Developing a predictive maintenance model for addressing underfill in Heineken Brewery's Share Company

By: Ruth Wondosen

A Master's Proposal/Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for Degree of Master of Science in Mechanical and Industrial Engineering (Industrial Engineering Stream)

Advisor: Dr. Ameha Mulugeta (PhD)
Co-Advisor: Natnael Mekonnen

Addis Ababa University
Addis Ababa Institute of Technology (AAiT)
School of Mechanical and Industrial Engineering (SMIE)
Industrial Engineering Chair

Oct 12, 2024, G.C

Addis Ababa University
Addis Ababa Institute of Technology
School of Mechanical and Industrial Engineering
Approval of M.Sc. Thesis

Submitted By:

Ruth Wondosen Student's name	---------------------- Date	---------------------- Signature

Recommended By:

Dr. Ameha Mulugeta Advisor	---------------------- Date	---------------------- Signature

Approved By:

Dr. Gezahegn Tesfaye Chair, Industrial Engineering	---------------------- Date	---------------------- Signature

Endorsed By:

Araya Abera (Ph.D.) Dean, School of Mechanical And Industrial Engineering	---------------------- Date	---------------------- Signature
Sosina Mengistu (Ph.D.) Associate Director of Postgraduate Program	---------------------- Date	---------------------- Signature

Declaration

I hereby declare that the work being presented in this thesis entitled "Developing a predictive maintenance model for addressing underfill in Heineken company" is original work of my own, has not been presented for a degree of any other university and all the resources of materials used for this thesis have been duly acknowledged.

____________________	____________________
Ruth Wondosen	Date

This is to certify that the above declaration made by the candidate is correct to the best of my Knowledge.

__________________________	__________________________
Dr. Ameha Mulugeta (Advisor)	Date

Acknowledgment

To begin with, I want to thank God Almighty for providing me with the courage, tenacity, and knowledge to complete this task. I've seen your hands. I owe a tremendous amount of thanks to my mom and my entire family. Your constant love, support, and encouragement have served as the basis for my path. Thank you for consistently believing in me. Thank you, friends I had and to the friends I made along the way, for your unwavering support and friendship. Your presence has made the voyage not only manageable but also pleasurable.

Special thanks to my Advisor, Dr. Ameha Mulugeta, and Co-Advisor, Natneal Mekonnen. Your invaluable advice, intelligent input, and ongoing encouragement have helped shape this research. I genuinely value your mentorship.

I would also like to express my heartfelt gratitude to Tsion Girma and her sister Mahider Tulu for their tremendous support at the beginning of this journey. Your kindness and generosity have had a huge impact on my journey. Finally, I want to acknowledge my classmates. Your camaraderie and collaborative spirit have significantly improved my academic experience. Thank you all for taking part in our journey.

Abstract

Predictive maintenance (PdM) is crucial for enhancing operational efficiency and reducing downtime in industrial processes. This study introduces a novel approach to developing a predictive maintenance model for addressing underfilling issues in the bottling processes of Heineken Company, which significantly impact revenue and product quality. Unlike previous works that rely solely on historical data, this research incorporates synthetic data generated using Conditional Generative Adversarial Networks (cGAN) to overcome data limitations and enhance model robustness. The methodology involved comprehensive data preprocessing, including imputation and feature engineering, to prepare the dataset for training a Random Forest classifier. The model development was refined through hyperparameter tuning via Grid Search and validated using cross-validation.

The results demonstrated a strong predictive capability, with a training accuracy of 90.67%, test accuracy of 90.21%, and cross-validation accuracy of 90.89%, indicating reliable generalization. The use of cGAN contributed to increased data variability, mitigating overfitting and ensuring realistic model training scenarios. This study advances the field of predictive maintenance by demonstrating how synthetic data can augment limited datasets to improve model accuracy and resilience. Integrating this model into the production line enables proactive maintenance scheduling, reducing disruptions and enhancing product consistency. Future work will focus on expanding the model's scope to incorporate real-time sensor integration for adaptive learning and exploring ensemble models, such as hybrid Random Forest and LSTM architectures, to handle temporal patterns and further optimize predictive performance.

KEYWORDS: Predictive Maintenance (PdM), Machine Learning, Underfill, Random Forest, conditional Generative Adversarial Network(cGAN)

Table of Contents

List of Figure

List of Tables

List of Abbreviations

AI	Artificial Intelligence
cGAN	Conditional Generative Adversarial Network
CM	Corrective maintenance
DT	Digital Twin
GAN	Generative Adversarial Network
IoT	Internet of things
ML	Machine learning
PM	Preventive maintenance
PdM	Predictive maintenance
RUL	Remaining useful life
SCADA	Supervisory Control and Data Acquisition
USD	United States Dollar

Chapter One

1. Introduction and Problem Justification

1.1 Introduction

Maintenance is a restoration to its ideal state for its intended use, a thorough process of technical, administrative, and managerial steps must be taken(Tran et al., 2021). System maintenance is an essential task because of the potential consequences of undetected faults, both financially and in terms of the company's reputation. Three main approaches are used in system maintenance; Corrective maintenance (CM), Predictive maintenance (PdM), and Preventive maintenance (PM) (Esteban et al., 2022). Corrective maintenance is a type of maintenance that takes place when the equipment is not working or operating (Costa & Balduíno, 2018). The purpose of corrective maintenance is to find and fix the root causes of a system's faults (Y. Wang et al., 2014). Preventive maintenance on the other hand is aimed at maintaining the physical integrity and functionality of a property. It involves putting a maintenance plan into action to reduce unexpected breakdowns and increase system reliability (Mrugalska et al., 2018).

Predictive maintenance is a maintenance strategy that is currently attracting attention, it is in the words of (Mobley, 2002) is "a system of monitoring critical industrial equipment by such techniques as noise analysis and thermal efficiency to predict machine failure as the basis for scheduling maintenance". In industrial processes, predictive maintenance a data-driven strategy based on machine learning and sensor technology that is becoming more and more crucial (Merkt, 2019). Predictive maintenance in the beverage industry used to detect and prevent underfill incidents.(Morelle et al., 2021). It can be done by using historical data, sensor readings, and advanced analytics, predictive maintenance models can forecast equipment failures and minimize unplanned downtime.

predictive maintenance has come a long way from utilizing visual inspection methods to automated systems using sophisticated signal processing techniques based on machine learning, pattern recognition, fuzzy logic, neural networks, etc., When human eyes or ears are no longer able to detect and gather sensitive information from equipment, primarily motors, automated technologies offer a workable option for many industries (Kamat & Dr.Sugandhi, 2020).

Many changes are currently occurring in the manufacturing sector, such as quicker innovation and development times, customized products, more flexibility in production and product development, fewer hierarchies, and improved resource efficiency.(Lasi et al., 2014). The beverage industry is no different. Its ever-changing landscape has bought change to traditional maintenance strategies to predictive maintenance. It has effectively implemented predictive maintenance, mostly using vibration analysis, to increase the dependability of vital equipment such as palletizers, empty bottle inspectors, and bottle packing(Ben et al., 2021a). This paper studies the development of a predictive maintenance model in the beverage industry.

1.2 Background and Justification

The wind and alcohol industry have experienced tremendous growth in the global beverage market, with alcohol sales surpassing $1.5 trillion in 2017(Jernigan & Ross, 2020). The main driver of this expansion is organizations that are being forced to adopt new competitive strategies because of the world's ongoing increase in competition. One of the strategies in the manufacturing industry is by minimizing the production loss due to machine breakdown (Ben et al., 2021a). The beverage industry being part of the manufacturing industry it expected the worldwide beverage industry to increase from 3.56 trillion USD in 2023 to 4.39 trillion USD by 2028 (Swennen, 2017).

At the heart of this intricate dance lies Heineken, a brewing behemoth renowned for its quality and innovation. Heineken's journey from humble Dutch beginnings to its current global reach is a testament to its commitment to excellence. However, navigating the fast-paced and competitive beverage landscape demands not just tradition, but also adaptation and optimization. Heineken has a robust infrastructure and stringent quality control, like any production powerhouse, is not immune to the lurking threat of equipment failures. These sudden breakdowns can have crippling consequences (Török et al., 2020).

On a production line, every minute of downtime means missed deadlines and lost profits(Costa & Balduíno, 2018). Temperature fluctuations, contamination, or improper mixing can happen from equipment failure or from the spoiling of costly components and finished products, making entire batches worthless. Inadequate equipment might put workers in danger and potentially result in accidents. Breaking news about production issues or safety hazards can damage a brand's reputation and quickly undermine consumer trust(Jernigan & Ross, 2020).

Underfilling is a common problem in the beverage industry, which occurs when a container is filled with less liquid than the specified volume(Nashrudin et al., 2020). Underfilling can have negative consequences for both the producer and the consumer, such as loss of revenue, damage to brand reputation, waste of resources, and reduced product quality. Therefore, it is essential to ensure that the filling process is correct and reliable and that any deviations or errors are detected and corrected promptly (Kumar & Shankar, 2011). Traditional reactive maintenance is no longer enough in this high-stakes game where a single misstep can be costly. The industry is shifting towards a proactive approach known as predictive maintenance (PdM).

PdM leverages the power of data analytics and sensor technology to predict equipment failures before they occur (Merkt, 2019). By analyzing historical data on sensor readings, maintenance logs, and production records, PdM algorithms can identify subtle patterns and anomalies that signal impending issues(Achouch, Dimitrova, Ziane, Karganroudi, et al., 2022). This early warning allows for timely interventions, preventing catastrophic breakdowns and their associated consequences(Tran et al., 2021).

Numerous academic investigations have investigated the utilization of predictive maintenance for diverse equipment and procedures in the beverage sector. For instance, (Zhang et al., 2019) developed a machine-learning model based on sensor data to predict pump failures in juice bottling lines. Predictive maintenance has recently been applied in different industries and has made a positive impact by increasing productivity. According to (Tran et al., 2021), Predictive maintenance tools can reduce a downtime by 30% to 50% and increase the remaining useful life (RUL) of the machine by 20% to 40%.

Predictive maintenance approaches provide more benefits than drawbacks. According to study done by (Arena et al., 2021) on industrial average savings, businesses who adopted predictive maintenance programs eliminated 70–75% of asset breakdown, cut maintenance costs by 25–30%, and boosted productivity by 20–25%. It was a suitable investment because the average return on investment (ROI) was ten times. For Heineken, implementing a robust PdM strategy is not just about minimizing maintenance costs or a return on investment, and reducing downtime it's about embracing a sustainable future. PdM aligns perfectly with Heineken's commitment to environmental responsibility, Product quality and consistency, Employee safety and well-being, and Resource optimization.

1.3 Problem Statement

The beverage industry relies heavily on accurate and reliable filling processes to ensure the quality and consistency of its products (Kulkarni & Elango, 2016). However, the filling process is subject to various sources of uncertainty and variability, leading to underfilling events where containers are filled with less liquid than the specified volume (Adarsh & Kulkarni, 2021). Underfilling is a significant issue that negatively impacts both producers and consumers by causing revenue loss, damaging brand reputation, and wasting resources (Adarsh & Kulkarni, 2021).

Heineken, a leading global beer producer, employs advanced high-speed and high-precision filling equipment in its packaging line. Despite this, the company faces persistent underfilling issues. Company data indicates that the B1200 line experiences the highest packaging loss, with a monthly underfill rate of 1.80%. The B1200 line produces 42,000 bottles per hour, and typically, 150 to 250 bottles are underfilled per 8-hour shift, translating to a fill volume below 325 ml for a 330 ml bottle. This resulted in significant losses, with approximately 200 bottles underfilled in 8 hours, escalating to 4,200 bottles weekly, 168,000 monthly, and over 2 million annually.

Figure 1 Heineken Extract loss deployment.

There is a pressing need to address the underfill loss to improve Heineken's operational efficiency and product quality.

1.4 Research Questions

- What are the key sensor data signals that are strongly correlated with underfilling incidents?
- How do to develop machine learning algorithms capable of accurately predicting underfilling occurrences?
- How to evaluate the performance of the machine learning model? And how do we integrate the predictive model into the existing maintenance strategy?

1.5 Objective of the Study

1.5.1 General Objective

This research aims to develop a predictive maintenance model that solves the problem of underfilling bottles.

1.5.2 Specific Objectives

- To collect and analyze the relevant sensor data signals associated with underfilling incidents in Heineken company.
- To develop machine learning algorithms capable of accurately predicting underfilling occurrences.
- To evaluate the performance of the model and the predictive maintenance strategy in improving Heineken Company’s underfilling of bottles.

1.6 Significance of the Study

The purpose of this paper is to develop a predictive maintenance model for addressing underfilling incidents in the bottling operations of the Heineken company. The significance of this research lies in its value, contribution to various stakeholders, practical implications, and theoretical advancements.

The Heineken company benefits directly from the research. Through optimizing the data collecting procedure and developing machine learning algorithms for underfilling prediction, the study aims to increase overall productivity, reduce product waste, decrease downtime, and improve operational effectiveness. This is important from a company standpoint because it can result in lower expenses, more reliable product quality, and happier and more devoted customers.

The study also adds to the body of knowledge in predictive maintenance, which advances theory. The study expands on the understanding of how predictive maintenance models can accurately forecast underfilling events by analyzing useful sensor data signals and developing machine learning algorithms. Researchers and academics can benefit from this theoretical contribution, which lays the groundwork for additional investigation and development of predictive maintenance strategies in the manufacturing industry.

1.7 Scope and Limitation of the Study

1.7.1 Scope of the Study

The scope of this study is focused on developing a predictive maintenance model for addressing underfilling incidents specifically within the bottling processes of Heineken Company. The study includes collecting and analyzing relevant sensor data signals linked to underfilling incidents, creating machine learning algorithms for accurate underfilling prediction, and assessing the effectiveness of the model and its potential to reduce underfilling incidents at Heineken Company. Underfilling incidents in other businesses or sectors, unrelated facets of Heineken's operations, and a thorough analysis of alternative maintenance strategies outside of the narrow focus of underfilling incidents are not going to be studied in this study.

1.7.2 Limitations of the Study

There are a few limitations that are expected when conducting this research within the specified scope of creating a predictive maintenance model for underfilling incidents in the bottling processes of Heineken Company. A few of these limitations are the data's quality and availability, underfilling incidents' inherent variability, the findings' limited generalizability to other contexts, the complexity of machine learning algorithms, and the findings' limited influence on alternative maintenance strategies. These limitations should be considered when interpreting the study's findings, as they may have an impact on its findings. Through recognition and resolution of these constraints, the study can yield significant knowledge and aid in the creation of efficient predictive maintenance plans for underfilling incidents within Heineken Company.

Chapter Two

2. Literature Review

2.1 Introduction

The incorporation of technology such as artificial intelligence (AI) and the Internet of Things (IoT) into daily life is what defines the fourth industrial revolution, which is a significant technological shift in 21st-century society (Ross & Maynard, 2021). This revolution presents both opportunities and challenges by disrupting traditional work concepts and driving the development of intelligent and connected infrastructure, which is integrated with analytical and cognitive technologies that enable machine-to-machine(M2M) and machine-to-human (M2H) communication (Compare et al., 2020). New technologies from Industry 4.0 integrate people, machines and products, enabling faster and more targeted exchange of information (Rauch et al., 2020)The large quantity of data collected industrial systems, contains information about processes, events and alarms that occur along an industrial production line. When this data analyzed and processed the data reveals an important knowledge in the manufacturing process and system dynamics. With this it is possible to find interpretive results for strategic decision making, by providing advantages such as, maintenance cost reduction, machine fault detection, repair stop reduction, spare part inventory reduction, spare part life increasing, increased production, improvement in operator safety, repair verification, overall profit, among others (Ross & Maynard, 2021b; Voronov et al., 2021; M. Wang et al., 2023). The mentioned advantages have a strong relationship with maintenance procedures. In industries, equipment maintenance is an important key and affects the operation time of equipment and its efficiency. Therefore, equipment faults need to be identified and solved, avoiding shutdown in the production processes *(*Wan et al., 2017). The industry 4.0 novel approach leads a company to face challenges in creating a much more dynamic environment (Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges Jovani Dalzochioa,)

Among other opportunities Industry 4.0 presents, predictive maintenance plays a fundamental role using condition monitoring data to detect anomalies in production process, manufacturing equipment’s, and products (compare.,2019).

