Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/AF07.20260003
Published: May 13, 2026

Semi-Automating Quality Testing Using Machine Learning: A Case Study of a South African FMCG Company

Mbidzo Ladzani and Gbeminiyi John Oyewole
School of Mechanical, Industrial, and Aeronautical Engineering
University of the Witwatersrand
Johannesburg, South Africa
mbidzodotladzani@gmail.com, gbeminiyi.oyewole@wits.ac.za

Abstract

Company X is facing the challenge of remaining competitive in an increasingly challenging economic environment. One way of doing this is to increase operational efficiency in manufacturing, to increase resource utilization and reduce waste. The quality testing process was identified as an opportunity area for semi-automation to improve the utilization of one of Company X's plants, and the purpose of this research was to investigate if machine learning would be effective in achieving this. Using historic data of batches produced in the plant, machine learning models were developed to predict the viscosity of batches made, to reduce the amount of time spent on testing and adjusting the product to achieve results that are within specification. Regression models were successful in achieving this, and the opportunity exists for further study through gathering more end-to-end data of the manufacturing process to refine the machine learning models even further.

Keywords
Quality testing, manufacturing, machine learning, regression models, classification models.

1. Introduction and Background

1.1. Introduction

Manufacturing companies are facing the challenge of trying to achieve profitability in a high-cost environment and are thus looking for ways to restore manufacturing competitiveness (Mirzaei, et al., 2021). In South Africa, the impact of the 2007-2008 global financial crisis (GFC) on consumers has led to limited disposable income and higher rates of unemployment. This crisis directly impacted their ability to afford manufactured goods and has led to reduced sales and revenue for local and international manufacturers. There has not been much improvement in the post-pandemic economic climate. Additional challenges for the South African manufacturing sector include load shedding, rising oil prices, and the risk of diminishing investment opportunities due to the country's recent placement on the Financial Action Task Force (FATF) Grey List (Makoni & Chikobvu, 2023). These challenges place an urgent mandate on manufacturing companies to find ways to increase their profits while remaining competitive in the market.

Operational efficiency has been found to be a key component in responding to increased business competition and can lead to increased sales and improved profit margins (Handoyo, et al., 2023). In a study of factors affecting operational efficiency, Dilshani, et al., (2019) posit that organizations need to fully utilize resources and achieve maximum productivity to maximize profits. In a manufacturing context, this productivity is measured for equipment using a quantitative measurement known as Overall Equipment Effectiveness (OEE). It is known as a 'best practice' measure for monitoring the effectiveness of manufacturing processes, and looks at availability, performance, and quality. A process that runs at 100% OEE is one that produces at its full design capacity with no defects (Singh, et al., 2018).

One way of improving operational efficiency is by applying lean manufacturing principles, a methodology that has been gaining prominence over the past two decades. Lean manufacturing focuses on eliminating waste to focus on value-adding activities. This methodology classifies waste into seven categories, namely, transportation, too much inventory, unnecessary motion, waiting, overproduction, overprocessing and defects (Kaneku-Orbegozo, et al., 2019). In addition to lean manufacturing, digitalisation also presents multiple opportunities for optimising manufacturing processes through increasing automation levels and a lower dependence on humans, thereby reducing variability in the process. By implementing lean manufacturing principles to identify the optimal activities to automate, digitalisation efforts can create more value (Buer, et al., 2021). The only time an equipment can be considered to be adding value, is when it is processing a part/material resource (Lee & Johnson, 2012), thus the aim should be to use digitalisation efforts to increase machine productive time.

1.2 Background and Research Objectives

Company X is a multinational Fast-Moving Consumer Goods (FMCG) company known for market-leading brands in household cleaning and laundry products, personal care products and food. With six manufacturing plants in South Africa, it is not exempt to the economic challenges prompted by the decreased spending power of consumers and has thus identified the need to reduce operational costs in its manufacturing plants to increase its gross margin without continuous price increases. Operational costs in their manufacturing plants consist of multiple factors, including but not limited to direct and indirect labour, factory overheads, raw material, and packaging material. Improving the efficiency of a plant and reducing waste produced in the manufacturing process has a substantial impact on reducing operational costs, as a higher efficiency means that more volume can be produced in a shorter period (thereby reducing labour and overhead costs), and there is less expenditure of raw and packaging materials.

One of Company X's manufacturing plants, named Plant A, produces cleaning products and there exists an opportunity for improvement by increasing efficiency and capacity utilization in the processing section of the plant. Plant A consists of four product categories, three of which are batch-manufactured processes and one continuous process. For each of the batch processes, quality inspections are conducted after each batch is made. These quality inspections ensure that each batch adheres to critical quality parameters specifications, namely viscosity, pH, available chlorine and specific gravity, and batches found outside of the specification limits are failed and disposed. The manufacturing process has the capability to produce 34 batches per day, and the process is stopped after the production of each batch for quality testing which takes a minimum of 15 minutes, consuming what could be eight hours of production time daily and possibly resulting in the rejection of 20 tons of semi-finished product should deviations be found. Therefore, a method that could reduce this quality inspection time and could still largely assure the quality of the batches would reduce the waiting time and increase the production time available. Machine learning could be a suitable method to improve the efficiency of the batch quality inspections process and, by extension, improve the efficiency of the processing section by increasing its capacity utilisation. Thus, this study set out to ascertain the effectiveness of machine learning in this context.

The critical research question that this study sought to address was: "Can machine learning enable semi-automation of the quality testing process?"
Based on this question, the primary objective of this research project was to determine the effectiveness of machine learning in predicting a batch's output quality parameters. The secondary objectives were to: (i) Determine what quality parameters would be sufficiently useful in predicting quality results, and (ii) Evaluate the improvement in the quality testing process in manufacturing.

The scope of this study was limited to one production process in Plant A. Furthermore, the quality results that the machine learning models were developed to predict were limited to one quality parameter to allow for a detailed exploration of that parameter. The assumption was that the same methodology could be followed for the other quality parameters. The study aimed to evaluate the effectiveness of the machine learning models theoretically, by calculating accuracy using the relevant metrics, and did not assess the effectiveness of the machine learning models in practice due to a lack of the necessary data architecture to host the models.

2. Literature Review

2.1 Quality Control in Manufacturing

The term 'quality' is used to refer to the conformance of a product or service to specifications and is a key component to customer satisfaction and product success (Mrugalska & Tytyk, 2015). One method of assessing product quality attributes during the production process is acceptance sampling. This method uses the information from a sample taken during manufacturing to determine whether a lot is acceptable or not. It allows manufacturing companies to reduce the quantity of product destroyed for inspection and testing, while increasing the frequency and effectiveness of testing (Judi, et al., 2011).Statistical Process Control (SPC) is another quality control tool that focuses on early detection and prevention of problems. It differs from acceptance sampling in that it allows for detection and correction in advance, where the latter detects issues after they have occurred. (Madanhire & Mbohwa, 2016).

2.2 Machine Learning and Quality Control

While SPC enables monitoring of process behaviour and identifying deviations, machine learning can be used for identifying patterns and for early detection of process deviations using time-series data of both process and product variables. This allows for preventative measures to be taken and for the stabilization of the production process (Gokalp, et al., 2016). A case study by Msakni, et al., (2023) investigates using machine learning to predict product quality, focussing on bumper beams. The goal of the case study was to realise an improvement in the quality control process through prediction of the quality of future products, enabling early adjustments, and a reduction in scrap (waste) and production downtime. Neural networks and random forests were used as the algorithms for the study and were trained using historical data of previously produced and measured parts.

In a review of the applicability of machine learning in quality control within manufacturing, Deshmukh, et al., (2022) detail a case study in which machine learning was used for quality assessment within a drilling procedure. In their study, sensors were used to obtain real-time values of the input parameters such as spindle speed and the depth of the cut, which are fused to find a relationship between the input and the output parameters such as Tolerance of the hole and surface roughness. Machine learning is thus used to determine if the input parameters are within an acceptable range to provide the required output, and if they are not the algorithm then works to change the input and give an instruction to actuators.

Arboretti, et al., (2023) conducted a similar experiment to determine the applicability of machine learning in detecting defects in the production of a fashion accessory made out of acetate. In their study, product and process features were used as inputs to the machine learning model. The product features used in the assessment of each batch were colour, design, and dimension, while the process features are materials used, process start and end time, batch size (quantity of products), and working time per step. The model provided the quality result as binary and made use of multiple classification algorithms such as logistic regression, neural networks, random forest, and support vector machines. The random forest algorithm emerged as the best one with the ability to minimize the rate of false negatives and provided robust results, achieving the optimal classification threshold of 0.2 (used to determine the likelihood of a new batch containing defects).

