**Enhancing Beverage Production Process Efficiency: A Machine Learning Approach**

Michelle M. Moran

A Thesis Submitted in Partial Fulfilment

of the requirements for the

Degree of

Master of Science in Data Analytics

Diagram

Description automatically generated with medium confidence

September 2023

Supervisor: Kislay Raj

**CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Capstone Project |
| **Assessment Title:** | Enhancing Beverage Production Process Efficiency: A Machine Learning Approach |
| **Supervisor Name:** | Kislay Raj |
| **Student Full Name:** | Michelle Moran |
| **Student Number:** | SBA22220 |
| **Assessment Due Date:** | Friday, 22 September 2023 |
| **Date of Submission:** | TBC |

Declaration

By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

**List of Contents**

**Declaration**

**List of Contents**

**List of Tables**

**List of Figures**

**Abstract**

**Acknowledgements**

**List of Tables**

**List of Figures**

**Abstract**

**Acknowledgements**

# Chapter 1 Introduction

## Problem Definition

In the modern beverage batch manufacturing industry, maintaining consistent production efficiency is crucial. However, there are evident challenges in the current production regime, especially concerning downtimes across different production tanks. This not only disrupts the flow of production but also signifies existing inefficiencies or gaps. The sheer complexity of batch data, coupled with its vastness, has made it difficult to derive actionable insights manually, leading to potential overlooked areas of optimization. Furthermore, without predictive mechanisms in place, anticipating these downtimes for better scheduling remains a challenge. This research intends to address these gaps through a comprehensive exploration of production downtimes, an efficiency-driven analysis using machine learning, and the development of predictive models to enhance scheduling processes. The overarching goal is to streamline the production process, minimize downtimes, and heighten overall production efficiency in the beverage batch manufacturing sector.

**Mucilage Containing Beverage Production Process**

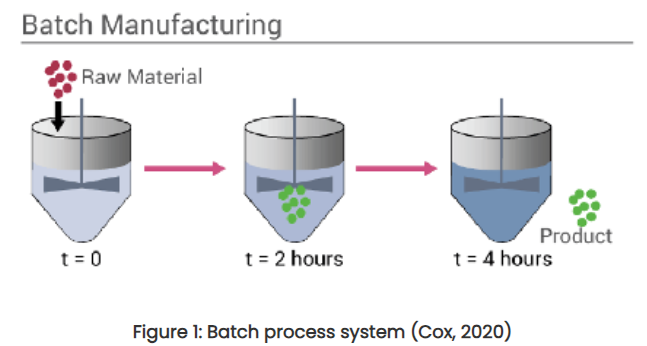
To produce a batch of Mucilage Containing Beverage, it involves on average 27 instruction steps. The production line consists of production tanks divided into systems of varying capacities and each system is denoted as 22, 23, 25 and 25 followed by MT (mobile tank), so in system 22 MT , there is 5 tanks , 1- 5 of capacity 20 tonne. These production tanks have two jobs, one is where batch production occurs and a destination tank where after certain instruction steps, the batch needs to be stored. This type of process can be termed continuous , as the end product is an order in volume, (Kang et al , 2020)

The production schedule determines the material batch to be produced, the quantity required, and this decides which production tank is to be used. The production process is mixed between automated and manual where some ingredients such as water is added via a tank delivery and gum addition is manual by a production operator. Each instruction step has parameters that is logged on a shopfloor system which is a batch data storage database. Instruction steps can be described as a recipe but in this organisation can also be known as phases. Each instruction step is monitored by various metrics, e.g Phase duration, this the length in minutes of how long each phase takes. There is also targe phase times for the duration and flowrate of ingredients , these targets have been determined based on historical batch data.

These beverage materials are so called because they contain an ingredient called mucilage or otherwise known as gum. The purpose of this ingredient is important to these materials as they add stability to the beverage batch but also is known to aid the enhancement of colour ingredient., (Chung et al, 2016, Benech et al , 2008)). But it’s this gum ingredient’s behaviour in the beverage production process which can impact the production process downtimes measured as either phase start delays which is the measure of the delay in the instruction step or phase overrun, which is the overrun time of the phase. For the purposes of this study, phase overrun will be the downtime measure for the mucilage containing beverage materials.



If the Gum addition instruction step overruns, creating downtime, it affects the following production process steps such as the mixing via agitation step, processing via deaeration step and the final texture of the beverage batch. The main reason the gum addition step in the production of beverages materials is problematic is due to length of time it takes to add the gum itself which is manual and to dissolve in the production tanks. After the initial addition, the rest of it tends to float on its surface. Therefore, it takes longer to mix into solution. The longer the agitation instruction step takes, the more gas that is created in the production tank resulting in a longer deaeration step. The agitators are switched off until deaeration is completed. All these impacts are monitored and recorded as phase overrun metric and logged in each batch details on the shopfloor system.



The current process of producing mucilage containing beverage batches involves collecting batch data from the various production line operations. This data is used to measure the effectiveness of a production process. The standard measurement is called overall Equipment Effectiveness (OEE). This is computed for the whole production line for each production tank at each production step. It looks at the availability, performance, and quality to determine production efficiency.

Lepenioti et al, 2020 states that this data, has the credentials to move beyond these OEE metrics and with the recent advancements of machine learning, predictive and prescriptive analytics using machine learning are possible with the aim of supporting the operator on the shopfloor and enhancing the production process. Another common term for this process is data mining which is essentially extracting valuable information from a vast amount of raw data and transforming into an understandable structure for future use. **(**Ge et al, 2017)

This research focuses on three primary objectives related to beverage batch manufacturing. First, it delves into the process to pinpoint and measure downtimes during mixing and deaeration in various production tanks, aiming to understand current inefficiencies. Second, it employs machine learning to rigorously analyse batch production data, emphasizing the benefits of such analysis for process optimization. Lastly, it aims to create predictive machine learning models to forecast phase overrun times for the agitation, addition of gum ingredient and deaeration, using these predictions to suggest enhanced batch scheduling to elevate overall production efficiency.

Following on from here, section 2 is the literature review, delving into existing research related to this this topic to provide an overview of key findings and methodologies used previously. Chapter 3 and 4 is concerned with the research methodology where the research methods used in this study are outlined in detail. Chapter 5 is the results and discussion where data and findings from the results are given and discussed. The discussion will connect with findings from the literature review, and highlight any implications, significance, and potential limitations of the results. Chapter 6 is the conclusion which will summarize the main findings, discuss their implications, and give suggestions for future work.

# Chapter 2: Literature Review

## 2.1 Introduction

The literature review will focus on a range of areas relating to production process optimization through the application of machine learning models for the downtime prediction in beverage containing mucilage production.

I will look at the importance and impact of efficient production in industry in general to examine how efficiency is achieved and changes to this process. Next to focus on production downtimes in batch manufacturing, define the different types of downtimes and the factors contributing to downtimes. To review methods and techniques that are used to quantify and monitor downtimes.

Following this, I will examine machine learning models applications in manufacturing process by first defining machine learning and giving a background detailing the advantages and challenges of implementing it in production. Further delve into using machine learning in production downtime predictions and its advantages for production optimization.

1. Efficiency-driven Analysis of Batch Data

The main motivation behind the analysis of batch data is to improve the efficiency of the production process through reducing downtimes, maximizing output, minimizing waste, optimizing resource utilization, and shortening the overall production cycle. There is an increasing importance to enhance the effectiveness and efficiency of decision making in a production process, through mining the production data both online and offline using more efficient techniques. (H., XIA et al, 2022). This is a methodical examination and evaluation of batch data and can involve inspecting, cleaning, transforming, and modelling data to discover useful information. This is more often referred to as data mining.

The world of manufacturing has the potential to utilize machine learning to extract patterns from existing datasets, which can serve as a basis for predicting future system behaviour, (Alpaydin et al 2010, Nilsson et al, 2005). Alpaydin also claims “Storage data only becomes useful when it is assessed and translated into knowledge that we may use, for example to develop predictions”. This is true of this research.

In the manufacturing context, "batch production" is a method where items are produced in groups or batches rather than in a continuous stream. "Batch data" would then refer to the data generated during these batch production processes. It could include variables like production start and end times, quantities produced, downtimes, error logs, equipment metrics, and any other relevant data points that can provide insights into the production process.

According to Kovalev et al, 2019, it highlights the importance of batch data as the head of the process of digital transformation part of the industry 4.0 revolution. Digital transformation is the approach used by the production industry undergoing this revolution for the optimization of production data.

Production Batch data is worthless on its own, the manufacturing industry requires efficient processes to be able to derive valuable information from it. The following are examples of processes of examining production batch data to uncover hidden patterns and correlations include data mining, machine learning, natural language processing.

## Traditional Process Optimisation Methods

Prior to machine learning, traditional methods used for improving production efficiency included manual inspection, statistical tools, expert systems, and mathematical modelling, (Wang et al, 2018) Other traditional statistical methods such as statistical control charts (SPC) are now deemed insufficient when it comes to enhancing production processes, (Ismail et al, 2021). The advantages to these control measures were their applicability and simplicity but are now not able to keep up with the increasing complexity of production and volume of data being gathered as a result. (Ismail et al 2022).