Maintenance costs account for 15% to 60% of total manufacturing costs (Mulders & Haarman, 2018). Maintenance strategies can be broadly categorized into four types: corrective, preventive, predictive, and prescriptive maintenance. Corrective maintenance is performed in response to evident degradation or faults. Studies have shown that 89% of failures in the aviation sector are not related to the age of the equipment, indicating that traditional preventive maintenance is often inadequate for addressing irregular failures (Nowlan & Heap, 1978). In contrast, predictive maintenance offers a proactive approach by enabling early detection of potential issues, thereby preventing malfunctions before they occur (Jimenez et al., 2020).

Predictive maintenance (PdM) gathers, cleanses, evaluates, and uses data from various manufacturing and sensors. To predict or foresee equipment failure before it occurs, it applies advanced algorithms to the data, automatically comparing the data it receives with information from prior cases. This helps to optimize equipment utilization and maintenance strategies, enhance performance and productivity, and prolong the life of the equipment(Tran et al., 2021).

2.2 Predictive Maintenance Approach

In literature, different terminology and groups of maintenance management strategies can be found. This paper considers the categories proposed by the works (Susto et al., 2012; Susto et al., 2015). They classify the maintenance procedures as follows:

Run-to-Failure (R2F) or Corrective maintenance happens only when an equipment stops working. This is the simplest maintenance strategy, since it is necessary both the stop on the production and the repair of the parts to be replaced, adding a direct cost to the process.

Preventive Maintenance (PvM), Time-based maintenance or Scheduled maintenance is a maintenance technique performed periodically with a planned schedule in time or process iterations to anticipate process/equipment failures. It is generally an effective approach to avoid failures. However, unnecessary corrective actions are taken, leading to an increase in the operating costs.

Predictive Maintenance (PdM) uses predictive tools to determinewhen maintenance actions are necessary. It is based on continuous monitoring of a machine or a process integrity, allowing maintenance to be performed only when it is needed. Moreover, it allows the early detection of failures thanks to predictive tools based on historical data (e.g. machine learning techniques), integrity factors (e.g. visual aspects, wear, coloration different from original, among others), statistical inference methods and engineering approaches.

Each of the maintenance classes has its role. But, by opting for R2F, industries delay maintenance actions and assume the risk of unavailability of their assets; on the other hand, PvM anticipates maintenance interventions, resulting in a spare part exchange with half-life. Thus, a good maintenance strategy should improve the equipment condition, reduce the equipment failure rates and minimize maintenance costs, while maximize the life of equipment. Due to this fact, the PdM strategy is the one that stands out most among the other strategies (Jezzini et al., 2013), and it is attracting attention in the era of Industry 4.0 due to its ability of optimizing the use and management of assets (Kumar et al., 2019). Its advantages include: maximizing time of use and operation of equipment, delaying/reducing maintenance activities, and reducing material and labor costs.

Predictive maintenance has become a promising maintenance approach as it increases the machine life span, minimizes unplanned downtime, and minimizes energy consumption (Ran et al., 2019). And as a result, of the growing complexity of the linkages between various production activities in progressively larger manufacturing ecosystems, it has assumed crucial importance for industries. Implementing PdM in the industry sector, benefits by reducing equipment failure, and maintenance costs, and increasing production(Arena et al., 2021)

(Meyer Zu Wickern, 2019), found that the reliability and success of PdM methods depend on the manufacturing company and purpose and that various factors impact their applicability. Typically, the predictive maintenance strategy has four phases: 1. Gathering sensor data, 2. Data preprocessing, 3. Faults diagnostics and prognosis, and 4. Decision-making on the maintenance strategy (Arena et al., 2021).

One widely accepted concept in predictive maintenance is fault detection. Early failure detection is crucial in preventing catastrophic machine failures, as highlighted by Amruthnath & Gupta, (2018). They categorize fault detection methods into process Knowledge-based methods, quantitative model-based methods, and qualitative model-based methods.

In the era of Industry 4.0, four primary predictive maintenance approaches are prevalent: the knowledge-based approach, the data-driven approach, the physical modeling approach, and the digital twin approach (Sajid et al., 2021). The knowledge-based approach leverages rules derived from prior experience or real-world scenarios, typically expressed in an "if-then" format. However, as Sajid et al. (2021)

Conversely, the digital twin approach, as described by Achouch et al. (2022), creates a virtual replica of a physical asset or product. This virtual copy is beneficial across numerous industries, enabling a seamless connection between digital and physical domains by combining models and data (Ferraz et al., 2022).

The choice of predictive maintenance approach depends on the specific objectives it aims to achieve. The knowledge-based approach relies on expert knowledge and deductive reasoning skills, making it suitable for traditional PdM methods such as expert systems and model-based reasoning. The digital twin approach, on the other hand, is ideal for integrating digital and physical domains. In environments where large quantities of data are available, such as the industrial sector, the data-driven approach is practical (Paolanti et al., 2018). This approach encompasses various machine learning techniques, including supervised, semi-supervised, unsupervised, and reinforcement learning.

2.2.1 **Machine Learning Approach**

Machine learning (ML) is a powerful technology capable of learning with minimal additional support (Amruthnath & Gupta, 2018). It has emerged as an effective tool for extracting meaningful insights and making informed decisions from large datasets. As a branch of artificial intelligence, machine learning focuses on developing models and algorithms for data analysis and prediction.

Machine learning (ML) enhances predictive maintenance (PdM) by analyzing large volumes of sensor data to predict equipment failures and optimize maintenance schedules. Techniques such as supervised learning, unsupervised learning, and reinforcement learning enable the creation of models that can learn from historical data, detect patterns, and make accurate predictions about future equipment performance (Singh, 2019). For instance, supervised learning algorithms, including Random Forest and support vector machines, are trained on labeled data to predict specific outcomes, such as equipment failures Gęca (2020). This integration allows for early detection of issues, minimizing downtime and maintenance costs. Different algorithms excel in different contexts, and choosing the right one can significantly enhance the accuracy and reliability of the model's predictions (Almamlook et al., 2022 and Ren, 2021)

Several algorithms have been proposed for predictive maintenance tasks, focusing on managing complex data and identifying temporal correlations. Nacchia et al. (2021) provide a comprehensive analysis of machine learning methods used in predictive maintenance, particularly supervised learning and the growing adoption of ensemble methods like Random Forest and Long Short-Term Memory (LSTM). These studies suggest that a combination of supervised and ensemble approaches, such as ARIMA, LSTM, decision trees, and Random Forest, is ideal for predictive maintenance tasks.

For example, ARIMA is a popular choice for time series forecasting due to its ability to capture trends, seasonality, and autocorrelation in data (Shivhare et al., 2021). LSTM, on the other hand, excels in processing and interpreting sensor data, making it highly effective for accurate Remaining Useful Life (RUL) prediction in IoT-based industrial systems. LSTM is particularly powerful in learning long-term dependencies in the data. Random Forest is a versatile algorithm suitable for handling both categorical and numerical data, commonly used in classification tasks (He et al., 2018).

Ultimately, the choice of algorithm depends on the specific characteristics of the data and the problem at hand. It is essential to carefully evaluate the strengths and weaknesses of each algorithm to select the one best suited for the task. This careful selection ensures that the predictive maintenance model will provide accurate and reliable predictions, thereby enhancing operational efficiency and reducing maintenance costs.

Table 1: literature summary on the strengths and weaknesses of some of the common types of ML algorithms

References	Model	Data type used	Strength of the algorithms	Weakness of the algorithms	Equipment
(Chazhoor et al., 2020)	Logistic regression with false positive rate	Real data	commendable accuracy with PCA and the ability to handle class imbalance problems.	assumes a linear relationship between the independent variables and the log odds of the dependent variable, is not suitable for complex relationships between the dependent and independent variables and cannot handle missing values effectively.	Semi-conductors manufacturing company
(Kizito et al., 2018.)	Random Forest	Real data	its ensemble learning approach, feature importance measurement, parallel data fitting, and precision in model development.	The article does not explicitly mention any disadvantages of random forest machine learning techniques.	Bearings
(Kaparthi & Bumblauskas, 2020)	Decision Tree	Not specified (232,662 records of sensor readings)	Scalability, Generalizability, Applicability to sustainable resource allocation by increasing uptime and utilization for expensive equipment, and Improved accuracy in predicting failure events	The article does not explicitly mention any disadvantages of decision tree-based machine learning techniques.	Study conducted with a wide range of equipment in mind

(Gawde et al., 2024)	multi-class classification algorithms such as Support Vector Machine (SVM), k-nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF)	Real data	Versatility, Interpretability, Performance Comparison, Established Techniques, Scalability, and Statistical Analytics.	The influence of the chosen feature representation on interpretation, the need to interpret the results from various techniques cautiously and in association with domain knowledge, and the inability of these techniques to guide the next steps to rectify the predicted output.	industrial rotating machines
(Le et al., 2014)	linear regression (Least square)	Not specified	Simplicity, Insightful description, Dimension reduction, Reduced variance, tradeoff between bias and variance, and Applicability to High Dimensional Data	the least squares estimator has the smallest variance among unbiased linear estimators when the input-output relationship is linear, it can still lead to a high overall mean squared error, assumes a linear input-output relationship, and the LS linear regression model may not perform optimally when the number of input parameters is large compared to the training size, leading to potential overfitting.	semiconductor etching chamber
(Amruthnath & Gupta, 2018)	Hierarchical Clustering	Real data	No need to specify the number of clusters in advance, Visualization of a dendrogram, and Flexibility in cluster formation.	Computationally intensive, Sensitivity to noise and outliers, and Lack of ability to revise previous decisions	Exhaust fan
(Andriani et al., n.d.)	Support Vector Classification (SVC) algorithm	does not specify	High Precision, Good Generalization, and Effective for Normal Class	Prediction Errors for Abnormalities, and Lower Recall.	cement processing manufacturing industry

			Prediction.		
(Alvarez Quiñones et al., 2023)	Support Vector Machine	Synthetic Data	SVM can handle complex data samples with non-linear relationships, it is effective in classifying whether a distribution transformer will fail or remain in good condition. Additionally, in the context of this article, SVM showed a lower percentage of error in the predictive capacity of failure in distribution transformers compared to other algorithms tested.	The information on transformer failures includes few predictive variables, which may limit the effectiveness of the SVM algorithm. Increasing the predictive variables, such as operating temperature, environmental humidity, oil level, climate monitoring, output current, and equipment operating time, could potentially enhance the performance of the SVM algorithm by capturing more trends in transformer failures.	Distribution transformer
(Allah Bukhsh et al., 2019)	Tree-based machine learning algorithms, namely decision trees, random forest, and gradient boosted trees.	Real data	Ease of Interpretability, efficient training, internal feature selection, no feature normalization required, and scalability	Longer training time, prone to overfitting, sensitivity to noise data, and lack of smoothness	Railway switch

Machine learning can be beneficial for predictive maintenance because it allows for analyzing historical data to predict when equipment is likely to fail or require maintenance. By using machine learning algorithms, patterns and trends in the data can be identified, enabling maintenance to be scheduled proactively rather than reactively. This can help reduce downtime, increase equipment lifespan, and ultimately save costs for businesses.

2.2.2 Random forest in predictive maintenance

Studies have demonstrated the efficacy of Random Forest in various predictive maintenance applications. For instance, Wu et al. (2019) utilized Random Forest to predict machine failures in a manufacturing environment, achieving high accuracy and reliability. The algorithm's ability to handle both categorical and numerical data makes it suitable for diverse maintenance scenarios. Moreover, Random Forest provides insights into feature importance, helping identify the most critical variables affecting equipment performance.

Random Forest has been extensively investigated in predictive maintenance, with research demonstrating its advantages over competing methods (Singh, 2019). This is further increased using synthetic data approaches, which can widen the model's training range while improving accuracy and dependability (Yang & Ren, 2020). Variable selection strategies, such as the Variable Depth Distribution, can also be used to simplify models and increase interpretability (Voronov et al., 2021). In the beverage industry, this strategy can be very useful in addressing issues like underfilling in beer production, guaranteeing uniform product quality, and reducing operational downtime (M. Wang et al., 2023).

Synthetic data, particularly that generated by Conditional Generative Adversarial Networks (cGANs), is becoming increasingly important in predictive maintenance (PdM). Its ability to simulate diverse failure modes and operational scenarios addresses one of the most pressing challenges in PdM—insufficient real-world data (Ma et al., 2024). This capability allows researchers and practitioners to develop and validate machine learning models in scenarios where real datasets are scarce or incomplete.

Building on this, Abassi et al. (2024) demonstrated how cGAN-generated data significantly enhances the predictive performance of machine learning models. Abood et al. (2023) further advanced this approach by introducing a hybrid model combining Convolutional Neural Networks (CNN) with cGANs, achieving even greater predictive accuracy. Complementary studies by Ranasinghe & Parlikad (2019) and Ma et al. (2024) explored specific applications, with the former focusing on generating real-valued failure data and the latter on augmenting corrosion datasets. Together, these works underscore the pivotal role of cGAN-generated synthetic data in strengthening the robustness and reliability of PdM systems.

2.3 Predictive Maintenance in the Beverage Industry

Numerous academic studies have emphasized the advantages of predictive maintenance within the beverage sector (Ben et al., 2021b). It was discovered that putting in place an autonomous maintenance program greatly decreased the number of machine breakdowns, increasing the dependability of crucial machinery. For example, (Lee et al., 2019) proposed a quality management ecosystem for predictive maintenance in the industry 4.0 era, using smart sensors, machine learning, big data analytics, and artificial intelligence. They presented five real-world cases of predictive quality management based on new technologies, such as detecting foreign objects in beer bottles, predicting the shelf life of beer, and optimizing the fermentation process. (Tercan & Meisen, 2022) conducted a systematic review of machine learning and deep learning-based predictive quality in manufacturing and identified the main challenges and opportunities for future research. They also discussed a case study of a bottle-filling line, where they used neural network methods to perform real-time predictive maintenance.

Although as discussed in the predictive maintenance approach, many different PdM approaches can be used as the foundation for prediction, the PdM tasks in this study are centered around using analytics software with historical data. Before a breakdown event occurs, equipment usually exhibits warning indications. These indications can include temperature increases above normal, an increase in vibration levels or a change in vibration spectrum, a decrease in performance, an increase in noise, changes in current and voltages, and a few other symptoms (Jimenez et al., 2020).

The beverage industry relies on predictive maintenance (PdM) to maintain high production standards while minimizing downtime (Schmidt & Wang, 2015).PdM in beer production lines has been shown to reduce underfilling and overfilling while maintaining product quality and optimizing maintenance schedules (Schmidt & Wang, 2015). In the context of Industry 4.0, PdM is critical for the dependability and availability of production systems. It can also improve machine availability and ensure efficient production processes (Nentwich et al., 2020). The use of PdM in the beverage industry can lower production costs by optimizing plant operations and reducing system downtime (Motaghare et al., 2018).

2.4 Underfill Prediction and Prevention in the Beverage Industry

Currently, inefficient maintenance practices (insufficient or excessive maintenance) are the reason behind the rising percentage of maintenance expenditure in the operating costs of businesses (Zhang et al., 2019). Consequently, it is imperative to devise a rational maintenance plan that not only minimizes waste resulting from over-maintenance but also guarantees proper maintenance and continued functionality of the equipment (Meddaoui et al., 2023).

Real-time monitoring of bottling lines in food, cosmetics, healthcare, and chemical product industries ensures quality control, including accurate filling, cap closure inspection of underfill and overfill, recycling plastic bottle sorting, label quality verification, and defect detection. Many aspects of bottle underfilling have been the subject of research, including detection, mechanical behavior, and filling time prediction(Anush et al., 2021).

(Anush et al., 2021) used image processing for automatic visual inspection of products before packaging and dispatch, including the detection of underfill and overfill. The image processor aims to check whether the bottle is being filled with the proper quantity of beverage or not (Pithadiya et al., 2011).