2.3 Machine Learning

Machine learning aims to learn patterns and relationships in data automatically through observations and examples (Bishop, 2006). Where artificial intelligence aims to automate intellectual tasks ordinarily performed by humans, machine learning is one of the specific methods used to do this. It differs from classical programming in that classical programming requires a dataset and an algorithm to create outputs, whereas machine learning requires a dataset and the associated outputs. It then creates the algorithm describing the relationship between inputs and outputs through the learning process, and the algorithm can then be used on future datasets (Choi, et al., 2020).

As noted by Janiesch, et al. (2021) machine learning techniques could be classified as supervised learning, unsupervised learning and reinforcement learning. (Kumar & Garg, 2018) noted that in supervised learning (regression or classification based), models are trained with datasets having inputs and corresponding outputs before predicting the target or output variables for new input variables. For unsupervised learning models (such as clustering, anomaly detection), the learning system detects patterns without any defined labels or target outputs. Reinforcement learning systems are not provided with inputs and outputs but are instead given a description of the current state of the system, a goal, a list of allowable or permitted actions, and their environmental constraints. The model then achieves the goal by itself, using the principle of trial and error with the aim of maximizing a reward or minimizing a risk (Kumar &

Garg, 2018). According to (Celik, 2018), machine learning algorithms commonly used are Artificial Neural Networks, Decision Trees, Support Vector Machines, Naïve Bayes, Logistic Regression and the K-Nearest Neighbour (KNN). We refer readers to their work for a comprehensive review of the working procedures of these algorithms. In order to improve the learning accuracy and reducing the learning time, feature selection must be adequately done. (Zhao, et al., 2010) indicates this as process of retrieving a subset from an original dataset according to specific criteria, with the aim of selecting only the relevant features from the dataset. Irrelevant and redundant features are removed, thus reducing the scale of the data to be processed.

Furthermore, in conducting a good machine learning study, Pruneski, et al. (2022) posits the following methodology: Problem identification (defining a problem and ask an appropriate question to determine why this problem must be solved), data collection and analysis (getting data that is sufficient, relevant and usable for decision making) , data splitting (split into training, validation and testing datasets such as random split of 70-80% for training, and 20-30% for testing (Figure 1). A subset of 10-20% of the training data is then set aside for validation), data exploration and preparation (understanding the data features and predictors and discovering correlations and redundancies), model development (addressing issues of underfitting and overfitting with training datasets and using appropriate validation techniques such as k-fold validation) model tuning and evaluation (hyper- parameter tunning of promising models using methods such as grid search with test data), Model deployment and monitoring (addressing how to make the model usable for intended purpose and also updating with new data for improved performance).

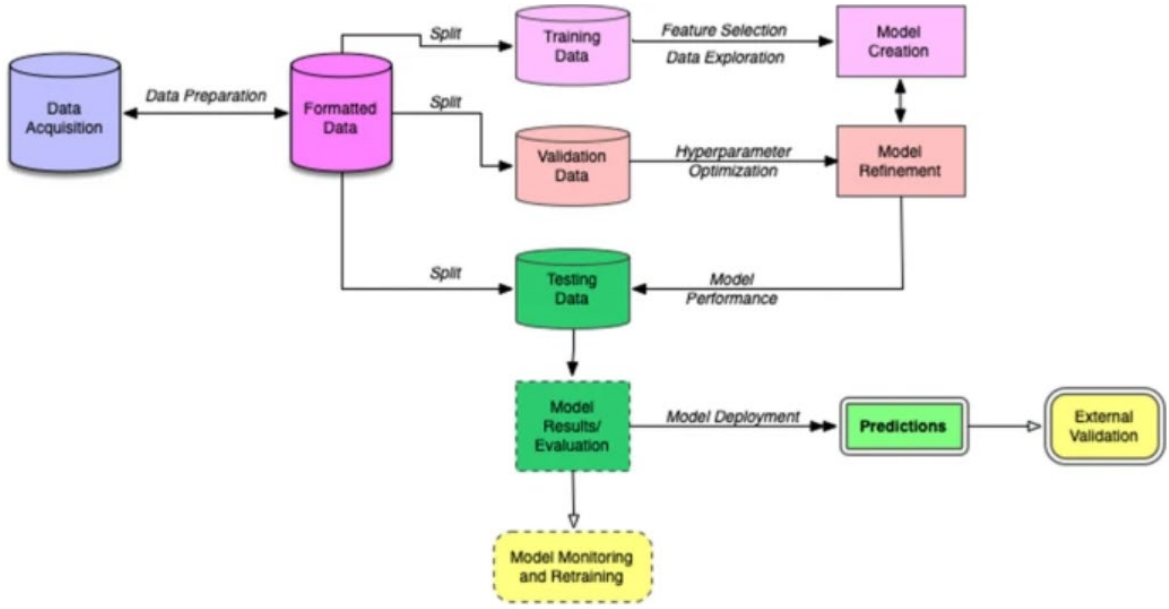

Figure 1. Machine learning model development process (Pruneski, et al., 2022)

2.4 Managing Imbalance in Data

Imbalanced learning occurs when there is one or more classes that have very little data, and others with adequate data, as in the case of fraud detection and medical diagnosis. Standard classification algorithms are designed to minimize error regardless of an imbalance in data, and in so doing they often sacrifice the minority class accuracy.

There are three approaches to managing data imbalance, namely, the algorithm level approach, the cost sensitive approach and the data level approach. The algorithm level approach modifies specifics in classification algorithms to account for the imbalance such as the K-Nearest Neighbour classifier, decision trees or support vector machines. These modifications are done in a way that focusses on the minority classes. The cost sensitive approach sets costs to misclassifications based on the degree of imbalance and the importance of the class as per a defined cost matrix.

The data level approach amends the data itself to rebalance the minority and majority classes by either under-sampling or over-sampling. Under-sampling refers to removing some majority class instances while over-sampling increases instances of the minority class (Elreedy & Atiya, 2019).

Managing data imbalance can be done through over-sampling with replacement, however this has been found to not yield a significant improvement in recognition of the minority class (Ling & Li, 1998). Synthetic Minority Oversampling Technique (SMOTE) is an approach that oversamples by creating synthetic, or artificial, instances. The minority class is over-sampled by taking each instance of the minority class and introducing synthetic examples along the line segments that join any or all of the k-nearest neighbours of the minority class (Chawla, et al., 2002).

2.5 Classification and Regression Model Evaluation

The validation of a machine learning model is done through the evaluation of its prediction performance. Metrics enable a quantification of the prediction performance and differ for classification and regression models.
The results of classification models can be summarized in a matrix, referred to as a confusion matrix. For binary classification, the confusion matrix classifies the samples into one of four categories depending on their true and predicted values. The metrics are useful when providing a global assessment of the performance of a model, as when selecting the "best model". However, they can be misleading when used in isolation. Thus, when assessing a classifier, all individual metrics should be employed, no single summary metric should be used in isolation and the prevalence in the target population must be considered (Varoquaux & Colliot, 2023). Table 1 shows some of the performance metrics used for evaluating the performance of classification models.

Table 1. Examples of performance metrics for evaluating binary classification models

Performance metrics	Formula	Description
Recall	$\frac{TP}{TP + FN}$	The fraction of positive samples that have been retrieved
***Specificity*:**	$\frac{TN}{TN + FP}$	The fraction of negative samples that were classified as negative
Precision (also known as Positive Predictive Value or PPV	$\frac{TP}{TP + FP}$	A fraction of the samples classified as positive that are actually positive.
Negative Predictive Value (NPV)	$\frac{TN}{TN + FN}$	A fraction of the samples classified as negative that are actually negative.
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$	A fraction of the samples that have been classified correctly
Balanced Accuracy	$\frac{Recall + Specificity}{2}$	A metric for accuracy that accounts for unbalanced samples
F1 Score	$\frac{2TP}{2TP + FP + FN}$	The harmonic mean of the recall and precision

In binary classification (1 and 0) or (yes and no), TP represents True Positives, samples for which both the true and predicted values are 1. TN represents True Negatives samples for which both the true and the predicted values are 0. FP implies False Positives for samples for which the predicted value is 1 while the true value is 0. False Negatives (FN) for samples which the predicted value is 0 while the true value is 1.
In addition, a summary of the performance of classification models could be shown using charts such as the area under the receiver operating characteristic curve (ROC), area under the precision-recall curve (AUC) (also known as average precision).

Bruce, et al. (2020) and Witten, et al. (2016) provide methods to evaluate the performance of regression models, by using metrics that are used to calculate the error size. The two most commonly used metrics for regression models are Root Mean Squared Error (RMSE) and Mean Squared Error (MSE). RMSE, often used together with Mean Absolute Error (MAE), is the selected metric used in multiple research papers that compare the performance of multiple models.