Ge et al, 2017 describes the importance of analysing batch production data for patterns and relationships between production variables leading to useful information can be extracted and used by Statistical models such as Operational equipment efficiencies (OEE). OEE can be developed for various applications such as process monitoring or fault diagnosis.

Another important aspect of batch data collection is highlighted in Arif et al ,2023, where it states most existing quality monitoring models only look at one manufacturing state and the batch data gathered is not processed until after the product is made or manufacturing process is over. This has a negative effect on resources, time, and production performance.

Three areas that make traditional methods obsolete and where utilising machine learning shines are:

* Information and communication technologies – mode of production has changed, large-scale tasks, operating performances and environments are more complex, (Wang et al, 2018a)
* Increased demand for real time dynamic self-adaptive and precise production management (Arashpour et al, 2018, Lamon et al, 2010.)
* The completion of various kinds of information systems deployed in manufacturing enterprises. E.g., CAPP, computer-aided process planning, (Papananias et al, 2019.)

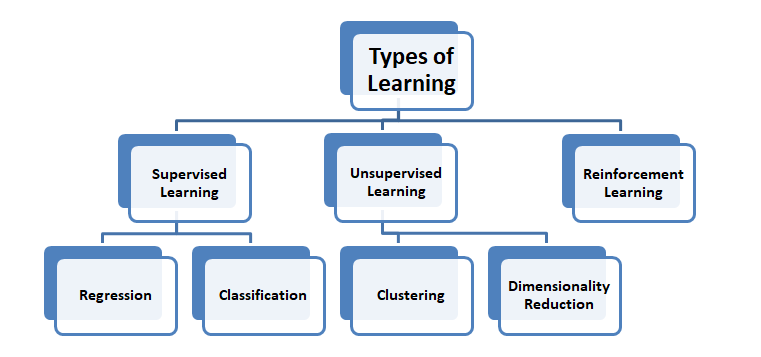
Machine learning cannot completely replace traditional methods. One can learn from the other while machine learning can aid the identification and modification of the parameters of the traditional methods to improve processes (H., XIA et al, 2022.)

Also, another consideration is the time and financial cost of developing and programming models of machine learning, this may be out of reach of small-scale manufacturers, however for large scale manufacturers and their various production lines, it is the varying complexities and lack of efficiencies these models can bring. It can be challenging to implement AI – machine learning in an entire organisation with existing processes and systems if the company lacks robust technology infrastructure and collected data. This is often a limitation It’s important that there is clear strategy stemming from top management to achieve goals otherwise AI will fail (Kang et al 2020, Heio et al, 2021).

## Machine Learning

Machine Learning model systems learn from data, identify trends or patterns from data, make decisions based on structure feedback and then perform tasks on their own, with continued improved performance and problem-solving skills without human intervention (Helo et al, 2019). It models the complex relationship between input and output data, (Wang et al, 2018b).

* Machine learning can be divided up into 3 types: unsupervised, supervised and reinforcement types, (Pugliese et al , 2021) as shown in fig. Its also shows the associated machine techniques. For this research, type of machine learning relevant is supervised as we have an input and output from a set of labelled training data and the technique is regression as the input value is numerical continuous valuable. These algorithms are used to optimize the coefficients of each independent variable to achieve a minimum error in the prediction.



The common methodology in using Machine learning algorithmsfor the improvement in manufacturing process and quality optimization are as follows, (Ismail et al, 2021, Kant et al.,2015, Aksa et all.,2021, Ahmad et al.,2018, Kulkarni et al.,2020, Koksal et al., 2011)

1. Data Collection – collect data on the performance of the chosen parameter under different operation conditions, e.g., phase overrun , phase duration ,flow rates, temperatures etc
2. Data Prepossessing – removing outliers and missing values ensure the data is clean and accurate
3. Feature selection – selecting the relevant features that can help prediction
4. Model development – develop the predictive model and evaluate it using metrics such as accuracy, precision, and recall
5. Model Optimization – to improve its performance using hyperparameters or ensemble learning techniques
6. Model Deployment – deploying the model in a real-world setting and validate its performance.

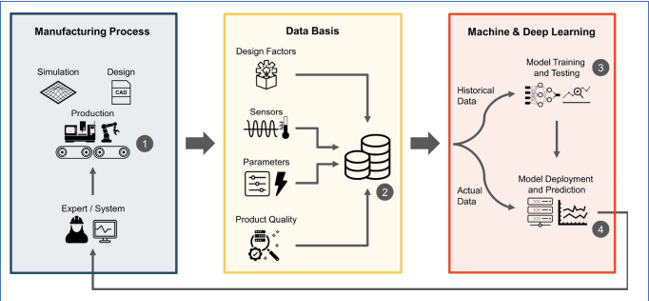
According to Kang el al, 2020, most machine learning applications in the manufacturing industry were concerned with supervised learning, due to the abundance of production data available. Regression is the main task applied for quality optimization problems while classification and anomaly detection are mainly applied for product failure detection.

Machine learning has successfully been applied in industry, these are summarized below, (Diez-Olivan et al ,2019, Aksa et al, 2021, Menezes et al, 2019):

* **Descriptive** – accounting and analysis of historical data. e.g. Fault detection and diagnosis
* **Predictive** – Considers near past data to predict coming future trends, biases tendencies and behaviours through causation and correlation.
* **Prescriptive** – finds or prescribes the best mode route manner or moves to operate based on given data (output) and models (inputs). help make decisions on what to do and how to minimize failure impact
* **Preventative** – determine the potential of failures happening.
* **Detective analytics** – makes diagnostics of collected data to eliminate and rectify inappropriate values used in predictive analytics.
* **Cognitive analytics** – automated predictions, prescriptions, and detections for smarter decisions over time

**4.6 Machine learning for Predictive analysis**

Through this literature review, there was no direct research on the process optimization using machine learning models by predicting production phases downtime in the beverage manufacturing. However, there was plenty of research reviews on other types of processes in the manufacturing industry using various machine learning (Monostori et al, 1996, Md et al , 2022, Paturi et al, 2021).



***Fig 2 Predictive quality approach for a selected manufacturing process***

Tercan et al, 2022, reviewed research publications between 2012 and 2022 dealing with predictive quality in manufacturing using machine learning models. The categorization was based on the manufacturing processes and machine models involved. None of the processes involved a beverage production process. Predictive quality uses machine learning methods in production to predict product-related quality base on process and product data. Most research papers reviewed by Tercan et al, 2022, involve having a base machine model and then doing a comparison with other machine models and these mainly are ensemble methods. Ensemble methods involve the combination of multiple learning models, thereby aggregating their decisions to make a prediction.



***Fig 3 Examples of baseline models used in Predictive quality processes (***Tercan et al, 2022)

In this research the initial part is focused on the exploratory data analysis where the goal is to understand the data and identify inefficiencies during the various instruction steps identified across the production tanks used. This is not a machine learning task but it is a crucial preliminary step before applying machine learning models, (Santos M et al,2023, Komorowski et al, 2016).

Examples of Machine learning models in production process optimization

Most manufacturing processes involve multistage steps to produce accurate products. Papananias et al, 2019 developed a Bayesian linear regression model to estimate part quality of and associated uncertainties given in process monitoring data. The predicted results compared well with the experimental measurements and further a neural network model was developed which also showed similar results. It highlighted challenges to the model of evolution of more complex products, big data, and manufacturing intelligence. It also gave the suggestion to look at self-organising maps (SOM, Lieber et al, 2013), or principal component analysis for the consideration of process variables such as high sampling rates or large measurement uncertainties to reduce them. A study on the injection moding process by Farahani et al, 2021, looked at over 10 different machine learning models. It was determined that Neural Networks is a perfect model for this type of prediction but the computional power, time and cost are factors need to be considered.

A research paper concerned with engine oil aeration process step looked at a gaussian regression model used to correlate the identified features to measure oil aeration. The results were successful in the prediction of oil aeration to an uncertainty of +/-0.02 from the measured oil aeration values. The model was trained using previous oil pressure data. The results also highlighted that importance of looking at sampling measure as the cases used showed overfitting. This was calculated from using the metrics of RMSE, root mean square deviation. (Kulkarni et al 2021)

Under Predictive tool performances, logistic regression as a machine learning model has been used for the reliability estimation for cutting tools using the variable vibration signals. It is using correlation analysis approach to estimate the reliability and failure time of the cutting tools. It concluded that by further looking at the mechanical knowledge and probability density functions of other variables could further enhance the model’s predictions on machine tool performances. (Chen et al, 2010)

IBM Research has developed a process and system regression optimization service for optimising set points for process controls. The Process and System Regression Optimization service consists of three main components -- (1) Regression component and (2) Single Process Optimization component, and (3) System-wide Optimization component. The three components are used to model the problem at hand with the regression component being used to train one or more regression models based on historical data from the process or asset, and one of the optimization components used in combination with the trained model(s) to optimize the control points. Datasets are trained on the API and then it uses various models of regression to make predictions requested. It uses Mean Squared error and R2 as metrics to compare different regression algorithms.

