Analyzing Critical Process Parameters Influencing Product Quality Defects using Machine Learning

A Real-World Case Study from a Foundry Line

Master project report in Production Engineering

Sukumaran Rajasekaran

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023
www.chalmers.se

Degree project report 2023

Analyzing Critical Process Parameters Influencing Product Quality Defects using Machine Learning

A Real-World Case Study from a Foundry Line

Sukumaran Rajasekaran

Department of Industrial and Materials Science
Chalmers University of Technology
Gothenburg, Sweden 2023

Analyzing Critical Process Parameters Influencing Product Quality Defects using Machine Learning A Real-World Case Study from a Foundry Line Sukumaran Rajasekaran

Academical supervisor: Sabino Francesco Roselli, Department of Electrical Engineering
Industrial supervisor Kanika Ghandi, Volvo Group Skövde
Examiner: Ebru Turanoglu Bekar, Department of Industrial and Materials Science

Master's Thesis 2023
Department of Industrial and Materials Science
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2023

Analyzing Critical Process Parameters Influencing Product Quality Defects using Machine Learning: A Real-World Case Study from a Foundry Line

Sukumaran Rajasekaran
Department of Industrial and Materials Science
Chalmers University of Technology

Abstract

Manufacturing has grown in importance with the goal of reducing production disruptions, minimizing unplanned downtime, minimizing quality defects, and increasing the efficiency of production systems. This project emphasizes quality control in manufacturing, where traditional quality control processes are reactive but modern processes utilizing Artificial Intelligence (AI) and Machine Learning (ML) techniques are proactive. These modern processes have the potential to reduce waste, improve efficiency, and decrease costs, thereby ensuring the success of a manufacturing company. AI/ML utilizes large datasets from multiple sources, such as machines, processes, and products, to predict deviations in the future and make better data-driven decisions, allowing organizations to stay competitive in this digitalized era.

This thesis employs CRISP-DM, a comprehensive tool used for data mining and ML projects, to provide a structured approach for solving business problems and facilitating data-driven decision-making. The case company specializes in producing castings for power-train components in heavy vehicles a process governed by a range of process parameters crucial to product quality. Poor control of these parameters can lead to casting defects. Leveraging ML models, this study aims to establish correlations between various process parameters in two core-making machines (Machine A & Machine B) along with their maintenance logs, with the goal of identifying root causes for different casting defects in the core-making process. The ML models utilized for identifying faulty cores attained accuracy rates of 66.2% for Machine A and 56% for Machine B. These levels of accuracy were deemed satisfactory by the domain experts within the company. The results from the thesis would aid human operators in determining whether to manually eliminate defective cores from the production line before casting. Therefore, this would effectively prevent the production of flawed castings and contribution to the production of high-quality products.

Keywords: Prediction of Quality Defects, Process Parameters, Maintenance Logs, Machine Learning, CRISP-DM, Manufacturing, Foundry

Acknowledgements

This master thesis has been written at the Production Engineering Master of Science Program at Chalmers University of Technology and in cooperation with project Trustworthy Predictive Maintenance (TPdM) which is an ongoing research project in the Department of Industrial and Materials Science at the Chalmers University of Technology. I would like to express my deep gratitude to my Academic Examiner Dr.Ebru Turanoglu Bekar who is a Senior Lecturer in the Department of Industrial and Materials Science, at Chalmers University of Technology for her patient guidance, enthusiastic encouragement and useful critiques of this thesis work.

I would also like to thank my Academical Supervisor Dr. Sabino Francesco Roselli who is a post-doctoral researcher at department of Department of Electrical Engineering, at Chalmers University of Technology for his advice and assistance in keeping our progress on schedule.

Last but not the least, I would also like to express very great appreciation to my Industrial Supervisor Kanika Gandhi who is the Analytics & Research Development Engineer at Volvo Trucks Operation in Skövde, Ida Eriksson who is a Quality Engineer at Volvo Trucks Operation in Skövde for their professional guidance and valuable support enabling me to visit the company and observe the operations.Thanks also to the Advanced and Innovative Digitalization Program by VINNOVA for their funding of the research project TPdM (Grant No. 2022-01710)

Sukumaran Rajasekaran, Gothenburg, August 2023

Contents

List of Figures

1
Introduction

Quality control is an essential component of manufacturing that helps to guarantee that products meet the necessary standards of quality, safety, and dependability. Without adequate quality control procedures in place, manufacturers run the risk of creating goods that could be faulty or unsafe for consumers, resulting in product recalls, legal liability, and harm to their brand's reputation[1].

Inspection and testing of products at various stages of production is a common practice in traditional quality control in manufacturing to make sure they adhere to the defined standards.To confirm that the product complies with specifications,visual inspection, measurements, and functional tests are done. Statistical process control, control charts, and Six Sigma are a few of the common techniques used in traditional quality control [2][3].

Though traditional quality control make sure that product complies to the standard, it frequently takes a lot of time and is expensive, which is a major drawback. Every product must be inspected, which can slow down production and raise manufacturing costs overall. Traditional quality control is also reactive, i.e., it finds flaws after they have already happened rather than preventing them from happening at all[4].

The rapid growth in newer technologies and digitization of manufacturing industries can get rid of the drawbacks from traditional quality control, using real-time monitoring,predictive maintenance, and automated quality control [5].Artificial Intelligence (AI) and Machine Learning (ML) could detect potential quality defects before they happen, enabling early intervention and prevention, by analyzing data coming from multiple data sources in manufacturing processes [6].

The Project Trustworthy Predictive Maintenance (TPdM), which is an ongoing research project in the Department of Industrial and Materials Science at the Chalmers University of Technology, aims to design human-centered decision support prototypes for Predictive Maintenance to achieve actionable decisions using advanced data science.One of the industrial partners of the project, is interested in analyzing and comprehending data derived from various sources whether an AI/ML system can predict casting quality issues from the data for a specific foundry line. This line's castings frequently have a variety of casting defects due to a variety of process-related factors. These costly and unacceptable defects drive up the cost of production. Therefore, it is crucial that these casting defects are found early on in the core-making process before being cast.Hence, this thesis has been generated within the scope of this research project, centering around the mentioned area of interest as a case study conducted in collaboration with the company.

1.1 Aim

This thesis aims to analyze the underlying causes behind the diverse types of quality defects occurring in a foundry line by combining multiple data sources (e.g., machine/process data, maintenance records, and quality information) and constructing a data-driven approach that utilizes ML. With the intention of achieving this aim, the following objectives are identified:

- Study and analyze the Value stream of certain production lines in the production process.
- Evaluate important machine parameters and maintenance actions influencing part quality defects.
- Build predictive model(s) to analyse the state of evaluated machine parameters, maintenance actions and quality defects.

1.2 Research Questions

The research questions which will help to clarify the aim and objectives are as follows:

1. What are the prominent quality issues pertaining in the production process in the foundry line?
2. Which process parameters/maintenance actions are influencing those prominent quality issues?
3. What ML methods can be applicable for predicting those quality issues?

1.3 Limitations

This thesis focuses solely on a dedicated foundry line that produces a certain class of parts. The research will not be including the melting and pouring and restricted to core making process and only two machines within the primary core making process due to time constraints and data availability. Any accompanying photographs have been adequately masked to maintain confidentiality, and the presented data accurately reflects current production circumstances. It's crucial to remember that the deployment of the model is outside the scope of this thesis due to time constraints, and that the Manufacturing Company's internal management will decide whether to do so.

1.4 Outline

The company specializes in casting components exclusively within a dedicated foundry line, specifically for power-train systems in heavy-duty vehicles. This foundry line plays a crucial role in producing these parts, requiring machines to compress green sand into the desired core shape for the final component. A conveyor system facilitates sand mixing in a hopper, which is then transported to the compaction machines. These machines are equipped with Programmable Logic Controller (PLC) systems that record data during various core-forming processes, from initiation to completion. Essential quality data is gathered during the final part inspection, while maintenance-related information is collected in the event of a breakdown. A detailed explanation of the foundry line is explained in Chapter 4. The structure of the Master's thesis is divided into several chapters, each serving a distinct purpose in presenting and analyzing the project. Chapter 1 introduces the project, with the aim of clarifying the thesis's purpose. Moving on to Chapter 2, the theoretical background is elaborated upon, offering insights into the aspects of casting, casting defects and insights on data driven application in manufacturing. Chapter 3 explains the methodology adopted for accomplishing thesis objectives. The results of the thesis are presented in chapter 4. Further discussion and implementation strategies for the companies are presented in chapter 5. Chapter 6 concludes the outcomes of this thesis.

2

Theoretical Background

This chapter offers a comprehensive introduction to pertinent terminologies and concepts associated with both foundries and quality defects. This understanding forms the foundation for comprehending the case study within this thesis. Additionally, the chapter provides a brief summary of prior research in the realm of data-driven approaches in manufacturing, particularly in the context of quality control.

2.1 Foundry

A foundry is a factory that produces metal castings through the process of melting metal and pouring it into a mold to create a desired shape. Foundries have a number of benefits over other types of production methods. These include inexpensive tooling costs, flexibility of the process, ability to cast almost all engineering alloy types, adaptability of design, capacity to manufacture a broad variety of shapes and intricate details [7].The production of castings in foundries is only constrained by the availability of molten metal and the equipment needed to lift and handle the output. When compared to the form tools needed for other processes, patterns used to build the molds are comparatively inexpensive, making foundries suitable for both one-off and mass production. Depending on the demands of the client, foundries might be small or bigger and more specialized [8].

2.1.1 Types of Metal Casting

The extensive history of the foundry industry may be traced back to the Bronze Age, roughly 3000 BC. Early metal casting methods were used, including investment casting, lost-wax casting, and sand casting. The invention of the blast furnace in the 12th century AD transformed the way that iron was produced, which prompted an increase in the foundry sector. The Industrial Revolution helped the sector grow even more by accelerating the need for metal castings, which was fueled by developments in technology. The foundry sector has thrived throughout the 20th and 21st centuries, adjusting to new materials and technology, and it continues to be an essential part of the worldwide economy [9].There are several types of metal casting technologies used in modern manufacturing processes. Each casting technique has benefits and drawbacks, and the choice of casting method is influenced by factors like the part size and complexity, the desired level of quality and precision, and the production volume. Sand casting is the most commonly used technique that uses sand as the mold material. This technique has a high degree of accuracy and can create large, complex shapes. Sand casting is widely used in the automotive, aerospace, and construction industries because of its affordability and adaptability. Investment casting also known as lost-wax casting, is a precision casting technique that uses a wax pattern to make a ceramic mold. Investment casting is widely used in the jewelry, aerospace, and medical industries because it can create intricate, high-quality parts [9]. Die casting casting technique that involves pushing hot metal into a mold cavity under intense pressure. Die casting is widely used in the production of consumer goods, electronics, and automobiles due to its high production efficiency and capability to create complex shapes [7]. Continuous casting is a casting technique that entails continuously pouring molten metal into a mold as it cools and solidifies. Continuous casting is widely used in the manufacture of steel and other metals because it can generate significant quantities of high-quality castings with little waste. Centrifugal casting uses centrifugal force to distribute molten metal into a mold cavity. Centrifugal casting is frequently used in the production of pipes, cylinders, and other hollow objects because it can create castings that are of a high caliber and homogeneity [8].

2.1.2 Sand Casting Process

The foundry line under study in Volvo AB for this project utilizes sand casting technique for producing cylinder heads owing to the advantages like high degree of accuracy and ability to create large and complex shapes as mentioned in subsection 2.1.1. Mold preparation, melting & casting, and finishing operations are the three main processes involved in sand casting process which are explained in below chapters [8].

2.1.2.1 Mold preparation

The mold-making process in sand casting involves pattern creation, molding sand preparation, and mold assembly. The first step in making a mold is making a pattern. Wood, plastic, or metal are the most common materials for patterns, which are exact replicas of the component that will be cast. The mold cavity, which will be filled with molten metal to produce the finished part, is created by using the pattern to create the cavity. The pattern must be created with care, taking into account any shrinkage or distortion that may occur during the casting process [8]. After the pattern has been made, the molding sand needs to be prepared. Molding sand is a mixture of sand, clay, and other materials that is used to create the mold cavity. In order for the sand to have the right strength, permeability, and thermal stability, it needs to be prepared with care. After the pattern and molding sand are prepared, the mold is assembled. To create the mold, the pattern is placed in a molding box and surrounded by sand. After the mold is divided in half, any necessary cores are added to create the internal features of the part [8].

2.1.2.2 Melting & Casting

The first step is setting up the furnace for melting, gas or electric furnaces are commonly used for melting in sand casting. The type of furnace used depends on

the size of the casting, the quantity of metal being melted, and the required casting quality [9]. When the metal is ready, it is charged into the furnace. Sand casting can be done with any alloy, including iron, steel, aluminum, and copper, that can be melted and poured. Typically, the furnace is charged with scrap metal or metal ingots. As the metal melts, it is essential to monitor the temperature and chemistry of the melt. To make certain that the temperature of the melt remains within the permitted range for the particular metal being melted, as well as to make sure that the chemistry of the melt complies with the requirements for the final casting [8].
When the metal is ready to be put into the mold after it has melted, casting is complete. Careful preparation is necessary to guarantee that the mold is error-free and that it will create a casting of the highest standard. The mold must be appropriately ventilated to permit the discharge of gases during the casting process. It is essential to control the temperature and pouring rate to ensure that the metal fills the mold completely and does not result in cold shuts or other defects. After being poured into the mold, the metal must cool and harden. The casting's cooling and solidification times are all influenced by the size, shape, and type of metal used. The casting is taken out of the mold after it has fully solidified [8].