For example, the *Mean Squared Error (MSE)*:

$$\frac{1}{n}\sum_{i=1}^{n}(p_i - a_i)^2$$

This metric squares errors to exaggerate them. MSE is useful for cases when larger errors are not desirable (Li & Shi, 2010).

Root Mean Squared Error (RMSE):

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - a_i)^2}$$

This is the square root of MSE, which ensures that the error and the predicted value have the same dimension. Both MSE and RMSE are more sensitive to error variance when compared to MAE.

Mean Absolute Error (MAE):

$$\frac{1}{n}\sum_{i=1}^{n}|p_i - a_i|$$

This metric handles all error sizes equally based on their magnitude as opposed to squaring errors. To avoid cancellation of positive and negative errors, errors are kept as an absolute value. However, this reduces the metrics ability to determine bias (Botchkarev, 2019).

2.6 Summary and Research Gap

Machine learning is one of the tools used in artificial intelligence and can also be used in predictive analytics as it uses historical data to predict future performance or outcomes. Machine learning algorithms work by using a training dataset and identifying patterns and relationships within the data to develop algorithms that can then be used for future, unseen data. The format of the training dataset, in terms of labelling, content and structure, determine whether the machine learning methods to be used will be supervised, unsupervised, semi-supervised or reinforcement learning. Each method has associated algorithms, and the algorithm can be selected based on the desired output (regression or classification), the computational resources required and the training dataset available.

Machine learning has been applied in various fields, ranging from Economics and Marketing to the medical field. However, few case studies of the application of machine learning to improving quality in the manufacturing industry are available. The case studies that were found indicated positive results, and the study conducted by Arboretti, et al., (2023) which showed how machine learning could be used to predict whether a batch will be defective or defect-free suggested that this would be applicable to this study as well. Based on the requirements stated in ISO standards 2958-1 and 2859-3, the ability to accurately predict the outcome of a batch (referred to as lots in the standards) can be grounds to reduce the size and frequency of sampling, thus the theory suggests that the objectives set in this study can be achieved and the quality testing process can be semi-automated.

Despite growing interest in applying machine learning to quality control, existing studies remain limited in their focus on real-world batch manufacturing environments, particularly within the FMCG sector. Prior research largely emphasizes defect detection or classification in discrete manufacturing, often relying on sensor-rich, real-time datasets, with less attention given to semi-automation of laboratory-based quality testing processes. Furthermore, there is insufficient exploration of regression-based approaches for predicting continuous quality parameters, such as viscosity, using routinely collected production data. This study addresses these gaps by investigating the feasibility of using machine learning to predict batch quality outcomes and reduce testing time in a practical industrial context

3. Methods

In this section, we provide description of both the process and quality control procedures. Additionally, we discuss how the research data was collected and analysed. Exploratory data analysis was also employed due to the limited research conducted regarding machine learning predicting quality output in a batch manufacturing process in the literature.

Data regarding the batch manufacturing process was gathered as secondary data from the manufacturing plant's process flow diagrams. Process flow diagrams, also known as process maps, make use of visualisation to depict process steps sequentially, showing starting and stopping points and showing branching points and possible successive

steps (O'Leary, et al., 2023). The process flow diagrams provided information on the inputs to the process (raw materials) and the order in which they are added to the mixer, and process parameters that are critical to manufacturing a batch such as agitation speed and temperature.

3.1 Process and Quality Description

A brief overview of the production process for Product 1 and Product 2 that form the basis for developing machine learning models for the quality testing is depicted in Figure 2.

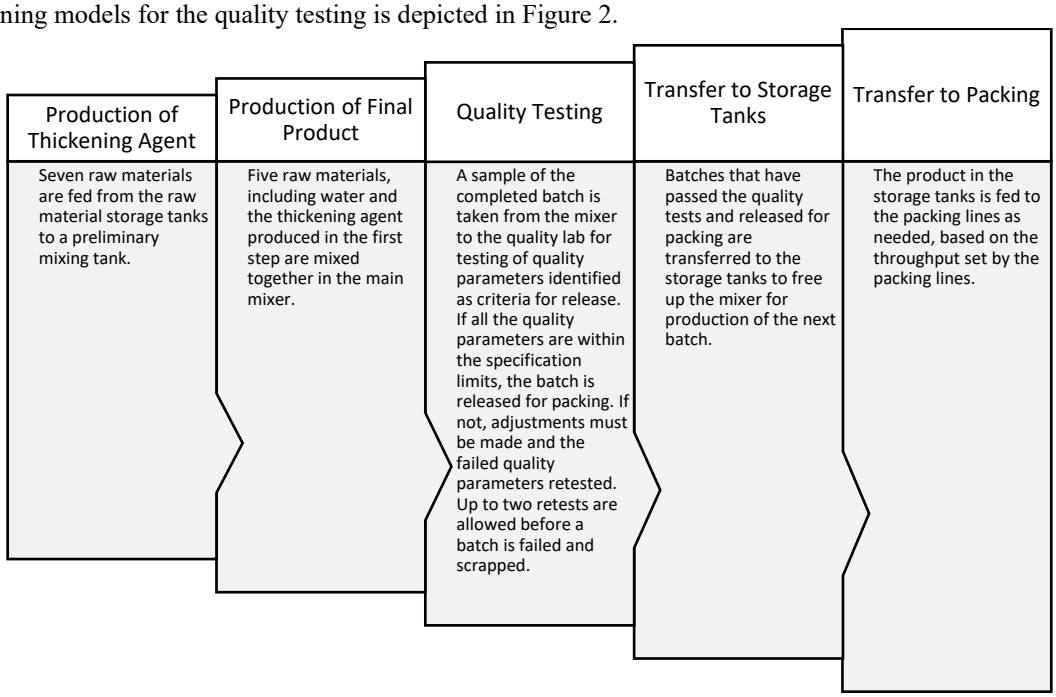

Figure 2. Overview of the batch production process

Due to the limitations in data availability for the production of the thickening agent, the machine learning model was developed for accepting or rejecting the final product only and the thickening agent was treated like the other raw materials over whose quality parameters Plant A has no control.

An observation and time study of the quality testing process done by quality technicians was conducted, and the findings are mapped in Appendix A. On average, the quality testing process takes 15 minutes from start to end. Deviations from this occur when one of the quality parameters is not within the specification limits, requiring adjustments and a retest. Informal conversation with the quality technicians observed revealed that the adjustment process sometimes takes longer when the operators dispute the test results. This was confirmed during one of the observations when the operator requested that the quality technician retest the sample before any adjustment was made to rule out the possibility of a false reading from the viscometer. A review of the testing results revealed that the quality parameters with the highest frequency of failing were viscosity and available chlorine. To narrow down to one parameter, the number of retests were compared. The need for more than one adjustment for a failed batch indicates a potential lack of clarity regarding the impact of the actions taken to adjust on the result. Based on the number of retests, viscosity was selected as the quality parameter to be predicted in this study.

Adjustment of a batch takes between 8 – 11 minutes, including the retesting. Batches that fail for viscosity were all adjusted by adding varying amounts of raw material T. Once adjustments are made, the batch needs to be retested for viscosity, pH and specific gravity regardless of which quality parameter had failed.

3.2 Instrumentation

The main instruments used for data collection and analysis is Azure Data Explorer (ADX) data warehouse which was queried making use of Kusto Query Language (KQL) to access the required data for analysis and building the machine learning model.

Data was queried from ADX into MS Excel for initial transformation tasks which included splitting columns and merging tables. Data from the different sources, ADX and Plant A's real time quality data collection interface, was collated on Excel to create one file that could be read into the machine learning model. Jupyter Notebook was used for data cleaning, preparation and the development and assessment of the resulting machine learning models. Python was the applied programming language for the above-mentioned tasks, and different libraries available within the software were used throughout the process.

3.3 Exploratory Data Analysis

Exploratory data techniques were used to determine the existence of outliers, missing data points and to make statistical assumptions. Descriptive statistics of the input data such as mean, median, mode and standard variation was done. Also, histograms and boxplots were utilized to visualize the distribution. Pair plot was plotted for insights on the relationships between variables, followed by a correlation matrix to assign a numeric value to the correlation between pairs of variables.

A summary of some of the results of the exploratory data analysis conducted is presented in Appendix B and Table B1 in the Appendix.

