 7-day windows. The results presented in Table 2 showed that each KPI is strongly interrelated to the others and together they provide a comprehensive view of the machine performance. For example, if the availability of a machine decreases, the downtime of that machine increases (Figure 6), or if the MTBF decreases, the downtime increases. Therefore, by tracking the trend of performance indicators, companies can take effective measures to curb downtime. The KPIs were fed to the predictive model as features and the model output was then used in a framework to categorize machines in high, medium and low-risk levels.

RQ3: How can predictive modelling and descriptive analysis be integrated to develop a decision-making framework for maintenance activities?

By combining predictive analysis with descriptive analysis, the study provides a unique solution to minimize downtime. Descriptive modelling was used to determine the likelihood of machine failure in the next week (Figure 9). In the final framework (Table 5), the failure probability was used to rank the machine in high, medium and low-risk. Descriptive analysis, on the other hand, was used to find out the dominant breakdown reason of each machine in the framework. The combination of these risk levels and dominant breakdown reason produces targeted recommendations. For example, for machine Q, where risk is high and the major issue is tooling fault, immediate inspection of tooling is recommended. This integration will help the company in transitioning from reactive maintenance to data-driven maintenance strategies.

5.2 Theoretical Implications

The study contributes to the existing theory in different ways. First, it extends the application of reliability engineering theory to low-data environments, especially in SMEs. Reliability engineering assumes that accurate MTBF, MTTR and availability derived from the sensor data or other advanced data sources (Müller et al., 2025). Some studies demonstrated that without detailed data, MTBF and MTTR can still be calculated using service records or condition monitoring data (Semotam, 2024). This study takes the application of the theory further by showing that, using only basic data like interruption logs and maintenance records, effective reliability metrics can be calculated for SME settings.

Second, the study highlights the importance and benefits of combining the descriptive and predictive approaches, whereas the previous studies only focused on these domains separately. Most studies in the field of predictive maintenance tend to anticipate the likelihood of a future failure. Pinciroli Vago et al. (2024) used different machine learning and deep learning models to find out the failure probability of three industrial machines. However, this thesis shows that prediction alone can not lead to a successful maintenance decision. For example, a prediction may indicate a 60% chance of machine failure in the upcoming week, but without knowing the underlying cause, it is hard to plan the maintenance strategy. Therefore, by combining the Pareto analysis of root cause with the random forest model, this study shows that a more practical and actionable maintenance strategy can be developed.

Third, this study emphasizes to consider the historical patterns rather than single performance metrics while making decisions about maintenance strategy. The feature importance analysis of this study shows that rolling downtime is more informative than the downtime itself, which means when the model makes a prediction about future failure, it learns more from the rolling downtime than from the total downtime. This supports the fact that failure occurrence derives from the gradual wear of a machine, not just from some random events. Although some researchers, like Carvalho et al. (2019) has already demonstrated the importance of evolving patterns over single events, but this study provides new empirical evidence from plastic extrusion SME.

Fourth, the implementation of a machine learning model in a real-world manufacturing setting comes with some challenges. Many research studies ignore the challenges, but this study tried to address them. Issues like imbalanced data (very few failures), missing information in the logs and irregular maintenance patterns often significantly affect the model performance. However, most of the research work in this domain, like M. Khan et al. (2022) developed and trained their model using clean and balanced dataset, but this is not always the case. This study shows that even if the model accuracy is very high, it still can be useless if the data quality is poor. Therefore, instead of striving to make the perfect algorithm, focus should be given to the other issues too, like data quality or operational constraints, while making a research model.

Finally, the study generates a simple decision-support framework that any SME can copy easily. Generally, design-based research makes very complex artifacts, which are difficult to implement in real industrial environments. Therefore, this study aimed to make a practical, step-by-step framework, which illustrates how can basic data, like interruption logs and fault reports, be used in making a useful maintenance plan. The steps involved in the framework are: (1) collect interruption logs, (2) calculate weekly MTBF/MTTR/Availability, (3) run a Pareto chart to find top causes, (4) train a Random Forest to rank failure risk, (5) combine risk score with root cause to decide action. This proves that the framework can be implemented in any resource-constrained SMEs.