Phan et al, 2021, used this IBM AutoAI toolkit for automated machine learning to search for the right learning algorithm and optimize its hyperparameters. They were able to restrict their models to decision tree, multivariate adaptive regression splines. Using cross validation to pick the best performance model.

Hassani et al, looked at the efficiency of the equipment rather than the product process in manufacturing, the methodology included a case study where the data was based on results from 2 machines. The applied methodology included data preparation, exploratory analysis of the data and creating training sets and applying various models. Including support vector regression, random forest. Model accuracy was calculated with and with cross validation. It is the same methodology for optimization of equipment as is process.

Deep Learning via Neural Networks

Deep learning provides advance analytics tools for processing and analysing big manufacturing data. It is seen as a breakthrough solution to the challenges of multimodal data, the high dimensionality of feature space and multicollinearity among data measurements. It has applications in speech recognition, image recondition, it allows automatically processing of data towards highly nonlinear and complex feature abstraction via a cascade of multiple layers, (Wang et al, 2018, Rivas et al, Manami et al, 2023, Trask, A.W., 2019, Chollet, et al., 2021). The following looks at its application in process optimizations in manufacturing.

Pfrommer et al, 2018, used ANN as a surrogate model to optimize the draping process of textiles. He used the surrogacy process to eliminate the costly process of trial experimentation that is often the approach for process optimization. The study achieved improved results using surrogacy with deep ANN but highlighted the importance of the training data and the sampling.

A combination of a neural networks with a genetic algorithm to predict critical parameters in a particle board and axial flow designs manufacturing process. The Genetic algorithm was then applied to the trained NN to determine the optimum values using the parameters successfully. (Cook et al, 2000, Liu et al 2023)

Fang et al, 2019, learned from Chen and Wang et al, 2013 studies to forecast the cycle time in a wafer fabrication factory using a fuzzy back propagation network. It shows that the NN models lacked generalization and fitting capabilities to deal todays big manufacturing data. It also highlighted the importance of considering expertise for feature extraction to reduce the input dimensionality. Fang et al, 2019 considered all this in their study of shop floor real time job remaining time prediction. They used a Deep stacked sparse Autoencoder to deal with every changing manufacturing floor. In its conclusion it suggested looking at LSTM, long short time models to analyse time series problems and find bottlenecks in the production line. It also suggests reinforcement learning to help with self-learning in the production control and optimization.

A Comparison between ARIMA, LSTM and GRU models was done on time series forecasting for bitcoin technology. ARIMA, Autoregressive integrated Moving average, turned out to be the better model, followed by the GRU model gated recurrent network whereas LSTM allows the tracking the dependencies of new observations with past ones. This study gave details on the methodology on creating and comparing types a recurrent network, with certain parameters and evaluation metrics. It highlighted the importance of data normalization, using Scikit-learn making it easier for the network model to learn. Another highlight from these studies was the importance of the architecture of the neural network itself and optimisation of these methods (Yamack et al, 2019, kuric et al, 2022, Sun et al, 2019).

The time series predictions is used to monitor time changes and monitor trends in the development of the examined parameter. The most used models in research and practice for time series prediction are linear autoregressive models (AR and ARX), LSTM neural networks, moving average mode (MA) and autoregressive moving average (ARMA), (Kuric et al.,2022).

Yamack et al, 2019 conducted a comparison between three different machine learning models in making a time series prediction. The three were ARIMA (Auto regressive integrate moving average), LSTM (Long Short-term Memory) and GRU (Gate recurrent Unit). The topic was predicting the price of Bitcoin. This research compared all three models by accuracy (measured by MAPE and RMSE) and time. It concluded that the outcomes could have been of effected by the parameters that were chosen and the total amount of data used.

Fang el al.,2019 examined the time predictions to complete a job in a discrete manufacturing system using a deep learning-based approach. A deep Stacked sparse autoencoder (S-SAE) model was designed to enable a machine algorithm to learn highly varying status of manufacturing for prediction. In this research the performance was compared to different models, such as linear regression, back propagation neural networks, multi- layer neural networks and deep brief neural networks. To do this, the above models were trained and tested using the same dataset via a fivefold cross-validation. He concludes that future work, includes using LSTM neural networks to analyse time series in production.

A study on a predictive maintenance of an industry machines by Geltz et al , 2021, looked at several machine models using firstly traditional models such as XGBoost , ridge regression and then neural networks using a multi-layer perceptron. The traditional methods produce good accuracies whilst the neural networks did not. They concluded that the lack of enough data points hampered neural networks thus highlight the importance of the data set. Kang et al , 2020 also noted that a larger dataset helps to achieve better results with neural networks , but they’re main advantage is ability to handle complexity which can be found in productions

**4.11 The Production process of manufacturing a mucilage contain beverage batches.**

A production line involves a series of factory operations that refine materials into end products (Kang et al., 2020). In this study, the focus is on a tank-based production line involving steps like agitation and deaeration. While machine learning excels at deciphering complex patterns, its effectiveness in this context is uncertain. A manufacturing company may operate multiple production lines, each with distinct challenges. Vast data from some lines can lead to effective machine learning solutions; however, applying the same solution to lines with limited data might not yield the desired results (Kang et al., 2020). The Production lines are classified as continuous where the target variables are continuously measured. According to Kang et al, 2020, most machine learning models applied to Quality optimization problems on production lines are regression type.

**Process Instruction Steps**

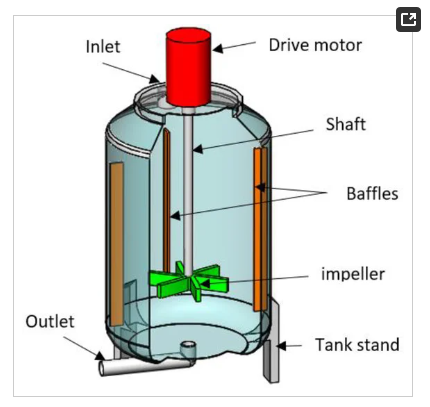
Three of the instruction steps in the process of producing mucilage containing beverage batches give rise to phase overrun downtimes. Each step is linked to each other.

Once the raw ingredients are added, the agitators are started, and the mixing begins. Agitators are equipment used to homogenise media inside a tank, they work by rotating immersed impellers at a controlled speed, call revolutions per minute. The fig below shows the various components of an agitator, which are present in the current production tanks that produce the mucilage containing beverage batches. From the production process batch data, agitation times are different for each tank which leads to varying production times and can lead to batch downtimes. The reasons for this is the dispersion of mucilage gum ingredient, once viscosity increase has started, agitation of the solution and therefore powder dispersion becomes increasingly difficult.

This results in longed mixing times are required to complete dispersion/hydration and creating more gas that needs to be dispersed during the deaeration phase.

It plays a crucial role in ensuring the quality and shelf life of the final product which was documented by Feilner et al. It involves the removal of dissolved gases such as oxygen from the beverage. The deaeration time must be carefully calibrated to achieve optimal results with the goal of preserving the desired sensory characteristic and stability of the beverage (Paquin et al, 2009).

There are many process parameters that can affect the length of deaeration time a beverage batch needs. Examples of process parameters include the type of final product to be produced, the equipment used, the initial methods of addition and nature of raw material added, temperature and pressure factors. Optimization of these parameters can lead to a lower deaeration time thus an increased efficient process without compromising the final product quality and stability.



* 1. **Summary**

The literature review underscores the application of machine learning in optimizing production processes, with an emphasis on predicting downtime in beverages containing mucilage. By analyzing batch data, the goal is to enhance efficiency, reduce waste, and streamline the production cycle. Industry 4.0 accentuates the importance of digital transformation and the leverage of batch data, with analytics tools like machine learning yielding invaluable insights from this data, , (Rai et al, 2021).. Traditional methods of optimization, like Statistical Process Control (SPC), are becoming outdated due to evolving technologies and the burgeoning integration of information systems. Machine learning, especially supervised learning, excels in identifying data patterns, continually refining its decision-making, and has found myriad applications in manufacturing, ranging from descriptive to cognitive analytics. In the broader context of manufacturing, machine learning is notably used for predictive analysis. Although the beverage industry hasn't been a focal point, other sectors have tapped into its potential, with various studies showcasing machine learning models, including the emerging deep learning techniques, for prediction and process efficiencies,, (Aksa et al, 2021).

Furthermore, a consistent theme across the literature is the adoption of multiple machine learning models to evaluate data and forecast outcomes. Equally pivotal is the data: its acquisition, refinement, and preliminary understanding of inherent trends and correlations. The nature of the data—be it labelled, structured, or unstructured—directly influences the choice of a machine learning model. There's a shared methodology across diverse research, encompassing distinct, replicable steps, which will significantly benefit this study. The proposed metrics for assessing predictions and trends will further fortify the research foundation. The crux of the present study is to assess the efficacy of machine learning models in predicting phase overrun times for mucilage-containing batches, employing a spectrum of production tanks with diverse capacities.