4. Data Collection

4.1 Data Pre-processing and Preparation

Data pre-processing ensures that the data is cleaned, integrated (if coming from different sources), transformed into the format required for consumption by machine learning algorithms, and consisting of necessary data only (Njeri, 2022). Initial data collected included Date, Shift, Product description, Batch identification number, Process parameter description, Process parameter set point and Process parameter actual. The data was cleaned following the below steps:

- Using the boxplots developed during the EDA phase, outliers were identified for the numerical variables. The outliers for process parameters and raw material dosing quantities were not removed or imputed, however outliers for viscosity were removed as they showed signs of data capturing errors.
- Missing data was found for a few of the process parameters and raw material data and were imputed with their medians making use of the *ColumnTransformer* and *SimpleImputer* functions from the scikit-learn library. This was however done after the data had been split into training and testing data to avoid data leakage.

The total sample size after data cleaning was 1139 batches. The data was converted into the necessary format for the machine algorithms to be used in what is called data preparation. The following data transformations were executed:

- The batch identification and set point columns were dropped from the dataset as they were deemed redundant and having no bearing on the outcome.
- The features (named "X") were defined as the shift, product, raw material actual quantities dosed and actual process parameter data while the viscosity result (named "y") was identified as the label.
- A correlation matrix was plotted (shown in Appendix C), in which the correlation between pairs of the input and target output label represented by a Pearson correlation coefficient with a value between negative and positive one. The features with the highest correlation to the label were torque and temperature, which had a positive and negative correlation respectively. The Date, Shift, batch number and products type had a low feature importance compared to the viscosity and were dropped from the features for the study.
- The data was then split into training and testing data with a ratio of 70:30 using the *train_test_split* function from the scikit-learn model selection library. Once the data was split, the below-mentioned tasks were done separately on both the testing and training datasets to avoid data leakage.
- One hot encoding was used to represent the categorical variables as binary vectors. This process assigns the category of a given observation the value one and sets all other to zero, and is widely used where categorical data needs to be transformed into a numerical input for machine learning algorithms (Samuels, 2024). This was achieved using the *get_dummies* function from the Python pandas library.
- Normalization of the data was done to transform how the numeric values are represented and have all numeric values appear on the same scale. This was done using the *MinMaxScaler* function from the scikit-learn preprocessing library, as some of the variables do not follow a Gaussian distribution.

The input data and output (label /target data) description needed for investigating the machine learning model development are provided in Table 2 below.

Table 2. Input and output data for classification and regression

Data for ML	Description	Data category (type)
C Actual	Quantity dosed of raw material C	Input (numeric)
P Actual	Quantity dosed of raw material P (perfume)	Input (numeric)
T Actual	Quantity dosed of raw material T	Input (numeric)
H Actual	Quantity dosed of raw material H	Input (numeric)
W Actual	Quantity dosed of raw material W (water)	Input (numeric)
Temp Actual	Actual temperature of mixer	Process input (numeric)
Torque Actual	Actual torque of mixer	Process input (numeric)
Viscosity Actual	Viscosity result	Output/ target outcome (numeric)
Viscosity	Batch quality (Acceptable or failed)	Output / target outcome (Categorical)

4.2 Machine learning model development (Regression and Classification)

Given the numeric values of the data, both regression and classification models could be used to predict the quality output of a manufactured batch.

For the regression linear models, a baseline prediction was determined using the median of the dataset and the performance evaluated using the Mean Absolute Error (MAE) metric. The baseline model returned a prediction of 661 with a MAE of 46,1.

Five regression models were used, and their performance compared against each other and against the baseline. The random forest regressor returned the best results with the lowest MAE of 40,2. Hyperparameter tuning was done to find the most optimal parameters for the model, and using the randomized search cross validation function the best parameters were determined as: Maximum depth = 6, Maximum features = None, Maximum leaf nodes = 6.

While the regression models were developed to predict the numerical viscosity of a batch, classification models were developed to determine whether a batch can be classified as released or failed. Six classification models were developed, namely K-Nearest Neighbors (KNN) Classification, Logistic Regression, Support Vector Machine (SVM), Random Forest, Decision Trees Classification and Gradient Boosting Classification. Multiple evaluation metrics were used to determine the models' performance including a confusion matrix summarizing comparing the true positives, false positives, true negatives, and false negatives returned by the model, a classification report assigning precision, recall, F1 and support scores to each model, and ROC-AUC curves.

5. Results and Discussion

5.1 Regression Model Results

The MAE yielded by each model is listed in Table 3, along with the MAE for the training dataset to evaluate for over- and or underfitting.

Table 3. Regression model results

Model	Mean Absolute Error (of test dataset)	Mean Absolute Error (of training dataset)	Over- or Underfitting? (Yes/No)
Linear Regression	43,5	41	No
Support Vector Regressor	41,3	39,6	No
Random Forest Regressor	40,2	37,1	No
Gradient Boosting Regressor	41,3	39,6	No
K-Nearest Neighbors Regressor	43,2	40,7	No

All five regression models proved to be successful in predicting the viscosity value of a batch and can therefore be used to semi-automate the quality testing process.

5.2 Classification Model Results

The results of each model are provided in Figure 3 to Figure 7. Data exploration revealed a high imbalance in classes, as the batches classified as failed represented only 2% of the entire dataset. To account for the imbalance and mitigate the impact on the model results, three rebalancing techniques were used and their performance compared namely, Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic Sampling Approach (ADASYN), and hybridization of SMOTE and Tomek.

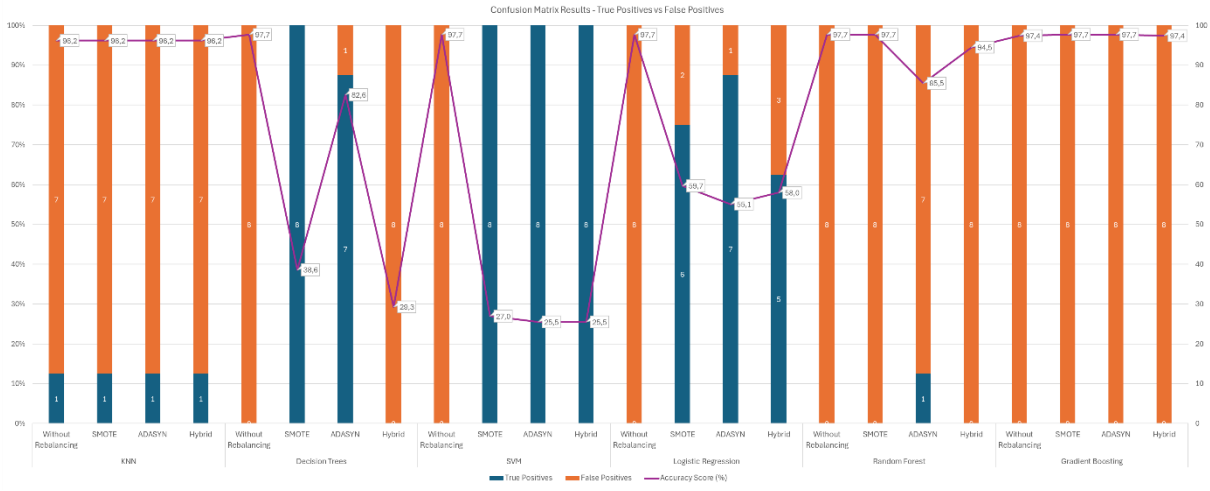

Figure 3. Confusion matrix - True Positives (TP) versus False Positives (FP)

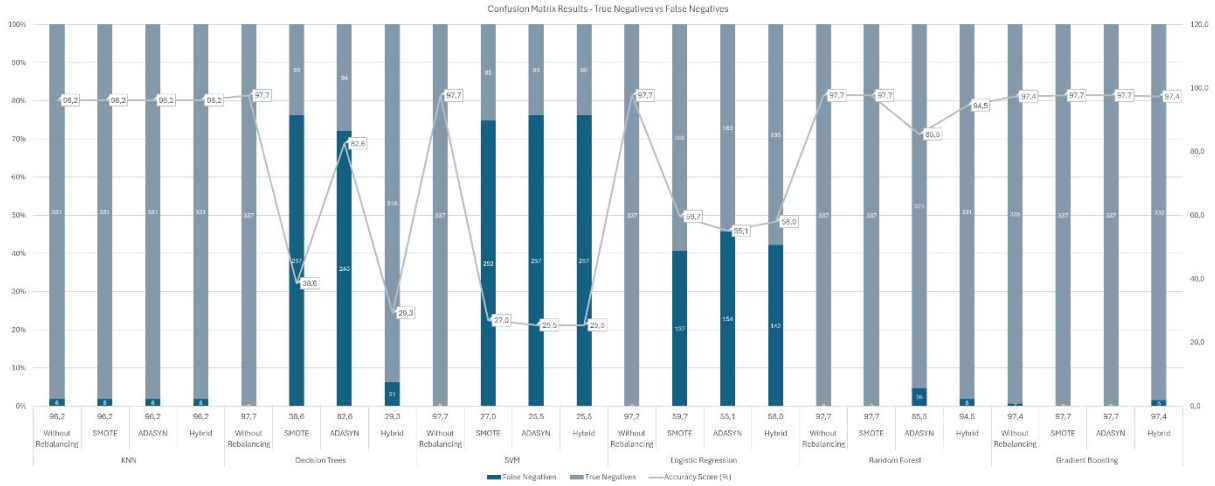

Figure 4. Confusion matrix - True Negatives (TN) versus False Negatives (FN)

The confusion matrix results showed that the Decision Trees, SVM, Logistic Regression, and Random Forest models had the highest accuracy when tested on the imbalanced data. However, it was assumed that the accuracy was biased and the accuracy for all models was considered to determine the best-performing model. Full confusion matrix results and ROC-AUC scores can be found in Appendix D.