5.3 Managerial Implications

A weekly risk review session should be incorporated into the current production planning cycle. The framework's output is only useful when it is applied to the practical field, rather than being used solely as an analytical tool. Management should allocate a 15-minute weekly review meeting where the production manager and the maintenance manager can discuss the output of the framework. The framework is not fully automatic; the machine risk and the underlying root cause are automatically extracted from the analytical results, but the recommendations based on these results have to be taken by humans. Therefore, the meeting will help the manager to make decisions about what kind of maintenance work each machine needs. This turns a technical output into an operational routine.

Currently, operators do not record any process parameters at the moment of failure. However, key process parameters, such as zone temperatures, melt pressure, motor current, and screw speed, tell a lot about the machine's health. Therefore, managers should require the operators to record these parameters; this will help the predictive model to learn early warning signs of machine failure. Also, managers should allow a pilot run of some low-cost condition-monitoring devices, such as handheld vibration meters or thermal imaging cameras. The occasional checking of the bearing temperature or vibration level can be a very good source of information for the predictive model. If the pilot is successful, the company can consider installing permanent sensors to collect data on machine health.

Management should implement a closed-loop feedback log for the maintenance activities, meaning that if the technicians find any issues when performing maintenance on a high-risk machine (based on the model's warning), they should note "issue found"; otherwise, "no issue found". This feedback will serve two purposes. First, it will help to measure the model's precision in real-world conditions, which means management can see how often a warning leads to a genuine problem. Second, the "no issue found" notes can be used to recalibrate the model's risk threshold (i.e., raising the thresholds to reduce the falls alarm) or to identify which features are causing this false positive. Over time, the staff will build trust in the model as they see more and more failures can be prevented by its warning. If not, the management should adjust the model.

For the framework to be used effectively, management should authorize a short session of training for the people who will use it. For example, from the training, the maintenance team can learn how to read the framework's output, interpret risk levels, and create a closed-loop feedback log. The production personnel can learn how to assign the right category for each downtime occurrence and how to spot early warning signs during the operation. Also, management should provide a one-page reference guide summarizing all the basic information, such as risk thresholds, the 22 downtime categories, and the closed-loop feedback procedure, and display it near all machines or on the notice board of the production floor.

5.4 Gradual AI adoption in manufacturing environments

In recent years, artificial intelligence (AI) has received significant attention in industrial sectors. Manufacturers are using this technology for various purposes, such as predictive maintenance, quality inspection and production optimization. However, despite its potential many manufacturers struggle to successfully implement AI in practice. AI hype is so prominent nowadays that companies are attempting to use it in all sorts of activities. One of the key challenges companies often face in adopting AI is the lack of necessary digital foundation. To be successfully implemented in a manufacturing setting, artificial intelligence needs an integrated data system, consistent data collection practice, and sufficient data quality. If these prerequisites are not met, AI may fail to provide meaningful insight or actionable outcomes.

Furthermore, studies showed that 80% companies couldn't adopt AI or failed to achieve the expected benefit ("Study reveals," 2025). The reason behind this failure is not the lack of AI technology itself, but the lack of organizational, infrastructural and strategic capability. Successful AI implementation not only depends on the changes of technologies, but also on the transformation of processes, skills and organizational culture. If the current process is not compatible with the AI system or the employees do not support such changes, it is unlikely that AI will bring good results. Organizations often disregard these issues and follow a 'big bang' approach while adopting AI technologies. This means that instead of implementing AI tools gradually, companies prefer to take a big jump and deploy complex AI systems all at once. Such approaches often lead to failed pilots, abandoned projects, and wasted investments.

These challenges suggest that AI implementation should follow a stepwise approach rather than a complete overhaul. Companies should first set the groundwork for AI deployment, like standardizing data collection processes, raising awareness among employees regarding AI benefits, and updating the existing processes. Also, companies should start with small, for example, first conduct a descriptive analysis to understand the existing process, then perform KPI-based monitoring, and after that, develop a basic predictive model. Once the maturity and organizational readiness improve, they can move to the next level of AI implementation. In this context, the framework developed in this study provides a practical step toward AI adoption. It needs no expensive infrastructure, only basic operational data, which ensures minimal risk of AI implementation. Also, it allows a smoother transition from traditional practice to a digital practice.