In terms of anticipating and avoiding bottle underfilling, research has advanced significantly. Both (Morelle et al., 2021) and (Anush et al., 2021) who concentrated on digital image processing and Morelle on foam dynamics, used machine learning techniques to identify and forecast underfilling. (Pavlović et al., 2020) presented numerical techniques for forecasting the mechanical responses of partially filled bottles to compression loadings, which may lead to underfilling. Using a predictive modeling technique(Dasnoy et al., 2022) adopted an alternative strategy that can help prevent underfilling by estimating the overfill amount needed for liquid-in-vial medicinal formulations. Together, these studies offer a thorough understanding of the elements causing underfilling as well as practical forecasting and preventative techniques.

2.5 Related Works

Predictive maintenance, or PdM, is a proactive strategy that uses data-driven methodologies to forecast asset failures and manage maintenance schedules efficiently(Kane et al., 2022). The application of machine learning in PdM has been a key focus, with a particular emphasis on the use of advanced algorithms to predict equipment failures (Trần et al., 2021).

In the context of Industry 4.0, a bibliometric analysis by (Grubisic et al., 2020) identified the most relevant studies on intelligent predictive maintenance, highlighting the importance of machine learning and data-driven approaches for predicting equipment failures. The study also identified potential gaps in the literature, such as the need for more research on the practical implementation of predictive maintenance in real-world industrial applications. For industrial systems, such as industrial gearboxes (Cardoso & Ferreira, 2020) proposed a predictive maintenance approach using machine learning algorithms to improve the availability of systems, reduce maintenance costs, increase operational performance and safety, and support decision-making regarding the ideal timing and actions for maintenance interventions. The study compared the performance of different machine learning models, such as Random Forest and Artificial Neural Networks, using the validation and test sets. The evaluation metrics used for comparison included Precision, Recall, and F1 Score for different components. The study also compared the correlation between different features in the dataset to analyze their relationship and relevance.

According to (Kane et al., 2022), Predictive maintenance and intelligent sensors are critical components of smart factories. By using predictive maintenance systems, a manufacturing facility's systems failures can be predicted, allowing for prompt maintenance to avoid total machine failure. Machine learning models that have been trained on sensor data are used to do this because they can predict parameter values over time. To generate precise predictions, the models use LSTM deep learning, regression, and classification algorithms. By predicting machine downtime based on variables like temperature, pressure, machine speed, and sensor data, the suggested system hopes to boost manufacturing productivity and save maintenance costs. However, this paper lacked a discussion of the challenges and limitations associated with the use of intelligent sensors and predictive maintenance systems.

Predictive maintenance, which utilizes machine learning techniques, has been extensively explored across various industries. In a notable study by Agostinelli & Cumo, 2022, they developed a predictive maintenance strategy specifically for building management systems. The study's focus was on a contemporary residential district in Rome, consisting of 16 buildings and 911 apartments. The investigation specifically targeted mechanical, electrical, lighting, and plumbing systems that catered to both the external and internal areas of the buildings. This strategy aims to ensure the functionality of buildings by optimizing the timing of maintenance operations and integrating data from building information modeling (BIM) and the Internet of Things (IoT) using machine learning algorithms. The study addresses the inadequacy of data integration for predictive maintenance, the absence of good predictive patterns, and the lack of description of predictive procedures found in previous research. The proposed methodology is based on a statistical evaluation of the deviation between the failure rate of a single device and a set threshold value, and it aims to minimize costs associated with scheduled maintenance activities.

The study conducted by Kaparthi & Bumblauskas in 2020 emphasizes the practical and social implications of a specific approach, highlighting its potential for sustainable resource allocation and efficient utilization across different systems. Their approach utilizes a decision tree-based machine learning technique to successfully predict system failures. It achieves this by analyzing data records that contain signals and their corresponding states. Importantly, this approach is scalable and can be applied to various systems, irrespective of the underlying physics governing them.

(Tiddens et al., 2022)) explores the implementation of predictive maintenance (PdM) in various industries in the Netherlands. The study identifies the choices made by companies in implementing PdM and the logical combinations of elements involved in the process. The findings suggest that successful companies combine various techniques, but still rely on those based on previous experiences. The research calls for better methods to guide the selection and use of suitable types of PdM and provides new insights into the application and selection of techniques for PdM in practice.

2.6 Summary of the literature review

The main points raised by the literature review are:

- Predictive maintenance (PdM) is a crucial approach that aims to prevent equipment failure by gathering, cleansing, evaluating, and utilizing data from various manufacturing and sensors.
- PdM uses advanced algorithms to predict or foresee equipment failure before it occurs, thereby optimizing equipment utilization and maintenance strategies, enhancing performance and productivity, and prolonging the life of the equipment.
- The industry 4.0 era has seen the integration of smart sensors, machine learning, big data analytics, and artificial intelligence into the management ecosystem for predictive maintenance, leading to real-world cases of predictive quality management in manufacturing.
- Various approaches to predictive maintenance exist, including knowledge-based, data-driven, physical modeling, and digital twin approaches, each with its advantages and disadvantages.
- Machine learning techniques such as logistic regression, decision tree-based techniques, and multi-class classification algorithms like Support Vector Machine (SVM), k-nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF) have been utilized in predictive maintenance strategies across different industries.

The literature review highlights the critical role of predictive maintenance (PdM) in reducing manufacturing costs and improving equipment longevity by predicting failures before they occur. Various PdM approaches, including knowledge-based, data-driven, physical modeling, and digital twin methods, leverage advanced algorithms and machine learning (ML) techniques to optimize maintenance schedules and enhance productivity. Machine learning models, particularly Random Forest, has shown significant promise in handling complex data and predicting equipment failures. Despite the proven benefits and increasing adoption of PdM, gaps remain in its practical implementation, data integration, evaluation of different PdM methods, scalability, and economic impact assessment. Addressing these gaps is essential for enhancing the reliability and efficiency of PdM solutions across diverse industrial applications.

Chapter Three

3. Methodology

3.1 Overview of the Methodology

The methodology begins with identifying the core industrial problem, which is the underfilling of bottles in the production line at Heineken. This problem is approached using a data-driven strategy that incorporates both real and synthetic data, with the latter being generated through cGAN and iteratively refined through expert validation. The methodology emphasizes process optimization by developing a predictive model using Random Forest, aimed at enabling proactive maintenance actions to minimize underfill occurrences. The process is inherently non-linear and iterative, with different stages interacting and feeding back into each other, ensuring continuous improvement of the model and its applicability. Throughout, the methodology is carefully aligned with Industrial Engineering principles, focusing on enhancing efficiency, maintaining quality control, reducing operational costs, and considering the overall system dynamics. This structured approach not only addresses the immediate problem but also aligns with broader industrial objectives.

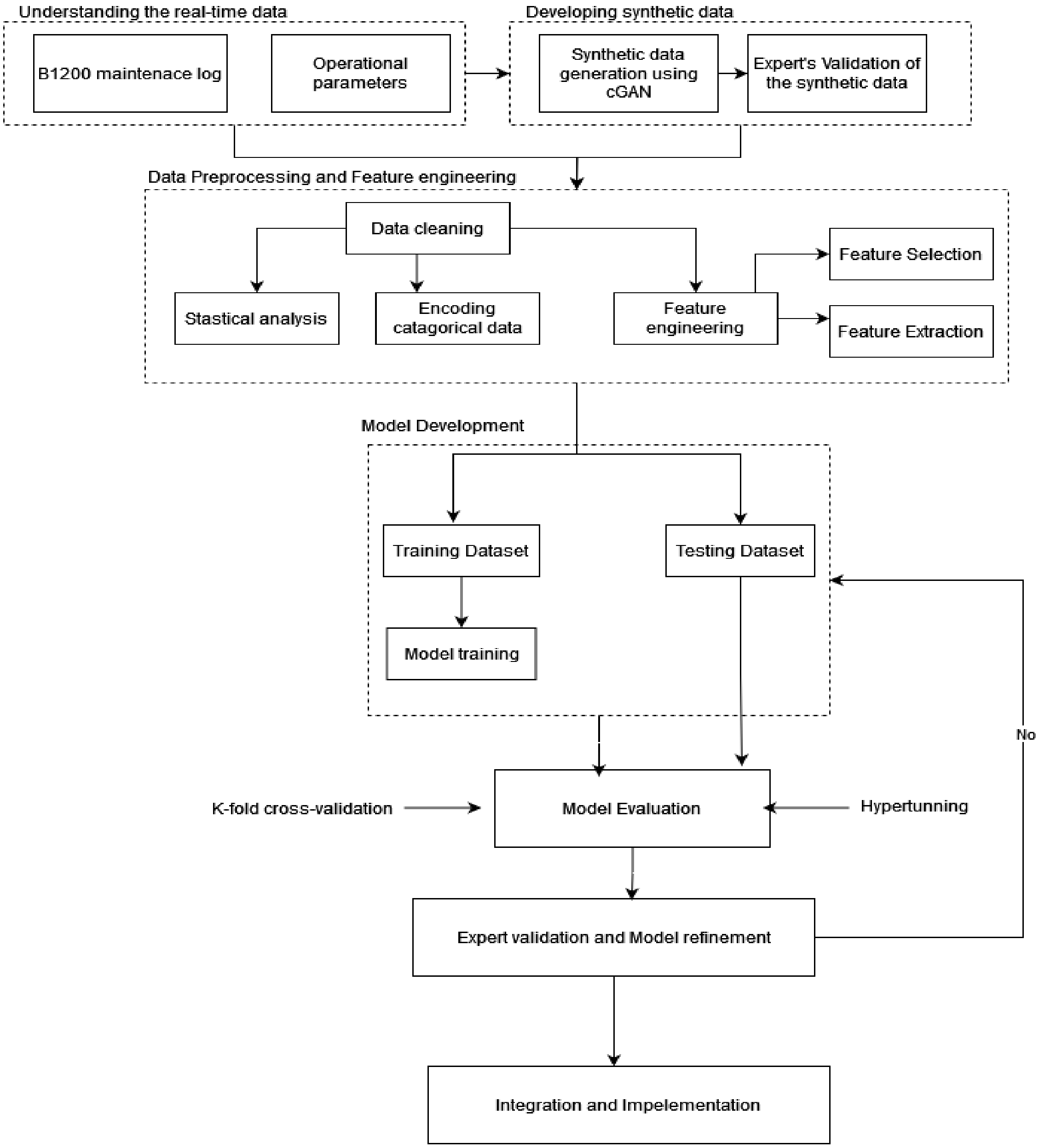


Figure 2 Conceptual framework.

3.2 Data Collection and Synthetic Data Generation

The data collection process commenced with the gathering of real-world data from the bottling line at Heineken. This data focused on key variables that influence fill levels, such as machine settings, environmental factors, and historical fill rates. These variables were chosen due to their critical impact on the accuracy and reliability of the predictive maintenance model. However, an immediate challenge arose due to the difficulty in obtaining data in large quantities.

Real-time data was collected from Heineken's B1200 line using advanced monitoring systems, as illustrated by the DisplayPLC interface in the production line control system. This interface provided comprehensive and precise real-time data on various critical parameters. Key elements of the collected data included:

- **Machine Logs:** Detailed records of operational events and status updates, providing insights into the performance and any anomalies in the production process.

- **Operating Parameters:** Data on fill level settings, speed settings, and other operational configurations that influence the filling process.

- **Production Rate:** Real-time data on the production rate, offering insights into bottling efficiency and correlating production speeds with underfill incidents.

- **Beer Temperature and Pressure:** Monitoring these parameters ensured consistent product quality and highlighted any deviations that could lead to underfilling.

- **Vacuum and Infeed Pressure:** Essential for maintaining uniform filling, these pressures were closely monitored to prevent underfilling.

- **Foam Sensitivity:** Data on foam formation, crucial for managing excessive foam that can cause underfilling issues.

The DisplayPLC, played a pivotal role in capturing this data. It interfaced with various sensors on the production line, continuously recording high-resolution, time-stamped data. This system ensured the collection of accurate and comprehensive data necessary for analyzing and predicting underfill events.

Despite its comprehensiveness, the data collection faced challenges such as ensuring data integrity amid sensor malfunctions, managing large data volumes, and securing sensitive production data. Regular calibration and maintenance of sensors were performed to ensure accuracy. Additionally, data storage and processing systems were optimized to handle large-scale data, and stringent data security protocols were implemented to protect sensitive information.

3.2.1 Data Augmentation using cGAN

To overcome this data limitation, the methodology incorporated the use of Conditional Generative Adversarial Networks (cGAN) to generate synthetic data. cGANs were selected for their ability to create realistic data points that mimic the properties of the real-world data. This synthetic data generation process was not a one-time event but rather an iterative process. Initially, synthetic data was generated based on the limited real-world data available. This data was then subjected to rigorous validation by domain experts who evaluated its accuracy and resemblance to actual bottling conditions.

Synthetic data generation involved using Generative Adversarial Networks (GANs) implemented with tools like NumPy, TensorFlow, and Pandas. GANs were trained on the real-time data to generate realistic synthetic data, validated by comparing statistical properties to ensure consistency. Once generated, the synthetic data was integrated with the real-time data, forming a comprehensive dataset for robust analysis and model training. This integration ensured a balanced dataset, representing normal operating conditions and underfill incidents, preventing model bias, and enhancing predictive accuracy.

By leveraging both real-time and synthetic data, the study developed a robust predictive maintenance model to effectively identify and address underfill issues in Heineken's production line. This combined approach ensured the model was trained on a diverse and comprehensive dataset, improving its reliability and performance in real-world applications.

The expert feedback played a crucial role in refining the synthetic data. Based on their insights, adjustments were made to the data generation process, and new iterations of synthetic data were produced. This cycle of generation, evaluation, and refinement continued until the synthetic data closely matched the characteristics of the real-world data. This iterative process ensured that the final dataset was both comprehensive and representative of the actual bottling process, thus providing a solid foundation for the subsequent model development.

The result of this comprehensive approach was a final dataset that combined both real and synthetic data. The synthetic data, validated and refined through expert input, supplemented the real-world data and compensated for the initial limitations in data quantity. This enriched dataset was then used in the development of the predictive maintenance model, ensuring that the model was not only based on sufficient data but also reflective of the complex dynamics of the bottling process. This approach not only addressed the immediate challenge of data scarcity but also aligned with the broader goals of creating a reliable and effective predictive maintenance system.

3.3 Data Preprocessing and Feature Engineering

We use a variety of analytical approaches to get insights from the gathered data during the data analysis phase. The following steps outline the methodology used for data analytics and feature engineering, utilizing various Python libraries and techniques to preprocess, analyze, and visualize the data.

Statistical Summary and Additional Analysis: A statistical summary was generated to provide an overview of the numerical features. This included calculating descriptive statistics such as mean, standard deviation, and range.

```
# Statistical summary
print(generated_df.describe())

# Check the proportion of each classification
print(generated_df['Classification'].value_counts(normalize=True))
```

Checking for Missing Values: The number of missing values in each column was calculated to ensure data completeness.

Skewness: The skewness of each numerical feature was calculated to understand the distribution of the data.