2.1.2.3 Finishing

During this procedure, extra material must be removed, the surface must be prepared, and any required coatings or finishes must be applied. A sand casting must first be finished by removing any extra material, including the gating system and any risers. In this technique, hand tools or mechanical cutting instruments like saws or grinders are usually employed. After the surplus material has been removed, the casting must be prepared for any necessary coatings or finishes. This requires cleaning the casting's surface to remove any remaining sand or other pollutants as well as any surface oxidation that may have formed during the casting process.Techniques like mechanical cleaning, chemical cleaning, and shot blasting, can be used for surface preparation. After the casting has been prepared, coatings and finishes improve its look or performance.The use of a coating could prevent corrosion or improve the casting's surface hardness. The casting's look can also be improved with finishes like polishing or painting [8].

2.2 Quality

Quality control is a system that maintains the desired level of quality, ensuring that products or services meet customer requirements and are of the highest quality. This process has evolved over time. Prior to 1900, individuals or groups were responsible for producing and controlling the quality of the entire product. Between 1900 and 1920, there was an industrial revolution in mass production, where workers were accountable for a specific portion of the product, and the foreman or supervisor assumed responsibility for product quality [10]. The inspection quality control period lasted from 1920 to 1940, during which production volume increased, and products and processes became more complex. Quality inspectors were included to check the product quality after specific operations. Standards were established, and products

were inspected against those standards. Non-conforming products were either reworked or discarded. According to the ANSI standard, a defect is a non-conformity, and a defective item is a non-conforming item. Statistical process control gained popularity from 1940 to 1960, which did not require 100% product inspection but used control charts [10].
Total quality control arose in the 1960s, emphasizing the importance of several departments in the quality control process. It preached that quality responsibility did not rest solely with people working on the shop floor or quality personnel. In the 21st century, there has been an information technology revolution that uses the internet and data to control quality [10].

Quality cost is the difference between the actual cost and the reduced cost if products and services were all conforming [10]. The cost of quality is divided into two categories: the cost of good quality and the cost of poor quality. Prevention costs and appraisal costs are associated with good quality, while failure costs result from poor quality. Prevention costs are associated with preventing defects and imperfections from occurring and can include training, process engineering, and quality planning. Appraisal costs are the direct costs of measuring quality, such as lab testing, inspection, and test equipment. Failure costs are associated with a product and can include scrap, rework, re-inspection, re-testing, and customer returns [11].

Preventive measures can help reduce operational costs by decreasing the chances of product failure.As the product progresses from design to production to distribution to the customer, the cost of quality associated with any issue increases exponentially. If problems are not addressed in the early stages, they can result in a tenfold increase in quality costs [11]. For example, Samsung's Galaxy Note 7 smartphone was recalled in 2016 due to a defect in its lithium-ion batteries. The phones were catching fire, causing damage to property, and Samsung issued a recall of 2.5 million devices, resulting in a financial loss of 5.3 billion. This incident damaged Samsung's reputation, and the company faced regulatory investigations and fines in several countries [12].
Therefore, improved quality leads to reduced costs, increased efficiency, and customer satisfaction that improves competitive position as shown in figure 2.1

2.3 Root Causes for Casting Defects

This section discusses the importance of understanding and analyzing root causes for casting defects in foundry industries.The section continues by describing several significant casting flaws in detail as well as their traits and offering corrective actions to reduce rejection rates and raise quality for various casting problems [8][9].

2.3.1 Blowholes

Blowhole defect occurs when air becomes trapped during the pouring of liquid metal, resulting in rounded or spherical cavities. It can be further classified as pinholes, endogenous (within the casting), and exogenous (on the surface). Pinholes, triangular

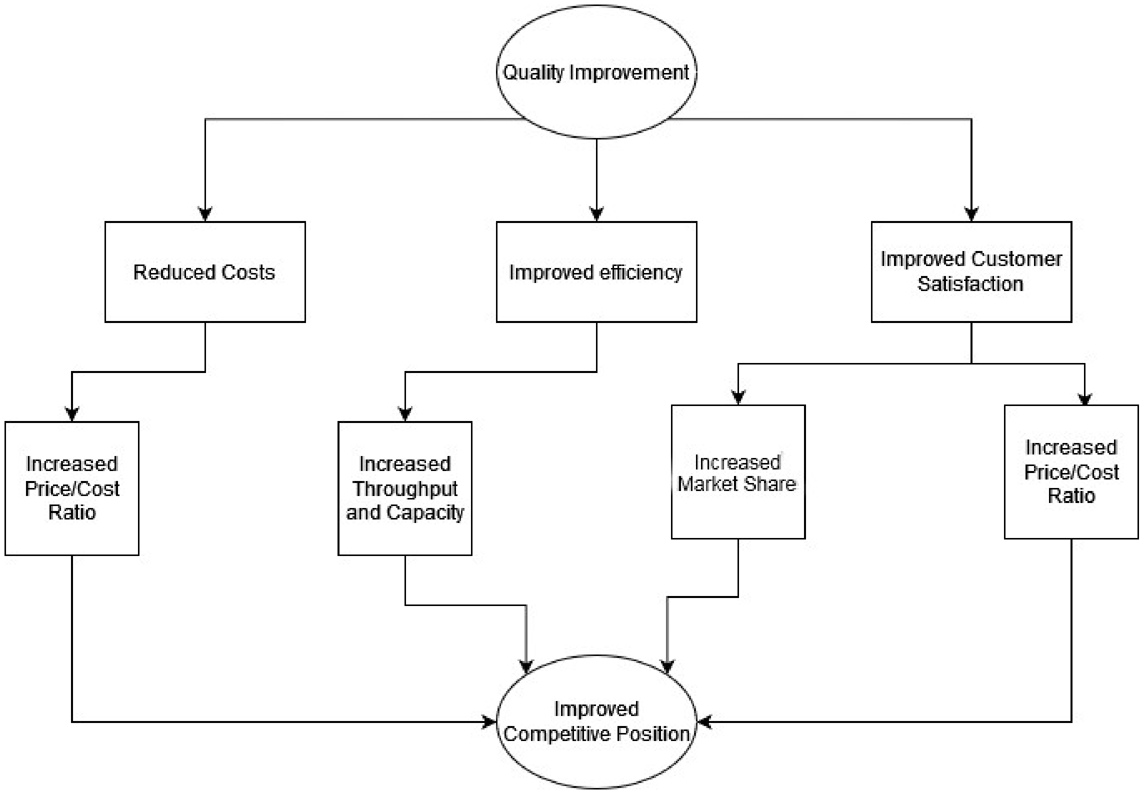

Figure 2.1: Impact of quality on competitive position (adapted from [10])

in appearance, are common in thinner castings and become visible after machining [9].

Blowholes in casting can be caused by a number of things. Blow holes can be caused by a variety of factors, including high moisture content on the chills, limited sand permeability, and inappropriate sand mixing. Blow holes can also result from improper venting in the mold, the use of uncured coatings, and high turbulence during filling. Additional causes include excessive gas entrapment in the liquid metal, insufficient metal degassing, and a high bentonite content in the molding sand. Furthermore, a high binder content can produce gases and aid in the development of blowholes [13].

There are numerous corrective actions that can be taken to avoid blowhole defect in casting. As excessive sand compaction and moisture on chills might result in blowholes, it's crucial to avoid overramming of the sand and only utilize dry chills. It's critical to maintain the ideal pouring temperature for the particular metal being cast. In order to reduce blowholes, it's also crucial to avoid using fine sand grains and to make sure the mold has enough ventilation. Use dried, properly dressed cores, and cut back on the amount of binder and additives in the molding sand [13][14].

2.3.2 Shrinkage

Shrinkage cavities is a result of poor design or inadequate feed metal. It happens when a closed loop forms during solidification and causes a cavity to form.Open (visible on the casting surface), closed (internal or blind), or axial (centerline shrinkage due to high pouring temperature) shrinkage are the three types of shrinkage [13].

Shrinkage defects in castings can result from a number of different factors. Low compressive strength, unevenly dried sand, and poor mold rigidity can all contribute to shrinkage problems. High metal pressure can also cause the mold wall to shift, which can result in uneven cooling and shrinkage. Chills are intended to encourage directed solidification, but when they are placed incorrectly, they can disturb the process and cause shrinkage flaws. Thermal gradients can be produced by abrupt changes in section thickness within a casting, which can cause shrinkage. Sharp interior corners that are too many can function as hotspots and result in localized shrinking. Additionally, casting problems due to shrinkage can be caused by using too much ferro silicate when metal charging [13].

There are numerous corrective actions that can be implemented to remedy shrinkage defects in castings. Using CO2 gas to dry the mold properly for at least 60 seconds will increase mold rigidity and reduce shrinkage. The proper placement of chills can promote directed solidification and minimize shrinkage flaws. Supplying suitable risers to feed sharp corners may prevent hot spots and encourage more uniform solidification. Riser placement, which serves as a reservoir for liquid metal, can help feed the casting as it solidifies and lessen shrinkage. Additionally, the application of inoculants can improve the microstructure and alter the solidification process to reduce shrinkage faults in castings [14].

2.3.3 Hot Tears/Cracks

A hot tear is a crack that develops as a result of an imbalance in temperature during solidification.At the casting's edges, they appear as angular, sharp lines.

Hot tears are more likely when coarse-grained sand is used in the mold, which may also reduce mold strength. Hot tears may form if the mold is disturbed in any way before it has fully solidified. Hot tear development can also be influenced by incorrect riser positioning, which has an impact on how the casting is fed during solidification. Internal tension and the development of hot tears can be caused by abrupt changes in the casting's shape or section thickness. Hot tears can occur as a result of a mold's inability to move freely during solidification due to a high binder concentration and excessive ramming density. Hot tears in castings can also result from high hydrogen content, a lack of eutectic cells at grain boundaries, excessive sulfur or phosphorus concentration in the metal [13].

In order to encourage uniform solidification and lower the possibility of hot tears, it is important to make sure the mold is sufficiently cooled. The possibility of hot tear development can be reduced and stress concentration can be reduced by avoiding abrupt bends or corners in the casting design. Hot tears can be prevented by creating a draft when the casting is being removed from the mold. Hot tear generation can be reduced by deoxidizing the metal to remove sulfur or phosphorus impurities. Casting integrity can be improved by using inoculants, such as a compound of manganese (Mn), silicon (Si), and magnesium (Mg), which can lower the sulfur or phosphorus content. Castings' sensitivity to hot tears can also be decreased by utilizing fine sand grains in the mold and adding coal dust to promote the creation of eutectic cells during solidification [14].

2.3.4 Sand Inclusion

Sand inclusion, typically occur when sand ruptures during mold preparation, is a common problem. Localized defects close to the casting edges are how it shows up. Abrasion of sand by hot metal can also result in inclusions, which are frequently accompanied by Carbon Monoxide bubbles and oxide particles.

Sand inclusions in castings can have a variety of reasons, and many things can lead to them. Directly orienting the metal stream onto the cores can lead to erosion and sand inclusions. Inadequate sand consolidation and the possibility of sand inclusions can result from molds that are unevenly compacted and have areas of varied density. Breakage of the mold during assembly may allow foreign objects to enter the mold, resulting in sand inclusions. Sand inclusions can arise as a result of uneven sand mixing and unequal sand distribution. Sand inclusions can be caused by inappropriate pouring procedures, such as excessive turbulence or poor pouring processes, which might upset the mold [13].

To address sand inclusions in castings, several remedial measures can be implemented. Improve sand cohesiveness and lessen the chance of sand inclusions by using molding sand with a higher bentonite concentration. Sand must be properly rammed during mold preparation to achieve uniform compaction and reduce voids that could result in sand inclusions. Mold boxes must be regularly cleaned to get rid of any loose sand or unwanted objects that can create inclusions. Sand inclusions must be avoided by using properly dressed cores that are free of loose sand or other contaminants. It's crucial to maintain the right sand-to-binder ratio in order to guarantee continuous sand quality and lessen the chance of inclusions.Sand inclusions can be avoided and mold disturbances can be minimized by using proper pouring techniques, which include managing the pouring time and maintaining an ideal pouring height[14].

2.3.5 Flash

Flash is a term used to describe extra material that protrudes perpendicular to the casting face as a thin metallic sheet. It often happens along the mold's splitting line and degrades quality.

There are a number of reasons why castings develop flash defects. Flash and molten metal leaks can occur if there is too much space between the top and bottom halves of the mold box. Flash defects and metal leakage during casting can both be caused by high pouring pressure. Flash formation can be caused by the use of molds with defective pattern designs, such as patterns with insufficient parting line clearance or inappropriate gating systems. Flash defects can also be caused by patterns with voids at the end that encourage the escape of molten metal. Flash can be caused by molten metal escaping from gaps or misalignments caused by improper clamping of the top and bottom halves of the mold [13].

There are numerous corrective actions that can be implemented to remedy flash defects in castings. Flash generation can be reduced and metal leakage can be avoided by sealing the mold box close to the separating line. It is possible to lessen the likelihood of flash defects and prevent molten metal from escaping from any end cavities in the patterns during casting. It's crucial to maintain control over the mold's dimensions and pattern design to prevent excessive gaps or misalignments that could cause flash. A tight seal can be maintained and flash generation prevented by properly aligning and clamping the mold assembly during careful setup. Reduced flash defects in castings can also be achieved through proper core setting, which involves making sure that cores are firmly in place and aligned [14].