5.5 Practical Implementation of the Framework

The decision-support framework generated by this study is not just an analytical tool; it's a practical solution that can be used in the industrial setting of the case company. To move the framework from a one-time analysis tool to an automated, repeatable process, the case company needs to integrate the framework into its existing system. This section outlines how the case company can operationalize the framework with minimal additional resources.

The main output of the framework is machine failure probabilities, risk level, dominant downtime categories, and targeted recommendations. A Power BI dashboard can be used to visualize the results from the framework. The dashboard will show the risk level of all machines in a traffic-light manner, where red means high-risk machine, yellow means medium-risk and green means low-risk machine. It can also show the trend lines of key KPIs (availability, MTBF, interruption frequency) and a Pareto chart of dominant downtime categories. This type of dashboard will help the production people and the maintenance team to identify the problematic machines quickly and take necessary steps to solve the issue.

Currently, the interruption data and the maintenance work orders are stored separately in the company's ERP system. A simple automated pipeline can be created using a Python script that will run on a regular basis, for example, once a week. The script will extract the data from the ERP system as CSV files and perform necessary steps, such as cleaning, preprocessing, calculating KPIs and running the Random Forest prediction model. The output of the model would be stored in a folder connected to Power BI. Power BI can then refresh its dataset directly from these files and update the dashboard. This way, the company won't need live data streaming, complex system integration, or expensive middleware.

One of the key challenges of this study was categorizing machine downtime, analyzing the clocking notes through a text mining process; it required both automatic and manual tasks. The company can incorporate the list of 22 downtime categories into its ERP system, so that in the future, when operators record downtime, they can choose the accurate category from a drop-down list. This will eliminate the need for TF-IDF and manual grouping, making the framework fully automatic and reducing the processing time significantly. Although the operator may need a short training on which category to select in which situation, it will make the framework more robust and easier to maintain.

The suggested model used a predefined risk threshold for the machine failure probability (i.e., probability > 0.7 = high-risk, 0.4 < probability < 0.7 = medium-risk, and probability < 0.4 = low-risk machine). These thresholds should be recalibrated as the company gets more maintenance records and the closed-loop feedback log gets more entries (i.e., “issue found”/”no issue found”). The main goal of the recalibration is to reduce the number of false positives (inspecting machines when there is no issue) and false negatives (missing a genuine failure). One simple method of recalibration is to raise the threshold level if too many false positives have been detected, or lower the threshold level if too many false negatives have been detected.

In the era of Industry 4.0, where the competition among companies is based on production efficiency, fast delivery and the use of AI, relying on traditional maintenance strategy means lagging behind in the global market. The framework demonstrates how easily organizations can adopt an advanced maintenance strategy using only the basic operational data. By combining automated data pipelines, standardized data entry, and user-friendly visualization tools, the framework helps the company to move toward a proactive and more structured maintenance management.

5.6 Limitations of the Predictive Model

The predictive model is mainly based on two datasets: interruption logs and maintenance report. Although the interruption log is comparatively complete, the maintenance report contains a lot of missing values, such as machine number, fault description, and component information. Therefore, not all the maintenance events could be matched with the corresponding machine number or time windows, which significantly affects the model's ability to predict.

For ease of analysis, all the calculations in each dataset are recorded in a 7-day window. However, maintenance events do not necessarily occur on a weekly basis. Therefore, in the dataset, it is observed that negative events are far more than positive events, which means that most of the 7-day windows contain 0 maintenance. As shown in Figure 8, only 17% of the test dataset contained maintenance occurrences. This imbalance forces the predictive model to favor the majority class, in this case, non-maintenance events, and limit its predictive capability of actual (maintenance) events.

This study used 17-month period datasets, which are sufficient for descriptive analysis, but relatively shorter for predictive analysis. More historical data would contain more information on maintenance events and eventually improve the model's recall and precision rate. In addition, the framework does not contain any economic factors, such as maintenance cost, failure cost, or production criticality, which could further enhance the decision-making.