Correlation Matrix: A correlation matrix was computed to identify the relationships between numerical features.

```
# Correlation matrix for numeric features
correlation_matrix = numeric_df.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```

Data Visualization: Various plots were created to visualize the data distributions and relationships between features. These are Histograms, correlation heatmap, and pair-plot.

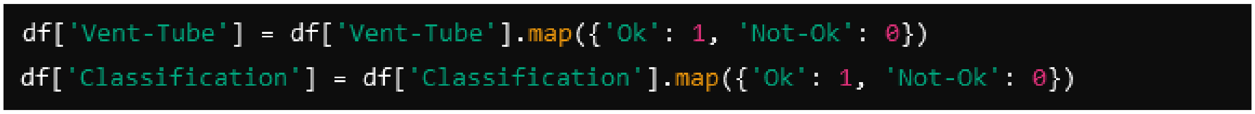
```
# Pairplot to see pairwise relationships
sns.pairplot(generated_df, hue='Classification')
plt.show()

# Histograms of each feature
numeric_df.hist(bins=30, figsize=(15, 10))
plt.show()

# Boxplot to visualize the distribution of volumes for each classification
plt.figure(figsize=(10, 6))
sns.boxplot(x='Classification', y='Actual-Volume', data=generated_df)
plt.title('Boxplot of Volume by Classification')
plt.show()
```

3.3.1 Data Preprocessing

Data preprocessing is a crucial step in developing a predictive maintenance model for underfilling. The primary activities involve cleaning and preparing the collected data for analysis, which includes handling missing values, outliers, categorical variables, creating new features, and dealing with inconsistent datasets. Missing values in the dataset are handled using imputation techniques such as mean, median, or most frequent value imputation to ensure data completeness. Rows with missing values resulting from lag and rolling window features are dropped to ensure a clean dataset for analysis. Outliers are identified and handled using methods like Z-score or Interquartile Range (IQR) to ensure data integrity. Categorical columns, such as 'Vent-Tube' and 'Classification,' are converted from categorical (string) values to numeric values to facilitate numerical analysis. Feature transformation techniques such as normalization and standardization are applied to bring features onto a similar scale, helping many machine learning algorithms perform better. Log transformation is applied to reduce skewness in the data distribution. New features are created based on domain knowledge or interactions between existing features (e.g., Temperature_Pressure_Ratio). Lag features are created to capture temporal dependencies, assuming the data is ordered sequentially. Rolling window features, such as rolling mean and standard deviation, are calculated to capture trends and variability over a window of past observations. Interaction features are created by combining existing features through multiplication, providing additional insights into the relationships between variables. Domain-specific features are derived based on expert knowledge, such as the difference between the actual volume and an expected volume.

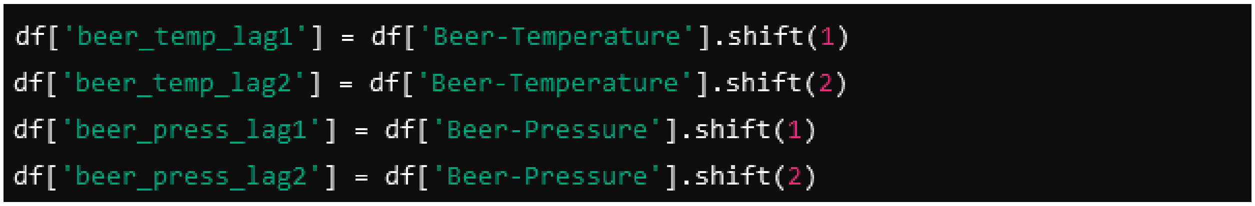
Convert Categorical Columns to Numeric: The Vent-Tube and Classification columns were converted from categorical (string) values to numeric values to facilitate numerical analysis.

```
# Convert 'Vent-Tube' to numerical values if they are categorical
if 'Vent-Tube' in synthetic_data_df.columns:
    synthetic_data_df['Vent-Tube'] = synthetic_data_df['Vent-
Tube'].map({'Ok': 1, 'Not Ok': 0, 'OK': 1, 'Not OK': 0})
```



```
df['Vent-Tube'] = df['Vent-Tube'].map({'Ok': 1, 'Not-Ok': 0})
df['Classification'] = df['Classification'].map({'Ok': 1, 'Not-Ok': 0})
```

Lag Features: Lag features were created to capture temporal dependencies, assuming the data is ordered sequentially.



```
df['beer_temp_lag1'] = df['Beer-Temperature'].shift(1)
df['beer_temp_lag2'] = df['Beer-Temperature'].shift(2)
df['beer_press_lag1'] = df['Beer-Pressure'].shift(1)
df['beer_press_lag2'] = df['Beer-Pressure'].shift(2)
```

Rolling Window Features: Rolling window features, such as the rolling mean and standard deviation, were calculated to capture trends and variability over a window of past observations.

Interaction and Domain-Specific Features: Interaction features were created by combining existing features through multiplication, providing additional insights into the relationships between variables. And domain-specific features were derived based on expert knowledge. For instance, the difference between the actual volume and the expected volume was calculated.

Handling Missing Values: Rows with missing values resulting from lag and rolling window features were dropped to ensure a clean dataset for analysis.

```
print("Missing Values:\n", data.isnull().sum())
```

3.3.2 Feature Engineering

Feature engineering involves extracting meaningful features from the collected data that can help in predicting underfill issues. This step includes feature selection and standardization. Features are selected based on their correlation with the target variable and their importance scores derived from models like Random Forest. Features with high correlation with the target variable and low correlation with each other are preferred. Features are standardized to ensure each feature contributes equally to the model's performance, preventing any single feature from dominating the model.

Feature Selection and Standardization: The features are separated from the target variable (Classification), and the features are standardized to ensure each feature contributes equally to the model performance.

```
# Select features to scale
features_to_scale = ['Beer-Temperature', 'Beer-Pressure', 'Vacuum-
Pressure', 'CO2', 'Actual-Volume']

# Scale the features
scaler = StandardScaler()
generated_df[features_to_scale] =
scaler.fit_transform(generated_df[features_to_scale])

# Display the scaled DataFrame
print(generated_df.head())
```

3.3.3 Data Splitting

The dataset is split into training and testing sets to evaluate the model's performance on unseen data. The data set will be classified as 80% to train the random forest model and 20 percent to test the model.

```
X = generated_df[features_to_scale]
y = generated_df['Classification']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
```

3.4 Model Development and Training

3.4.1 Model Training and Hyperparameter Tuning

A pipeline is created to streamline preprocessing and model training. Hyperparameter tuning is performed using Grid Search with cross-validation to find the optimal parameters for the Random Forest classifier.

```
from sklearn.ensemble import RandomForestClassifier


model = RandomForestClassifier()

model.fit(X_train, y_train)
aram_grid = {
    'n_estimators': [100, 300, 500, 1000],
    'max_depth': [10, 15, 20, None],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4],
    'max_features': ['sqrt', 'log2', None]
}

grid_search = GridSearchCV(estimator=rf, param_grid=param_grid, cv=5,
n_jobs=-1, verbose=2)
grid_search.fit(X_train, y_train)

print(f"Best Parameters: {grid_search.best_params_}")
best_rf = grid_search.best_estimator_
```

3.5 Model Evaluation

Specific evaluation metrics will be used to evaluate the model performance, such as:

Cross-Validation: Cross-validation is performed to evaluate the model's generalizability across different subsets of the data.

```
best_cv_scores = cross_val_score(best_rf, X, y, cv=5)
print(f"Best Cross-Validation Scores: {best_cv_scores}")
print(f"Mean Best CV Score: {best_cv_scores.mean()}")
```

Accuracy Calculation: The model's performance is assessed on both the training and testing sets to determine accuracy and identify potential overfitting.

```
accuracy_dt = accuracy_score(y_test, y_pred_dt)
print(f"Decision Tree Accuracy: {accuracy_dt:.2f}")
print("Classification Report:\n", classification_report(y_test,
y_pred_dt))
```

Feature Importance:

The feature importance step is crucial for interpreting the model, understanding the significance of different features, and ensuring the model's reliability and transparency. It is positioned at the end of the modeling process to leverage the final, well-tuned model, providing a meaningful analysis of which features most impact the model's predictions.

The importance of each feature is assessed to understand the contribution of each variable to the model's predictions.