2.3.6 Mismatch

A defect known as mismatch is brought on by the top and bottom halves of a mold moving above or below the centerline. This displacement causes castings to be defective.

There are several causes of mismatch defects in castings. Misalignment and mismatch between the top and bottom components of the mold might come from clamping the cope and drag with worn-out dowels. Mismatch defects can also be caused by poor pattern designs for the top and bottom sections, such as insufficient parting line clearance or improper dimensions. Additionally, casting misalignment and mismatch might result from not applying enough weight to the mold's cope portion [13].

Mismatch problems can be reduced by using an appropriate gating system that guarantees a balanced flow of molten metal. During the mold clamping procedure, placing a large weight on top of the casting can help guarantee that the coping component stays in place and prevent misalignment. Before clamping, it's crucial to check that the mold's top and bottom halves are correctly aligned. To guarantee accurate alignment, worn-out dowel pins should be replaced with new ones if they are the root of the mismatch. Accurately aligning the top and bottom sections of the mold can be accomplished with the aid of locators such alignment pins or markings. Additionally, maintaining alignment and avoiding mismatch flaws can be achieved by securely clamping the mold box with C-clamps.

2.3.7 Misrun

Misrun is a defect brought on by the molten metal's lack of fluidity. It has a smooth and unfinished cavity and is more prevalent in castings with a high surface area to volume ratio. Lower pouring temperatures can cause fluid streams to break up, which can cause misruns.

Lower pouring temperatures can make the molten metal less fluid, which can cause the mold cavity to not fill completely and cause misrun defects. When molten metal is poured in erratic bursts, this intermittent pouring might disrupt the flow and leave the mold partially empty. Pouring that is delayed past the point at which it is most effective might cause the metal to solidify too early, limiting appropriate filling and producing misruns. The flow of molten metal can be hampered and misrun defects can result from back pressure during pouring that is brought on by impediments or limitations in the gating system. The mold's insufficient venting might retain air or gases, obstructing the flow of metal and leading to misruns [13].

Several corrective actions can be taken to remedy casting misrun problems. To keep the molten metal fluid and encourage full filling of the mold, it is essential to maintain an adequate pouring temperature. It is possible to avoid pauses and guarantee correct filling by designing an effective gating system that enables a smooth and continuous flow of the metal. Before pouring, it is crucial to properly clean the mold box to get rid of any dirt or particles that can hinder the metal flow and cause misruns [14].

2.3.8 Surface Defects

Due to the flow of molten metal, streaky lines develop as a pattern of lines on the casting surface. These lines, which resemble tiny channels, show a flawed surface.

There are several causes of defective casting surfaces. Surface flaws may develop from the formation of oxide layers on the casting surface. Defects can also be brought on by outside contaminants, such as dirt or debris, that flow onto the casting surface during the pouring process. Low mold temperatures can cause inappropriate solidification and surface flaws. In addition, molten metal with a high slag content may generate slag inclusions on the casting surface, producing surfaces with imperfections [13].

The likelihood of surface imperfections can be decreased by preheating the mold before pouring to ensure that it is at the proper temperature for the solidification process. Slag inclusions on the casting surface can lead to flaws, therefore regular monitoring for slag production during the charging process and proper action to reduce slag concentration will assist prevent problems. The occurrence of surface imperfections can be decreased by lowering the pouring temperature. The amount of foreign impurities that are introduced into the casting surface can be reduced by controlling the pouring parameters, such as the pouring rate and turbulence[14].

2.4 Data-driven Manufacturing

Traditional manufacturing, also known as model-based manufacturing, is based on experts who apply their understanding of physical phenomena to construct theoretical, experimental, and numerical models[15].To grasp the mechanisms underlying manufacturing processes, these models are used. These model-based techniques do, however, have limitations in terms of their accuracy and useful range[15]. They frequently contain oversimplifications and assumptions, and it's possible that human experts aren't always totally impartial or mentally stable while sharing their experiences.Thus, modern production is characterized by data-driven manufacturing [16].

Understanding the shift to data-driven production requires an understanding of the evolution of manufacturing data across time. Limited information based on human experience was verbally transferred during the handicraft era. In the machine age, worker-related and machine-related data emerged, but data handling remained manual. The information age brought about information technologies, which increased the amount of data but frequently created data silos. In the era of big data, manufacturers have improved capabilities for data collecting, storage, and processing thanks to IoT and AI [16].

Quality, adaptability, and autonomy of production can all be enhanced by the data created by industrial processes. The manufacturing layer, the data layer, the knowledge layer, and the decision layer are the four layers that make up the data-driven manufacturing structure.The manufacturing layer encompasses various processes, while the data layer integrates sensors for monitoring and inspection. Raw data is processed in the knowledge layer to extract meaningful features, and intelligence is applied in the decision layer to transform knowledge into decisions [17]. However, accuracy loss can happen at each layer as a result of things like sensor specifications, erroneous knowledge extraction, incorrect data processing, and assumptions made during data correlation. It is important to address these challenges to ensure the effectiveness and reliability of data-driven manufacturing [17].

2.4.1 Data-driven Applications of Product Quality Control

AI and the development of sophisticated sensors have helped to make manufacturing processes more data-driven. Smart sensors allow for the direct measurement of very valuable data, avoiding the need for time-consuming data processing and enhancing accuracy. In order to help with decision-making, cutting-edge machine intelligence and data analysis tools extract important insights from low-value data. The study examines the most recent advancements in data gathering and analysis in data-driven manufacturing processes and explores potential future directions [18].

Data-driven techniques are being developed for product quality control. Numerous quality data, including geometry parameters, location parameters, and machining parameters, are gathered by sensors, Radio Frequency Identification (RFID)s, and machine vision applications. Big data analytics provide for thorough quality control, early defect discovery, and rapid root cause analysis. For the binary classification of quality conditions and the prediction of their association with defects, historical and process condition data are used. Finding influential factors and failure causes is made easier with the aid of weighted association rule mining and Bayesian inference. Hidden causes of production problems can be discovered by combining data and using data mining techniques, enabling the control or elimination of elements that cause defects. With ML, big data analytics facilitate case-based reasoning and transfer of lessons learned to prevent similar problems in the future, embedding quality management into every step of the manufacturing process.Big data analytics with ML enable case-based reasoning and the transfer of learned lessons to avert similar issues in the future, integrating quality control into each stage of the production process [16].

Data driven techniques can help in identifying defects in additively manufactured parts. In [19] the author described Laser Powder Bed Fusion (LPBF) is a complex manufacturing process influenced by multiple factors like laser power, scanning speed, and beam shape. Traditional optimization methods fall short due to limited parameter sets and uncertainties. LPBF's data-rich nature further complicates matters. ML offers a solution by processing vast data, extracting crucial insights, and understanding the entire process-thermal dynamics-structure-property relationship. Techniques like Gaussian process, artificial neural network, logistic regression, random forest, gradient boosting, and deep belief network effectively optimize LPBF processes, control part quality, and accommodate uncertainties.In [20] the author introduces two ML techniques to automatically identify and categorize wheel defects on railway wagons. These methods utilize the wheel vertical force data collected by a permanent sensor system on the railway network. The proposed methods can detect defects like flat spots, shelling, and non-roundness as trains pass measurement sites at full operational speed. The study highlights the effectiveness of the Support Vector Machine (SVM) and Deep Neural Network (DNN) models in comparison to classical methods, demonstrating their superiority in detecting and classifying defects. Use of these ML techniques offers an efficient and cost-effective means to enhance wheel defect detection, minimizing damage, noise, and reliability issues in railway infrastructure.

ML also has extensive use in foundry for process optimization and predicting quality defects in cast components. In [21], the author investigates the use of ML techniques to identify defects in high-precision metal casting production, specifically focusing on micro shrinkage and ultimate tensile defects. Various methods like decision trees, boosting, and neural networks are employed to detect these defects. The evaluation is based on how accurately these methods predict defects and how close their predictions are to actual values. Bayesian networks and neural networks are found to be the most successful in defect identification. The study conducts experiments on two types of defects: micro shrinkage and ultimate tensile strength production defects. For micro shrinkage, Bayesian probability theorem and ANN achieve the highest accuracy rates of 96.6% and 94.1%, respectively. For ultimate tensile defects, cat boosting technology exhibits the highest accuracy of 99.2%. The comparison of ML algorithms utilizes performance metrics like MAE, MSE, RMSE, and RMSLE.

In another article [22] the author explores how ML can predict defects in casting surfaces within the foundry industry, a crucial matter due to its potential economic and environmental consequences. They gathered data from a steel and cast iron foundry and used it to train six ML models, including Gradient Boosting, CatBoost, Extremely Randomized Trees, XGBoost, Random Forest, and Ridge Regression. These models aimed to anticipate the number of components that would be rejected by optimizing production processes. The evaluation used metrics like MAE, MSE, RMSE, R2 score, and RMSLE. The research showcased how ML can efficiently enhance processes and decrease waste in foundry operations. The Extremely Randomized Trees model stood out with the best prediction performance. The authors discovered that factors like the content of organic binders, fines, and gas permeability in the mold material had a significant impact. This insight is important for the industry's sustainability, economic efficiency, and waste reduction efforts.

2.5 ML Algorithms

ML is a branch of AI that focuses on creating algorithms and systems capable of learning from data to make predictions or decisions. This enables computers to automatically recognize patterns and relationships in data, using that knowledge to make predictions or take actions without explicit programming. ML finds use across diverse domains, including image recognition, language processing, personalized marketing, fraud detection, manufacturing optimization, healthcare diagnosis, financial prediction, and autonomous systems. The potential applications of ML are vast, holding the capacity to transform industries by enhancing efficiency, accuracy, and decision-making processes [23]. There are three main types of ML models: supervised learning, unsupervised learning, and reinforcement learning.
Supervised learning models are trained on labeled data, which means that the data has been classified into categories. The model learns to predict the category of new data points. Common supervised learning techniques include classification and regression. Unsupervised learning models are trained on unlabeled data. The model learns to find patterns in the data and group the data points together into clusters. Common unsupervised learning techniques include clustering. Reinforcement learning models are trained by trial and error. The model learns to make decisions by interacting with its environment and receiving rewards for good decisions and punishments for bad decisions [24]. The choice of ML algorithm depends on the specific problem that is being solved. It is important to consider the type of input data, the desired outcome, and the limitations of each algorithm.

2.5.1 ML Algorithm used in this Thesis

Supervised learning is a fundamental ML approach where algorithms learn from labeled data to make predictions or classifications.Since the data used in this thesis are labelled, supervised learning is adapted. Common models include linear and logistic regression for predicting numerical and binary outcomes, decision trees for both classification and regression, K-nearest neighbors for assigning new samples based on nearest neighbors, support vector machines for optimal class separation, Naive

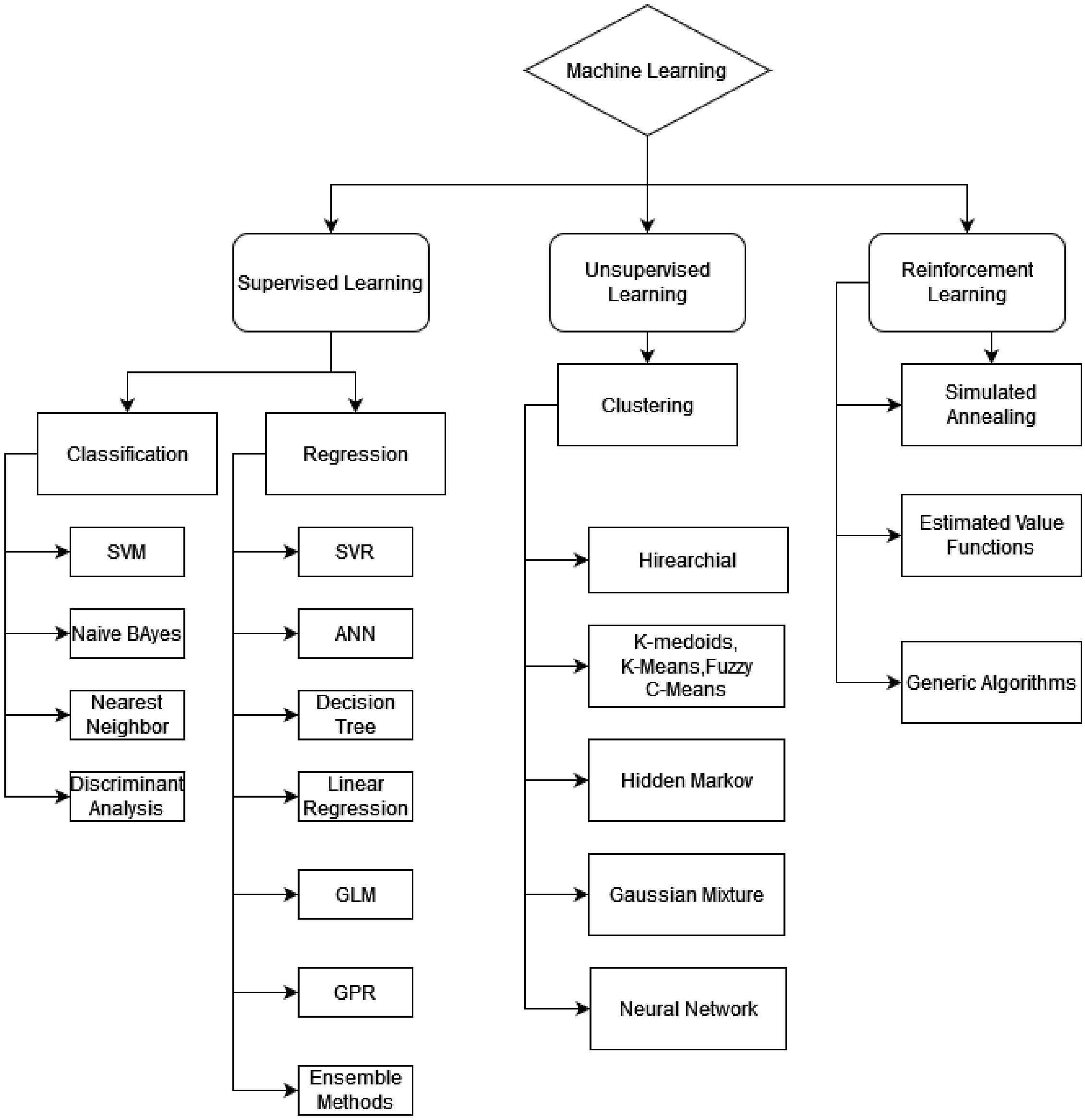

Figure 2.2: Classification of ML Algorithms.