# Chapter 3: Methodology

* **Research Methodology**

Research methodologies typically fall into two main categories: primary and secondary data collection (Saunders, Lewis, and Thornhill, 2016). For the research questions poised at the start of this paper, a combination of primary and secondary data collection will be utilised.

Within both categories, various methods of data collection can be utilized. This section aims to outline the chosen collection strategy or strategies that, in the author's view, best serve to answer the research questions presented in Section 1 above.

**Primary Research Data Collection**

The author determined that employing a qualitative and quantitative approach for collecting primary research data would be the most suitable technique to gather first-hand insights, approaches as observing and interviewing individuals knowledgeable in beverage process production and data analytics.

Factors considered by the author when choosing this primary research method encompassed:

* Research Objectives – the appropriateness of the research method involved.
* Author status – role/access in the organisation if applicable
* Expertise of the Selected Participants – availability and validity
* Timeframe – how long this research method could take.
* Ethical Considerations – gaining consent of the participants to partake and the use the data collected.
* Bias - ensuring a diverse range of inputs by involving individuals with distinct roles in the beverage production process.

Two approaches are as follows:

**Data Collection through Observation**: **Qualitative Approach**

The author works full-time in a beverage production company and had the opportunity to be a participant as observer in the production area. This qualitative method was deemed appropriate as it helped develop a better understanding of the production process which is where the root of the research questions originated.

To gain access to the production area, an informal request to observe the production process was made to the appropriate production manager and associate. It was clearly stated as to the reasons why this request was made. The author explained in detail the research that was been conducted and the potential benefit to this observation would make to the research. Such benefit for example would be a clearer picture for the author of what is going on in the process.

The Author also checked prior to the visit that it would be ok to write notes on any observations or informal conversations had. It was also highlighted to all participants that they would remain anonymous. These notes will be in the appendices.

Potential errors in this type of research approach considered were:

* Observer Error could occur here due to the authors lack of understanding of the process area. There is importance on the interpretation of the data collected.
* Observer Effect where the author could affect the behaviour of those being observed thus posing a threat to the reliability and validity of the data collected. This also known as the Hawthorne effect.

The observational data collected, and its potential errors outlined above would be further clarified in further primary method of in person interviews.

The observational approach proved successful as the author gained invaluable insight into the physical workings of the process and made connections with experts in the process area which went on to prove useful in the next primary research approach – in person interviews. One such insight was the terminology used in the process area and how it related to the secondary data collected.

**Data Collection through in person interviews:**

In-person interview can take many forms but the one used for this research is an unstructured informal interview, and for practicality and ease of transcription, these interviews were conducted online via the Microsoft Teams platform. By employing an informal interview using open- ended questions as this data collection method, it facilitated a more comfortable scenario allowing for in depth discussions surrounding the beverage production process and the potential use of data analytics. Chosen Participants were able to freely talk about the process and the author was able to observe the participant reactions and adapt the interview flow, accordingly, as suggested by Saunders et al. (2012). This is a more flexible approach.

To address potential respondent bias, the author ensured a broad range of perspectives by engaging individuals from various roles within the beverage production process. Each interviewee contributed a distinct viewpoint: one oversaw the production process, another actively worked in the production area, and the third played a role in data analytics within the organization. Through interviews with this diverse group, the author aimed to achieve a comprehensive understanding of the existing beverage production process, encompassing both common practices and individual variations.".

Transcripts from the interviews can be seen in appendix XXX.

**Secondary Research Data Collection**

The author determined that employing a quantitative approach for collecting secondary research data would be the most suitable technique.

Factors considered by the author when choosing secondary data research method encompassed:

* Research Objectives – the purpose and compatibility of the secondary data
* Author status – role/access in the organisation if applicable
* Expertise of the Selected Participants – availability and validity
* Timeframe – how long this research method could take.
* Ethical Considerations – gaining consent of the participants to partake and the use the data collected.
* Bias - ensuring a diverse range of inputs by involving individuals with distinct roles in the beverage production process.

The secondary data source used was internal to the organisation, from various industrial information systems which is integrated. The support systems are FactoryTalk batch software and SQL Server Management Studios.

This data can be unstructured, and need to be prepared, by screening for duplicate data, missing data, irrelevant data records. Extraction of indicators and features by labelling the data that will be needed in the learning/training process.

The data for this research was originally stored on the manufacturing shopfloor database system. For each batch produced, each production process time step was recorded. By selecting the production time, the author was able to request via SQL query all material batch details that had a deaerating step in their process. Therefore, reducing the amount of data was sent and stored as a CSV file.

**Research Validity ( Kang et al, 2020)**

To determine the extent of which the primary and secondary data collection above accurately assesses the research problem objectives is the basis of validity. We need to validate if the data collected for this research answers the research questions poised and are these answers trustworthy and meaningful.

The most relevant components of validity:

**Internal Validity**: Addressing potential biases and threats to the reliability of the collected data. The use of diverse participants from various roles within the beverage production process helps mitigate bias, and employing both observation and interviews allows for cross-validation of findings, enhancing the robustness of the research outcomes.

**External Validity**: The extent to which the findings can be generalized to other contexts. By involving individuals from different roles within the production process and considering multiple data sources such as SCADA and MES, the research aims to capture a broad representation of the beverage production domain, increasing the potential for generalizability.

**Ecological Validity**: Ensuring that the research settings and conditions closely resemble real-world situations. The use of actual production environments and interactions with experts in the beverage production area through observation and interviews enhance the ecological validity of the study.

**Content Validity**: Ensuring that the data collected accurately represent the content domain of the research. The qualitative approach of observing the production process and conducting unstructured interviews with knowledgeable individuals helps capture a holistic and nuanced understanding of the beverage production process and its data analytics.

**Face Validity**: The extent to which the research methods appear to measure what they intend to measure. The use of multiple data collection methods, combined with clear communication with participants about the research objectives, helps establish the face validity of the research approach.

By addressing these components of validity, the research aims to enhance the credibility and reliability of its findings, contributing to a more robust and meaningful exploration of the beverage production process and data analytics. The research's data collection methodologies are rooted in thoroughness and adaptability. The combination of first-hand observation, in-depth interviews, and empirical data analysis provides a comprehensive view of the beverage production process. However, consistent checks for potential biases, coupled with methodological rigor, are essential to maintain the highest levels of validity.

**Research Ethics**

Ethics in research is important because it ensures that the participants are treated fairly and respectfully. It will also help protect their privacy and well-being. It also ensures that the data gathered is honest and trustworthy.

In the context of the primary data collection, the expert participants will be invited to join the research with the freedom to withdraw consent or exclude their data contributions at any point before the submission date. During the interviews a professional approach will be maintained with confidentiality and anonymity of the participants and the data information given preserved for research purposes. A master file containing the responses will be kept by the author and will not be shared. This master file will be securely stored, and password protected as an enhanced security. Additionally, if the participant has any queries during or after the interview, they will be answered promptly and sensitively.

The secondary data source used was internal to the organisation and it is stored on the relational database system called shopfloor system. The author followed proper protocols when gaining access to the data by engaging in the relevant channels within the organisation. Such channels include the organisations legal counsel and ethics point software where you can determine if a non-disclosure is relevant. However, since the data received will be swiftly anonymized to prevent linkage to the individual participants and organisation, there was no legal or ethical worries. The safeguarding of the data will be ensured through secure storage and the implementation of passwords.

Furthermore, the study acknowledges the importance of adhering to the General Data Protection Regulation (GDPR) and is committed to implementing necessary measures to uphold compliance throughout all stages of the research process. This demonstrates a commitment to ethical research practices and the protection of participants' privacy and rights.

1. **Sampling Strategy**

An important part of any research is the collected data and how to determine how much if it is needed to answer the questions posed by the research. The data collected through the primary and secondary data collection as described above constitutes the population. Depending on the research’s objectives, it may not be necessary to analyse the entire population. A sampling strategy may be needed to be employed to reduce the population but still maintain its representativeness.

The type of sampling strategy will be different for the data collected through primary and through secondary. Primary data was collected through in-depth unstructured interviews which is qualitative research and thus nonprobability sampling will be applied. For Secondary data collection, the type of sampling examined would be probability sampling where every item of the population has an equal chance of being included in the sample.

**Primary Data**

In-depth non structured interviews were used as the primary data collection methodology and therefore only a non-probability sampling strategy could be applied. The participants were purposely selected because of their expert knowledge and distinct perspectives in the research area. The type of non-probability sampling strategy used here is known as purposive or judgemental sampling. This strategy is convenient and appropriate for exploratory research design, but unlike random sampling it doesn’t try to represent the whole of the population but focuses on the depth of the information provided.