```
importances_rf = random_forest.feature_importances_
indices_rf = np.argsort(importances_rf)[::-1]
```

3.6 Expert Validation and Model Refinement

An integral component of the methodology was the expert validation and refinement process. After the initial development and evaluation of the predictive model, its predictions were periodically reviewed by domain experts. This expert feedback loop was crucial in ensuring that the model not only performed well statistically but also met practical, real-world expectations and requirements. Experts assessed the model's predictions in the context of actual bottling operations, providing valuable insights that informed necessary adjustments.

The feedback from experts could lead to several actions, including adjustments to the model itself, re-generation of synthetic data to better capture real-world conditions, or refinements in feature engineering. This iterative process ensured that the model remained aligned with practical needs, addressing any discrepancies between predicted and actual outcomes. By incorporating expert input, the methodology maintained a strong focus on practical relevance, ensuring that the model effectively supported the operational goals of Heineken.

The iterative, non-linear nature of this expert validation process underscored the methodology's alignment with Industrial Engineering principles. It highlighted the focus on process optimization and system efficiency, ensuring that the model was not only technically robust but also pragmatically useful in the industrial context.

3.7 Integration and Implementation

The final phase of the methodology involved integrating the predictive model into Heineken's existing maintenance systems. The goal was to enhance decision-making by providing operators with actionable insights to prevent underfill issues. This integration was carefully planned with an emphasis on usability and real-time decision support, which is critical in an industrial setting where timely and accurate information can significantly impact operational efficiency.

In addition to the integration, the methodology incorporated a continuous improvement feedback loop. The model's performance in the real-world setting was monitored on an ongoing basis, allowing for adjustments as needed. This continuous monitoring and refinement process is in line with Industrial Engineering's emphasis on continuous process improvement, ensuring that the model evolves and adapts to changing conditions and emerging challenges.

Chapter Four

4. Result and Discussion

4.1 Identifying the Problem in the Heineken B1200 Production Line

4.1.1 Identifying the Underfill Issue in B1200 Filler

The underfill issue in the B1200 filler was identified through various diagnostic and operational checks as part of the quality control measures during production. The key indicator was the frequent rejection of bottles due to underfilling, a critical defect detected by both automated sensors and manual inspections. Through a series of inspections using tools like the Sniffing and Filler Checkmate, which involved checking vacuum levels and product presence in the filling bowl. Another key method was Leakage Detection and Pressure Checks, where fluctuations in the product pump and CO2 pressure levels were monitored. Any deviation from standard levels indicated a potential issue with the filler, leading to underfilling.

Additional causes of underfilling were traced to specific operational conditions such as over foaming, which occurred when improper filling angles or temperature variations affected the flow of beer into the bottles. These operational inefficiencies resulted in the filler either prematurely stopping or not dispensing the correct amount of product.

Figure 3 Worn out O-ring

Restoration to basic machine conditions, such as replacing worn-out O-rings and recalibrating filling parameters, was necessary to bring the equipment back to standard operational performance. The preventive measures implemented included predictive maintenance and regular inspections to detect and mitigate potential underfilling issues before they became significant, ensuring the B1200 filler operated efficiently with minimal defects.

4.1.2 Current Maintenance Practices for the B1200 Filler

The current maintenance practice for the B1200 filler at Heineken primarily revolves around a Preventative Maintenance (PM) strategy. This approach is designed to ensure the reliability and efficiency of the bottling line by scheduling regular maintenance activities aimed at preventing equipment failures before they occur.

The PM practices for the B1200 involve a range of activities tailored to the unique operational needs of the filler. These activities include:

- **Routine Inspections**: Scheduled inspections are conducted to monitor the condition of critical components, such as filling valves, pumps, and sensors. Technicians assess these components for wear and tear, ensuring that any potential issues are identified early.

- **Lubrication and Adjustments**: Regular lubrication of moving parts is performed to reduce friction and wear, while adjustments to machine settings are made to maintain optimal performance levels.

- **Cleaning**: Routine cleaning procedures are implemented to prevent contamination and ensure that the filling process remains sanitary. This includes cleaning the filling nozzles and surrounding areas to prevent build-up.

- **Replacement of Worn Parts**: Components that show signs of wear, such as seals, gaskets, and O-rings, are replaced as part of the maintenance routine to avoid breakdowns during production.

The maintenance schedule for the B1200 filler is structured to minimize disruption to production while ensuring the equipment remains in optimal condition. Key elements of the schedule include:

- **Weekly Maintenance**: Weekly tasks typically involve thorough inspections, lubrication, and minor adjustments. These tasks are designed to catch any early signs of wear and ensure that the filler operates smoothly.

- **Mini-Overhauls**: Every two to three days, the B1200 undergoes a mini-overhaul where more extensive maintenance activities are performed. This may include checking and replacing filters, recalibrating sensors, and ensuring that all systems are functioning correctly.

- **Major Overhauls**: Scheduled major overhauls occur periodically, which may involve comprehensive checks and replacements of critical components. This is typically done during planned production downtimes to minimize impact on overall operations.

The maintenance practices at Heineken's B1200 line are classified under preventative maintenance, aimed at maintaining and improving the reliability of the filling equipment. By adhering to a structured maintenance schedule, Heineken seeks to:

- **Enhance Operational Efficiency**: Regular maintenance helps to reduce the risk of unexpected breakdowns, thereby improving the overall equipment effectiveness (OEE) of the B1200 filler.

- **Maintain Product Quality**: Preventative measures ensure that the filling process remains consistent, and that product quality is not compromised due to equipment failure.

- **Optimize Resource Utilization**: By preventing failures and minimizing downtime, Heineken can maximize its production output and resource usage, thereby improving overall profitability.

4.1.3 **Analysis of Bl200 malfunctions on Operation**

The malfunction of the B1200 filler is a significant operational issue that impacts the overall performance, efficiency, and profitability of the production line. The key areas where the B1200 malfunction creates inefficiencies include product rejections, extract loss, downtime, and energy inefficiency. Analyzing these malfunctions involves understanding the underlying technical failures and their ripple effects on production.

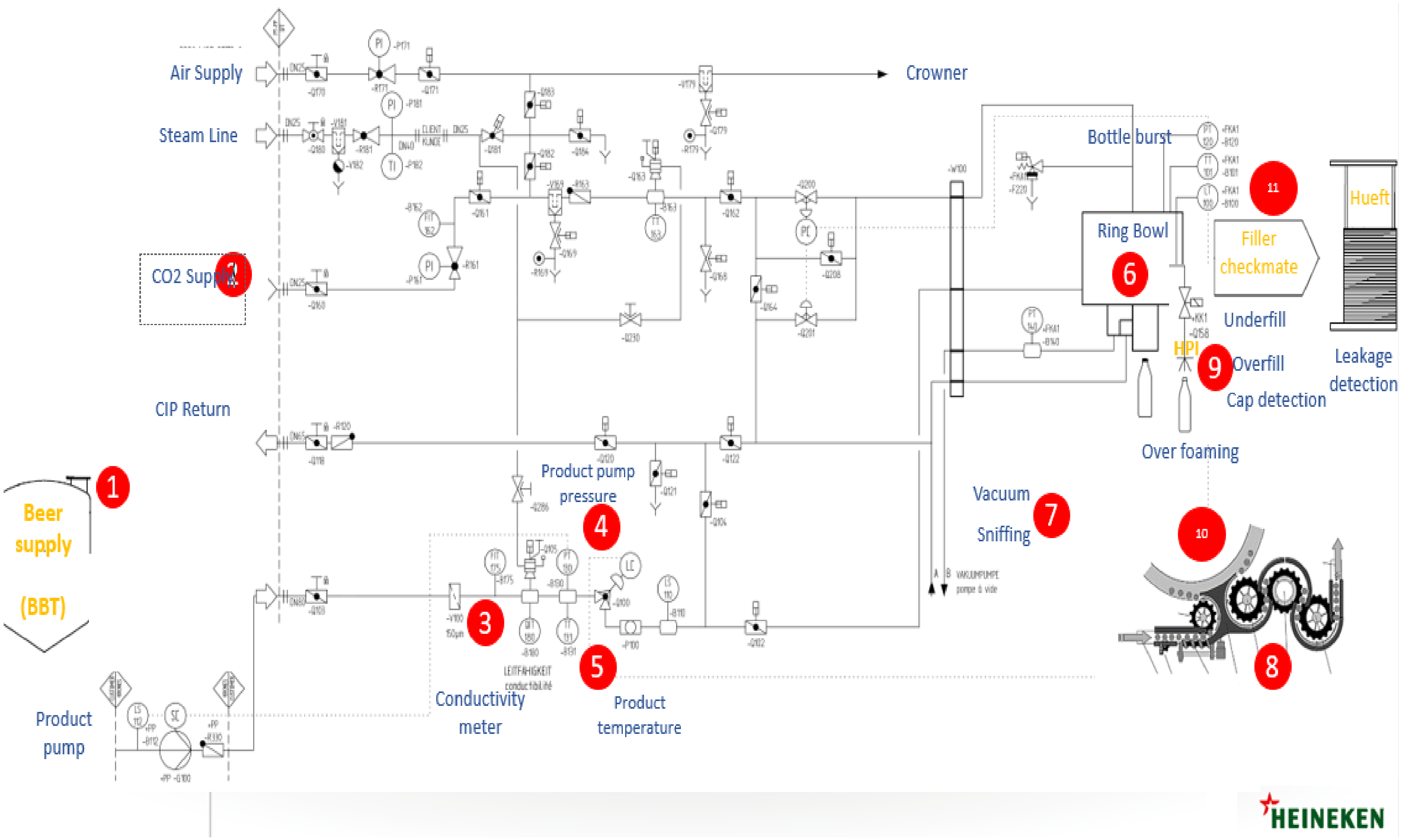


Figure 4 Loss point and source of extract loss on B1200 filler

Product Rejection Due to Inconsistent Filling

The B1200 filler experiences frequent malfunctions, leading to inaccurate fill levels in bottles. One of the primary consequences is the high rate of underfilled or overfilled bottles. Bottles that do not meet the required fill level are rejected by the quality control systems. This rejection rate has been documented at 6.87% for the B1200 filler. With a production capacity of 42,000 bottles per hour, this translates to nearly 2,887 rejected bottles every hour, resulting in significant waste over time.

The cause of these rejections is often linked to mechanical issues such as improper valve functioning, pressure drops in the CO2 system, or problems with the product pump. These mechanical failures affect the precision of the filling operation, leading to variability in the amount of liquid dispensed into each bottle. Over time, the continued occurrence of such malfunctions exacerbates the issue, causing a cumulative loss of product and requiring rework or disposal of rejected bottles.

1. **Root Causes of Malfunction:**

- CO2 Pressure Drops: The B1200 filler relies on controlled pressure to fill bottles accurately, and any drop in pressure leads to inconsistent product flow. This malfunction causes the filler to either overfill or underfill bottles.
- Product Pump Failure: Another root cause of malfunction is the inconsistent operation of the product pump, which may suffer from leakage or wear. When the pump does not maintain a stable flow of product, the filler fails to deliver the required amount of liquid into the bottles.
- Valve and Sensor Issues: The filler valves, which control the opening and closing of the filling nozzles, may malfunction due to wear and tear. Similarly, the sensors used to monitor product levels and flow rates may provide inaccurate data, causing the filler to make incorrect adjustments during the process.

2. **Downtime and Unscheduled Maintenance:** When the B1200 filler malfunctions, it leads to unscheduled downtime, as technicians must stop the production line to address the issues. Each time a malfunction occurs, the production must be paused for troubleshooting, cleaning, and recalibrating the machine. This downtime not only reduces the overall

equipment effectiveness (OEE) but also delays the fulfillment of production targets. Given that the B1200 produces 42,000 bottles per hour, any downtime translates to significant lost productivity. Throughout a typical 8-hour shift, even an hour of downtime can result in a loss of 42,000 bottles, disrupting the entire bottling process and causing bottlenecks in the supply chain.

Furthermore, the reactive nature of the maintenance process in response to the B1200 malfunctions increases the wear and tear on the machine. Repeatedly stopping and restarting the filler for repairs leads to more frequent breakdowns, shortening critical components' operational lifespan, such as valves, pumps, and seals. As a result, maintenance and repair costs increase over time, adding to operational expenses.

3. **Energy and Resource Inefficiency:** Each malfunction of the B1200 filler increases energy consumption due to the need for additional filling cycles and rework. The repeated stops and starts, combined with the need to reprocess rejected bottles, lead to increased energy usage in the form of electricity, CO2 for carbonation, and water for cleaning. The inefficiency not only raises operational costs but also has a negative environmental impact, as more resources are consumed for the same level of output. Over time, these inefficiencies reduce the sustainability of the production line, as more energy and materials are required to compensate for the filler's malfunctions.

The analysis of the B1200 filler malfunction reveals a range of operational and financial impacts, from product rejections to increased downtime and resource inefficiencies. The root causes of these malfunctions, such as pressure drops, pump failures, and valve issues, need to be addressed through a more proactive maintenance strategy. By implementing predictive maintenance and real-time monitoring systems, the B1200 filler's performance can be stabilized, minimizing malfunctions and their associated costs. In doing so, the overall efficiency of the production line can be improved, leading to higher yields, lower operational costs, and a more sustainable bottling process.

4.1.4 Justification for Implementing Predictive Maintenance in the B1200 Filler

Implementing predictive maintenance for the B1200 filler is essential to enhance operational efficiency, reduce downtime, and minimize costs associated with product rejections and resource wastage. Traditional maintenance strategies, such as reactive or scheduled maintenance, often lead to unexpected equipment failures and unscheduled downtimes, resulting in significant losses in production capacity and increased operational costs. By adopting predictive maintenance, which leverages real-time data and analytics to anticipate equipment failures before they occur, the company can proactively address potential issues, such as pressure drops or valve malfunctions, that lead to underfilling or overfilling of bottles. This approach not only reduces the incidence of product rejections—thereby conserving valuable resources and minimizing extract loss—but also optimizes the overall equipment effectiveness (OEE) of the production line. Furthermore, predictive maintenance fosters a culture of continuous improvement by utilizing data-driven insights to inform maintenance decisions and refine operational practices. Ultimately, this results in increased production reliability, improved product quality, and enhanced sustainability, making predictive maintenance a strategic choice for the B1200 filler operation.

4.2 Overview of the Results

This section presents the results of applying a Random Forest model to the predictive maintenance dataset, focusing on evaluating the model's performance using various metrics. The dataset was preprocessed and augmented using a Conditional Generative Adversarial Network (cGAN) to improve the robustness and the quantity of the training data.

4.2.1 Random Forest Model Architecture

The Random Forest model, trained on the preprocessed hybrid data, displayed a robust performance with a training accuracy of 90.67%, a test accuracy of 90.21%, and a cross-validation accuracy of 90.89%. Additionally, the model achieved a weighted F1 score of 89.71%, indicating a well-balanced model with minimal overfitting. These metrics reflect the model's capability to generalize well to unseen data while maintaining high predictive accuracy.

To delve deeper into the workings of the Random Forest, we examined the architecture of one of the decision trees within the forest. This exploration provides insights into the decision-making process and the key factors influencing the model's predictions.

4.2.2 Model Performance Summary

The performance of the Random Forest model was evaluated using several metrics: training accuracy, test accuracy, cross-validation accuracy, training F1 score, and test F1 score. The results are summarized in Table 3.

Table 2: Model of performance metrics for the Random Forest model.

Metric	**Results**
Training Accuracy	0.9067 (90.67%)
Test Accuracy	0.9021 (90.21%)
Cross-Validation Accuracy	0.9089 (90.89%)
Training F1-Score Accuracy	0.8971 (89.71%)
Testing F1-score Accuracy	0.8903 (89.03%)

The confusion matrix offers a detailed breakdown of the model's classification performance by displaying true positives, false positives, true negatives, and false negatives for each class. The confusion matrices without and with normalization are presented below.

The confusion matrix (without normalization) reveals the following

Class 0 (No-underfill-or-overfill): The model accurately classified 104 instances, with 0 misclassified as Class 1 (Underfill), and 1 as Class 2 (Overfill).

Class 1 (Underfill): The model correctly identified 66 instances, with 6 misclassified as Class 0, and 1 as Class 2.

Class 2 (Overfill): The model perfectly classified 9 instances with 2 misclassifications as class 0 and 5 as class 2.

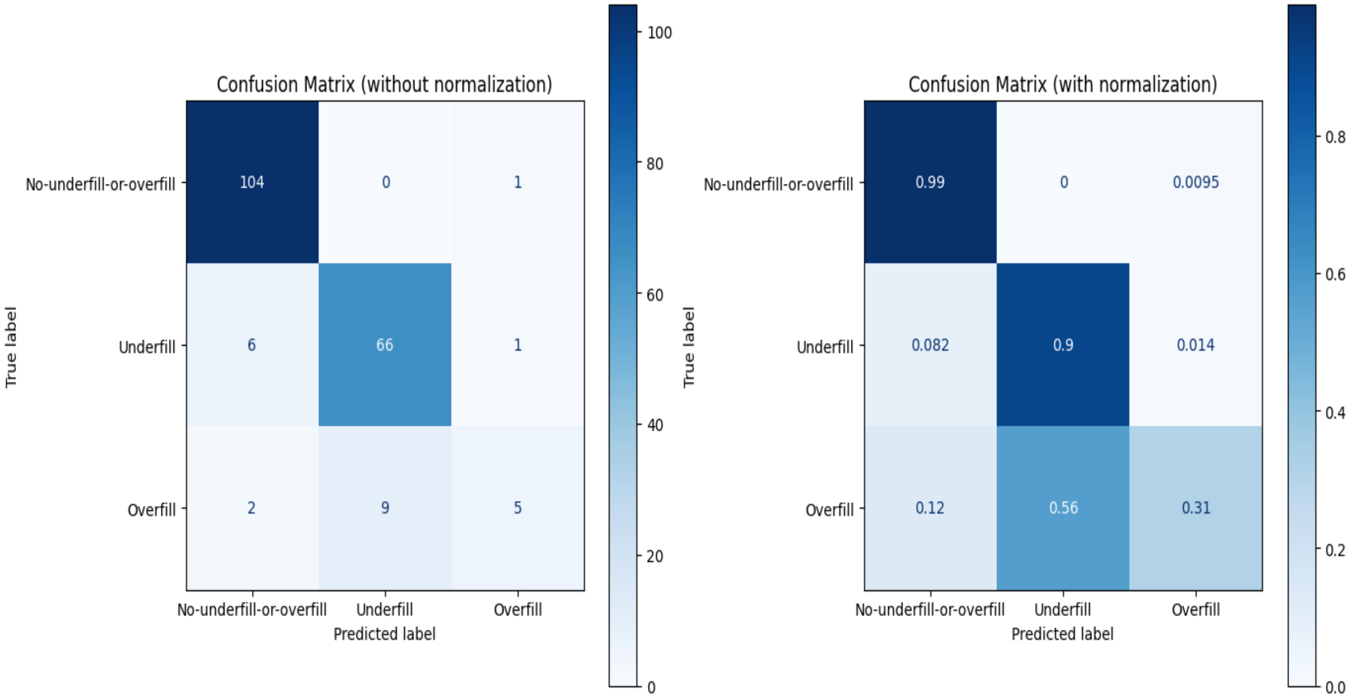


Figure 5 Confusion Matrix without normalization (on the left side). And the confusion matrix with normalization (On the right)

The normalized confusion matrix shows:

Class 0(No-underfill-and-overfill): A precision of 99% and a misclassification rate of zero as Class 1.

Class 1(Underfill): A precision of 90% with 8.2% of instances misclassified as Class 0 and 1.4% of misclassified as class 2..

Class 2(Overfill): A precision of 31% with 1.2% of instances misclassified as Class 0 and 56% of misclassified as class 1.

4.2.3 Interpretation of the random forest performance metric