The classification report, detailing a summary of metrics calculated from the above data, provided further information about the performance of the different models. The data from the confusion matrices and classification reports of all the models showed that none performed well in predicting when a batch would fail, while most performed well in predicting when a batch could be released despite the techniques employed to address the imbalance in data. Therefore, classification machine learning models are not effective in predicting the viscosity output of a batch in order to semi-automate the quality testing process.

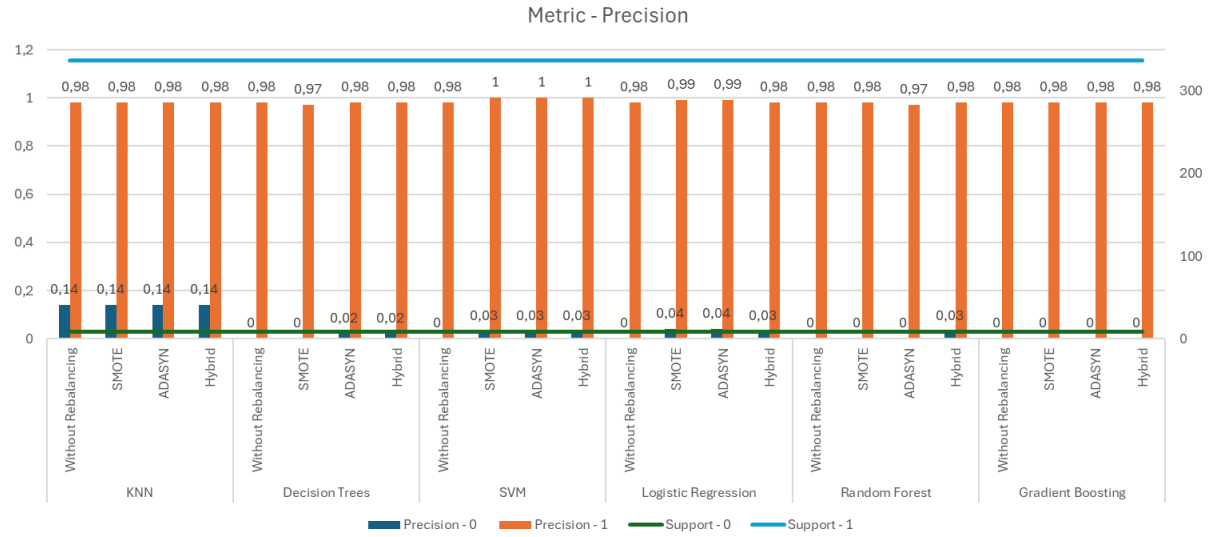

Figure 5. Precision scores

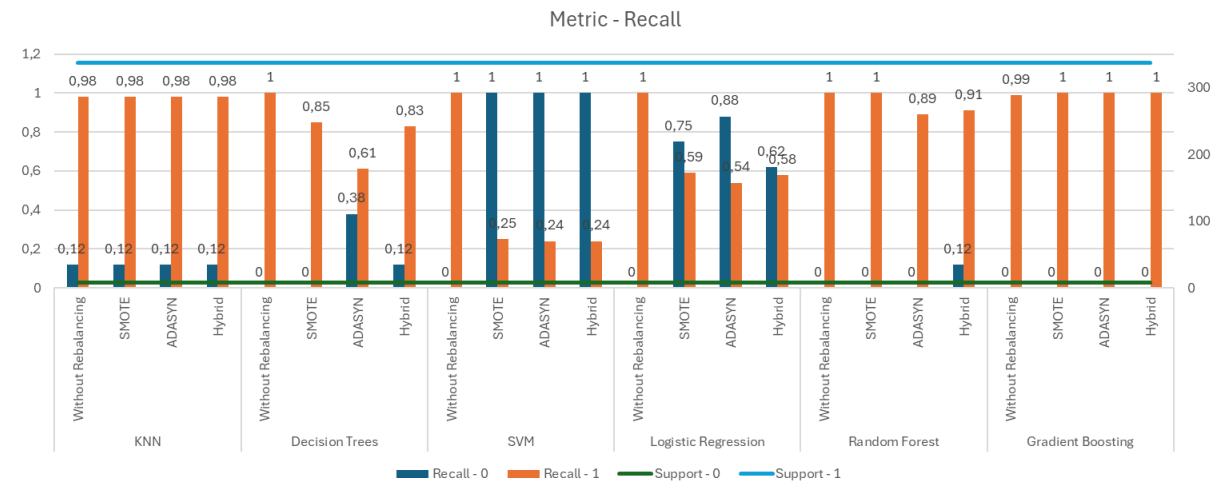

Figure 6. Recall scores

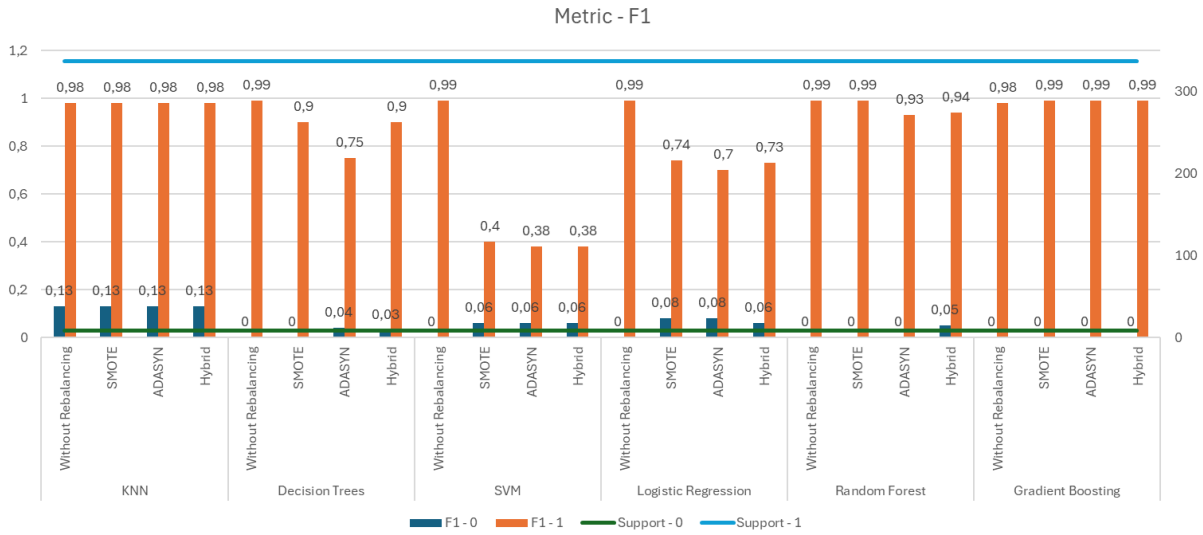

Figure 7. F1 scores

5.3 Proposed Improvements

The regression machine learning models developed proved to be effective in predicting the viscosity of a batch based on the raw material quantities dosed and the torque and temperature of the mixer for that batch. Using a metric of mean absolute error, all five models outperformed the baseline model's error of 46,1. No over- or under-fitting was detected on any of the models. The random forest regression model returned the lowest mean absolute error for both the training and testing dataset with 37,1 and 40,2 respectively.

The classification models were not effective in predicting whether a batch would be classified as released or failed. This was due to the high imbalance of data representing instances where batches failed and attempts to address the imbalance improved the models' prediction but not sufficiently. Despite the performance of the classification models, the first objective was met, and it was determined that machine learning regression models can be used to predict batch quality results for viscosity.