5.7 Recommendations for Future Research

Future research should first address the data limitations of this study. The quality and completeness of the data greatly affect the model's predictive performance. All the shortcomings of the model, like low recall rate or minimal precision rate, are derived from the use of small datasets. Therefore, future research should consider using maintenance records over a longer period (e.g., three to five years). Researchers could also use different time windows (e.g., 14 days) for the prediction model. In this study, it is observed that maintenance events happened infrequently in 7-day time windows, which creates a huge class imbalance in the dataset. Longer time windows or oversampling techniques (such as SMOTE) can balance the dataset artificially. Additionally, the use of sensor-based monitoring data, such as temperature, vibration, and pressure, would add valuable features and help improve the model's performance.

Second, future researchers should validate the framework by applying it to different industries, for example, injection molding, metal forming, or packaging lines. The study has been conducted on a single case company, so applying it to other manufacturing settings will improve the generalizability and adaptability of the framework. Furthermore, future research could also incorporate economic factors, such as downtime cost per hour, maintenance cost, and spare part prices, into the framework. This enables the framework to generate recommendations not only based on risk but also based on financial return.

6 Conclusion

The study tried to find out how unplanned downtime of a traditional plastic extrusion machine can be reduced using only basic data, like the interruption log and maintenance report. The research was conducted on a case company; it's a manufacturing company that makes plastic profiles using traditional extrusion machines. The study was guided by three research questions related to (1) the identification of major downtime causes, (2) the evaluation of machine health using KPIs, and (3) the integration of descriptive and predictive analysis into a decision-making framework.

The main goal of the research is to develop a framework that can support the company's decision-making regarding maintenance activities without needing any advanced sensor data or expansive Industry 4.0 infrastructure. To reach the goal, it follows a structured methodological approach. First, the notes from the interruption log were processed using text mining and converted into 22 downtime categories. Then, a descriptive analysis (Pareto analysis and KPIs calculation) was conducted to identify the root causes and evaluate machine health. After that a predictive model was developed using a Random Forest Classifier to estimate the probability of machine failure in the next week. Finally, by combining the descriptive analysis and the predictive analysis, a decision support framework was created that can generate targeted recommendations for each machine.

The study used historical data from multiple sources, including interruption logs, maintenance work orders, and production reports. The data were collected over a period of approximately 17 months. 800 interruption notes were organized into 22 downtime categories through successful text mining. Pareto analysis identified that among the 22 categories, equipment failure, tooling fault, component failure, calibration issue, and material issue are the five that are responsible for 70% of total downtime. The KPI evaluation showed that machine performance indicators such as MTBF, MTTR, and availability are some effective measures of machine reliability and maintenance efficiency. Also, the performance indicators are highly interconnected; therefore, they should be studied together, not separately.

The predictive analysis of the study was conducted by developing a Random Forest predictive model. The goal of the model was to estimate the probability of a machine failure in the upcoming week. Although the accuracy of the model was quite good (approximately 80%), because of the class imbalance and incomplete maintenance records, the recall rate was low (approximately 14%). Despite this limitation, the risk probabilities generated by the model were very useful in prioritizing the machines in the framework. The model is based on six key features, including total downtime, failure frequency, MTBF, MTTR, availability, and rolling downtime (two-week average). Feature importance analysis identified that rolling downtime (30%) is the most important feature of the model, followed by MTBF (20%) and availability (17%).

The main contribution of the thesis was the development of a decision-making framework for the company. The framework combined the results from the descriptive analysis and the predictive analysis to generate targeted recommendations for each machine. The failure probabilities achieved from the predictive model help prioritize the machine risk and the Pareto chart from the descriptive analysis helps identify the dominant category of downtime in a machine. High-risk machines were assigned to immediate intervention strategies, whereas medium and low-risk machines were assigned to monitoring and routine maintenance strategies, respectively. The framework is very helpful for resource-constrained SMEs in reducing unplanned machine downtime. As SMEs often lack the sensor data or computational capabilities to build an advanced maintenance strategy, this framework offers a practical alternative by creating a smart maintenance strategy only using basic operational data.

In addition, the study proposed a detailed practical implementation pathway for the company. Without expensive modifications of the existing system, the company can apply the framework to its manufacturing setting by using only simple Python scripts and Microsoft Power BI. The Python scripts will automate the KPI monitoring, while the Power BI dashboard will help visualize the framework’s output. The standardization of downtime categories and embedding them into the ERP clocking interface will further automate the framework. The study also makes some theoretical contributions, such as extending the reliability engineering theory to low-data environments, demonstrating the value of combining descriptive and predictive analysis, and addressing the challenges of implementing predictive models in real-world settings.