The Random Forest model applied to the underfill prediction task on the preprocessed hybrid data demonstrates robust performance across various metrics. The model's results show strong predictive capability, which is essential for ensuring that the underfilling of bottles can be detected and addressed efficiently during production. Here's what the key results indicate in the context of predicting underfill:

1. ***Training Accuracy*** *(0.9067):* This indicates that approximately **90.67%** of the training dataset was correctly classified by the Random Forest model. This high training accuracy suggests that the model has learned the patterns in the training data effectively.
2. ***Test Accuracy*** *(0.9021):* The test accuracy of **90.21%** shows that the model performs well on unseen data, indicating good generalization capabilities. This slight drop from the training accuracy is expected and suggests that the model is not overfitting.
3. ***Cross-Validation Accuracy*** *(0.9089):* The cross-validation accuracy of **90.89%** reinforces the reliability of the model's performance across different subsets of the data. This value being close to both training and test accuracy indicates consistent performance.
4. ***Training F1 Score*** *(0.8971):* An F1 score of **89.71%** on the training set reflects a balance between precision and recall. This score indicates that the model effectively identifies the positive class while maintaining a low false positive rate.
5. ***Test F1 Score*** *(0.8903):* The F1 score of 89.03% on the test data demonstrates that the model maintains a strong balance between precision and recall on unseen data as well. The slight decrease from the training F1 score suggests that the model's performance is stable and not overly biased towards the training data.

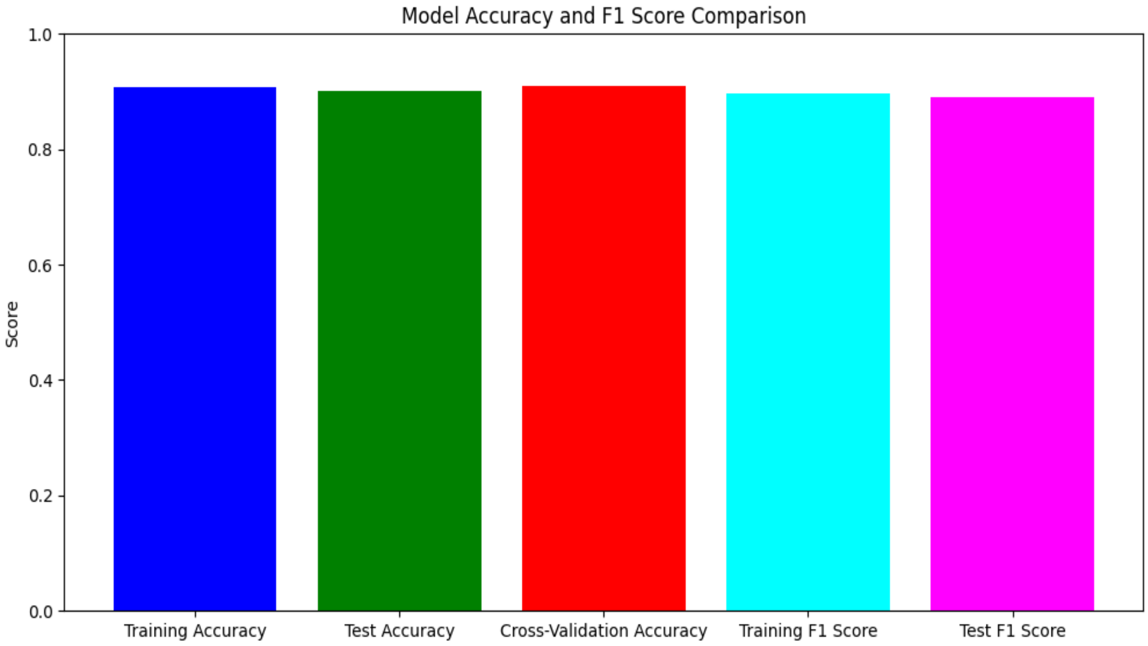


Figure 6 Random Forest Model Accuracy and F1 Comparison

4.2.4 Baseline Performance Comparison

The performance of the Decision Tree, Random Forest, and Naive Bayes models was evaluated using training and testing accuracy, Cross-Validation accuracy, and F1 score. The results are summarized in Table 4.

Table 3: Performance metrics for Decision Tree, Random Forest, and Naive Bayer

Metric	**Decision Tree**	**Random Forest**	**Naive Bayes**
Training Accuracy	0.9430	0.9067	N/A
Test Accuracy	0.8299	0.9021	0.8144
Cross-Validation Accuracy	0.8208	0.9089	0.8063
Training F1 Score	0.9413	0.8971	0.7842
Test F1 Score	0.8350	0.8903	0.8034

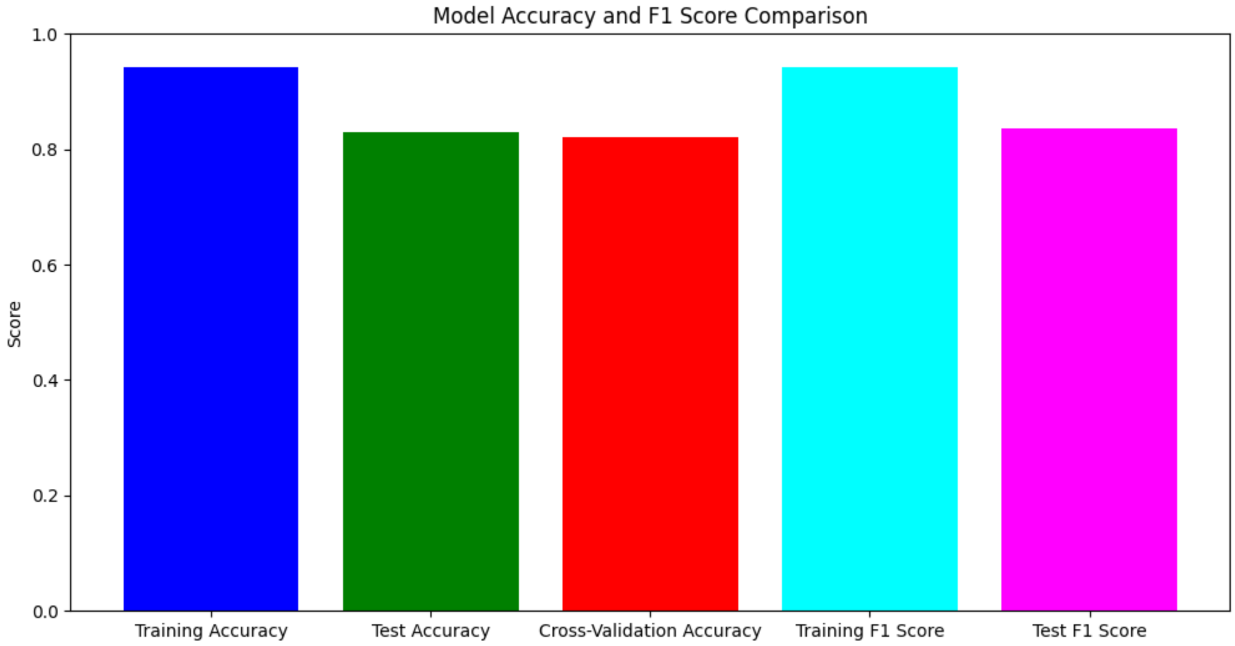


Figure 7 Decision Tree Model Accuracy and F1 Score Comparison

The Decision Tree model achieved a high training accuracy of 94.30%, indicating that it fits the training data well. However, the test accuracy of 82.99% reveals a significant drop in performance on unseen data, suggesting that the model is overfitting. This overfitting is further confirmed by the discrepancy between the training F1 score (0.9413) and the test F1 score (0.8350). Despite its simplicity and interpretability, the Decision Tree model's tendency to overfit makes it less reliable for generalizing to new data.

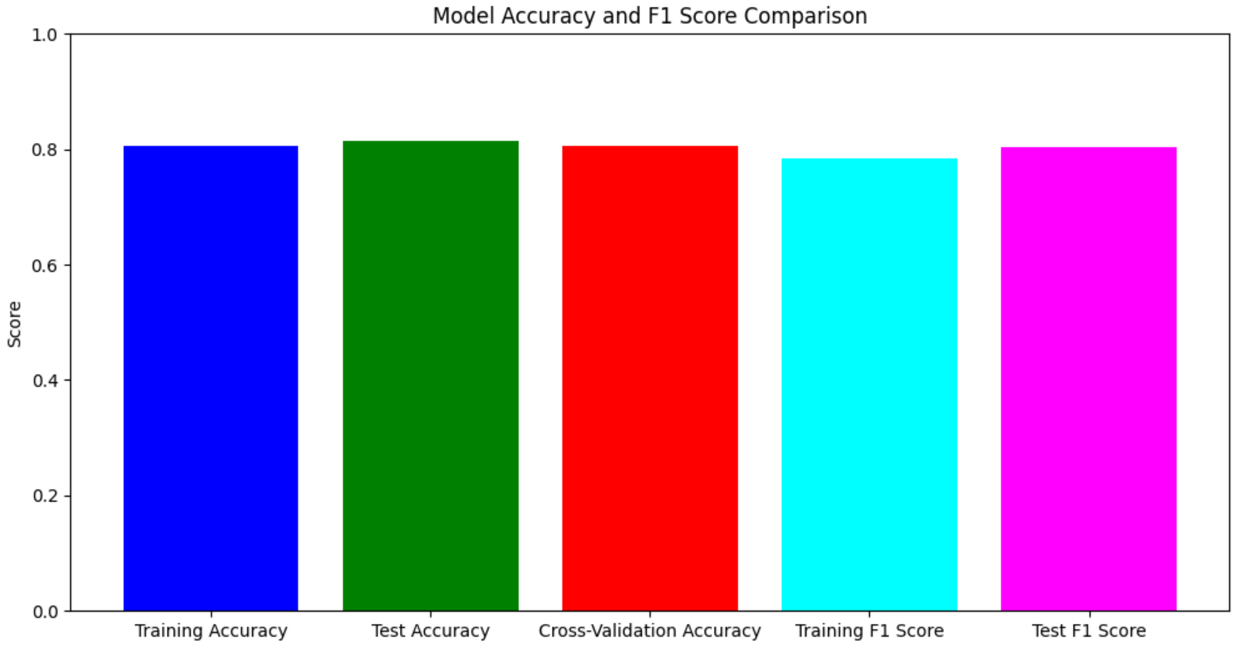


Figure 8 Naive Bayer Model Accuracy and F1 Score Comparison

The Naive Bayes model, while simpler and computationally efficient, performed the least well among the three models. It achieved a test accuracy of 81.44% and a cross-validation accuracy of 80.63%. The training F1 score (0.7842) and test F1 score (0.8034) are relatively consistent but lower compared to the other models. This consistency indicates that Naive Bayes is less prone to overfitting but also less powerful in capturing complex patterns within the data. The lower performance metrics suggest that Naive Bayes may not be the best choice for the predictive maintenance task.

In conclusion, the Random Forest model, augmented with cGAN-generated data, demonstrates superior performance in predictive maintenance tasks compared to the Decision Tree and Naive Bayes models. Its balanced accuracies and F1 scores across training, test, and cross-validation datasets indicate that it generalizes well, making it the most suitable model for proactive maintenance strategies in industrial applications.

4.2.5 Implications for Underfill Prediction

- **Effective Prediction of Underfill Events**: The high accuracy and F1 scores suggest that the model can reliably predict underfill events before they occur. This means that production managers can use the model's predictions to intervene early, preventing defective bottles from continuing down the production line and reducing waste.
- **Reducing Downtime and Waste**: The model’s ability to generalize well to new data is key for reducing downtime caused by underfilling incidents. By accurately predicting when underfill is likely to occur, maintenance and quality control teams can address potential issues before they escalate, minimizing production halts and reducing the volume of wasted product.
- **Optimizing Preventative Measures**: Since the Random Forest model is able to capture the complex, non-linear relationships between various factors (e.g., machine settings, environmental conditions, sensor readings) and underfilling, it provides a data-driven foundation for optimizing machine parameters. This could lead to better calibration of the filling machinery and more efficient maintenance schedules, further minimizing the risk of underfill.
- **Improving Overall Product Quality**: The model reliably predicts and prevents underfill, helping ensure that the final product meets quality standards. This has a direct impact on customer satisfaction, as consistently filled bottles contribute to maintaining brand reputation and reducing customer complaints related to underfilled products.

4.3 Discussion

4.3.1 Evaluation of Random Forest Model Performance

The training accuracy of 90.67%, test accuracy of 90.21%, and cross-validation accuracy of 90.89% indicate a high level of consistency in model performance. This consistency across training and testing datasets shows that the Random Forest model has a good generalization capability. In predictive models, such balance between training and test accuracy is a key indicator that the model is neither overfitting nor underfitting. From a production standpoint, this implies that the model can reliably identify underfill issues under varying production conditions, even when introduced to new batches of data.

The F1 scores of 89.71% on the training set and 89.03% on the test set are particularly important for understanding how well the model handles both false positives (incorrectly predicting underfill) and false negatives (failing to predict actual underfill). The relatively high F1 score indicates that the model maintains a strong balance between precision and recall. This means the model is not only efficient at detecting real underfill incidents but also ensures that it does not flag too many false alarms. In a practical sense, this reduces the number of unnecessary machine stoppages and product rejections, which would otherwise disrupt production flow.

A cross-validation accuracy of 90.89% further strengthens the reliability of the model by demonstrating its robustness across different subsets of the data. This is a significant finding because, in the context of industrial applications, a model needs to perform well across various operational conditions such as fluctuations in environmental factors, machine wear, or varying bottle sizes. Cross-validation shows that the model is not sensitive to the specificities of any one dataset but can generalize across different production scenarios.

4.3.2 Integration of the Predictive Model with the Current Maintenance System

Integrating the predictive Random Forest model into the existing system, particularly for the B1200 filler, can significantly enhance the efficiency of maintenance and quality control processes. The goal of this integration would be to seamlessly blend the predictive model with the current maintenance strategy, which primarily relies on Preventative Maintenance (PM), while gradually transitioning towards a more Predictive Maintenance (PdM) approach.

1. Enhancing the Existing Preventative Maintenance System

The current maintenance strategy for the B1200 filler focuses on scheduled preventive maintenance activities, including inspections, lubrication, and part replacements, which are carried out at fixed intervals regardless of the equipment's real-time condition. By integrating the Random Forest predictive model, this system can be enhanced to dynamically adjust maintenance schedules based on real-time predictions, making the process more data-driven and less reliant on predefined time intervals.

- **Real-Time Monitoring**: The Random Forest model can be integrated with the existing **SCADA (Supervisory Control and Data Acquisition)** systems, which monitor various parameters such as machine pressure, temperature, and sensor readings. The predictive model can analyze this real-time data to predict potential underfill events before they occur, allowing maintenance teams to intervene early and prevent issues.

- **Dynamic Maintenance Scheduling**: Instead of adhering strictly to fixed schedules, maintenance activities can be triggered when the model predicts an increased likelihood of underfill events. This dynamic approach ensures that parts are replaced or recalibrated based on actual machine conditions, reducing unnecessary maintenance actions and avoiding machine failures.

2. Integration into the SCADA System

The SCADA system, which already monitors the production line, can be leveraged to feed real-time data directly into the predictive model. The integration process would involve:

- **Data Collection**: Data from existing sensors on the B1200 filler, such as those monitoring CO2 pressure, temperature, fill level, and machine vibrations, will be continuously fed into the predictive model.

- **Alerts and Automation**: When the model detects conditions that indicate a high probability of underfill or filler malfunction, it can automatically trigger alerts to operators via the SCADA interface. These alerts can also be tied to automated maintenance workflows, such as sending notifications to the maintenance team or initiating corrective actions (e.g., adjusting machine pressure).

- **Preventing Downtime**: By predicting issues before they cause significant downtime, this integrated system can reduce the frequency of unexpected breakdowns and keep the production line running smoothly.

3. Predictive Maintenance (PdM) Evolution

The integration of the Random Forest model moves Heineken’s maintenance strategy from being purely preventative to **predictive**. Over time, the maintenance process can shift away from scheduled mini-overhauls every two to three days toward a condition-based maintenance system where interventions occur only when the model identifies a need.

- **Proactive Maintenance**: The predictive model can forecast the remaining useful life of critical machine components, allowing for targeted interventions just before failure is likely. This helps to optimize resource use, reducing unnecessary part replacements and extending the life of components.

- **Reducing Downtime and Waste**: With accurate predictions of underfill incidents and equipment degradation, the system can minimize downtime associated with unexpected breakdowns. By addressing issues before they escalate, the production line experiences fewer disruptions, leading to reduced waste from rejected bottles and improved overall equipment effectiveness (OEE).

4. Gradual Implementation and Feedback Loop

To fully integrate the predictive model into the existing system, it would be important to implement it gradually, ensuring alignment with Heineken's operational processes. The steps include:

- **Pilot Testing**: Implementing the model in parallel with current preventive maintenance processes during a pilot phase. This would allow the team to evaluate its effectiveness and accuracy in predicting underfill events without fully disrupting the existing system.

- **Feedback Loop**: The model should be continuously refined based on data from actual production runs. As more data is collected from the B1200 filler, the model can be retrained and fine-tuned to improve prediction accuracy. This feedback loop ensures that the predictive system remains aligned with the evolving conditions of the production line.

4.3.3 Action Plan for Integration of the Random Forest Model

To effectively integrate the Random Forest model into the B1200 filling line, the following action plan outlines the necessary phases, tasks, durations, and expected outcomes:

Phase 1: Data Integration

- **Task**: Collaborate with the IT and data engineering teams to access historical and real-time data from the B1200 filling system. This involves identifying key data sources, such as sensor outputs, PLC data, and production logs.

- **Outcome**: Establish a robust data pipeline that feeds relevant data into the Random Forest model for processing. This pipeline will ensure that the model has access to high-quality data for accurate predictions.

Phase 2: Model Deployment

- **Task**: Deploy the Random Forest model within the production environment. This involves integrating the model with existing systems, ensuring it can access real-time data, and setting up necessary APIs or interfaces for data retrieval and prediction output.

- **Outcome**: The model becomes operational, providing real-time underfill predictions based on live data. This allows operators to take proactive measures in response to predicted underfill scenarios.

Phase 3: Dashboard Development

- **Task**: Develop a user-friendly monitoring dashboard that visualizes model predictions, alerts, and performance metrics for operators. This dashboard should be intuitive, allowing operators to quickly understand the state of the filling line and respond accordingly.

- **Outcome**: Operators can easily visualize data and receive alerts for potential underfill issues, enhancing their ability to monitor production quality and intervene when necessary.

Phase 4: Training and Change Management

- **Task**: Conduct training sessions for operators and maintenance staff to familiarize them with the new system. This training should emphasize the importance of predictive maintenance, how to interpret dashboard alerts, and the appropriate actions to take in response to predictions.

- **Outcome**: Staff are confident in using the new system and understand how to act on predictions made by the model. This enhances operational efficiency and promotes a culture of data-driven decision-making.

Phase 5: Monitoring and Continuous Improvement

- **Task**: Establish a feedback loop for continuous monitoring of the model's performance and accuracy. This includes regularly reviewing prediction outcomes, gathering user feedback on the dashboard's usability, and identifying areas for improvement.

- **Outcome**: Continuous improvement of the model based on real-world usage leads to more accurate predictions over time. This process ensures that the model adapts to changing production conditions and incorporates new data as it becomes available.

This action plan provides a structured approach to integrating the Random Forest model into the B1200 filling line at Heineken. By following these phases, Heineken can leverage predictive analytics to enhance operational efficiency, reduce waste, and improve product quality. The integration of real-time monitoring and data-driven insights will empower operators and maintenance teams to make informed decisions that optimize the filling process.

4.3.4 Practical Implications

The research presented offers valuable insights into enhancing the operational efficiency and quality control practices within the bottling industry, specifically focusing on the B1200 filler at Heineken. By effectively utilizing a Random Forest model for predicting underfill incidents, the study demonstrates significant advancements in maintenance.

One of the primary practical implications of this research is the improvement in predictive maintenance practices. By accurately predicting underfill events before they occur, companies can shift from reactive maintenance strategies to proactive, data-driven approaches. This transition minimizes unexpected downtime, reduces operational costs associated with equipment failures, and optimizes the use of resources. Maintenance teams can now anticipate and address potential issues based on real-time data insights, ensuring the continuous operation of the production line without compromising on product quality.

Furthermore, the integration of machine learning models into existing production systems promotes a culture of continuous improvement and innovation. Operators and managers can leverage advanced analytics to make informed decisions, optimize production processes, and respond swiftly to changing operational conditions. This promotes efficiency gains across the entire production chain, from filling operations to packaging and distribution.