**Non-Probability Sampling Strategy: Purposive sampling**

**Validity and Generalizability:**

The use of purposive sampling might not provide a representative sample of the entire population, it does offer high validity for addressing specific research questions. The insights gained from participants with specialized knowledge contribute to a deeper understanding of the topic. However, the findings may not be easily generalizable to broader populations.

**Advantages and Disadvantages:**

Purposive sampling is advantageous for its ability to provide rich and targeted insights from individuals with expertise. It is particularly useful when the goal is to explore complex or specialized topics. However, its limitations include potential selection bias and reduced generalizability to broader populations.

**Conclusion:**

Purposive sampling is a valuable non-probability sampling strategy that allows researchers to gather in-depth and contextually relevant insights. By intentionally selecting participants with specific expertise, researchers can obtain valuable information that addresses the research objectives. While the findings may not be easily generalizable, the depth and quality of insights obtained make purposive sampling an important approach in research design.

**Secondary Data**

The secondary data collection source used was internal to the organisation and it is stored on the relational database system called shopfloor system. Here all data relating to production of all batches in the organisation are stored. This type of data would be described more recently as Big -Data due to the shear volume and velocity of it. This is a massive population size and not feasible to work with so therefore a sampling strategy is needed to be applied.

The type of strategy initially examined was probability sampling where every item in the population which was the database has an equal chance of being included in the sample. But this was too much data so to narrow down the population, based on the research objectives, the type of materials specifically those containing mucilage and a time of 2 years was chosen. This changed the sampling strategy to purposive or judgemental sampling, an example of non-probability sampling. A focus on a specific time range could yield possible trends, patterns, or changes during that period by examining variables involved in the production of the batches.

**Non-Probability Sampling Strategy: Purposive sampling**

**Sampling Frame:**

The sampling frame is ‘a complete list of all the data in the population from which the sample will be drawn’. So, in the case of the secondary data collection, the complete list is the organisation database of all the production batches produced and the sample is the filtered production data containing only mucilage containing materials, it will be necessary to consider the different production tanks used.

**Sample Size:**

Based on the research objectives, only mucilage containing materials produced in the last 2 years were chosen to be filtered from the main organisations databases. These materials, 46 in total, can only be produced in a range of 16 production tanks varying in capacity. The scheduling of these materials depends on customer demands. A total number of 347 batches were produced in this time over various tanks. The number of batches per tank produced were in the range from as low as 6 to 54, this allowed all batches to be included in the sample.

**Sample Technique:**

Under Purposive sampling, 6 techniques were reviewed including extreme case is only applicable to more unusual research. Homogenous sampling was deemed appropriate where focus is given to a particular subgroup in this case is the production tanks used in which all the materials produced are similar in that there are mucilage containing. The characteristics of tanks are similar which allows them to be explored in a greater depth.

**Check for Representation:**

The Author carefully chooses production batch material data with specific ingredient such as mucilage component. This approach enhances the depth and breadth of the insights gained.

**Validity and Generalizability:**

Like primary data collection, the use of purposive sampling might not provide a representative sample of the entire population, it does offer high validity for addressing specific research questions. The insights gained from the mucilage containing material production data chosen will contribute to a deeper understanding of the research topic. However, the findings may not be easily generalizable to broader populations.

**Advantages and Disadvantages:**

Purposive sampling is advantageous for its ability to provide rich and targeted insights from individuals with expertise. It is particularly useful when the goal is to explore complex or specialized topics. However, its limitations include potential selection bias and reduced generalizability to broader populations.

**Applying the Sampling Strategy**

Purposive sampling is a valuable non-probability sampling strategy that allows researchers to gather in-depth and contextually relevant insights. By intentionally selecting participants with specific expertise, researchers can obtain valuable information that addresses the research objectives and selecting mucilage containing materials over 2 years, the author can obtain invaluable insight in the production of these materials in various production tanks under various conditions.

While the findings may not be easily generalizable, the depth and quality of insights obtained make purposive sampling an important approach in research design.

The sampling strategies outlined above will be applied to primary data collection and secondary data collection.

**Experimental Methodology**

This chapter offers a detailed explanation of the experimental methodology adopted for answering the research objectives of Exploration and Quantification of Production Downtimes, applying machine learning models and predictive modelling of production downtimes.

Below is a breakdown of the main goals and how they were approached.

Research Objective 1: Understanding Production Downtimes

First, we wanted to understand the current process better. We investigated how often and why there were pauses (downtimes) in the various phase stages in the production manufacturing tanks. This gave us a clear picture of where there were production phase overruns for each tank.

Research Objective 2: Using Machine Learning to Analyse Data

Next, we turned to machine learning to analyse the production batch data. By studying the data, we wanted to point out where the process could be made more efficient. This step showed the value of using advanced tools and methods to analyse production data.

Research Objective 3: Predicting and Planning for Downtimes

Lastly, we created machine learning models to predict when these downtimes might happen. With these predictions, we aimed to plan production schedules better, reducing the number of pauses and making the whole process faster.

The comprehensive approach that follows integrates data collection, pre-processing, model selection, training, evaluation, and optimization.

**3.2 Data Collection**

**3.2.1 Source of Data**

The author acquired historical production process batch data through the Microsoft SQL Server Management Studio (SSMS). A SQL request was created with the following requirements:

* Time period – 2 years
* Production batch Material – Mucilage beverage containing batches that require a deaeration phase.
* All production phase details including time duration details.

This information was received in the excel format and was converted to a CSV file for better management and transferring of data because it preserves the original data values. More importantly CSV format is easily read and written by python programs.

**Data Acquisition**

These are the most important activities of building a machine learning model. It is important to collect the relevant data and create a proper dataset. Data sources can be external and internal. In this research, support for data and information communications come from various industrial information systems which is integrated. Support systems are Manufacturing execution systems (MES), supervisory control and data acquisition (SCADA) system and the programmable logic controller system (PLC) which directly controls the reaction parameters of machines and warehouse management system (WMS). The SCADA system is where production data is acquired, and it is stored on the relational database system called shopfloor system. (Min et al, 2019)

This data can be unstructured, and need to be prepared, by screening for duplicate data, missing data, irrelevant data records. Extraction of indicators and features by labelling the data that will be needed in the learning/training process. Data needs to be mapped depending on the knowledge of what it is going to be used for. (Lee et al, 2019, Min et al , 2019).

The data for this research was originally stored on the manufacturing shopfloor database system. For each batch produced, each production process time step was recorded. By selecting the production time, we were able to request via SQL query all material batch details that had a deaerating step in their process. Therefore, reducing the amount of data was sent and stored as a CSV file. This data will need to be prepared as stated above by Lee et al ,2019 before it can be passed to a machine learning model.

The Production batch data in the organisation is acquired from software called FactoryTalk® Batch. It provides a flexible batch control system. It allows you to specify procedures or recipes and enforce their execution in production.

The FactoryTalk Batch Server operates FactoryTalk® Batch software and coordinates the following functions:

* Creating a Batch: Transforms the configured recipe into an executable.
* working recipe.
* Executing a Recipe: Communicates with the process-connected.
* devices to execute phases.
* Arbitrating Equipment: Allocates resources based on recipe and
* operator requirements.
* Collecting Data: Gathers and stores production information for
* reporting and archiving.
* Performing Client Communications: Transfers data between the
* process-connected devices (PCDs), operator displays, Human Machine
* Interfaces (HMIs), databases, and various other software packages.

This FactoryTalk® Batch software uploads batch data from each of the production activity known as phases per batch to a Microsoft SQL database which is managed by SQL Server Management Studios. One of the components of this software is the FactoryTalk Event Archiver which translate the FactoryTalk batch tab-delimited ASCII electronic batch record files to a user-specified file type. This organisation uses Microsoft SQL database which is managed by SQL Server Management Studios to maintain files for each batch created.

Microsoft SQL Server Management Studio (SSMS) is a software application developed by Microsoft that is used for configuring, managing, and administering all components within Microsoft SQL Server. A central feature of SSMS is the Object Explorer, which allows the user to browse, select, and act upon any of the objects within the server.

**3.2.2 Features and Target Variables**

The features integrated into the model include:

* **Quantity** (Kg)
* **Phase\_duration** (min)
* **Phase\_start\_delay** (L/min)
* **Flowrate\_KGMIN (KGMIN)**

The primary target variable is **Phase\_overrun (min)**

**3.3 Data Preprocessing, (Xu et al , 2015)**

**3.3.1 Software , Libraries, Web Applications**

For this research, the programming language employed was Python. Its selection stemmed from its user-friendly nature and the ease with which the researcher could learn and utilize it. The vast availability of open-source libraries in Python further enhances its adaptability, allowing customization based on the programmer's requirements, (McKinney,2012). One of the major advantage of using python and its libraries, is that they can be used together, in tandem for data analysis, scientific computing and machine learning. The following are examples of the open-source python libraries that were used:

* Pandas – provides data structures like DataFrames and Series that make it easy to handle structured data. It allows for quick data cleaning, visualisation, and statistical analysis. In this research its main use was in the handling of missing data, filtering rows and aggregation of data.
* NumPy -employed for its basic mathematical and array operations.
* Matplotlib – A plotting library – creates and display graphs and visualisations of the data working alongside other libraries such as pandas.
* TensorFlow – Deep learning frameworks – used for building and training neural network models.
* Seaborn – works along with Matplotlib to create statistical graphics.
* Scikit-Learn - allows a programmer to quickly implement a range of machine learning algorithms in conjunction with other libraries such as pandas and matplotlib.
* Scipy – a advanced scientific computing library built on NumPy for scientific computational tasks. It adds additional functionality.