Finally, the thesis acknowledges some of its limitations. The low recall and precision rate of the predictive model are derived from the class imbalance and the incomplete dataset. The framework is not a replacement for fully sensor-based predictive maintenance; it's just an intermediate solution for resource-constrained SMEs. To address these limitations, the study also suggested directions for future research, such as collecting longer historical data, using a longer time window (e.g., 14 days), and incorporating low-cost condition monitoring data. Nevertheless, this study demonstrates that even with basic operational data (interruption log and fault report), a smart maintenance strategy can be developed. Therefore, although the framework seems lightweight and simple, it provides a realistic pathway to reduce downtime, improve machine reliability and build a foundation for advanced maintenance in the future.

References

Abeykoon, C., McMillan, A., & Nguyen, B. K. (2021). Energy efficiency in extrusion-related polymer processing: A review of state of the art and potential efficiency improvements. *Renewable and Sustainable Energy Reviews*, *147*, 111219. https://doi.org/10.1016/j.rser.2021.111219

ALMashaqbeh, S., & Hernandez, E. M. (2024). Evaluation and Improvement of a Plastic Production System Using Integrated OEE Methodology: A Case Study. *Management Systems in Production Engineering*, *32*(3), 450–463. https://doi.org/10.2478/mspe-2024-0042

Basri, E. I., Abdul Razak, I. H., Ab-Samat, H., & Kamaruddin, S. (2017). Preventive maintenance (PM) planning: a review. *Journal of Quality in Maintenance Engineering*, *23*(2), 114–143. https://doi.org/10.1108/JQME-04-2016-0014

Birolini, Alessandro. (2017). *Reliability engineering : theory and practice*. Springer.

BMW Group Corporate Communications. (2023, November). *Smart maintenance using artificial intelligence*. BMW Group.

Cao, H., Yu, J., & Duan, F. (2025). Condition-Based Maintenance in Complex Degradation Systems: A Review of Modeling Evolution, Multi-Component Systems, and Maintenance Strategies. *Machines*, *13*(8), 714. https://doi.org/10.3390/machines13080714

Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, *137*, 106024. https://doi.org/10.1016/j.cie.2019.106024

Ciocoiu, L., Siemieniuch, C. E., & Hubbard, E.-M. (2017). From preventative to predictive maintenance: The organisational challenge. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *231*(10), 1174–1185. https://doi.org/10.1177/0954409717701785

Duarte, A. L. C. M., & Santiago Scarpin, M. R. S. (2023). Maintenance practices and overall equipment effectiveness: testing the moderating effect of training. *Journal of*

Quality in Maintenance Engineering, *29*(2), 442–459. https://doi.org/10.1108/JQME-04-2021-0033

Fadeyi, J. A., Okwu, M. O., Mgbemena, C. O., & Ezekiel, K. C. (2016). The Pareto principle and a hazard model as tools for appropriate scheduled maintenance in a manufacturing firm. *African Journal of Science, Technology, Innovation and Development*, *8*(2), 173–177. https://doi.org/10.1080/20421338.2016.1147203

Fathi, K., & Wernher van de Venn, H. (2024). *Data, Models, and Performance: A Comprehensive Guide to Predictive Maintenance in Industrial Settings* (pp. 81–104). https://doi.org/10.5772/intechopen.1005511

GE Aerospace. (2024). *GE Aerospace expands scope of artificial intelligence to support engine monitoring and maintenance*. AINonline.

Gholipour, Y. (2025). A Comprehensive Review of Maintenance Strategies: From Reactive to Proactive Approaches. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5349871

Gogos, C. G. ., & Tadmor, Zehev. (2014). *Principles of polymer processing* (2nd ed.). Wiley-Interscience.