Bayes for probabilistic classification, and neural networks mimicking human brain structure. These models offer benefits like accuracy, interpretability, and versatility, finding applications in predictive maintenance, customer segmentation, medical diagnosis, and more. They excel in making accurate predictions and classifications across various domains, making them a valuable tool with wide-ranging practical applications [23].

2.5.1.1 Decision Trees

A Decision Tree (DT) is a supervised ML algorithm that mimics the structure of an actual tree, consisting of a root node, branches, and leaf nodes. The root node acts as the starting point and is considered the parent of all other nodes. Each branch in the tree represents a decision or rule based on input features, leading to a leaf node as the final outcome. This outcome can be either a categorical or numerical value. The DT algorithm learns by iteratively making simplified decisions through a sequence of if-else clauses, eventually assigning labels to data points. This algorithm is particularly adept at grouping data, making it valuable for tasks like clustering and regression. Its interpretability and applications in diverse fields, such as medical diagnosis, voice and character recognition, finance, and logistics, contribute to its widespread adoption [25].

2.5.1.2 Ensemble Learning

Ensemble learning combines multiple models to enhance accuracy and robustness. It comes in two forms: bagging (training different models on subsets of data) and boosting (sequentially training models to correct errors). This helps prevent overfitting and strengthens predictions.

2.5.1.3 Random Forest

Random Forest uses a collection of decision trees to predict outcomes. It creates these trees using random subsets of data and features. By averaging predictions from individual trees, it improves accuracy and is especially useful for datasets with many features [23].

2.5.1.4 Gradient Boosting Methods

Gradient Boosting combines weak models in sequence to form a strong model. Each model focuses on correcting errors made by its predecessor. It is effective for regression and classification tasks and finds applications in various domains. For further information how decision trees, ensemble learning, Random Forest, and Gradient Boosting methods contribute to the field of ML, their mechanisms, advantages, and applications, the readers are referred to the reference [23].

2.5.2 Evaluation Metrics

Evaluation metrics play a pivotal role in the realm of ML pipelines, offering insights into the model's progress and efficiency. Constructed around the principle of constructive feedback, ML model development entails receiving feedback from metrics, leading to iterative enhancements until desired outcomes are attained. A key attribute of evaluation metrics is their capacity to effectively distinguish between different model results [26]. Predictive models can be categorized as regression models, yielding continuous outputs, or classification models, producing nominal or binary outcomes. Each model type employs distinct evaluation metrics. In classification problems, two algorithm types are utilized based on output nature: those generating class outputs, such as Support Vector Machine (SVM) and k-Nearest Neighbor (KNN), and those yielding probability outputs, including Logistic Regression, Random Forest, Gradient Boosting, and Adaboost. Conversely, regression problems entail continuous outputs with no inherent inconsistencies, requiring no additional treatment [26].

2.5.2.1 Confusion Matrix

In this section, a detailed explanation of the evaluation and enhancement of classification problems is provided, employing various matrices. The Confusion Matrix stands as a crucial tool, offering a comprehensive overview of the model's predictions and their accuracy. Presented in a tabular format, it places actual ground truth labels against the model's predictions. Each column signifies instances within a predicted class, while each row corresponds to instances within an actual class. Although not a performance metric by itself, the Confusion Matrix serves as a fundamental framework for subsequent evaluation metrics [27]. The matrix's cells encompass the following components:

1. True Positive (TP): Signifies correctly predicted positive class samples.
2. True Negative (TN): Indicates accurately predicted negative class samples.
3. False Positive (FP): Pertains to negative class samples erroneously predicted as positive, corresponding to a Type-I error.
4. False Negative (FN): Denotes positive class samples mistakenly predicted as negative, analogous to a Type-II error in statistical terminology.

The different metrics to evaluate the performance of a ML model are explained in the below sections

2.5.2.2 Accuracy

Accuracy is a metric that quantifies the proportion of predictions made by a ML model that are correct. It's widely used as a fundamental measure for evaluating a model's performance. However, in scenarios where the classes being predicted are imbalanced, accuracy can be misleading since it doesn't consider the distribution of the classes [28].

2.5.2.3 Precision

Precision is a metric that calculates the ratio of correctly predicted positive instances to the total number of instances predicted as positive. It is particularly valuable when the cost of false positives is high. For example, in medical diagnosis, precision is crucial. If a model predicts a patient has a disease, it's important to be confident in that prediction to avoid causing unnecessary stress or treatments for a patient who is actually healthy [28].

2.5.2.4 Recall

Recall, also known as sensitivity or true positive rate, assesses the fraction of actual positive instances that were correctly predicted as positive by the model. It's beneficial when the cost of false negatives is high. In scenarios like fraud detection, where missing a fraudulent transaction has significant consequences, a high recall is essential to capture as many actual positives as possible [28].

2.5.2.5 F1 score

F1 score is a composite metric that strikes a balance between precision and recall. It's the harmonic mean of precision and recall, offering a single value that considers both aspects. This metric is particularly valuable when both precision and recall are of equal importance. For instance, in predicting exam pass rates for students, a high F1 score ensures that students who will pass aren't missed while also minimizing incorrect predictions [28].

3
Methodology

This chapter gives an organized overview and comprehension of the thesis project's methodology. It will detail the steps used to develop the analysis, as well as the resources and methods used. The thesis adopts the CRoss Industry Standard Process for Data Mining (CRISP-DM), as a reference model for its execution. This model offers a standardized structure for the case study planning and management. CRISP-DM entails six key phases in the data mining process: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. This established framework is tailored to suit the case specific application. For instance, the deployment phase is not considered in the scope due to the limited time as well as the decision lies on the company's internal management. Instead of deploying the built model in the production environment, a report and presentation are given to the company to clearly communicate the findings and knowledge gained through the thesis.

3.1 CRISP-DM

Organizations are rapidly realizing the importance of efficiently processing and interpreting massive amounts of data to support their decision-making processes in today's data-driven environment. Data science has become a vital field for gaining important insights from data using models and applications based on math and analysis. The adoption of project management and process approaches has been noted as an important success element for data science projects. However, using such approaches creates significant difficulties in data science teams.

The CRISP-DM as garnered attention within the industry since it is a universally applicable process and presently widely adopted as the standard framework for conducting data mining projects.

Despite proven process models having been around for a while, only limited use of explicit process models in modern data science projects is noted.[29] includes a thorough literature review that focuses on the use of CRISP-DM in research studies as a means of bridging this gap, to find the best practices and process stages that can more effectively aid data analysts.The subsequent sections of this study will delve deeper into the CRISP-DM seeking to harness the potential of data processing and mining methodologies for informed decision-making processes in diverse domains.

In data analytics projects in the engineering sector, the CRISP-DM process contains six steps that explain the business understanding, data understanding, data preparation, modeling and assessment, and deployment phases.[29][30]

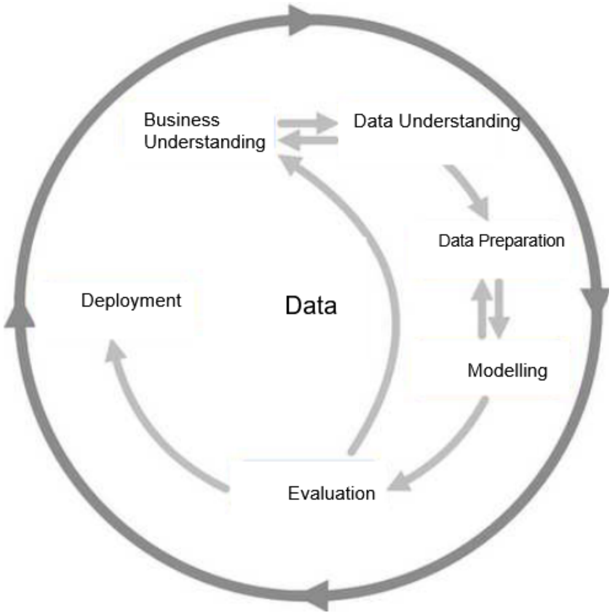

Figure 3.1: Methodology of CRISP-DM.
(adapted from [31]

3.1.1 Business Understanding

The initial stage of the CRISP-DM process model is known as Business Understanding. During this phase, the project's aims and requirements are defined. The primary focus is on comprehending the business issue at hand, pinpointing the data mining objectives, and devising a project strategy. This phase encompasses several key tasks: firstly, determining the business objectives by grasping the business problem and establishing project goals; secondly, evaluating the context by identifying available resources, limitations, assumptions, and potential risks; thirdly, establishing data mining objectives, encompassing identifying the key questions to be answered and specifying the data for analysis; and finally, formulating a project plan detailing the tasks, schedules, and resources necessary for project completion [31].

3.1.2 Data Understanding

The second phase of the CRISP-DM process model is Data Understanding. During this stage, the primary focus lies in collecting, examining, and preparing data for subsequent modeling activities. The emphasis is on achieving a comprehensive comprehension of the data, pinpointing potential quality concerns, and selecting pertinent data for analysis. This phase involves several essential tasks: firstly, gathering initial data from various sources; secondly, providing an overview of the data, including its types, distributions, and any existing quality issues; thirdly, delving into data exploration through visualization, pattern identification, and outlier detection; and lastly, ensuring data quality by addressing missing values, inconsistencies, and errors [31].

3.1.3 Data Preparation

The third phase of the CRISP-DM process model is Data Preparation. This phase focuses on getting the data ready for modeling by cleaning, transforming, and preprocessing it. The key tasks here involve picking the right data for analysis, eliminating irrelevant data, and ensuring data quality. This is followed by cleaning the data to remove duplicates, fill in missing information, and fix errors. Furthermore, new variables are created, data is combined or summarized, and information is transformed to better suit the modeling process. Lastly, data from various sources are combined and any differences are resolved to create a consistent dataset for analysis.[31]

3.1.4 Modelling

The fourth phase of the CRISP-DM process model is Modeling, where the data is transformed using techniques like decision trees, neural networks, and logistic regression. The primary focus lies in constructing and validating models, picking the most suitable one, and thoroughly documenting the entire modeling process. This phase involves a series of tasks: first, selecting the right modeling technique according to the data mining objectives and data attributes; next, devising a test plan to assess the model's performance; then, building the model by training it on the available data and adjusting parameters as needed; and finally, evaluating the model's effectiveness using test data and recording all steps followed during modeling [31].

3.1.5 Evaluation

The Evaluation phase, which is the fifth step in the CRISP-DM process model, involves thoroughly assessing and testing the developed models to ensure their alignment with the intended business goals. The primary focus here is to gauge how well the models perform, validate their underlying assumptions, and ascertain their practical utility. This phase encompasses several tasks. Initially, it entails evaluating the model's outcomes by comparing its predictions with the actual observed values. Subsequently, a comprehensive review of the modeling process takes place, aiming to analyze and validate the assumptions made during the model's creation. Lastly, based on the insights gained from the evaluation, informed decisions are made about the subsequent course of action. These decisions could encompass refining the model, considering alternative modeling techniques, or progressing to the deployment stage [31].

3.1.6 Deployment

The sixth and concluding phase of the CRISP-DM process model is Deployment. In this phase, the model is put into action within the business context. The focus lies in effectively implementing the model, keeping a close watch on its performance, and ensuring alignment with the business goals. The phase encompasses several key tasks: initially planning the deployment, determining necessary resources, and creating a deployment strategy; subsequently, devising a monitoring and upkeep plan, including identifying metrics to assess the model's performance; then proceeding to integrate the model into existing systems and processes within the business environment; lastly, continuously monitoring the model's accuracy, reliability, and overall effectiveness to ensure it functions optimally [31].

4
Results

This chapter provides an elaborate elucidation of the progression of cores produced by Machine A and Machine B through the casting process. Furthermore, it presents the outcomes yielded by two ML models, adhering to the systematic phases of the CRISP-DM methodology.