In conclusion, the research underscores the transformative impact of predictive maintenance and quality control practices in the bottling industry. By leveraging advanced data analytics and machine learning, companies like Heineken can achieve operational excellence, reduce waste, and maintain a competitive edge in the global marketplace.

4.3.5 Comparison with Existing Research

Several studies have explored machine learning models for predictive maintenance in manufacturing. For instance, Zhang et al. (2019) developed a machine-learning model for pump failure prediction in bottling lines, achieving an accuracy of **85.2%** using sensor data. Similarly, Gęca, (2020) applied Random Forest in industrial predictive maintenance, reporting an average accuracy of **88.5%**. In comparison, the Random Forest model in this study achieved a training accuracy of **90.67%**, test accuracy of **90.21%**, and cross-validation accuracy of **90.89%**, demonstrating improved predictive performance. The higher accuracy can be attributed to the enriched dataset, which included synthetic data generated via cGAN, thereby addressing the issue of limited real-world data.

A growing body of research highlights the advantages of synthetic data in predictive modeling. Abassi et al. (2024) demonstrated how cGAN-generated data significantly improved fault detection accuracy in industrial systems. Similarly, Abood et al. (2023) integrated synthetic data with Convolutional Neural Networks (CNNs) to enhance predictive maintenance models. This study further validates the role of synthetic data by showing that cGAN-generated underfill data contributed to a more diverse and realistic dataset, improving model generalization and mitigating overfitting.

While predictive maintenance has been widely applied in industries such as aerospace and manufacturing (Tiddens et al., 2022), its application in the beverage industry remains limited.

Morelle et al. (2021) applied machine learning for underfill detection but primarily relied on historical sensor data without synthetic augmentation. This study extends their work by incorporating real-time monitoring and synthetic data to create a more robust predictive model.

This research advances the field of predictive maintenance in three keyways:

1. **Enhanced Model Performance:** By leveraging synthetic data, the predictive model demonstrated superior accuracy compared to traditional approaches.
2. **Application of cGAN in the Beverage Industry:** Unlike previous studies primarily focused on physical sensors or traditional machine learning, this study introduces cGAN-based synthetic data augmentation for predictive maintenance in the beverage sector.
3. **Practical Implementation Strategy:** The proposed model aligns with Heineken's operational goals by providing a structured integration plan, ensuring feasibility for real-world deployment.

4.3.6 Economic, Environmental, and Social Benefits of the Predictive Maintenance Model

Integrating the predictive maintenance model into the B1200 filler's packaging process addresses underfill issues while exemplifying a data-driven approach to operational excellence. This model not only enhances production efficiency but also promotes sustainability by aligning maintenance strategies with real-time data. The resulting benefits span environmental, economic, and social dimensions.

Economic Impacts

The economic benefits of the model are grounded in its potential to optimize the cost-efficiency of bottling operations:

- By replacing traditional preventive schedules with condition-based interventions, unnecessary maintenance actions are reduced. This leads to lower costs for spare parts, lubricants, and labor associated with routine checks that may not always align with equipment needs.

- Downtime due to unplanned failures is a key bottleneck in packaging operations. By preemptively addressing potential issues, the model ensures higher OEE (Overall Equipment Effectiveness), a critical KPI in industrial engineering. The economic gains from improved throughput and minimized stoppages are substantial.
- The model's ability to ensure consistent fill accuracy minimizes costs associated with non-conformance, including product recalls, rework, and potential loss of market reputation. This has a direct impact on maintaining customer trust while controlling operational expenses.

Environmental Impacts

The integration of the predictive maintenance model into the B1200 filler represents a targeted approach to environmental sustainability by addressing inefficiencies at the operational level:

- The model's ability to detect underfill events in advance reduces the rejection of defective bottles. This directly conserves raw materials such as water, CO2, and packaging components, which are critical inputs in the bottling process. By optimizing fill levels, resource utilization is maximized, minimizing waste disposal challenges.
- Enhanced machine reliability translates to fewer energy-intensive unplanned shutdowns and restarts. This not only reduces the plant's energy footprint but also aligns with industry goals for lower greenhouse gas emissions.
- Predictive maintenance prevents accelerated wear-and-tear by identifying degradation early, thereby extending the lifecycle of critical components and reducing the frequency of replacement parts. This reduces the indirect environmental burden associated with manufacturing spare parts.

Social Impacts

The model contributes to social value creation as follows:

- The integration of predictive analytics into SCADA systems provides operators with actionable insights into equipment performance. This reduces reliance on guesswork, improves decision-making capabilities, and enhances job satisfaction by allowing personnel to focus on value-adding tasks rather than firefighting breakdowns.
- Proactive identification of equipment anomalies reduces the likelihood of catastrophic failures, which can pose safety risks to employees working near high-speed filling lines.
- Adoption of data-driven, sustainable practices strengthens the organization’s position as an industry leader committed to social and environmental stewardship. This fosters goodwill among consumers and within the communities where the company operates.
-

Chapter Five

5. Conclusion and Recommendations

5.1 Conclusion

This research focused on developing a predictive maintenance model using Random Forest algorithms to address the pressing issue of underfilling in the B1200 filler line at Heineken. Through a systematic approach, the study aimed to enhance operational efficiency and product quality within the bottling process.

The methodology involved comprehensive data collection and preprocessing from various sources, including historical operational data and real-time sensor readings. Feature engineering techniques were applied to identify key variables that influence underfilling, ensuring the model was trained on relevant and informative data. The Random Forest model was selected for its ability to handle complex, non-linear relationships, which are typical in industrial applications.

The results of the model were promising, demonstrating a training accuracy of 90.67%, a test accuracy of 90.21%, and a cross-validation accuracy of 90.89%. These metrics indicate that the model is capable of generalizing well to unseen data, suggesting it can reliably identify underfill incidents under varying production conditions. Furthermore, the F1 scores of 89.71% on the training set and 89.03% on the test set reflect a strong balance between precision and recall, minimizing false positives and false negatives in predictions.

The successful implementation of this predictive maintenance model has significant implications for Heineken's production processes. By integrating the model into the existing B1200 filler system, operators can leverage real-time predictions to proactively address potential underfill issues before they escalate. This integration facilitates a shift from traditional reactive maintenance practices to a more efficient predictive maintenance strategy.

An action plan has been outlined for the seamless integration of the model, which includes phases such as data integration, model deployment, dashboard development, staff training, and ongoing monitoring for continuous improvement. This structured approach will ensure that the model is effectively utilized in a production environment, enhancing the overall operational efficiency and product quality.

In conclusion, the research not only addresses the specific problem of underfilling but also highlights the broader potential of predictive maintenance models in industrial settings. The findings underscore the value of data-driven decision-making in optimizing production processes and resource utilization. Future work could expand on these results by exploring the integration of additional machine learning algorithms and real-time sensor data, further enhancing the model's predictive capabilities and adaptability to changing production conditions.

5.2 Recommendations

Based on the findings of this research, it is recommended that Heineken implements continuous evaluation and updates of the Random Forest model to adapt to changes in production processes and machine behavior, ensuring its accuracy and reliability over time. Additionally, integrating more data sources, such as advanced sensor data and operational metrics, will enhance the model's predictive capabilities for underfill incidents. Investing in real-time monitoring technologies is crucial, as it will enable operators to respond promptly to potential issues, further optimizing the filling process.

Furthermore, extending the predictive maintenance framework to other critical processes within the production line can lead to significant improvements in overall efficiency and product quality. Ongoing training and support for staff are essential to ensure they are proficient in utilizing the new tools and understanding the insights generated by the model. Establishing a feedback loop among operators, maintenance teams, and data scientists will help refine the predictive maintenance process based on practical experience. Finally, further research into alternative machine learning techniques should be encouraged to identify additional methods that could enhance underfill prediction and other quality control applications. Implementing these recommendations will foster a proactive approach to quality assurance and operational excellence throughout Heineken's production processes.

REFERENCE

Abassi, A., Bakkas, B., Jai, M. El, Arid, A., & Benazza, H. (2024). A Multi-Split Cross-Strategy for Enhancing Machine Learning Algorithms Prediction Results with Data Generated by Conditional Generative Adversarial Network. *Journal of Computer Science*. https://api.semanticscholar.org/CorpusID:269034347

Abood, A. M., Nasser, A. R., & Al-Khazraji, H. (2023). *IAES International Journal of Artificial Intelligence (IJ-AI)*. https://api.semanticscholar.org/CorpusID:260595941

Achouch, M., Dimitrova, M. H., Ziane, K., Karganroudi, S. S., Dhouib, R., Ibrahim, H., & Adda, M. (2022). On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges. *Applied Sciences*. https://api.semanticscholar.org/CorpusID:251612466

Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., & Adda, M. (2022). On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges. In *Applied Sciences (Switzerland)* (Vol. 12, Issue 16). MDPI. https://doi.org/10.3390/app12168081

Adarsh, C. K., & Kulkarni, S. D. (2021). Process capability improvement of an eight-head liquid filling machine by the reduction in volume variation. *IOP Conference Series: Materials Science and Engineering*, *1033*. https://api.semanticscholar.org/CorpusID:234189670

Agostinelli, S., & Cumo, F. (2022). MACHINE LEARNING APPROACH FOR PREDICTIVE MAINTENANCE IN AN ADVANCED BUILDING MANAGEMENT SYSTEM. *WIT Transactions on Ecology and the Environment*. https://api.semanticscholar.org/CorpusID:250107811

Allah Bukhsh, Z., Saeed, A., Stipanovic, I., & Doree, A. G. (2019). Predictive maintenance using tree-based classification techniques: A case of railway switches. *Transportation Research Part C: Emerging Technologies*, *101*, 35–54. https://doi.org/https://doi.org/10.1016/j.trc.2019.02.001

Almamlook, R., Alshammaria, N., & Albalawi, O. (2022). Machine Learning for Predictive Maintenance in the Manufacturing Sector. *Proceedings of the International Conference on Industrial Engineering and Operations Management*. https://api.semanticscholar.org/CorpusID:258753650

Alvarez Quiñones, L. I., Lozano-Moncada, C. A., & Bravo Montenegro, D. A. (2023). Machine learning for predictive maintenance scheduling of distribution transformers. *Journal of Quality in Maintenance Engineering*, *29*(1), 188–202. https://doi.org/10.1108/JQME-06-2021-0052

Amruthnath, N., & Gupta, T. (2018). A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*, 355–361. https://doi.org/10.1109/IEA.2018.8387124

Andriani, A. Z., Kurniati, N., & Santosa, B. (n.d.). *Enabling Predictive Maintenance Using Machine Learning in Industrial Machines With Sensor Data*.

Anush, C., Yashwanth, K. L., Shashank, S., Reddy, M. V., & Kumar, A. (2021). Bottle Line Detection using Digital Image Processing with Machine Learning. *Journal of Physics: Conference Series*, *1998*. https://api.semanticscholar.org/CorpusID:237286733

Arena, F., Collotta, M., Luca, L., Ruggieri, M., & Termine, F. (2021). Predictive Maintenance in the Automotive Sector: A Literature Review. *Mathematical and Computational Applications*. https://api.semanticscholar.org/CorpusID:245721621

Ben, J., Mohamed, A., & Muduli, K. (2021a). Effect of Preventive Maintenance on Machine Reliability in a Beverage Packaging Plant. *Int. J. Syst. Dyn. Appl.*, *10*, 50–66. https://api.semanticscholar.org/CorpusID:234366108

Ben, J., Mohamed, A., & Muduli, K. (2021b). Effect of Preventive Maintenance on Machine Reliability in a Beverage Packaging Plant. *Int. J. Syst. Dyn. Appl.*, *10*, 50–66. https://api.semanticscholar.org/CorpusID:234366108

Cardoso, D. S. E., & Ferreira, L. (2020). Application of Predictive Maintenance Concepts Using Artificial Intelligence Tools. *Applied Sciences*. https://api.semanticscholar.org/CorpusID:234444779

Chazhoor, A., Mounika, Y., Sarobin, M. V. R., Sanjana, M. V, & Yasashvini, R. (2020). Predictive Maintenance using Machine Learning Based Classification Models. *IOP Conference Series: Materials Science and Engineering*, *954*(1), 12001. https://doi.org/10.1088/1757-899X/954/1/012001

Compare, M., Baraldi, P., & Zio, E. (2020). Challenges to IoT-Enabled Predictive Maintenance for Industry 4.0. *IEEE Internet of Things Journal*, *7*(5), 4585–4597. https://doi.org/10.1109/JIOT.2019.2957029

Costa, A. L. B., & Balduíno, Â. R. (2018). The Importance of Preventive and Corrective Maintenance in Works. *International Journal of Advanced Engineering Research and Science*, *5*, 237466. https://api.semanticscholar.org/CorpusID:169550624

Dasnoy, S., Simonin, L., Radulovic, S., White, A., Decoster, J.-F., & Denis, L. (2022). A Predictive Modeling Approach to Support the Overfill Volume Definition of Liquid-in-Vial Drug

Products. *PDA Journal of Pharmaceutical Science and Technology*, *76*, 384–394. https://api.semanticscholar.org/CorpusID:245963193

Esteban, A., Zafra, A., & Ventura, S. (2022). Data mining in predictive maintenance systems: A taxonomy and systematic review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*. https://api.semanticscholar.org/CorpusID:250005733

Ferraz, R., Da Silva, F. R. G., Fontes, C. E., Contessi, B. A., Campagnac, L. A., & Daoud, I. L. A. (2022). Simulation-based digital twin for predictive maintenace. *Rio Oil and Gas Expo and Conference*, *22*(2022), 315–316. https://doi.org/10.48072/2525-7579.rog.2022.315

Gawde, S., Patil, S., Kumar, S., Kamat, P., Kotecha, K., & Alfarhood, S. (2024). Explainable Predictive Maintenance of Rotating Machines Using LIME, SHAP, PDP, ICE. *IEEE Access*, *12*, 29345–29361. https://doi.org/10.1109/ACCESS.2024.3367110

Gęca, J. (2020). PERFORMANCE COMPARISON OF MACHINE LEARNING ALGORITHMS FOR PREDICTIVE MAINTENANCE. *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*. https://api.semanticscholar.org/CorpusID:222480748

Grubisic, V. V. F., Aguiar, J. P. F., & SIMEU-ABAZI, Z. (2020). A Review on Intelligent Predictive Maintenance: Bibliometric analysis and new research directions. *2020 International Conference on Control, Automation and Diagnosis (ICCAD)*, 1–6. https://api.semanticscholar.org/CorpusID:227221876