Grooss, O. F. (2024). Digitalization of maintenance activities in small and medium-sized enterprises: A conceptual framework. *Computers in Industry*, *154*, 104039. https://doi.org/10.1016/j.compind.2023.104039

He, Y., Gu, C., Chen, Z., & Han, X. (2017). Integrated predictive maintenance strategy for manufacturing systems by combining quality control and mission reliability analysis. *International Journal of Production Research*, *55*(19), 5841–5862. https://doi.org/10.1080/00207543.2017.1346843

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–106. https://doi.org/10.2307/25148625

Hyvärinen, M., Jabeen, R., & Kärki, T. (2020). The Modelling of Extrusion Processes for Polymers—A Review. *Polymers*, *12*(6), 1306. https://doi.org/10.3390/polym12061306

Jamwal, A., Agrawal, R., & Sharma, M. (2025). Challenges and opportunities for manufacturing SMEs in adopting industry 4.0 technologies for achieving sustainability:

Empirical evidence from an emerging economy. *Operations Management Research*, *18*(2), 718–743. https://doi.org/10.1007/s12063-023-00428-2

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, *20*(7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012

Keshta, I., Majeed, M. E., Al-Shaikhli, T. R., Adel, A. H., Jabbar, K. A., & Soni, M. (2025). AI-Based Predictive Maintenance for Machine Vibration and Condition Monitoring in Industrial Applications. *2025 4th OPJU International Technology Conference on Smart Computing for Innovation and Advancement in Industry 5.0, OTCON 2025*. https://doi.org/10.1109/OTCON65728.2025.11070934

Khan, M., Ahmad, A., Sobieczky, F., Pichler, M., Moser, B. A., & Bukovsky, I. (2022). A Systematic Mapping Study of Predictive Maintenance in SMEs. *IEEE Access*, *10*, 88738–88749. https://doi.org/10.1109/ACCESS.2022.3200694

Khan, U., Cheng, D. S., Setti, F., Fummi, F., Cristani, M., & Capogrosso, L. (2026). A Comprehensive Survey on Deep Learning-based Predictive Maintenance. *ACM Transactions on Embedded Computing Systems*, *25*(2), 1–43. https://doi.org/10.1145/3732287

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, *104*, 799–834. https://doi.org/10.1016/j.ymssp.2017.11.016

Molęda, M., Małysiak-Mrozek, B., Ding, W., Sunderam, V., & Mrozek, D. (2023). From Corrective to Predictive Maintenance—A Review of Maintenance Approaches for the Power Industry. *Sensors*, *23*(13), 5970. https://doi.org/10.3390/s23135970

Muchiri, P., & Pintelon, L. (2008). Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion. *International Journal of Production Research*, *46*(13), 3517–3535. https://doi.org/10.1080/00207540601142645

Müller, I., Zhang, Z., Tordeux, A., Pietruschka, J., Bracke, S., Julitz, T. M., Schlüter, N., Löwer, M., & Fricke, N. (2025). System Reliability Engineering in the Age of Industry

4.0: Challenges and Innovations. In *Safety Engineering* (pp. 75–123). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-47473-7_4

Nasirinejad, M., Afshari, H., & Sampalli, S. (2024). Challenges and Solutions to Adopt Smart Maintenance in SMEs: A Literature Review and Research Agenda. *IFAC-PapersOnLine*, *58*(19), 917–922. https://doi.org/10.1016/j.ifacol.2024.09.164

Nasirinejad, M., Afshari, H., & Sampalli, S. (2025). Empowering SMEs in the Fourth Industrial Revolution: A Framework for Maintenance 4.0 Adoption. *Procedia Computer Science*, *253*, 1382–1391. https://doi.org/10.1016/j.procs.2025.01.200

Nunes, P., Rocha, E., & Santos, J. (2024). Adaptive Framework for Maintenance Scheduling Based on Dynamic Preventive Intervals and Remaining Useful Life Estimation. *Future Internet*, *16*(6), 214. https://doi.org/10.3390/fi16060214

Nwanya, S. C., Udofia, J. I., & Ajayi, O. O. (2017). Optimization of machine downtime in the plastic manufacturing. *Cogent Engineering*, *4*(1). https://doi.org/10.1080/23311916.2017.1335444

Phongmoo, S., Suedumrong, C., Kuensaen, C., Sinthavalai, R., & Leksakul, K. (2025). Predictive maintenance in semiconductor manufacturing: Comparative analysis of machine learning models for downtime reduction. *Computers and Industrial Engineering*, *205*. https://doi.org/10.1016/j.cie.2025.111211