All the libraries above were used in Jupiter notebook which is another open-source web application that allows the creation and sharing of documentation that contain live code, equations, visualisations, and narrative text.

* + 1. **Data Cleaning**:
       1. **Handling missing values**:

Missing values are entries in the dataset, such as 0 or not a number (NaN). Using the panda’s library in python, the amount and type of missing values should be determined, (S.Xu et al , 2015). What columns in the data contain missing values and will it have an impact on the model used are questions that will be answered. Missing data can cause bias in estimating model parameters and loss of information ( Ismail et al , 2022, Lee et al , 2019).

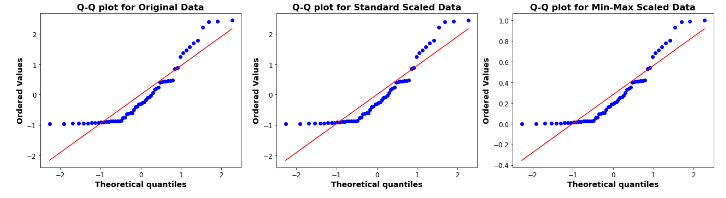
* + - 1. **Removing duplicates**

To protect data integrity, the dataset needs to be checked for duplicate rows of data. Duplicated data can distort the actual data analysis and give inaccurate results, skew distributions. Using panda’s library, duplication can be determined, and before removing by drop function, the data results need to be check in case they are valid repetitions.

* + - 1. **Data Normalization**

This pre-processing technique that involves the transformation of all numeric variables in the dataset to a standard scale. This is necessary when the data features in the dataset have different units or varying scales which can give skew interpretations from machine learning models.

An investigation using the Quantile-Quantile (Q-Q) Plot was looked at to determine the effect of the different types of normalisations such as standard scaling and minmax scaling have on the data. Using the data from the production tank 22MT – reference ProductionTank22\_df2, the following graphs shows results of the distribution of the phase overrun data. Using two examples of standard scalar and min-max scaler , there is a S – Shaped curve which indicates that the data may be following a logistic or s curve distribution rather than a normal distribution.



From the data above , there was no difference between the two scalar methods and below table gives details of the machine models and their sensitivity to scaling.



The StandardScaler function from Scikit-learn was used to execute Z-score normalization, ensuring each feature contributes equally to model efficiency. This a solid general-purpose choice and had broad applicability,

* 1. **Exploratory Data Analysis (EDA)**

Exploratory Data Analysis is an approach of analysing data sets to summarize their main characteristics. It is a critical process of performing initial investigations on data. It is used to discover trends or patterns, to spot anomalies and check statistical assumptions with the help of summary statistics and graphical representations. This is aided by the python libraries: Pandas/ NumPy/ Matplotlib/ Scikit-Learn/ Seaborn.

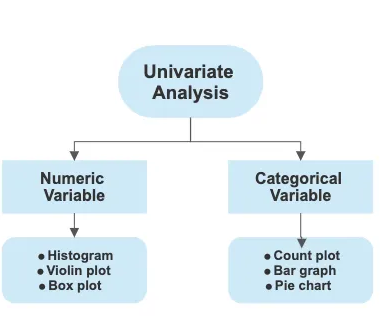
As Phase overrun is the target variable, the EDA was used to gain understanding of its distribution and characteristics in relation to the other feature variables for each production tank in the DataFrame.

Further EDA is examined on defined instruction process steps for the following using phase overrun as the target variable:

* Ingredient Addition – gum material addition
* Agitation Phase times
* Deaeration phase times

EDA was examined under the following headings:

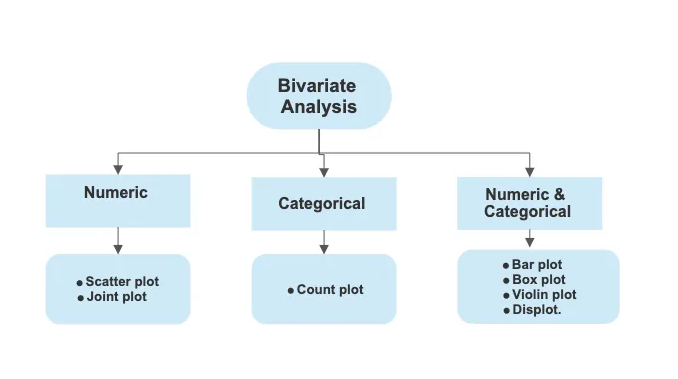
**3.4.1 Univariate Analysis:**



**Numerical Features**: Histograms, box plots, and descriptive statistics determined to understand the central tendency, spread, and shape of the distributions of the target and feature variables.

**Categorical Features**: Use frequency tables or bar plots to understand the distribution of each category.

**Bivariate and Multivariate Analysis:**



**Correlation Analysis**: Compute correlation matrices for numerical features to determine if there were any linear relationships between the variables. It also highlights if there is any multicollinearity detected.

This is important because it creates difficulty in estimating regression coefficients accurately, (kumar et al 2020).

**Scatter Plots**: For visual inspection of relationships and trends between pairs of numerical features.

**6. Identification of Outliers**:

**Visual Methods**: Use box plots to visually inspect for outliers.

**Statistical Methods**: Using the techniques IQR method can help in a more formal identification.

**9. Reporting and Visualization**:

Leverage visualization tools like Matplotlib, Seaborn, or Plotly in Python for plotting.

Documentation: Summarize the findings, insights, and any patterns observed during the EDA in a clear and concise manner, preferably complemented with visual aids.

* 1. **Machine Learning Model Application**

An evaluation of multiple machine learning models was performed, see table for details of models examined and their advantages and disadvantages.

For each production tank group data, the following steps were completed to apply each model to the data.

**3.5.1 – Data loading and Processing**

The data was loaded into a pandas dataframe. The features and target variable were defined.

**3.5.1 Dataset Split**

Once the dataset is established, it needs to be split into three subsets: Training, validation, and test sets. There is no optimal proportional for splitting, according to research, the rule of thumb was 70% for training and 15% for both validation and testing. It depends on the size of the dataset, (Barkov, 2019., Rajasekaran et al, 2022) The training set is used to build the model, validation set will be used to choose the machine learning algorithm and find the best values for hyperparameters, and test set will be used to assess the model,

To train the machine learning model, the collected data must be divided up into training and test sets, which are randomly divided by 70:30 ratio of the entire data set. (Cavalcante et al, 2019, Goli et al, 2019).

Using the train\_test\_split function in Scikit-learn, the dataset was partitioned into 80% for training and 20% reserved for testing.

* + 1. **Models Initialization**
* The model is initialized.
* The model is trained using fit on the training data.
* Predictions are made on both the training and test datasets.
* Performance metrics (MSE and R2 score) are calculated for both training and test datasets.
* Results are stored in the results\_df.
  + 1. **Cross -Validation**

for each model in the table, a 5- fold cross-validation was performed, (CV=5). The mean and standard deviation of the mean square error was calculated.

A list of the machine models is creating so that a loop can iterate over each model to perform the cross validation. The model is trained and evaluated 5 times (CV  
=5), each time with a different split of data in training and validation sets.

The scoring parameter is neg\_mean\_squared\_error which means the mean squared error will be calculated for each of the 5 crosses.

* + 1. **Evaluation Metrics** –

The predictions from the models are compared to the true values to calculate the Mean Squared Error (MSE) and the R-squared (R2) score for both training and testing datasets. The MSE provides a measure of the prediction error, whereas the R2 score provides a measure of the proportion of the variance in the dependent variable that is predictable from the independent variables.



* + 1. **Hyperparameter Tuning** –

Hyperparameter tuning is performed using GridSearchCV. This ensures that the best parameters are chosen for the model to enhance its performance.