He, L., Levine, R. A., Fan, J., Beemer, J., & Stronach, J. (2018). Random Forest as a Predictive Analytics Alternative to Regression in Institutional Research. *Practical Assessment, Research and Evaluation*, *23*, 1. https://api.semanticscholar.org/CorpusID:158670739

Jernigan, D. H., & Ross, C. S. (2020). The Alcohol Marketing Landscape: Alcohol Industry Size, Structure, Strategies, and Public Health Responses. *Journal of Studies on Alcohol and Drugs. Supplement*, 13–25. https://api.semanticscholar.org/CorpusID:211233099

Jimenez, V. J., Bouhmala, N., & Gausdal, A. H. (2020). Developing a predictive maintenance model for vessel machinery. *Journal of Ocean Engineering and Science*, *5*(4), 358–386. https://doi.org/https://doi.org/10.1016/j.joes.2020.03.003

Kamat, P., & Dr., R. S. (2020). Anomaly Detection for Predictive Maintenance in Industry 4.0- A survey. *E3S Web of Conferences*. https://api.semanticscholar.org/CorpusID:219816281

Kane, A. P., Kore, A. S., Khandale, A. N., Nigade, S. S., & Joshi, P. (2022). Predictive Maintenance using Machine Learning. *ArXiv*, *abs/2205.09402*. https://api.semanticscholar.org/CorpusID:248887450

Kaparthi, S., & Bumblauskas, D. (2020). Designing predictive maintenance systems using decision tree-based machine learning techniques. *International Journal of Quality & Reliability Management*, *37*(4), 659–686. https://doi.org/10.1108/IJQRM-04-2019-0131

Kizito, R., Scruggs, P., Li, X., Kress, R., Devinney, M., & Berg, T. (n.d.). *The Application of Random Forest to Predictive Maintenance*.

Kulkarni, S., & Elango, M. (2016). *DEVELOPMENT OF PLC BASED CONTROLLER FOR BOTTLE FILLING MACHINE*. https://api.semanticscholar.org/CorpusID:212518575

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, *6*(4), 239–242. https://doi.org/10.1007/s12599-014-0334-4

Le, T., Luo, M., Zhou, J., & Chan, H. L. (2014). Predictive maintenance decision using statistical linear regression and kernel methods. *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, 1–6. https://doi.org/10.1109/ETFA.2014.7005357

Lee, S. M., Lee, D., & Kim, Y. S. (2019). The quality management ecosystem for predictive maintenance in the Industry 4.0 era. *International Journal of Quality Innovation*, *5*(1), 4. https://doi.org/10.1186/s40887-019-0029-5

Ma, H., Geng, M., Wang, F., Zheng, W., Ai, Y., & Zhang, W. (2024). Data Augmentation of a Corrosion Dataset for Defect Growth Prediction of Pipelines Using Conditional Tabular Generative Adversarial Networks. *Materials*, *17*. https://api.semanticscholar.org/CorpusID:268165287

Meddaoui, A., Hain, M., & Hachmoud, A. (2023). The benefits of predictive maintenance in manufacturing excellence: a case study to establish reliable methods for predicting failures. *The International Journal of Advanced Manufacturing Technology*, *128*(7), 3685–3690. https://doi.org/10.1007/s00170-023-12086-6

Merkt, O. (2019). Predictive Models for Maintenance Optimization: An Analytical Literature Survey of Industrial Maintenance Strategies. *AITM/ISM@FedCSIS*. https://api.semanticscholar.org/CorpusID:215878595

Meyer Zu Wickern, V. (2019). *Challenges and Reliability of Predictive Maintenance*. https://doi.org/10.13140/RG.2.2.35379.89129

Mobley, R. K. (2002). 4 - Benefits of Predictive Maintenance. In R. K. Mobley (Ed.), *An Introduction to Predictive Maintenance (Second Edition)* (pp. 60–73). Butterworth-Heinemann. https://doi.org/https://doi.org/10.1016/B978-075067531-4/50004-X

Morelle, E., Rudolph, A., McHardy, C., & Rauh, C. (2021). Detection and prediction of foam evolution during the bottling of noncarbonated beverages using artificial neural networks. *Food and Bioproducts Processing*, *128*, 63–76. https://doi.org/https://doi.org/10.1016/j.fbp.2021.03.017

Motaghare, O., Pillai, A. S., & Ramachandran, K. I. (2018). Predictive Maintenance Architecture. *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 1–4. https://api.semanticscholar.org/CorpusID:199440442

Mrugalska, B., Zasada, B., & Wyrwicka, M. K. (2018). Preventive Approach to Machinery and Equipment Maintenance in Manufacturing Companies. *Advances in Manufacturing, Production Management and Process Control*. https://api.semanticscholar.org/CorpusID:169056468

Mulders, M., & Haarman, M. (2018). *Predictive Maintenance 4 . 0 Beyond the hype : PdM 4 . 0 delivers results September 2018*. https://api.semanticscholar.org/CorpusID:208653932

Nacchia, M., Fruggiero, F., Lambiase, A., & Bruton, K. (2021). A Systematic Mapping of the Advancing Use of Machine Learning Techniques for Predictive Maintenance in the Manufacturing Sector. *Applied Sciences*. https://api.semanticscholar.org/CorpusID:233669331

Nashrudin, M. N., Abas, M. A., Abdullah, M. Z., Ali, M. Y. T., & Samsudin, Z. (2020). Study of Different Dispensing Patterns of No-Flow Underfill Using Numerical and Experimental Methods. *Journal of Electronic Packaging*. https://api.semanticscholar.org/CorpusID:229461771

Nentwich, C., Benker, M., Ellinger, J., Zhai, S., Kleinwort, R., Reinhart, G., & Zäh, F. M. (2020). *Predictive Maintenance in der Produktion/Predictive Maintenance within the industrial value chain*. https://api.semanticscholar.org/CorpusID:234989154

Nowlan, F. S., & Heap, H. F. (1978). *Reliability-Centered Maintenance*.

Paolanti, M., Romeo, L., Felicetti, A., Mancini, A., Frontoni, E., & Loncarski, J. (2018). Machine Learning approach for Predictive Maintenance in Industry 4.0. *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, 1–6. https://doi.org/10.1109/MESA.2018.8449150

Pavlović, A., Fragassa, C., Vegliò, L., de Camargo, F. V., & Minak, G. (2020). Modeling Palletized Products: The Case of Semi-Filled Bottles under Top-Load Conditions. *Applied Sciences*. https://api.semanticscholar.org/CorpusID:213572629

Pithadiya, K. J., Nagar, V. V, Modi, C. K., & Chauhan, J. D. (2011). *Selecting the Most Favourable Edge Detection Technique for Liquid Level Inspection in Bottles*. https://api.semanticscholar.org/CorpusID:14469295

Ran, Y., Zhou, X., Lin, P., Wen, Y., & Deng, R. (2019). *A Survey of Predictive Maintenance: Systems, Purposes and Approaches*. http://arxiv.org/abs/1912.07383

Ranasinghe, G. D., & Parlikad, A. K. (2019). Generating Real-valued Failure Data for Prognostics Under the Conditions of Limited Data Availability. *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 1–8. https://api.semanticscholar.org/CorpusID:201814645

Rauch, E., Linder, C., & Dallasega, P. (2020). Anthropocentric perspective of production before and within Industry 4.0. *Computers & Industrial Engineering*, *139*, 105644. https://doi.org/https://doi.org/10.1016/j.cie.2019.01.018

Ren, Y. (2021). *Optimizing Predictive Maintenance With Machine Learning for Reliability Improvement*. https://api.semanticscholar.org/CorpusID:234271963

Ross, P., & Maynard, K. (2021a). Towards a 4th industrial revolution. *Intelligent Buildings International*, *13*(3), 159–161. https://doi.org/10.1080/17508975.2021.1873625

Ross, P., & Maynard, K. (2021b). Towards a 4th industrial revolution. *Intelligent Buildings International*, *13*(3), 159–161. https://doi.org/10.1080/17508975.2021.1873625

Sajid, S., Haleem, A., Bahl, S., Javaid, M., Goyal, T., & Mittal, M. (2021). Data science applications for predictive maintenance and materials science in context to Industry 4.0. *Materials Today: Proceedings*, *45*, 4898–4905. https://doi.org/https://doi.org/10.1016/j.matpr.2021.01.357

Schmidt, B., & Wang, L. (2015). *Predictive Maintenance : Literature Review and Future Trends*. https://api.semanticscholar.org/CorpusID:106821367

Shivhare, N., Rahul, A. K., Dwivedi, S. B., & Dikshit, P. K. S. (2021). ARIMA based daily weather forecasting tool: A case study for Varanasi. *MAUSAM*. https://api.semanticscholar.org/CorpusID:244078809

Singh, P. (2019). Supervised Machine Learning. *Learn PySpark*. https://api.semanticscholar.org/CorpusID:202780184

Swinnen, J. (2017). Beer Consumption and Trade in a Era of Economic Growth and Globalization. *Choices. The Magazine of Food, Farm, and Resources Issues*, *32*. https://api.semanticscholar.org/CorpusID:198668017

Tercan, H., & Meisen, T. (2022). Machine learning and deep learning based predictive quality in manufacturing: a systematic review. *Journal of Intelligent Manufacturing*, *33*(7), 1879–1905. https://doi.org/10.1007/s10845-022-01963-8

Tiddens, W., Braaksma, J., & Tinga, T. (2022). Exploring predictive maintenance applications in industry. *Journal of Quality in Maintenance Engineering*, *28*(1), 68–85. https://doi.org/10.1108/JQME-05-2020-0029

Török, Á., Szerletics, Á., & Jantyik, L. (2020). Factors Influencing Competitiveness in the Global Beer Trade. *Sustainability*, *12*(15). https://doi.org/10.3390/su12155957

Tran, N. T., Trieu, H. T., Tran, V. T., Ngo, H. H., & Dao, Q. K. (2021). An overview of the application of machine learning in predictive maintenance. *Petrovietnam Journal*, *10*, 47–61. https://doi.org/10.47800/PVJ.2021.10-05

Trần, N. T., Triệu, H. T., Trần, V. T., Ngô, H. H., & Đào, Q. K. (2021). An overview of the application of machine learning in predictive maintenance. *Petrovietnam Journal*. https://api.semanticscholar.org/CorpusID:244845887

Voronov, S., Jung, D. E., & Frisk, E. (2021). A forest-based algorithm for selecting informative variables using Variable Depth Distribution. *Eng. Appl. Artif. Intell.*, *97*, 104073. https://api.semanticscholar.org/CorpusID:229345180

Wang, M., He, J., Feng, Z., & Xiao, W. (2023). An Online Predicting Method for Product Quality Based on Energy Consumption Data Using Random Forest. *2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP)*, 977–980. https://api.semanticscholar.org/CorpusID:262075748

Wang, Y., Deng, C., Wu, J., Wang, Y., & Xiong, Y. (2014). A corrective maintenance scheme for engineering equipment. *Engineering Failure Analysis*, *36*, 269–283. https://api.semanticscholar.org/CorpusID:18221652

Yang, F., & Ren, J. (2020). Reliability Analysis Based on Optimization Random Forest Model and MCMC. *Computer Modeling in Engineering & Sciences*. https://api.semanticscholar.org/CorpusID:243146260

Zhang, W., Yang, D., & Wang, H. (2019). Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. *IEEE Systems Journal*, *13*(3), 2213–2227. https://doi.org/10.1109/JSYST.2019.2905565

APPENDIX

I. Understanding The Basic Working Parameters

Brand - Heineken

Line B1200

Understanding the basic condition

		Current	Standard/ Method	What Has been Done	Set point/standard	Remark
Loss point	Vent tube					
8	Vent tube height	ok	Yes(I5)	-	115mm	
8	Straigntness	Not ok	Yes(I5)	Replace	Stright	Inspection tool in place
8	Vent tube spreader	Not ok	Yes(I5)	Cut	Not worn /cut	CILT checklist in place(Two monthly)
8	Gas escape hole diameter	ok	No(CO)	-	4mm	Added
8	Surface finish	Not ok	Yes(I5)	-	Smooth	CILT checklist in place
8	O-ring	ok	Yes(I5)	-	Avaliablity	To be added

I	Inspection standard
Q	Quality standard
C	Cleaning standard
T	Tigtning standard
CO	Changeover standard
M	Maintainace

		Centering bell cap seal	Not ok	Yes(I23)	Worn	Not worn/cut	Not replaced according to the yearly service time
	10	Handling part	ok	Yes(I3)	-	Proper fitting	Changeover/PM
	10	Breakshoe	ok	Yes(I3)	-	Not worn	Weekly PM
	10	Transfer plate	ok	Yes(C13)	-	Proper adjustment	
		Valve block					
	7	Vacuume channal	-0.67	Yes(Y)	-	No air leakage/unblocked (-1)	Yearly
	7	Snifting channal	ok	Yes(Y)	-	No air leakage/unblocked	Yearly
	2	Supply pressure	5 bar	Yes(M)	Adjusted	6bar	
	2	Lifting cylinder air pressure	ok	Yes(I25)	-	>=3.2 bar	

II. The Building Code For The Machine Learning

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split, GridSearchCV,
StratifiedKFold, cross_val_score
from sklearn.metrics import accuracy_score, f1_score, confusion_matrix,
ConfusionMatrixDisplay
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import Pipeline
from sklearn import RandomForestClassifier
import joblib
import matplotlib.pyplot as plt

# Load the synthetic data from the provided file
file_path = '/content/generated_synthetic_dataOJN.csv'
synthetic_data_df = pd.read_csv(file_path)

# Convert 'Vent-Tube' to numerical values if they are categorical
if 'Vent-Tube' in synthetic_data_df.columns:
    synthetic_data_df['Vent-Tube'] = synthetic_data_df['Vent-
Tube'].map({'Ok': 1, 'Not Ok': 0, 'OK': 1, 'Not OK': 0})

# Convert 'Classification' to numerical values if they are categorical
if 'Classification' in synthetic_data_df.columns:
    synthetic_data_df['Classification'] =
synthetic_data_df['Classification'].map({
        'No-underfill-or-overfill': 0,
        'Underfill': 1,
        'Overfill': 2
    })

# Ensure all columns are numeric
for col in synthetic_data_df.columns:
    if synthetic_data_df[col].dtype == 'object':
        synthetic_data_df[col] = pd.to_numeric(synthetic_data_df[col],
errors='coerce')

# Handle missing values in 'Vent-Tube' column
imputer = SimpleImputer(strategy='most_frequent')
synthetic_data_df['Vent-Tube'] =
imputer.fit_transform(synthetic_data_df[['Vent-Tube']])
```

```
# Drop any remaining rows with NaN values
synthetic_data_df.dropna(inplace=True)

# Drop 'Vent-Tube' column as it has zero importance
synthetic_data_df.drop(columns=['Vent-Tube'], inplace=True)

# Separate features and target variable
X = synthetic_data_df.drop(columns=['Classification'])
y = synthetic_data_df['Classification']

# Save the feature names before standardizing
feature_names = X.columns.tolist()

# Standardize the features
scaler = StandardScaler()
X = scaler.fit_transform(X)

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42, stratify=y)

import seaborn as sns
import pandas as pd

# Train a Random Forest Classifier
clf = RandomForestClassifier(n_estimators=100, random_state=42)
clf.fit(X_train, y_train)

# Define the parameter grid
param_dist = {
    'n_estimators': [100, 200, 300, 400, 500],
    'max_features': ['auto', 'sqrt', 'log2'],
    'max_depth': [10, 20, 30, 40, 50, None],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4],
    'bootstrap': [True, False]
}

# Initialize RandomForestClassifier
clf = RandomForestClassifier(random_state=42)

# Initialize RandomizedSearchCV
random_search = RandomizedSearchCV(clf, param_distributions=param_dist,
```

```
                                   n_iter=100, cv=3, verbose=2,
random_state=42, n_jobs=-1)

# Fit RandomizedSearchCV
random_search.fit(X_train, y_train)

# Make predictions
y_pred = clf.predict(X_test)

# Evaluate the model
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))

# Calculate accuracy
train_accuracy = accuracy_score(y_train, y_train_pred)
test_accuracy = accuracy_score(y_test, y_test_pred)

# Calculate F1 scores
train_f1 = f1_score(y_train, y_train_pred, average='weighted')
test_f1 = f1_score(y_test, y_test_pred, average='weighted')

print(f"Training accuracy on preprocessed hybrid data: {train_accuracy}")
print(f"Test accuracy on preprocessed hybrid data: {test_accuracy}")
print(f"Cross-validation accuracy: {cv_scores.mean()}")
print(f"Training F1 score on preprocessed hybrid data: {train_f1}")
print(f"Test F1 score on preprocessed hybrid data: {test_f1}")

# Compare training and test accuracies to identify overfitting
if train_accuracy > test_accuracy:
    print("The model is likely overfitting the training data.")
else:
    print("The model is not overfitting.")

# Calculate the confusion matrix without normalization
cm = confusion_matrix(y_test, y_test_pred)

# Calculate the confusion matrix with normalization
cm_normalized = confusion_matrix(y_test, y_test_pred, normalize='true')

# Plot the confusion matrix without normalization
fig, ax = plt.subplots(1, 2, figsize=(14, 6))

disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=['No-
underfill-or-overfill', 'Underfill', 'Overfill'])
disp.plot(ax=ax[0], cmap=plt.cm.Blues)
```

```
ax[0].set_title('Confusion Matrix (without normalization)')

# Plot the confusion matrix with normalization
disp_normalized = ConfusionMatrixDisplay(confusion_matrix=cm_normalized,
display_labels=['No-underfill-or-overfill', 'Underfill', 'Overfill'])
disp_normalized.plot(ax=ax[1], cmap=plt.cm.Blues)
ax[1].set_title('Confusion Matrix (with normalization)')

plt.tight_layout()
plt.show()

# Print the confusion matrix without normalization
print("Confusion Matrix (without normalization):")
print(cm)

# Print the confusion matrix with normalization
print("Confusion Matrix (with normalization):")
print(cm_normalized)

# Plotting the accuracy and F1 score comparison
metrics = [train_accuracy, test_accuracy, cv_scores.mean(), train_f1,
test_f1]
labels = ['Training Accuracy', 'Test Accuracy', 'Cross-Validation
Accuracy', 'Training F1 Score', 'Test F1 Score']

plt.figure(figsize=(12, 6))
plt.bar(labels, metrics, color=['blue', 'green', 'red', 'cyan',
'magenta'])
plt.ylim(0, 1)
plt.ylabel('Score')
plt.title('Model Accuracy and F1 Score Comparison')
plt.show()

# Plotting the heatmap
corr_matrix = data_numeric.corr()
corr_matrix = data.corr()
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt='.2f',
linewidths=0.5)
plt.title('Correlation Heatmap')
plt.show()
```