Based on the objective to determine the parameters that are useful in predicting the outcome of quality release checks, the feature importance function revealed that the features with promising relevance in predicting viscosity were temperature, product type, the quantity of raw material H dosed, and the quantity dosed of raw materials C, P, T and

W. The torque and the temperature features however showed very high correlation (positive and negative respectively) with viscosity. The relevance of temperature in predicting viscosity was expected, as a change in temperature can result in an exponential change in viscosity (Wenhao, 2021). An increase in temperature results in a decrease in viscosity (Rao, 1999). This was also identified during data exploration where the correlation matrix showed a negative correlation between temperature and viscosity. The feature with the next highest importance was torque. Torque is the force applied to rotate an object (Özkaya & Nordin, 1999), and it has been established that there is a positive correlation between torque and viscosity (Daqiqshirazi, et al., 2014). Thus, the torque required to mix a batch could be used as an indicator of the viscosity of that batch.

Whether the product was produced in the morning or night shift had little importance to the prediction of the model, meaning that factors such as human error or ambient temperature could be ruled out.

The machine learning models developed create opportunities for improvement, as they can improve the quality release checks through eliminating the need for retests, ensuring that no more than eight hours are spent daily on quality testing. Having a prediction of the quality results before the batch is sent for testing allows operators to make the necessary adjustments beforehand and the tests conducted in the lab can confirm the predicted values. It enables the proactive management of quality parameters and eliminates disputes between operators and quality technicians over the results. Furthermore, they can be used to reduce the sampling frequency; to increase production capacity in the factory. Predictions from the machine learning models can be used in place of actual tests, and tests can be conducted on every second batch to confirm the model predictions. The dataset, comprising of data for a four-month period, showed that non-conformities accounted for only 3% of the total batches produced. With stable and continuous production, and a high conformance rate, the plant may be eligible to apply for ISO skip-lot (ISO, 2005) inspection qualification.

Based on the success of the machine learning models in predicting viscosity, the third objective was achieved as they will be effective in improving the quality release checks process. Improving the quality release process would directly impact the effectiveness of the manufacturing process, by improving both the availability and the quality of the plant. This would also reduce waste, namely the waiting and defect waste types defined in lean methodology.

6. Conclusion

This study aimed to determine the effectiveness of machine learning in semi-automating the quality testing process. Furthermore, the study sought to determine if machine learning would be effective in predicting batch quality results, the parameter useful for predicting the outcome of quality release checks and to evaluate the improvement in the quality release checks process.

The regression models not only predicted the viscosity results successfully, but they also revealed insights into the relationships between parameters and the output. When compared with a baseline model, all five models outperformed the baseline when measured on error meaning that they can be used to predict the viscosity of batches. This would fulfil the purpose of semi-automating the quality testing process, by letting the operator know and rectify the quality parameter results before they are sent for testing in the lab. It would also provide the factory with the chance to reduce the sampling frequency. The feature importance highlighted the relationship between features and viscosity and, along with the correlation studies conducted during the exploratory data analysis, highlighted the parameters that are key to achieving the desired viscosity.

The accuracy achieved by the regression models means that this could be used to predict the quality results of a batch, and semi-automate the process by automating the adjustment and troubleshooting tasks. However, given the importance of conducting quality testing and conformance requirements of manufacturing organisations to quality standards such as the ISO (ISO,1999), semi automation of quality testing could be implemented should an ISO skip-lot qualification (ISO,2005) be granted. This will reduce the time spent on quality testing, and increase the capacity of the plant, with the latter providing a higher benefit.

The machine learning models were developed using viscosity as the quality result to be predicted as it was the one that had the highest frequency of failing and retesting. The same models can be trained to predict available chlorine results as it had the second-highest frequency, however, a larger dataset taken over a longer period will be required to train the models to predict all the quality parameters tested in the lab.

The machine learning models were developed using the data available for process parameters and raw materials per batch. However, the quality of the raw materials dosed was not included when training the models and determining the relationships that exist between the quality of raw materials and the resulting batch quality results. Daqiqshirazi, et al. (2014) proved that there is a relationship between the concentration of raw materials and viscosity. Therefore, it would be value-adding to train the model on not only the quantity of raw materials dosed but the quality of the materials as well. The temperature of the batch, which had the highest correlation with the viscosity and was the feature with the highest relative importance, could also be expanded to include the temperature of the raw materials in the bulk storage tanks as that would have an impact on the temperature of the batch. This could allow for predictions of a batch's quality results before production of the batch even begins, enabling proactive mitigation of risks. This would also ensure that predictions are accurate even with supplier or raw material recipe changes.

The final recommendation is to use the learnings from this model to develop process control measures. Using torque as an indicator of viscosity, a closed-loop control system could be implemented. A closed-loop control enables the system to be adjusted dynamically to match a specific set point and minimize the impact of external disturbances (Giraud & Giraud-Audine, 2019). The control could use the torque set point to be referenced and have the process keep dosing raw materials until the set point is achieved. To mimic the current practice of using raw material T to adjust batches with low viscosity, once the other raw materials set points are reached the process could then dose only raw material T until the desired torque is achieved. This would further help the plant with their qualification for skip-lot inspection, as it can give indication that non-sampled batches perform as well as the sampled batches. The findings and proposed process control measures are applicable to batch and continuous manufacturing beyond the FMCG sector. By leveraging machine learning to identify relationships between raw materials, process parameters, and quality outcomes, the approach enables semi- or fully automated process adjustments to ensure products meet specifications prior to testing.

References

Arboretti, R., Product quality control forecast using machine learning algorithms: A case study, *Proceedings of the 5th International Conference on Statistics: Theory and Applications*, vol. 118, pp. 1–4, 2023.

Bishop, C., *Pattern Recognition and Machine Learning (Information Science and Statistics)*, New York, Springer-Verlag, 2006.

Botchkarev, A., A new typology design of performance metrics to measure errors in machine learning regression algorithms, *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 14, pp. 45–76, 2019.

Bruce, P., Bruce, A. and Gedeck, P., *Practical Statistics for Data Scientists*, 2nd ed., O'Reilly Media, Inc., 2020.

Buer, S.-V., Semini, M., Strandhagen, J.O. and Sgarbossa, F., The complementary effect of lean manufacturing and digitalisation on operational performance, *International Journal of Production Research*, vol. 59, no. 7, pp. 1976–1992, 2021.

Celik, O., A research on machine learning methods and its applications, *Journal of Educational Technology and Online Learning*, vol. 1, no. 3, pp. 25–40, 2018.

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., SMOTE: Synthetic minority oversampling technique, *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

Choi, R.Y., Introduction to machine learning, neural networks and deep learning, *Translational Vision Science & Technology*, vol. 9, no. 14, pp. 1–12, 2020.

Daqiqshirazi, M., Riasi, A. and Nourbakhsh, A., Numerical study of flow in side chambers of a centrifugal pump and its effect on disk friction loss, Pune, India, 2014.

Deshmukh, H., Jadhav, V., Deshmukh, S. and Rane, K., Towards applicability of machine learning in quality control – A review, *The Electrochemical Society*, vol. 107, no. 1, pp. 14729–14737, 2022.

Elreedy, D. and Atiya, A.F., A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance, *Information Sciences*, vol. 505, pp. 32–64, 2019.

Giraud, F. and Giraud-Audine, C., Control in the rotating reference frame, in *Piezoelectric Actuators: Vector Control Method*, Butterworth-Heinemann, pp. 137–160, 2019.

Gokalp, M. *et al.*, Big data for Industry 4.0: A conceptual framework, *International Conference on Computational Science and Computational Intelligence*, pp. 431–434, 2016.

Handoyo, S., Suharman, H., Ghani, E.K. and Soedarsono, S., A business strategy, operational efficiency, ownership structure, and manufacturing performance: The moderating role of market uncertainty and competition intensity, *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 9, no. 2, pp. 1–14, 2023.

ISO, *ISO 2859-1: Sampling procedures for inspection by attributes – Part 1*, 2nd ed., 1999.