The following steps will be used for hyperparameter tuning:

* A grid of potential hyperparameter values will be defined.
* **GridSearchCV** is used to search over this grid and find the best hyperparameters for the model based on 5-fold cross-validation.
* The best model (with the optimal hyperparameters) is then used for predictions and the same metrics as before are computed.
* The hyperparameters for each model examine are outlined below in table :

S. Farahani et al, 2021 suggested the importance to have a framework for the model adaptation and hyperparameter tuning for better model implementation. The following table also explains the reasoning behind the hyperparameters chosen.

|  |  |  |
| --- | --- | --- |
| **Breakdown per Machine Learning Model for Hyperparameter Tuning** | | |
| **Model** | **Hyperparameters** | **Comments** |
| Linear Regression | None | No hyperparameters are tuned since it's a straightforward algorithm. |
| Ridge Regression | alpha: [0.01, 0.1, 1.0, 10.0] | A common range for the regularization parameter, covering a spectrum from light to strong regularization. |
| Lasso Regression | alpha: [0.01, 0.1, 1.0, 10.0] | Like Ridge, this is a common range for regularization intensity. |
| Random Forest Regressor | n\_estimators: [100, 200, 300], max\_depth: [None, 10, 20] | A reasonable starting range. Consider testing more n\_estimators based on computational constraints. |
| Gradient Boosting Regressor | n\_estimators: [100, 200, 300], learning\_rate: [0.01, 0.1, 0.2], max\_depth: [3, 4, 5] | The interplay between learning\_rate and n\_estimators is important. Depending on resources, a broader range for n\_estimators might be considered. |
| Decision Tree Regressor | max\_depth: [None, 10, 20] | Offers a choice between a shallow tree, a deeper tree, and a fully grown tree. |
| Bagging Regressor | n\_estimators: [50, 100, 200], max\_samples: [0.5, 0.7, 1.0], max\_features: [0.5, 0.7, 1.0] | Provides variability in base estimators and how much of the dataset and features they should consider. |
| K-Nearest Neighbors | n\_neighbors: [3, 5, 7, 9], weights: ['uniform', 'distance'], algorithm: ['auto', 'ball\_tree', 'kd\_tree', 'brute'] | Covers various configurations of neighbors and algorithms to account for dataset's characteristics. |
| Support Vector Machine | C: [0.1, 1, 10], kernel: ['rbf', 'linear', 'poly'], degree: [2, 3], gamma: ['scale', 'auto'] | Incorporates various kernel functions and regularization strengths. Note that some combinations might not be meaningful (e.g., degree with rbf). |
| Simple Neural Network (Keras) | dense1\_neurons: [32, 64, 128], dense2\_neurons: [16, 32, 64], epochs: [30, 50], batch\_size: [16, 32, 64] | Covers various neuron counts for two dense layers and different training strategies. |
| LSTM (Keras) | lstm\_neurons: [30, 50, 70], batch\_size: [16, 32, 64], epochs: [30, 50, 100] | Tuning for LSTM models can be computationally intensive. This setup offers flexibility in LSTM neuron counts and training parameters. |
| Simple Dense Neural Network | Layers: [128, 64, 32], epochs: 50, batch\_size: 32 | A basic three-layer feedforward neural network. |
| Dense Neural Network (Optimized via RandomizedSearchCV) | batch\_size: [16, 32, 64], epochs: [20, 50, 100], neurons\_layer1: [64, 128, 256], neurons\_layer2: [32, 64, 128], neurons\_layer3: [16, 32, 64] | Random grid search over different configurations, sampling 10 combinations. The goal is to find an optimal architecture and training configuration for the neural network. |

* + 1. **Individual model training –**

For predictive modelling, each model goes through a structured process of training, prediction, and evaluation. The process ensures that models are validated on unseen data to provide a reliable measure of their expected performance on new data.

**Training Process:**

Linear Regression, Ridge Regression, Lasso Regression, Random Forest Regressor, Gradient Boosting Regressor, Decision Tree Regressor, Bagging Regressor, AdaBoost Regressor, SVR, KNN, and Neural Networks:

**Training:** The model is trained using the fit() method with the scaled training dataset.

**Prediction**: Once the model is trained, predictions are made on both the training dataset and the testing dataset. This helps in understanding

**Evaluation:** This was facilitated using the metrics module in Scikit-learn the predictions from the models are compared to the true values to calculate the Mean Squared Error (MSE) and the R-squared (R2) score for both training and testing datasets. The MSE provides a measure of the prediction error, whereas the R2 score provides a measure of the proportion of the variance in the dependent variable that is predictable from the independent variables.

**Neural Networks require additional steps:**

**Architecture Definition**: The architecture of the neural network is defined, including the number of layers, neurons in each layer, and activation functions.

**Compilation**: The neural network model is compiled specifying the loss function, optimizer, and evaluation metrics.

**Training**: The model is then trained using the fit() method. Hyperparameters such as the learning rate, batch size, and number of epochs can be adjusted based on performance requirements.

**Summary:**

Through this structured training process, each model is systematically trained, tested, and evaluated. The results provide insights into the performance of each model and help in selecting the best model for predictions on new, unseen data.

**3.8 Limitations**

The study's potential limitations encompass:

A strong reliance on the accuracy and consistency of data extracted from the XYZ production plant through the PlantLog V2.3 system.

The inherent assumption that the highlighted features have a linear impact on the product quality.

**3.9 Conclusion**

This chapter presented a comprehensive methodology for leveraging machine learning to enhance production parameters. The following chapter will discuss the results obtained from implementing this methodology.

**5.1 Introduction**

This chapter explains the impact of the primary data and secondary data on the research objectives, through the information received from the three expert participants and the production batch data downloaded from the organizations database system.

The author examines various machine models to determine if the production downtime as phase overrun variable can be predicted for the various production tanks used to produce mucilage containing beverage materials. It leverages historical data and relationships between features and a target variable to determine if predictions about phase downtime in the future is possible. The downtime was examined for three phases of the production: Agitation, Deaeration and the addition of Gum ingredient.

**5.2 The Primary Data**

From the in-depth interviews, responses from the participants with expert knowledge and distinct perspectives in the research area was reviewed and summarised here. The actual interview transcripts are in appendices. .

**5.2.1 Participant no. 1: Data Scientist**

Participant number 1, a Data Scientist, played a crucial role in enhancing the existing system's objectives. Initially, the system solely focused on executing batches, with data collection lacking a structured approach for analytics. To rectify this, the participant was involved in refining tracking mechanisms by developing a stored procedure that summarized batch data comprehensively, covering aspects like start times, consumption, problems, weights, agitation times, deaeration times, and homogeneity. They further established targets for different phases through historical data analysis, providing clear objectives for each batch and aiding in issue identification when batches fell short. Resource management was addressed by quantifying departmental needs for shared resources like mobile tanks. Operational Efficiency (OE) calculations, focused on equipment effectiveness rather than usage, were employed to pinpoint areas for improvement. The participant also highlighted the company's emphasis on data analytics, mentioning the creation of a Digital Performance Management (DPM) system for data visualization and future consideration of machine learning where it genuinely adds value. Moreover, they stressed the significance of preventative maintenance, citing examples of early issue detection through thermal balance tracking and the importance of team awareness and training in efficient problem resolution.

The key insights from the Data Scientist's contributions revolved around enhancing batch process tracking, target setting, issue resolution, and emphasizing the role of data analytics and preventative maintenance in optimizing organizational processes.

**5.2.1 Participant no. 2: Production Manager**

Participant 2 highlighted the company's data-driven approach, where they utilize operational data to calculate Overall Equipment Effectiveness (OE) figures for each batch, which are then reviewed during daily 9:15 meetings involving cross-functional teams. They stressed the importance of acting based on these figures, with diligent follow-up to ensure completion. OE calculations are performed at the phase level, considering prerequisite times, enabling a granular assessment of phase efficiency. The interviewee discussed various batch phases, such as agitation, deaeration, and ingredient additions, outlining their timing and criteria for proceeding to the next phase. Quality checks before batch finalization were emphasized to meet quality standards. Initiatives to reduce downtime were mentioned, including pre-weighing bulk powders, and optimizing temperature and pump speed. Downtime challenges like liquid filling delays and breakdowns were addressed. Recording downtime reasons and their impact on OE data were explained. Setting targets for parameters like flow rate involved collaboration between different teams. Tank assignments, numbering, and capacity details were provided, with some batches requiring multiple tanks. Different systems, ranging from 20-tonne to medium-to-small systems, were discussed based on production volume needs. An efficiency improvement example involving juice barrel handling was presented. The interviewee reiterated the importance of continuous collaboration, data analysis, and process enhancement to minimize downtime and boost productivity.

**5.2.1 Participant no. 3: Production Operative**

From the participant's point of view, the conversation primarily revolved around production downtimes and efficiency in a manufacturing process, specifically related to gum mixing. The participant discussed the challenges of interruptions during production, which could result in extended downtimes, impacting operational efficiency (OE). They emphasized the importance of accurately recording extra downtime for batches where gum mixing takes longer due to poor agitation, as this data is crucial for analysis. Additionally, the participant acknowledged that target times for production phases should be more realistic to account for batch-to-batch variations and highlighted potential solutions like improving agitation and pre-mixing highly concentrated gum batches. The conversation also touched on the phases of a typical batch, including ingredient addition, agitation, and transitioning to the high-pressure phase.

The key insight from the participant's point of view is the focus on addressing production downtimes and operational efficiency challenges in gum mixing, emphasizing the need for accurate data recording, realistic target times, and potential solutions to improve the manufacturing process.