Pietsch, D., Matthes, M., Wieland, U., Ihlenfeldt, S., & Munkelt, T. (2024). Root Cause Analysis in Industrial Manufacturing: A Scoping Review of Current Research, Challenges and the Promises of AI-Driven Approaches. *Journal of Manufacturing and Materials Processing*, *8*(6), 277. https://doi.org/10.3390/jmmp8060277

Pinciroli Vago, N. O., Forbicini, F., & Fraternali, P. (2024). Predicting Machine Failures from Multivariate Time Series: An Industrial Case Study. *Machines*, *12*(6), 357. https://doi.org/10.3390/machines12060357

Preethi, E., Ahmed, A. S., Shyamala, G., Vasukidevi, G., Sunil, T., & Atheeswaran, A. (2024). Optimizing Predictive Maintenance in Smart Factories Using Random Forests and Internet of Things Sensors. *2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 321–326. https://doi.org/10.1109/ICECA63461.2024.10801060

Resanovic, D., & Balc, N. (2024). *Literature Review on the Current State-of-the-Art in Research and Technological Advancements in the Field of Machine Learning Applied to Predictive Maintenance* (pp. 103–119). https://doi.org/10.1007/978-3-031-71633-1_8

Schröder, C. (2016). The challenges of Industry 4.0 for small and medium-sized enterprises. Friedrich-Ebert-Stiftung.

Semotam, P. (2024). A Data-Driven Approach to Reliability and Fault Analysis in Industrial Maintenance. *IFAC-PapersOnLine*, *58*(9), 97–102. https://doi.org/10.1016/j.ifacol.2024.07.378

Setiawan, B., Al Latif, F., & Rimawan, E. (2022). Overall Equipment Effectiveness (OEE) Analysis: A Case Study in the PVC Compound Industry. *IJIEM - Indonesian Journal of Industrial Engineering and Management*, *3*(1), 14. https://doi.org/10.22441/ijiem.v3i1.12066

Singh, S., Khamba, J. S., & Singh, D. (2021). *Analyzing the Role of Six Big Losses in OEE to Enhance the Performance: Literature Review and Directions* (pp. 411–421). https://doi.org/10.1007/978-981-33-4320-7_37

Smith, D. J. (2011). Understanding Terms and Jargon. In *Reliability, Maintainability and Risk* (8th ed., pp. 13–27). Elsevier. https://doi.org/10.1016/B978-0-08-096902-2.00002-7

Study reveals why 80% of companies fail to adopt AI successfully. (2025, January 21). *Helsinki Times*. helsinkitimes.fi

Surendran Pillai, R., Denny, P., O'Connell, E., Dooley, A., & Penica, M. (2026). Optimizing Predictive and Prescriptive Maintenance Using Unified Namespace (UNS) for Industrial Equipments. *Journal of Experimental and Theoretical Analyses*, *4*(1), 13. https://doi.org/10.3390/jeta4010013

Tsallis, C., Papageorgas, P., Piromalis, D., & Munteanu, R. A. (2025). Application-Wise Review of Machine Learning-Based Predictive Maintenance: Trends, Challenges, and Future Directions. *Applied Sciences*, *15*(9), 4898. https://doi.org/10.3390/app15094898

Weidner, T. J. (2023). Planned maintenance vs Unplanned maintenance and facility costs. *IOP Conference Series: Earth and Environmental Science*, *1176*(1), 012037. https://doi.org/10.1088/1755-1315/1176/1/012037

Wolniak, R. (2019). Downtime in the Automotive Industry Production Process – Cause Analysis. *Quality Innovation Prosperity*, *23*(2), 101–118. https://doi.org/10.12776/qip.v23i2.1259

Zhao, Q., Lolli, F., Balugani, E., & Gamberini, R. (2025). A tertiary review of condition-based maintenance literature: Insights and future directions. In Proceedings of the 30th Summer School Francesco Turco.

Zhu, T., Ran, Y., Zhou, X., & Wen, Y. (2026). A Survey on Intelligent Predictive Maintenance (IPdM) in the Era of Fully Connected Intelligence. *IEEE Communications Surveys & Tutorials*, *28*, 633–671. https://doi.org/10.1109/COMST.2025.3567802

Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering*, *150*, 106889. https://doi.org/10.1016/j.cie.2020.106889