4.1 Business Understanding

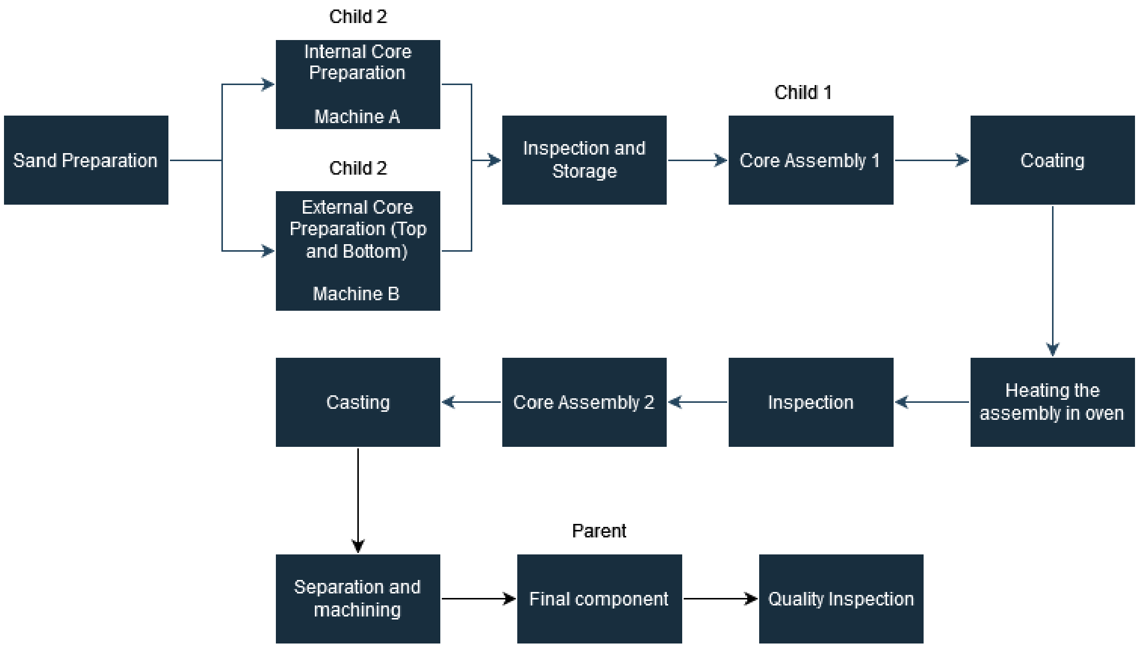

Figure 4.1: Foundry line flowchart.

The above Figure 4.1 illustrates a comprehensive flow chart depicting the foundry line that has been taken into consideration for the project at a Swedish Automotive Industry that casts a power-train component. The flow chart represents the sand casting process as explained in 2.1.2, The initial stage involves the preparation of sand, where sand with desired properties is carefully mixed with the appropriate amount of binder and additives to enhance its strength. This sand mixture is then transported via a conveyor to Machine A and Machine B, where it is compacted to form the shape of the core. Machine A and B are responsible for crafting the

internal (inside) and external cores(Top and Bottom cores), respectively. Following their creation, these cores undergo thorough inspection.
Next in the core assembly 1, the bottom core and internal cores are assembled with the help of robots that precisely position the internal core within the bottom core, utilizing a camera to accurately identify the correct placement. Subsequently, the core assembly undergoes a coating and heating process within an oven. This step is crucial as it yields a superior surface finish, enhances thermal insulation, and minimizes any undesirable interaction between the core and the metal.
After the coating and heating, a meticulous manual inspection is conducted to detect any potential core damages as any flaw in the core could result in a defective component. Next stage is Core Assembly 2 where the Top core and the core Assembly 1 are matched and sent to the casting bay.Here, molten metal from the furnace is carefully transferred to a ladle, and the ladle automatically fills the cores with the molten metal.
The filled cores are allowed to cool, and subsequently, they are moved to a separation area where the core sand and the solidified metal are separated. Any excess metal parts are machined and removed. Finally, a thorough inspection is performed to ascertain the quality of the component, ensuring that it meets the required standards before being deemed acceptable for use.

4.2 Data Understanding

In this master thesis project, four distinct data sets was utilized, each collected continuously from January 2020 until May 2023 from the foundry line. The essential attributes were extracted from these individual data sets and subsequently merged to create the final data set, which served as the foundation for applying machine learning techniques.

The first data set comprises *Product data* from Manufacturing Execution System (MES), which encompasses all the components generated in the production line. In order to cast the component, several elements are necessary: a top core, a bottom core that maintains the outer shape of the component, and a series of internal cores positioned between these two halves. These internal cores bring forth the intricate internal structures required for the component. The MES data has 9 attributes that encompasses information about each individual core and the sequential operations performed on them. With the aid of the MES data, it becomes effortless to trace which cores were utilized to create the core assemblies, which, in turn, was used to manufacture the final component. Additionally, the MES data includes timestamps for the various operations conducted on these cores.

The *Quality Deviation Data* as shown in 4.2 is obtained from the Quality System which is maintained by the Quality Control Division of the foundry line. As the name suggests, this dataset tracks deviations to facilitate the identification of nonconformities, root cause analysis, process improvement, regulatory compliance, and the maintenance of efficient quality control procedures.

The Quality Deviation Data, consisting of 26 attributes corresponds to defective components produced in the line. This data provides answers to several important questions, including the product's serial number, component name, type of deviation observed in the component, specific production line where the deviation occurred, cost incurred by the defective component,origin of the deviation (problem owner), casting date of the product, and registration date of quality deviation in the casting.

	Serial no	Komponentnamn	Deviation	Department	Problem owner	Cost	View	Casting time	Deviation reg. time	Action reg. time	YYYYWWD
0	[illegible]	[illegible]	Solid burning	Rensbur3	Casting	2000	Top section	2022-05-13 18:41:47	2022-05-17 12:17:05	2022-05-17 12:17:06	2022202
1	[illegible]	[illegible]	Solid burning	Rensbur3	Casting	2000	Top section	2022-05-13 18:40:21	2022-05-17 12:18:37	2022-05-17 12:18:38	2022202
2	[illegible]	[illegible]	Solid burning	Rensbur3	Casting	2000	Top section	2022-05-13 18:01:21	2022-05-17 12:32:32	2022-05-17 12:32:32	2022202
3	[illegible]	[illegible]	Solid burning	Rensbur3	Casting	2000	Top section	2022-05-13 17:53:12	2022-05-17 12:30:55	2022-05-17 12:30:56	2022202
4	[illegible]	[illegible]	Solid burning	Rensbur3	Core making	2000	Top section	2022-05-13 17:59:56	2022-05-17 12:34:16	2022-05-17 12:34:17	2022202
5	[illegible]	[illegible]	Solid burning	Rensbur3	Casting	2000	Top section	2022-05-13 18:00:38	2022-05-17 12:28:43	2022-05-17 12:28:44	2022202
6	[illegible]	[illegible]	Solid burning	Rensbur3	Casting	2000	Top section	2022-05-13 17:02:37	2022-05-17 12:27:26	2022-05-17 12:27:27	2022202
7	[illegible]	[illegible]	Solid burning	Rensbur3	Casting	2000	Top section	2022-05-13 16:52:48	2022-05-17 12:24:47	2022-05-17 12:24:48	2022202
8	[illegible]	[illegible]	Solid burning	Rensbur3	Casting	2000	Top section	2022-05-13 16:13:29	2022-05-17 12:29:29	2022-05-17 12:29:30	2022202
9	[illegible]	[illegible]	Solid burning	Rensbur3	Casting	2000	Top section	2022-05-13 12:47:20	2022-05-17 12:35:52	2022-05-17 12:35:52	2022202

Figure 4.2: Registered deviation Data from Quality System

The *process data* obtained from the Manufacturing Execution System (MES) is specifically focused on two machines within the production line, namely Machine A and Machine B, as illustrated in Figure 4.1.These machines are responsible for producing internal and external cores. The Programmable Logic Controller (PLC) system integrated into Machine A and Machine B systematically logs every process step carried out within these machines. These logged data from both machines are vital features for the analysis, aiming to investigate their potential impact on the quality of the final component. Figure 4.3 visually presents the organization of the process data concerning Machine A, providing a clear representation of how this data is systematically stored and managed within the MES infrastructure. The process data from both machines consists of 11 attributes that furnish essential information about the production process. These attributes encompass the work area, unique identifying codes for specific operations, process descriptions, start time and date of each operation, product ID associated with the process, the actual value achieved during the process, the target value set for optimal performance, and the allocated tolerances for the process

Maintenance data pertaining to machines A and B in the foundry line is also considered. This maintenance data is sourced from the Computerized Maintenance Management System (CMMS) of the foundry line. The maintenance data encompasses relevant information, including the number of instances the machines experienced breakdowns, the type of breakdowns encountered, the time of breakdown

ProcessDescription	StartTime	ProductUnit	UpperTolerance	LowerTolerance	ActualValue	TargetValue	Day_Month_Year
[illegible]	[illegible]	[illegible]	0.0	0.0	6.00	0.0	2020-01-07
[illegible]	[illegible]	[illegible]	0.0	0.0	2.00	0.0	2020-01-07
[illegible]	[illegible]	[illegible]	0.0	0.0	10.00	10.0	2020-01-07
[illegible]	[illegible]	[illegible]	0.0	0.0	0.00	2.0	2020-01-07
[illegible]	[illegible]	[illegible]	0.0	0.0	1.90	1.6	2020-01-07
...	...	...	...	...	...	...	...
[illegible]	[illegible]	[illegible]	0.0	0.0	5.04	5.0	2020-01-31
[illegible]	[illegible]	[illegible]	0.0	0.0	0.00	0.0	2020-01-31
[illegible]	[illegible]	[illegible]	0.0	0.0	6662.35	2400.0	2020-01-31
[illegible]	[illegible]	[illegible]	0.0	0.0	5844.56	2400.0	2020-01-31
[illegible]	[illegible]	[illegible]	0.0	0.0	0.00	0.0	2020-01-31

Figure 4.3: Sample MES Process Data for Machine A.

occurrences, and the total downtime experienced by each machine.Maintenance data plays a crucial role in understanding and identifying potential factors that may impact product quality and overall operational efficiency within the foundry line.

4.2.1 Data Description and Visualization

This phase includes various activities to get a deeper understanding and familiarity with the data using visualization techniques. During this period between January 2020 and may 2023, there were a total of 571.673 Product units were produced.

As per Figure 4.4 quality deviations generated from the line is along Y axis and their corresponding value counts along X axis. Around 54% of the problem occurs in the casting process and 38% in the core making process, together these two process account for more than 90% of all deviations.By going a step further in finding out what kind of deviations occur the most it is evident from the figure 4.5 that five out of all 16 types of deviation such as pore holes, wrongly assembled cores, blow holes, core bit, and disintegration contribute to 88% of defective components. Components that has no quality deviation are marked as 'none'.Only the top 5 significant occurences of defects are considered for the analysis.

Considering the maintenance operations there were a total of 350 maintenance operation performed for Machine A and 210 maintenance operation performed for Machine B.

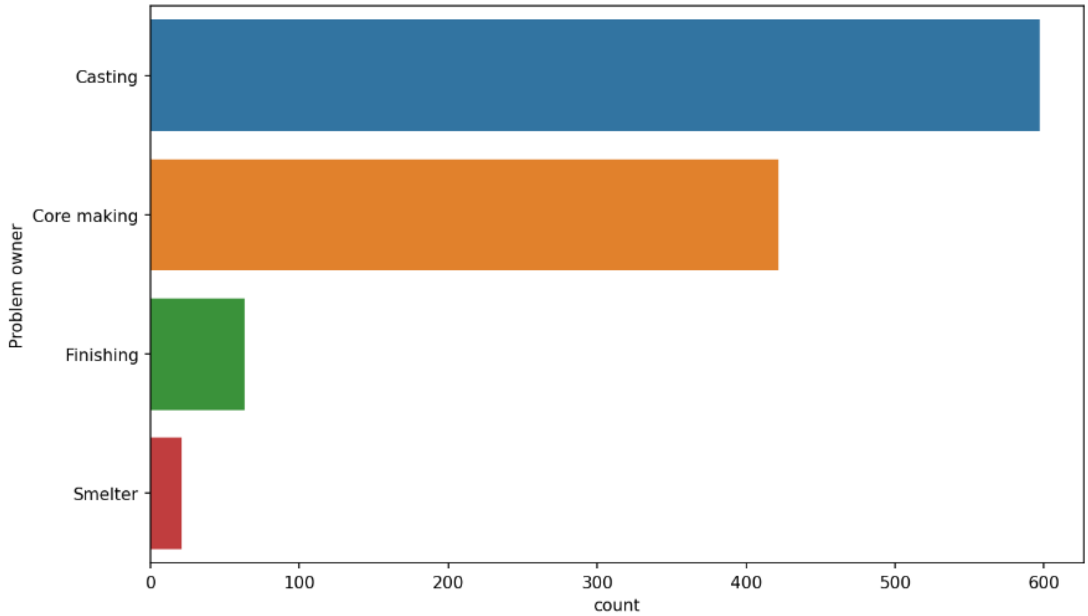

Figure 4.4: Distribution of deviation across different processes in casting cylinder head.