ISO, *ISO 2859-3: Sampling procedures for inspection by attributes – Part 3*, 2nd ed., 2005.
Judi, H.M., Jenal, R. and Devendran, G., Quality control implementation in manufacturing companies: Motivating factors and challenges, in *Applications and Experiences of Quality Control*, pp. 495–508, 2011.
Kaneku-Orbegozo, J., Martinez-Palomino, J., Sotelo-Raffo, F. and Ramos-Palomino, E., Applying lean manufacturing principles to reduce waste and improve process in a manufacturer: A research study in Peru, *IOP Conference Series: Materials Science and Engineering*, vol. 689, pp. 1–6, 2019.
Kumar, V. and Garg, M., Predictive analytics: A review of trends and techniques, *International Journal of Computer Applications*, vol. 182, no. 1, pp. 31–37, 2018.
Lee, C.-Y. and Johnson, A.L., Operational efficiency, in *The Handbook of Industrial and Systems Engineering*, 2012.
Ling, C. and Li, C., Data mining for direct marketing problems and solutions, *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, New York, AAAI Press, 1998.
Madanhire, I. and Mbohwa, C., Application of statistical process control (SPC) in manufacturing industry in a developing country, *Procedia CIRP*, vol. 40, pp. 580–583, 2016.
Makoni, T. and Chikobvu, D., Assessing and forecasting the long-term impact of the global financial crisis on manufacturing sales in South Africa, *Economies*, vol. 11, no. 6, p. 158, 2023.
Mrugalska, B. and Tytyk, E., Quality control methods for product reliability and safety, *Procedia Manufacturing*, vol. 3, pp. 2730–2737, 2015.
Msakni, M.K., Risan, A. and Schutz, P., Using machine learning prediction models for quality control: A case study from the automotive industry, *Computational Management Science*, vol. 20, no. 1, 2023.
Njeri, N.R., Data preparation for machine learning modelling, *International Journal of Computer Applications Technology and Research*, vol. 11, no. 6, pp. 231–235, 2022.
O'Leary, M.C. *et al.*, Optimizing process flow diagrams to guide implementation of a colorectal cancer screening intervention, *Cancer Causes & Control*, vol. 34, no. 1, pp. 89–98, 2023.
Özkaya, N. and Nordin, M., Moment and torque, in *Fundamentals of Biomechanics*, New York, Springer, pp. 29–46, 1999.
Pruneski, J.A. *et al.*, The development and deployment of machine learning models, *Knee Surgery Sports Traumatology Arthroscopy*, vol. 30, pp. 3917–3923, 2022.
Rao, M.A., *Rheology of Fluid and Semifluid Foods: Principles and Applications*, 1st ed., Aspen Publishers, 1999.
Samuels, J.A., One-hot encoding and two-hot encoding: An introduction, 2024.
Singh, R.K., Clements, E.J. and Sonwaney, V., Measurement of overall equipment effectiveness to improve operational efficiency, *International Journal of Process Management and Benchmarking*, vol. 8, no. 2, pp. 246–261, 2018.
Varoquaux, G. and Colliot, O., Evaluating machine learning models and their diagnostic value, in *Machine Learning for Brain Disorders*, Humana Press, pp. 601–630, 2023.
Wenhao, Z., Influence of temperature and concentration on viscosity of complex fluids, *Journal of Physics: Conference Series*, vol. 1965, art. 012064, pp. 1–6, 2021.
Witten, I., Frank, E., Hall, M. and Pal, C., *Data Mining*, 4th ed., Morgan Kaufmann Publishers, 2016.
Zhao, Z., Advancing feature selection research, *ASU Feature Selection Repository*, pp. 1–28, 2010.

Biographies

Mbidzo Ladzani is a seasoned Supply Chain professional with extensive experience in managing end-to-end supply chain operations across both local and multinational fast-moving consumer goods (FMCG) organizations. Her professional journey has cultivated a strong interest in automation, optimization, and machine learning, with a particular focus on leveraging digital technologies to enhance operational efficiency. She recently completed a Master of Engineering in Industrial Engineering at the University of the Witwatersrand, where her academic work further deepened her commitment to integrating advanced analytics and digitalisation into supply chain practices.

Gbeminiyi John Oyewole worked as a senior lecturer at the University of the Witwatersrand and is currently an honorary senior lecturer involved with research and supervision of doctoral and master's students. His research interests are in Facility location problems, modeling and simulation, data analytics, applied optimization, operations research, machine learning, and supply chain engineering.

Appendix A

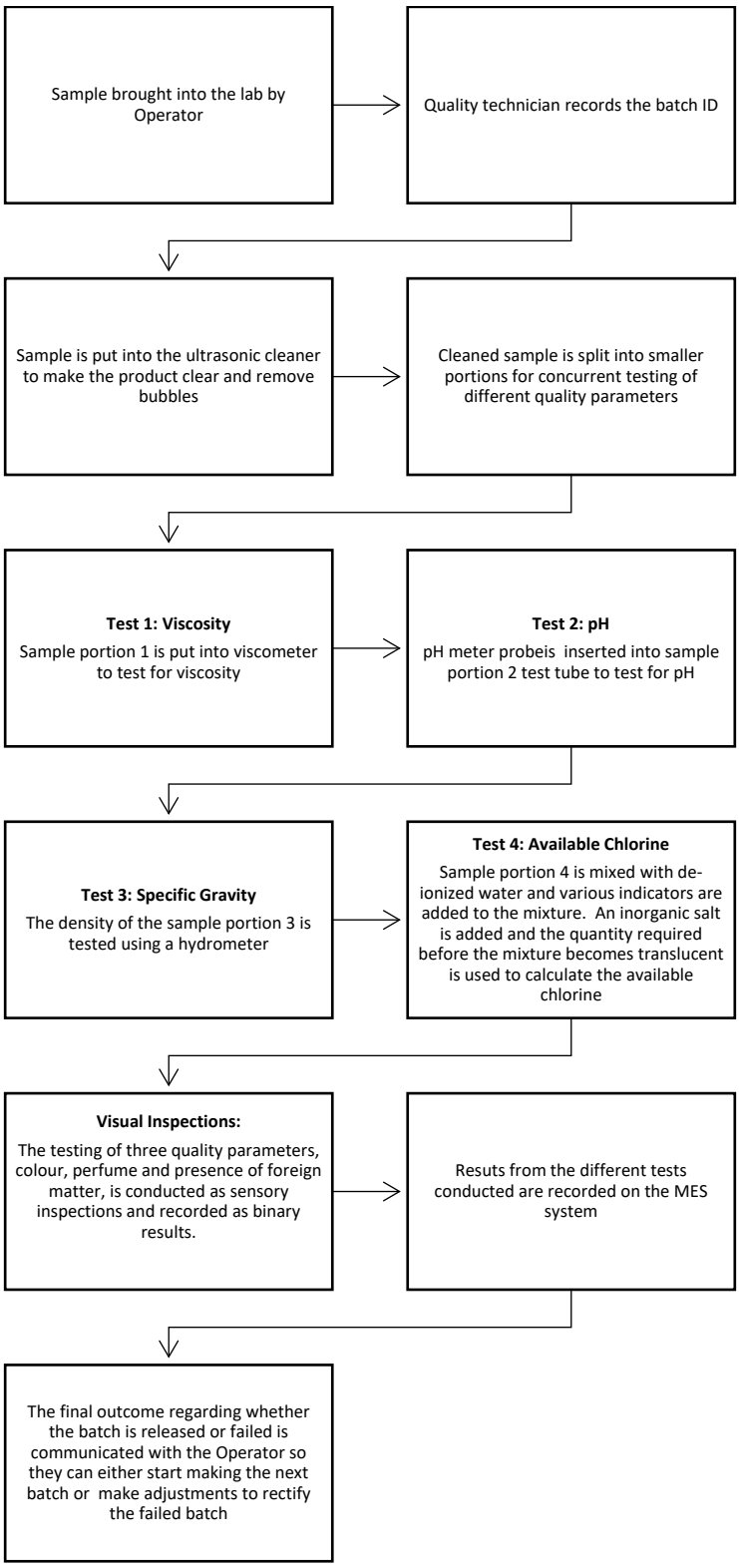

Figure A1: *Quality testing process done by quality technicians was conducted*

Appendix B

	C Set Point	C Actual	H Set Point	H Actual	P Set Point	P Actual	T Set Point	T Actual	W Set Point	W Actual	Temp Set Point	Temp Actual	Torque Set Point	Torque Actual	Viscosity Result
Count	1297	1297	1256	1256	1289	1289	1289	1289	1289	1289	1241	1241	1241	1241	1150
Mean	181	176.4	2249	2211	28.4	28.4	2230	2233	5346	5316	25	27	13.2	13.4	647
Std	0	1.3	71.3	75.1	22.9	22.9	43	42.6	45	46	0	3.1	0.3	0.3	74.4
Min	181	173.2	1801	1770	3	2.7	2223	2205	4681	4650	25	20.1	13.2	10.7	0
25%	181	175.5	2223	2173	4	4	2223	2224	5351	5320	25	24	13.2	13.2	617
50%	181	176.3	2223	2183	50	49.1	2223	2225	5351	5323	25	27	13.2	13.3	661
75%	181	177.2	2223	2210	50	49.8	2223	2227	5351	5324	25	29	13.2	13.5	687
Max	181	183	2572	2556	50	53	2572	2574	5500	5463	25	34	13.2	15	1401