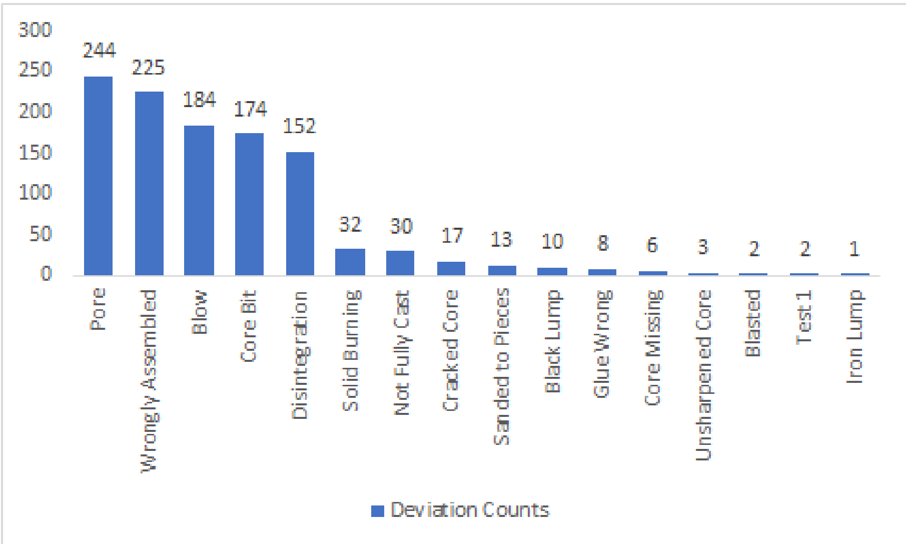

Figure 4.5: Distribution of deviation across different processes in casting cylinder head.

4.3 Data Preparation

The process of Data Preparation was done using Python, which is open source software. The preparation involved using libraries like numpy and pandas. For visualization, seaborn and matplotlib were used. Most of the data in the dataset is in Swedish. Therefore, to enhance the understanding of the data, the initial data preparation steps involved translating the Swedish words used in the dataset. However, not all of the data was translated. The most important attributes for

	WORK_TYPE	Day_Month_Year	Breakdown	WorkArea
4454	DWO	2020-01-24	1	[illegible]
11766	DWO	2020-02-01	1	[illegible]
11767	DWO	2020-02-04	1	[illegible]
11774	EWO	2020-02-05	1	[illegible]
11768	EWO	2020-02-06	1	[illegible]
...	...	...	...	
199223	DWO	2022-05-26	1	[illegible]
203063	EWO	2022-05-29	1	[illegible]
201773	DWO	2022-05-29	1	[illegible]
203061	EWO	2022-05-30	1	[illegible]
198305	EWO	2022-05-31	1	[illegible]

350 rows × 4 columns

Figure 4.6: Machine A maintenance counts.

	WORK_TYPE	Day_Month_Year	Breakdown	WorkArea
4466	EWO	2020-01-14	1	[illegible]
11788	EWO	2020-02-04	1	[illegible]
8264	DWO	2020-02-16	1	[illegible]
12931	DWO	2020-02-17	1	[illegible]
11793	DWO	2020-02-26	1	[illegible]
...	...	...	...	
203098	EWO	2022-05-20	1	[illegible]
203090	DWO	2022-05-20	1	[illegible]
204476	EWO	2022-05-25	1	[illegible]
201295	EWO	2022-05-26	1	[illegible]
203089	EWO	2022-05-27	1	[illegible]

210 rows × 4 columns

Figure 4.7: Machine B maintenance counts.

common readers to comprehend, such as the different types of deviations, problem owners, and the deviation view of the quality deviation data, were translated.
The subsequent phase involved integrating quality deviation data with product data to identify cores utilized in defective components. However, due to the intricacies within the MES data, pinpointing the cores in the faulty component posed challenges. The MES Product data showcases parent and child attributes, as illustrated in Figure 4.1. Nevertheless, second and third-order child products were not explicitly specified. Multiple iterations were carried out on the product data to trace the ultimate child item. This enabled the identification of individual cores produced by both Machine A and B. To facilitate better comprehension, Figure 4.8 presents

the product architecture. In this representation, the final component assumes the role of the parent, while Core Assembly 2 stands as the child component. The task involves extracting Child Component 1 along with the subsequent individual components. For this purpose, Core Assembly 2 is promoted to the parent component status, and its child, Core Assembly 1, is extracted. In subsequent iterations, Core Assembly 1 assumes the parent role, thus allowing the extraction of individual cores as its child components. After completing the iterations, the data was merged with

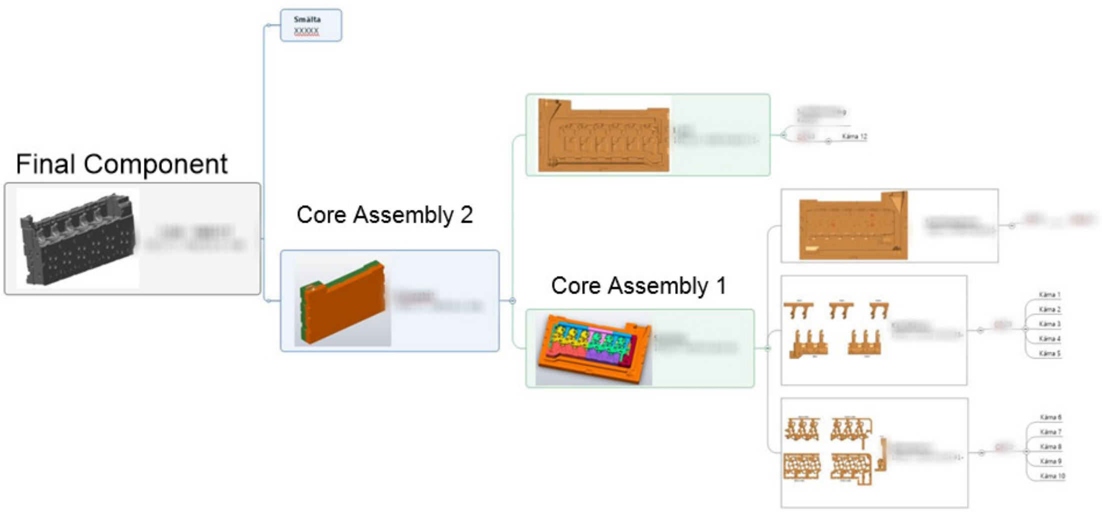

Figure 4.8: MES Product data architecture.

the individual process data. The merging was performed seperately for Machine A and Machine B. These separate comprehensive datasets enable us to identify which processes in both machines have an impact on the final product quality.
Finally, two pivot tables were generated for Machine A and Machine B to transpose the process descriptions into attributes and their corresponding maintenance data is merged to find out the impact of the maintenance onf the final product quality in the analysis.

4.4 Algorithm Modeling and Evaluation

Modeling was conducted using Python with the scikit-learn library. Considering one machine for the instance, machine B. The determination of importance features for each of the top five deviation classes, namely "Pore," "Wrongly Assembled," "Blow," "Core Bit," and "Disintegration," is a critical aspect of this study, as depicted in Figure 4.5. However, it is noteworthy that the distribution of these deviation classes is imbalanced, as illustrated in Figure 4.9. Particularly, the class labeled "None" significantly dominates the dataset, exhibiting an exponential abundance compared to the other classes. This imbalance arises due to the majority of individual core components produced by the machines do not have quality deviation, with only small fraction exhibiting deviations. The dataset is divided into a training dataset, where 80% of the data was used to train the model, and a testing dataset, which was used to evaluate the model, consisting of 20% of the data.

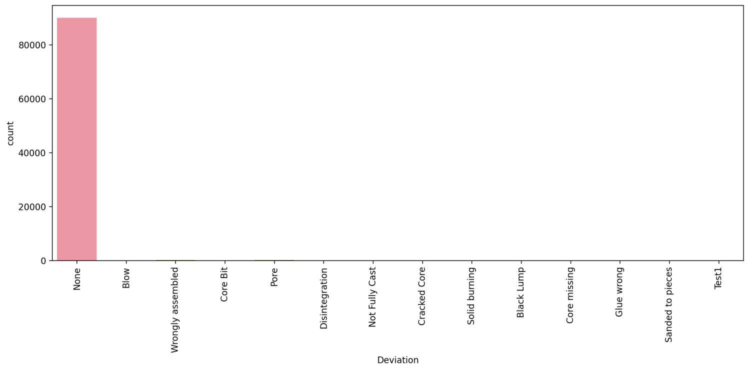

Figure 4.9: Unbalanced class distribution of deviation data.

4.4.1 Data Balancing

The presence of class imbalance poses a challenge in constructing an unbiased and effective machine learning model. A model trained on imbalanced data may be biased towards the majority class, leading to sub optimal performance on the minority classes. To address this issue, it becomes imperative to balance the class distribution in the dataset. Thus, a division of the dataset into two distinct DataFrames is performed based on the target variable, 'Deviation.' The DataFrame labeled as 'majorityclass' comprises samples where the 'Deviation' is labeled as 'None,' which denotes that the product do not have any deviations. while the DataFrame denoted as 'minorityclass' consists of samples where there is a 'Deviation'. The first minority class considered is 'Blow' as shown in figure 4.10. Further iterations will be carried on with other four deviations as the minority class.

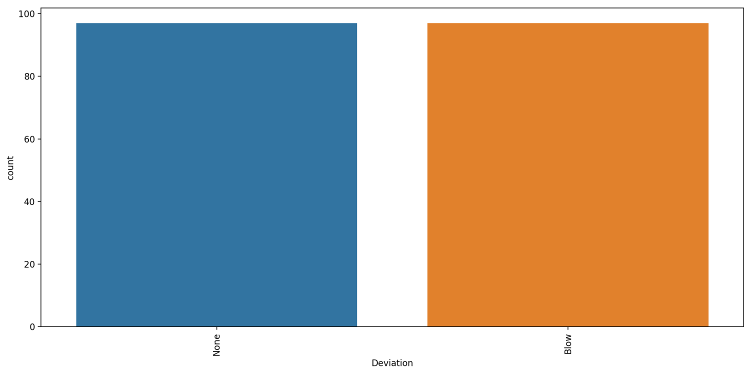

Figure 4.10: Balanced class distribution of none and blow.

One minority deviation class is considered at a time to avoid bias in the analysis. If all five minority deviation classes are combined, the undersampled majority class would still dominate the dataset as shown in the figure 4.11, potentially leading to biased results. Additionally, oversampling the minority class may result in overfitting the model. To address this issue, each minority deviation class is separately analyzed to ensure fair representation and avoid biases. This approach enables us to focus on specific deviations and derive meaningful insights without compromising the model's performance due to class imbalance.

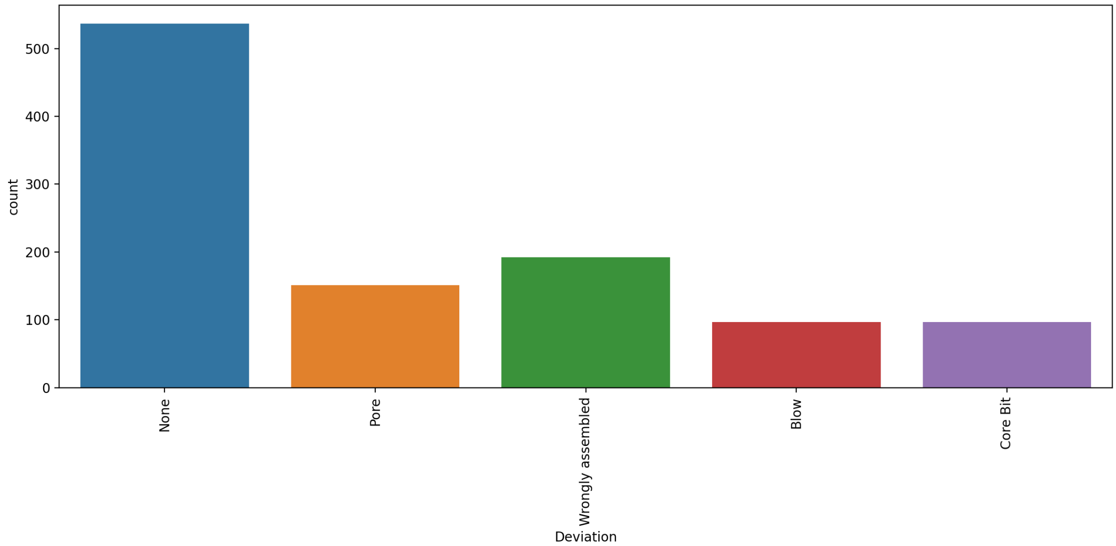

Figure 4.11: Undersampled classes with imbalance.

4.4.2 Model Evaluation

The scikit-learn ensemble random forest classifier was employed to perform machine learning on the data after under-sampling the majority class and scaling the features. The classification report provided an in-depth evaluation of the model's performance in classifying the two classes.
For the "Blow" class, the model achieved a precision of 0.70, indicating that 70% of the instances predicted as "Blow" were correctly classified. The recall, also known as sensitivity or true positive rate, was 0.64, suggesting that the model accurately identified 64% of the actual "Blow" instances. The F1-score, which balances precision and recall, was 0.67, representing an overall measure of the "Blow" class's classification performance.
Similarly, for the "None" class, the model achieved a precision of 0.58, meaning that 58% of the instances predicted as "None" were correctly labeled. The recall for the "None" class was 0.65, indicating that the model correctly identified 65% of the actual "None" instances. The F1-score for the "None" class was 0.61, reflecting the harmonic mean of precision and recall for this class.
The overall accuracy of the model, measuring the correct classification of all instances, was 0.64, indicating that 64% of all instances were classified accurately.

Similarly, for the other minority classes, namely Random Forest and Gradient Boosted Classifier methods, were also utilized. The identical procedure was applied to Machine A, and the corresponding classification results are presented in the figures 4.13. & 4.14

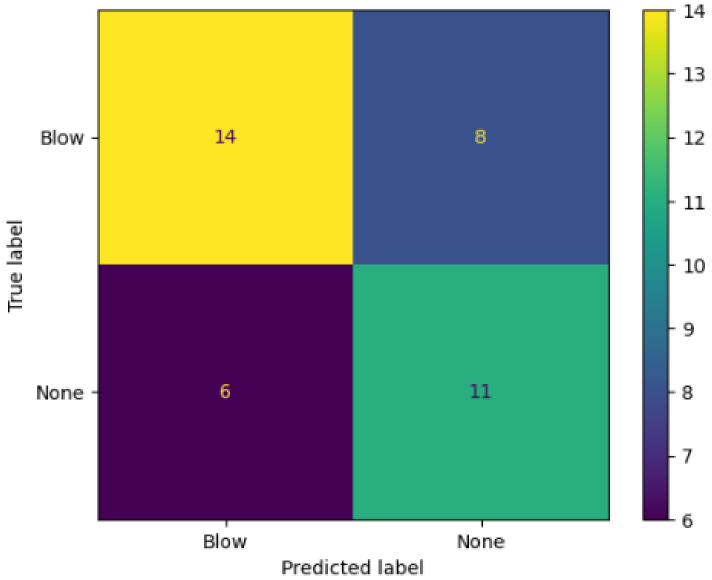

Figure 4.12: Confusion matrix of Blow and none class in Random forrest classifier

Machine A		Precision		Recall		F1		Accuracy
			none		none		none	
Blow	RF	0,58	0,78	0,86	0,44	0,7	0,56	0,64
	GB	0,62	0,79	0,84	0,54	0,71	0,64	0,68
Pore	RF	0,59	0,76	0,89	0,37	0,71	0,49	0,63
	GB	0,6	0,63	0,69	0,53	0,64	0,58	0,61
Core Bit	RF	0,69	0,85	0,9	0,59	0,78	0,7	0,74
	GB	0,61	0,65	0,69	0,56	0,65	0,6	0,63
Disintegration	RF	0,64	0,69	0,81	0,47	0,72	0,56	0,65
	GB	0,59	0,55	0,69	0,45	0,63	0,49	0,57
Wrongly Assembled	RF	0,63	0,7	0,83	0,45	0,72	0,55	0,65
	GB	0,51	0,81	0,8	0,53	0,62	0,64	0,63

Figure 4.13: Machine A classification report for Random Forest and Gradient Boosted Models

Machine B		Precision		Recall		F1		Accuracy
			none		none		none	
Blow	RF	0,7	0,58	0,64	0,65	0,67	0,61	0,64
	GB	0,68	0,55	0,59	0,65	0,63	0,59	0,62
Pore	RF	0,42	0,4	0,34	0,48	0,38	0,44	0,41
	GB	0,43	0,39	0,38	0,45	0,4	0,42	0,41
Core Bit	RF	0,71	0,52	0,45	0,76	0,56	0,62	0,59
	GB	0,67	0,47	0,27	0,82	0,39	0,6	0,51
Disintegration	RF	0,58	0,54	0,58	0,54	0,58	0,54	0,56
	GB	0,6	0,53	0,46	0,67	0,52	0,59	0,56
Wrongly Assembled	RF	0,56	0,66	0,69	0,51	0,62	0,58	0,6
	GB	0,55	0,62	0,61	0,56	0,58	0,59	0,58

Figure 4.14: Machine B classification report for Random Forest and Gradient Boosted Models

5 Discussion

This chapter addresses a common challenge encountered by companies, which involves translating insights derived from results into effective business actions. The primary focus here is to discuss and interpret outcomes, equipping stakeholders with the necessary motivation to implement future measures. Additionally, the chapter delves into research questions, enhancing the understanding of acquired results.

The goal of the quality prediction model is to provide insights about potential issues in the core making process and to efficiently identify defective cores before they are casted during the final manual inspection, prior to being sent to the casting area. The foundry line exhibits several quality deviations, including Pore holes, Wrongly assembled cores, Blow holes, Core Bit, and Disintegration. These deviations address Research Question 1, which pertains to the major quality issues during production.

The model evaluation comprises two datasets: one for training the model and the other for testing its performance. During data preparation for ML, only the top 5 deviations were considered due to their prevalence over other, less frequent deviations. An issue arises from the data's class imbalance, with the 'none' class, representing products without deviations, being significantly higher compared to the other classes. This class imbalance could bias the ML model's accuracy towards the majority class.

To address this imbalance, it is vital to utilize appropriate techniques such as resampling methods and to employ evaluation metrics like precision, recall, and F1-score to assess the model's performance across all classes. This approach helps prevent bias towards the majority class and ensures the model effectively identifies and predicts various deviations, including critical ones.

Erroneously predicting a good core as bad could lead to unnecessary inspections and increased time spent on non-existent damage investigation. Conversely, wrongly predicting a bad core as good could result in defective cores entering production undetected, leading to casting defects and compromising the final product's quality. To mitigate these risks, the majority class 'none' is undersampled to match the minority class deviations, followed by transforming the data into a binary classification. The results indicate that the model's performance is not biased towards the majority class.

Evaluating the model's performance when used in the foundry line is crucial. For

classification models, accuracy, precision, recall, and F1-score are vital performance metrics. Two ML models, Random Forest and Gradient Boosting, were employed. Results reveal that the Random Forest model has better accuracy in predicting cores with or without defects in both machine A and machine B, this answers to the RQ3. Feature importance analysis of the Random Forest model highlights critical operations significantly contributing to specific defects in the production process.
The analysis of feature importance conducted on the Random Forest model vividly underscores the vital operations that substantially contribute to specific defects within the production process. Illustrated in Figure 5.1, the assessment for Machine A brings forth key insights. For instance, concerning blow holes, Operations 1, 5, 11, and 7 emerge as highly influential factors. Likewise, Core Bit deviations are significantly shaped by Operations 10, 1, 5, and 11. Meanwhile, Disintegration is prominently influenced by Operations 1, 2, 5, 10, and 3. The contribution of Operations 1, 10, and 5 plays a significant role in Pore defects, while wrongly assembled products are primarily influenced by Operations 1, 2, 5, 3, 11, and 10. Moving on

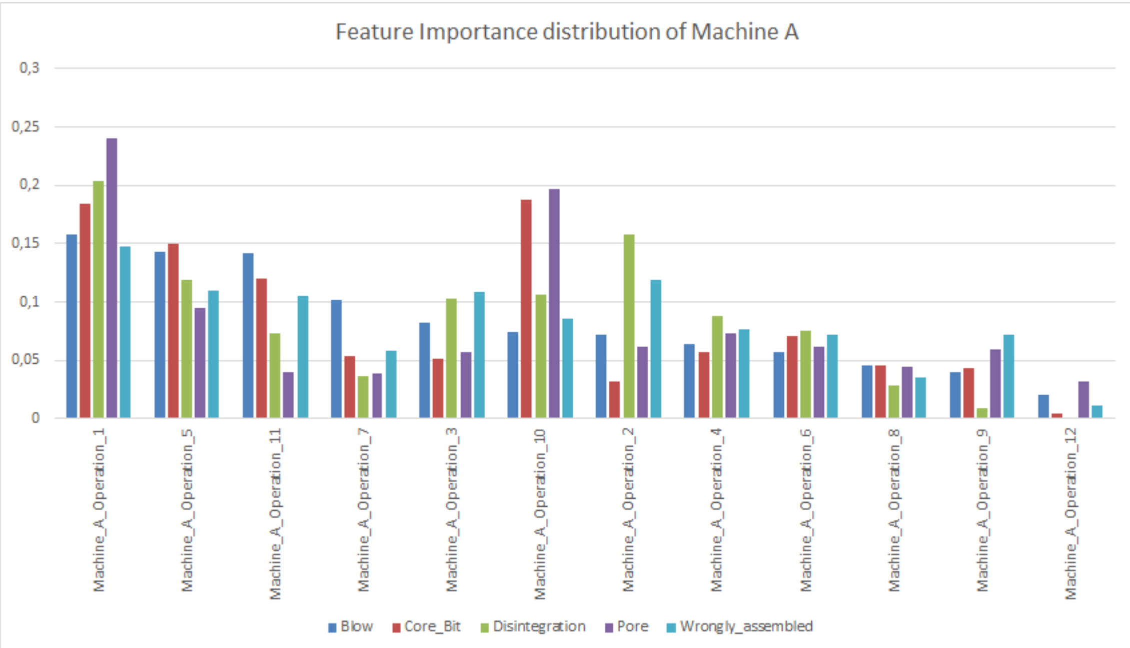

Figure 5.1: Feature Importance distribution of Machine A.

to Machine B and detailed in Figure 5.2, the analysis emphasizes the crucial roles of distinct operations in shaping defects. Operations 1, 2, 3, 4, and 5 stand out for blow holes, while Core Bit deviations are influenced by Operations 5, 3, 4, 1, and 2. Disintegration's primary contributors are Operations 2, 11, 9, 3, and 5. Notably, Operations 1, 10, 3, 9, 10, and 12 drive Pore defects, whereas wrongly assembled products are primarily impacted by Operations 5, 4, 2, and 3. Moreover, it's evident that features 14, 15, 16, 17, 18, and 19 hold no significant role in these deviations.

This comprehensive exploration of feature importance addresses Research Question 2, shedding light on the pivotal process parameters and maintenance factors influencing notable quality issues. Importantly, this analysis underscores that maintenance

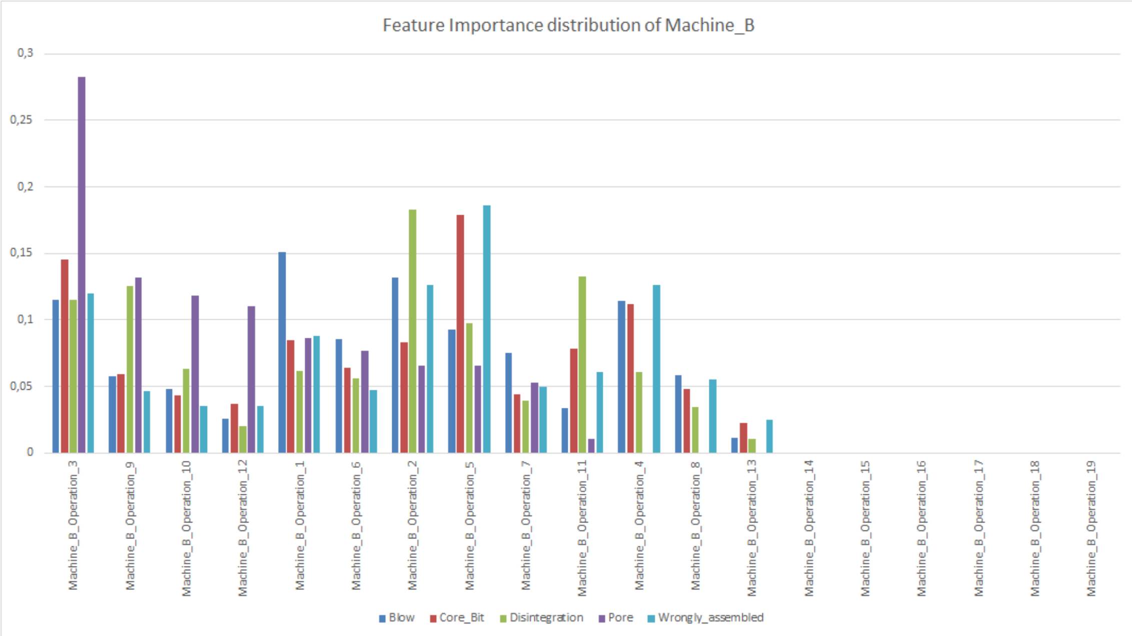

Figure 5.2: Feature Importance distribution of Machine B.

has no discernible impact on the final product quality of the components.

5.1 Recommendation and Further Development

The necessity for recommendations arises from the potential of ML models to enhance core-making processes and predict product quality. Integrating these models can revolutionize manufacturing by offering proactive insights.

Recommendations for the company:

- Gradual Integration: Start by employing the ML model alongside manual inspection to accumulate real-world data and foster trust in the model's predictions.
- Alignment Verification: Ensure model predictions align well with actual outcomes before reducing reliance on manual inspection.
- Efficient Data Preprocessing: Streamline data preprocessing by ensuring easy extraction of child product units from product data.
- Balancing Techniques: Implement better data balancing methods to improve model performance.
- Explore Alternative Models: Experiment with various ML models beyond decision trees to identify optimal predictive performance.
- Comprehensive Data Utilization: Include all available machine and process data in the core-making process to enhance prediction accuracy.

5.2 Contributions

The paper serves as a pivotal analysis of critical process parameters impacting product quality defects within a foundry line, employing the capabilities of ML. The notable contributions of this work encompass several dimensions: Firstly, it underscores the paramount significance of quality control within manufacturing, shedding light on how conventional quality control processes can be enriched and refined through the application of AI and ML methodologies. Secondly, the paper presents a compelling demonstration of how ML models can be adeptly harnessed to unravel intricate correlations between diverse process parameters within a foundry line. This holistic approach enables the identification of root causes underlying distinct casting defects in the core-making process, thereby enabling targeted interventions for quality enhancement. Furthermore, the paper introduces a systematic and structured approach to problem-solving by integrating the CRISP-DM framework, a comprehensive tool that orchestrates data mining and ML endeavors, thereby fostering data-driven decision-making in a coherent and effective manner. Lastly, the empirical results manifest in an accuracy rate of 66.2% for Machine A and 56% for Machine B, illustrating the practical efficacy of the ML model in discerning defective cores. This achievement holds considerable promise as it equips human operators with an indispensable tool to preemptively avert the generation of flawed castings, thus substantiating the overall value and impact of this research.