Figure B1: Descriptive statistics of data used for the research

Table B1: Exploratory data analysis

1. *C Actual*: While the dosing of raw material C was mostly below the set point, the data showed inconsistency with the quantity dosed varying between 174 and 183. The data was positively skewed, and the outliers were close in value which could be an indication that they represent actual process values.
2. *H Actual*: Raw material H has highly inconsistent dosing. The histogram of actual quantity dosed was bimodal, and the box plot displayed data that was not highly dispersed but had a high number of outliers. Based on this, the outliers were not removed from the dataset.
3. *P Actual*: The set point for raw material P varies largely for different products. The box plot showed highly dispersed data, caused by the difference in the requirements per product. There are no outliers in the dataset.
4. *T Actual*: The thickening agent, referred to as raw material T, showed a low dispersion of data. There are multiple outliers, so these could be indicative of variability of the process rather than a misreading or erroneous value.
5. *W Actual*: The data for raw material W dosing was consistent for all products except one, suggesting a fairly stable process. The distribution of data was negatively skewed but showed a low data dispersion, which was reinforced by the box plot as well. The outliers were not removed from the dataset as they are actual data points for one product.
6. *Temp Actual*: The temperature feature dataset showed variation, varying even for the same product. The histogram reinforces the insights, as there is a large dispersion of data resulting in a bimodal distribution. There were no outliers for this variable.

7. *Torque Actual*: The torque variable showed some instability, with all products having data points both above and below the set point. There were visible outliers on all data plots and further analysis was done to understand if they are an accurate reflection of reality before any action could be taken to remove them.
8. *Viscosity Result*: The dataset for the label, viscosity, initially showed a slight negatively skewed distribution with outliers. The extreme outliers were deemed as unrepresentative of reality based on the process capability and were removed from the dataset.
9. *Set Points*: The set point variables were removed from the dataset after the initial exploratory data analysis, as they do not affect the viscosity result.

Appendix C

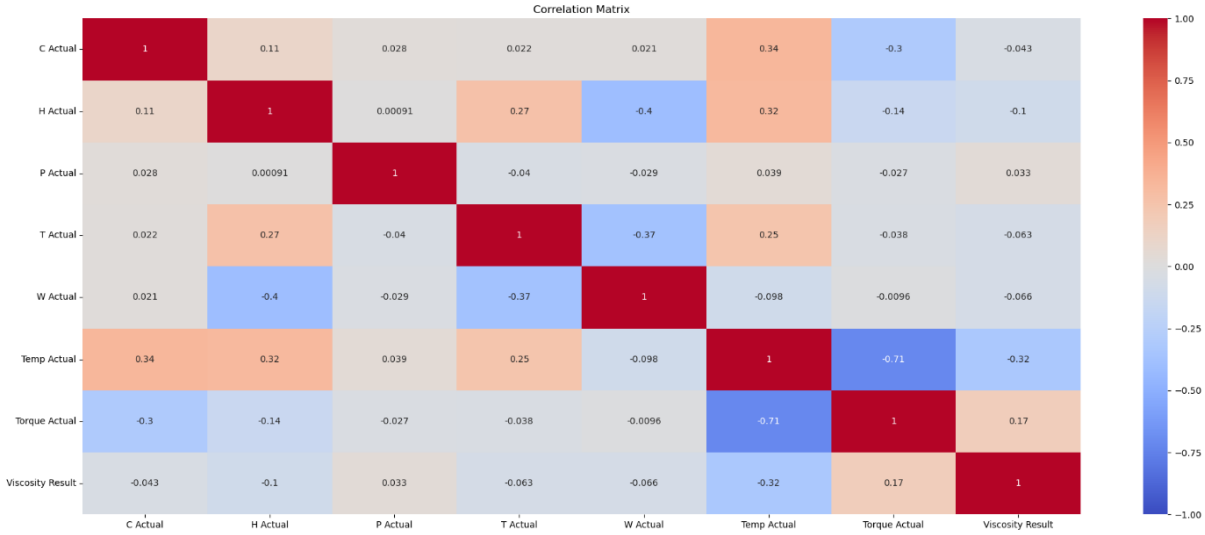

Figure C1: Feature correlation matrix

Appendix D

Table D1: Confusion matrix results for classification models

Model		**Accuracy Score (%)**	**True Positives**	**False Positives**	**False Negatives**	**True Negatives**
KNN	*Without Rebalancing*	96.2	1	7	6	331
	SMOTE	96.2	1	7	6	331
	ADASYN	96.2	1	7	6	331
	Hybrid	96.2	1	7	6	331
Decision Trees	*Without Rebalancing*	97.7	0	8	0	337

	SMOTE	38.6	8	0	257	80
	ADASYN	82.6	7	1	243	94
	Hybrid	29.3	0	8	21	316
SVM	*Without Rebalancing*	97.7	0	8	0	337
	SMOTE	27	8	0	252	85
	ADASYN	25.5	8	0	257	80
	Hybrid	25.5	8	0	257	80
Logistic Regression	*Without Rebalancing*	97.7	0	8	0	337
	SMOTE	59.7	6	2	137	200
	ADASYN	55.1	7	1	154	183
	Hybrid	58	5	3	142	195
Random Forest	*Without Rebalancing*	97.7	0	8	0	337
	SMOTE	97.7	0	8	0	337
	ADASYN	85.5	1	7	16	321
	Hybrid	94.5	0	8	6	331
Gradient Boosting	*Without Rebalancing*	97.4	0	8	2	335
	SMOTE	97.7	0	8	0	337
	ADASYN	97.7	0	8	0	337
	Hybrid	97.4	0	8	5	332

Table D2: Classification report results for classification models

Model		**Precision**		**Recall**		**F1**		**Support**	
		0	**1**	**0**	**1**	**0**	**1**	**0**	**1**
KNN	*Without Rebalancing*	0.14	0.98	0.12	0.98	0.13	0.98	8	337
	SMOTE	0.14	0.98	0.12	0.98	0.13	0.98	8	337
	ADASYN	0.14	0.98	0.12	0.98	0.13	0.98	8	337

	Hybrid	0.14	0.98	0.12	0.98	0.13	0.98	8	337
Decision Trees	*Without Rebalancing*	0	0.98	0	1	0	0.99	8	337
	SMOTE	0	0.97	0	0.85	0	0.9	8	337
	ADASYN	0.02	0.98	0.38	0.61	0.04	0.75	8	337
	Hybrid	0.02	0.98	0.12	0.83	0.03	0.9	8	337
SVM	*Without Rebalancing*	0	0.98	0	1	0	0.99	8	337
	SMOTE	0.03	1	1	0.25	0.06	0.4	8	337
	ADASYN	0.03	1	1	0.24	0.06	0.38	8	337
	Hybrid	0.03	1	1	0.24	0.06	0.38	8	337
Logistic Regression	*Without Rebalancing*	0	0.98	0	1	0	0.99	8	337
	SMOTE	0.04	0.99	0.75	0.59	0.08	0.74	8	337
	ADASYN	0.04	0.99	0.88	0.54	0.08	0.7	8	337
	Hybrid	0.03	0.98	0.62	0.58	0.06	0.73	8	337
Random Forest	*Without Rebalancing*	0	0.98	0	1	0	0.99	8	337
	SMOTE	0	0.98	0	1	0	0.99	8	337
	ADASYN	0	0.97	0	0.89	0	0.93	8	337
	Hybrid	0.03	0.98	0.12	0.91	0.05	0.94	8	337
Gradient Boosting	*Without Rebalancing*	0	0.98	0	0.99	0	0.98	8	337
	SMOTE	0	0.98	0	1	0	0.99	8	337
	ADASYN	0	0.98	0	1	0	0.99	8	337
	Hybrid	0	0.98	0	1	0	0.99	8	337

Table D3: ROC-AUC Scores for classification models

Model	**Variants**	**ROC-AUC Score**
Decision Trees	ADASYN	0,58
	Hybrid	0,58
	SMOTE	0,54
	Without Rebalancing	0,5

Gradient Boosting	ADASYN	0,55
	Hybrid	0,5
	SMOTE	0,49
	Without Rebalancing	0,5
KNN	ADASYN	0,55
	Hybrid	0,55
	SMOTE	0,55
	Without Rebalancing	0,55
Logistic Regression	ADASYN	0,56
	Hybrid	0,57
	SMOTE	0,58
	Without Rebalancing	0,5
Random Forest	ADASYN	0,5
	Hybrid	0,5
	SMOTE	0,5
	Without Rebalancing	0,5
SVM	ADASYN	0,61
	Hybrid	0,68
	SMOTE	0,64
	Without Rebalancing	0,5