6
Conclusion

In light of the comprehensive findings presented within this thesis, a definitive inference emerges that underscores the remarkable efficacy of ML models in scrutinizing critical process parameters, culminating in the identification of underlying causes for distinct casting defects within a foundry line.

The study demonstrates the profound potential of ML models as indispensable tools for the intricate analysis of manufacturing intricacies. Through the lens of the CRISP-DM framework, a structured and methodical approach for data mining and ML projects is unveiled, subsequently fortifying the capacity for data-driven decision-making within the manufacturing domain. This, in turn, imparts a palpable enhancement to the efficiency of production systems. It is noteworthy, however, that the accuracy of the employed ML models could be further refined, thus engendering opportunities for heightened predictive precision. Furthermore, the translation of these models into real-world manufacturing settings might encounter multifaceted challenges, necessitating a nuanced approach to implementation.

In sum, this study delivers a substantial contribution to the ongoing landscape of quality control in manufacturing, serving as a potent testament to the transformative capabilities of AI and ML techniques. These techniques exhibit a profound potential to ameliorate resource utilization, streamline operations, and mitigate costs within the intricate fabric of the production process, ultimately yielding a positive trajectory for the evolution of manufacturing practices.

Bibliography

[1] F. Han, Y. Wang, and Y. Cheng, "Quality control in the manufacturing process," *IEEE Industry Applications Magazine*, vol. 26, no. 5, pp. 22–30, 2020.

[2] D. K. Kim, H. C. Lee, S. S. Kang, M. C. Kim, and K. B. Park, "Quality control of automotive parts using control charts and failure mode and effects analysis," *Journal of Mechanical Science and Technology*, vol. 34, no. 2, pp. 853–862, 2020.

[3] N. Singh and S. Kumar, "An integrated approach to quality control using six sigma and taguchi methods: A case study in a manufacturing company," *Journal of Industrial Engineering International*, vol. 17, no. 3, pp. 573–586, 2021.

[4] S. A. A. Razak, D. Kurniawan, and R. A. Sari, "Industry 4.0 technologies for quality control and inspection in the automotive manufacturing industry: A review," *Journal of Manufacturing Systems*, vol. 60, pp. 146–162, 2021.

[5] N. Jain, P. Jain, and A. Sharma, "Industry 4.0: A review on industrial automation and digitalization," *International Journal of Engineering and Advanced Technology*, vol. 10, pp. 273–279, 2021.

[6] Y. Liu, H. Jiang, L. Yang, and Y. Liu, "Real-time defect prediction for manufacturing processes based on multisensor data fusion and machine learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3286–3295, 2021.

[7] P. Jain, *Fundamentals of foundry technology.* Laxmi Publications, 2013.

[8] P. Beeley, *Foundry Technology.* Butterworth-Heinemann, 2001.

[9] R. Srinivasan, *Foundry engineering.* Khanna Press, 2011.

[10] A. Mitra, *Fundamentals of Quality Control and Improvement.* Hoboken, New Jersey: John Wiley & Sons, Inc., 2019.

[11] S. Teli, V. Majali, U. Bhushi, L. Gaikwad, and V. Surange, "Cost of poor quality analysis for automobile industry: A case study," *J. Inst. Eng. India Ser. C*, vol. 94, no. 4, pp. 373–384, October–December 2013.

[12] E. Thomas, A. Zeron, and J. S. O'Rourke, *Samsung Electronics Company, Ltd.: Galaxy Note 7 Crisis.* The Eugene D. Fanning Center for Business Communication, Mendoza College of . . . , 2017.

[13] S. Chaudhari and H. Thakkar, "Review on analysis of foundry defects for quality improvement of sand casting," *International Journal of Engineering Research and Applications*, vol. 4, no. 3, pp. 615–618, 2014.

[14] A. Juriani, "Casting defects analysis in foundry and their remedial measures with industrial case studies," *IOSR Journal of Mechanical and Civil Engineering*, vol. 12, no. 6, pp. 43–54, 2015.

[15] K. Ehmann, S. Kapoor, R. DeVor, and I. Lazoglu, "Machining process modeling: A review," *Journal of Manufacturing Science and Engineering*, vol. 119, no. 4B, pp. 655–663, Nov. 1997.

[16] F. Tao, Q. Qi, A. Liu, and A. Kusiak, "Data-driven smart manufacturing," *Journal of Manufacturing Systems*, vol. 48, pp. 157–169, 2018, special Issue on Smart Manufacturing. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278612518300062

[17] A. Kusiak, "Smart manufacturing must embrace big data," *Nature*, vol. 544, no. 7648, pp. 23–25, Apr 2017.

[18] I. Hashem, I. Yaqoob, N. Anuar, S. Mokhtar, A. Gani, and S. Khan, "The rise of "big data" on cloud computing: Review and open research issues," *Information Systems*, vol. 47, pp. 98–115, 07 2014.

[19] P. Wang, Y. Yang, and N. S. Moghaddam, "Process modeling in laser powder bed fusion towards defect detection and quality control via machine learning: The state-of-the-art and research challenges," *Journal of Manufacturing Processes*, vol. 73, pp. 961–984, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1526612521008446

[20] G. Krummenacher, C. S. Ong, S. Koller, S. Kobayashi, and J. M. Buhmann, "Wheel defect detection with machine learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1176–1187, 2018.

[21] BramahHazela, J. Hymavathi, T. R. Kumar, S. Kavitha, D. Deepa, S. Lalar, and P. Karunakaran, "Machine learning: Supervised algorithms to determine the defect in high-precision foundry operation," *Journal of Nanomaterials*, vol. 2022, p. 1732441, 2022.

[22] D. Blondheim, "Improving manufacturing applications of machine learning by understanding defect classification and the critical error threshold," *International Journal of Metalcasting*, vol. 16, no. 2, pp. 502–520, 2022. [Online]. Available: https://doi.org/10.1007/s40962-021-00637-0

[23] M. W. Berry and M. Verleysen, *Supervised and Unsupervised Learning for Data Science: Unsupervised and Semi-Supervised Learning.* Springer, 2020.

[24] S. Kaur and S. Jindal, "A survey on machine learning algorithms," *IJIRAE International Journal of Innovative Research in Advanced Engineering*, vol. 3, 2016. [Online]. Available: https://www.ijirae.com/volumes/Vol3/iss11/02.NVAE10084.pdf

[25] H. H. Patel and P. Prajapati, "Study and analysis of decision tree based classification algorithms," *International Journal of Computer Sciences and Engineering*, vol. 6, pp. 74–78, 2018. [Online]. Available: https://www.ijcseonline.org/full_paper_view.php?paper_id=2984

[26] Tavish. (2019) 11 important model evaluation metrics for machine learning everyone should know. [Online]. Available: https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/

[27] A. Bajaj. (2022) Performance metrics in machine learning [complete guide]. [Online]. Available: https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide/

[28] K. P. Shung. (2018) Accuracy, precision, recall or f1? [Online]. Available: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

[29] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying crisp-dm process model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021, cENTERIS 2020 - International Conference on ENTERprise Information Systems / ProjMAN 2020 - International Conference on Project MANagement / HCist 2020 - International Conference on Health and Social Care Information Systems and Technologies 2020, CENTERIS/ProjMAN/HCist 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050921002416

[30] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "Dmme: Data mining methodology for engineering applications – a holistic extension to the crisp-dm model," *Procedia CIRP*, vol. 79, pp. 403–408, 2019, 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 18-20 July 2018, Gulf of Naples, Italy. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212827119302239

[31] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," *DaimlerChrysler Research & Technology FT3/KL PO BOX 2360 89013 Ulm, Germany*, 2000.

A

Appendix 1

The pseudo code provided in this Appendix represents the Data prepossessing for Machine A.

Input: - Data source
Output: - Preprocessed data
Step 1: Import Required Libraries - Import a suitable data manipulation library
Step 2: Load Data - Load data from a specific source
Step 3: Prepare Deviation Data - Read and preprocess deviation data - Create mappings for values
Step 4: Replace Deviation Values - Replace original deviation values with new mapped values
Step 5: Prepare Problem Owner Data - Create mappings for problem owner values
Step 6: Replace Problem Owner Values - Replace original problem owner values with new mapped values
Step 7: Prepare View Data - Create mappings for view values
Step 8: Replace View Values - Replace original view values with new mapped values
Step 9: Rename Columns - Rename specific columns as required
Step 10: Merge DataFrames - Perform an outer merge of dataframes
Step 11: Select Columns - Choose relevant columns for analysis
Step 12: Handle Missing Values - Define columns with missing values - Replace missing values with a common value
Step 13: Rename Columns - Rename specific columns as required
Step 14: Filter Data - Create filtered DataFrame by selecting specific rows
Step 15: Filter Data - Create filtered DataFrame by selecting specific rows
Step 16: Merge DataFrames - Perform a left merge of dataframes
Step 17: Rename Columns - Rename specific columns as required
Step 18: Rename Columns - Rename specific columns as required
Step 19: Filter Data - Remove rows with missing values
Step 20: Load Data - Load data from a CSV file
Step 21: Convert Data Types - Convert specific columns to string type
Step 22: Replace Symbols and Convert - Define columns to convert - Replace symbols and convert columns to float
Step 23: Round Data - Define columns to round - Apply rounding function to specified columns
Step 24: Rename Columns - Rename specific columns as required
Step 25: Convert Date and Time - Convert 'StartTime' column to datetime format
Step 26: Create Date Column - Create a new column with formatted date

Step 27: Split and Extract - Split and extract information from column
Step 28: Merge DataFrames - Perform an inner merge of dataframes
Step 29: Drop Duplicates - Remove duplicate rows from DataFrame
Step 30: Load Data - Load data from an Excel file
Step 31: Rename Columns - Rename specific columns as required
Step 32: Prepare Maintenance Data - Create the Maintenance DataFrame by selecting specific columns - Filter rows based on specific conditions
Step 33: Filter Maintenance Data - Filter maintenance data based on specific criteria
Step 34: Filter Maintenance Data - Filter maintenance data based on production unit and station object
Step 35: Convert Date and Time - Convert 'REG-DATE' column to datetime format
Step 36: Create Date Column - Create a new column with formatted date
Step 37: Add Column - Add a new column with constant value
Step 38: Sort Data - Sort the DataFrame based on date
Step 39: Extract Work Area - Create 'WorkArea' column based on a specific column
Step 40: Modify Work Area - Modify 'WorkArea' values as required
Step 41: Filter Data - Filter rows based on 'WorkArea'
Step 42: Drop Columns - Remove specified columns from DataFrame
Step 43: Drop Duplicates - Remove duplicate rows from DataFrame
Step 44: Create a new DataFrame - Create a new DataFrame by selecting specific columns
Step 45: Drop Duplicates - Remove duplicate rows from DataFrame
Step 46: Merge DataFrames - Perform a left merge of dataframes
Step 47: Fill Missing Values - Fill missing values in DataFrame
Step 48: Drop Duplicates - Remove duplicate rows from DataFrame
Step 49: Create Pivot Table - Create a pivot table using the DataFrame
Step 50: Drop Columns - Remove specified columns from pivot table
Step 51: Fill Missing Values - Fill missing values in pivot table
Step 52: Create a new DataFrame - Create a new DataFrame by selecting specific columns
Step 53: Drop Duplicates - Remove duplicate rows from DataFrame
Step 54: Merge DataFrames - Perform a left merge of dataframes
Step 55: Fill Missing Values - Fill missing values in DataFrame
Step 56: Drop Duplicates - Remove duplicate rows from DataFrame
Step 57: Create a new DataFrame - Create a new DataFrame by selecting specific columns
Step 58: Merge DataFrames - Perform an inner merge of dataframes
Step 59: Save Data - Save the DataFrame to a CSV file

B

Appendix 2

The pseudo code provided in this Appendix represents the Random Forrest ML applied to Machine A Blow hole deviations.

Step 1: Import Libraries Import data manipulation library (e.g., pandas)
Step 2: Load Data Load data from CSV file
Step 3: Prepare Data - Drop columns ('A', 'B') - Filter rows by condition
Step 4: Undersample Data - Separate majority and minority classes - Undersample majority class - Combine minority and undersampled majority - Shuffle dataset
Step 5: Split Data - Separate features (X) and target (y) - Split data (80% train, 20% test)
Step 6: Scale Data - Initialize StandardScaler - Fit scaler to training data - Transform training and testing data
Step 7: Initialize Model Import model (e.g., RandomForest)
Step 8: Hyperparameter Tuning - Import GridSearchCV - Define parameter grid - Create model instance - Initialize GridSearchCV - Fit GridSearchCV to scaled data
Step 9: Access Best Model Params Access best model params from grid search
Step 10: Model Evaluation - Predict using best model - Calculate accuracy - Generate classification report - Plot confusion matrix
Step 11: Feature Importance Access feature importances

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

AVANCEZ
1829
CHALMERS
UNIVERSITY OF TECHNOLOGY