**5.2 The Secondary Data**

**5.2 The Dataset**

The original production csv file contained data on beverage batches that contain a mucilage ingredient produced in various production tanks with differing capacities over time of 2 years. In compliance with the organisations data usage agreement, specific data points have been omitted to maintain the confidentiality. Table 1 below provides an overview of the dataset’s column data and explanations. The dataset is called. ProductionDateupdated1



**Table 1 Dataset Column Headings and Detail**

Over 46 different type of beverage mucilage containing materials with a total of 367 beverage batches were produced across 16 production tanks of varying capacities. Which tank they were produced in; was dependent on the type of material and Quantity required. For the purposes of this data analysis, The batches will be examined, as not all materials would be produced in each tanks.

The dataset, ProductionDataupdated1is loaded into the relevant notebook, and depending on the production Tank group and instruction step investigated, the dataframe will be transformed via aggregation to show the relevant data for the specific data analysis and is named according to the production tank grp reference number. e.g., productionTank22\_df1

The following analysis is performed on all Production tanks groups for all phases, Deaeration, agitation and gum addition:

1. Exploratory Data Analysis
   * Univariate Analysis:
   * Bivariate and Multivariate Analysis:
   * Handling Outliers:
   * Data Standardization: Standardized data using Z-score normalization.
2. Model Evaluation for Predicting Phase Overrun. Production Tanks
   * Data Details
   * Model Performance

Overall, the analysis provided insights into production tanks, with variations in phase metrics for different materials. Different machine learning models were evaluated for predicting phase overrun for various phases, with each phase presenting varied model performance.

**5.2.1 Beverage Process Production Tanks (MT- Mobile Tanks)**

Table 2 shows the division of materials and their batches produced per production tank over 2 years. Some tanks are utilised more than others, this is determined based on the production schedule. Some of the production tanks are used as destination or holding tanks and may not be used for beverage production.

The production tanks were grouped together according to their capacity and the number of batches produced. The number of batches produced per tank over the 2-year time selected was low and would be an issue for the performing any accurate machine model on prediction. It would limit the training of the model and therefore affect the accuracy of the prediction.

For example, for production tank 22MT02, the number of batches produced was 13 which is a small number to perform any data analysis on, whereas for all the tanks in the group 22MT, the total number of batches were 47.



**Table 2 Production Tank /Materials/Batch Dataset Details**

**5.2.2 Production Phases (Instruction Steps)**

Table 3 below details what each Production Phasesmeans in the production of mucilage containing batches. The list of instruction steps is common to all batches detailed for each of the production tanks. So, for some materials, an extra HP (homogenisation may be required). This list of instructions or recipe is automatically activated at the start of the batch production.

For each of the instruction steps, the phase start time and finish time is logged. This data is converted into the total phase duration time, the start of the phase time, the phase overrun time which is all calculated based on the Target times logged into the system in the background. These target times for the phase duration and overrun times are historically determined and applied to each material. Each ingredient addition (Step1\_cons), the quantity and flowrate are also logged.

**Table 3 Instruction Steps**



|  |  |  |
| --- | --- | --- |
| **Instruction\_Step/Phase Step** | **Details** |  |
| S3\_BATCH\_IN\_PROGRESS | Batch process is in progress inside the tank. |  |
| STEP1\_CONS | 1st step in the consumption process where ingredients or materials are added to the tank: Treated Water |  |
| PLEASE VERIFY BULK ADDITION | A prompt to check and verify the bulk addition of materials or ingredients. |  |
| STEP1\_CONS |  |
| STEP1\_CONS |  |
| STEP1\_CONS |  |
| STEP1\_CONS |  |
| STEP1\_CONS |  |
| STEP1\_AGITATION | The agitation or mixing process that takes place after the first set of ingredients is added to ensure uniform distribution. |  |
| STEP2\_AGITATION |  |
| STEP3\_AGITATION |  |
| HP | High-Pressure phase or operation: homogenize the batch. |  |
| SELECT\_DESTINATION\_TANK | Indicates a phase where the finished or semi-finished product is transferred to a different tank, |  |
| S4\_BATCH\_COMPLETE\_QA\_PENDING | Signifies that the batch processing is complete and is now pending Quality Assurance (QA) checks. |  |
| TAKE A SAMPLE AND SUBMIT FOR QA. | An instruction for the operator to take a sample from the batch and submit it for quality assurance testing. |  |
| SAMPLE TO LAB. RESULTS OK? (NO TO HOMOGENISE) | A prompt suggesting that the sample has been sent for testing. |  |
| STEP8\_AGITATION | The agitation or mixing process |  |
| S7\_RELEASED\_TO\_FILLING | Indicates the batch has passed all processing stages and is now ready or has been released for filling |  |

**5.2.3 Ingredients and their Quantities**

The ingredients and their quantities are important to the phase overrun times, any delay in their addition reflects in the phase duration times. The table below lists the main ingredient components in the batches in the dataset. The quantities are approximated average, they depend on the recipe for each of the batch produced. The main ingredients that have the most quantities are treated water which is delivered via the automated bulk delivery system, and the mucilage /gum ingredient. The dry ingredients are all added manually via bags through the manifold on the top of the tank. So, for the gum ingredient, this is very labour intensive. Not every batch has a colour ingredient addition, but for those that do, there is a significant quantity to be added.



**Table 4. Ingredients Categories /Quantities**

**5.4 Data Preparation and Preliminary Analysis**

To prepare the data in a way that maximises the performance of the subsequent machine learning models using data pre-processing and feature selections.

The Original DataFrames which given actual details of the beverage containing mucilage batch production was reviewed, and some data was removed as it was deemed unnecessary for the purpose of answering the objectives of this study. But most importantly removed to preserve the anonymity of the data. Column heads were updated for easy of manipulation. Once this was performed, the dataset was saved and used for the research, reference Productiontankupdated 1. The total number of columns are 22 and rows at 9487. Initially exploratory data analysis was performed to determine the relationships and any trends between the variables.

The beverage batch data from each group of production tank were examined under the following instruction step/phases and these investigations were saved as follows also:

* Production Tanks: All Phases
* Production Tanks: Agitation Phase/
* Production Tanks: Deaeration Phase/
* Production Tanks: Gum Addition Phase

Each section above was prepared, and preliminary analysis was performed using pythons’ pandas library. The data processing step includes data cleansing and transformation, to ensure the quality and accuracy of the data.

* Data Cleansing was completed and involved removing or correcting errors, inconsistencies, and duplications in the dataset to ensure its accuracy and completeness. ensuring the quality of the data used for analysis.
* Data transformation was completed and involves converting the data into a usable format and structure for analysis. This includes aggregating data, splitting data, or merging data from different sources to create a unified dataset relative to the production tank and phase was being examined.

**5.4.1.1 Handling Missing Values:**

There were missing values in some of the rows of data. This was to be expected as the rows were for each of the instruction steps/phases for each batch produced. So, for example for rows relating to agitation, there wouldn’t be any values for ingredient addition, or quantities. The entries were in the form of Nan (not a number) or just 0. For each of the notebook investigations, the missing values were checked, and values changed to a 0. as the missing value signified that there were no occurrences/events, and zero is an appropriate replacement.

* + - 1. **Duplicate Values**
* There were no duplicate values in the dataset.

Here is a summary of the steps taken : The dataset, ProductionDataupdated1is loaded into the relevant notebook, and depending on the production Tank group and instruction step investigated, the dataframe will be transformed via aggregation to show the relevant data for the specific data analysis and is named according to the production tank grp reference number. e.g., productionTank22\_df1

The following analysis is performed on all Production tanks groups for all phases, Deaeration, agitation and gum addition:,( Komorowski et al , 2016, Mukhiya et al , 2020)

1. Exploratory Data Analysis
   * Univariate Analysis:
   * Bivariate and Multivariate Analysis:
   * Handling Outliers:
   * Data Standardization: Standardized data using Z-score normalization.
2. Model Evaluation for Predicting Phase Overrun. Production Tanks
   * Data Details
   * Model Performance

**5.4.2. Exploratory Data Analysis : productiontankupdated1**

**5.4.2.1 Univariate Analysis:**

* **Numerical Features**:

1. The seaborn library(‘sns’) was used to visualise distribution of batches via a count plot overall production tank, (Fig 1). The height of each bar corresponds to the number of records for each tank. The 25MT03 and 04 tanks produced the most batches, they have a capacity of 10 tonne.

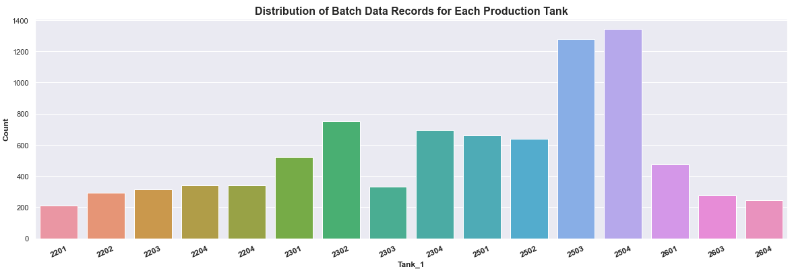
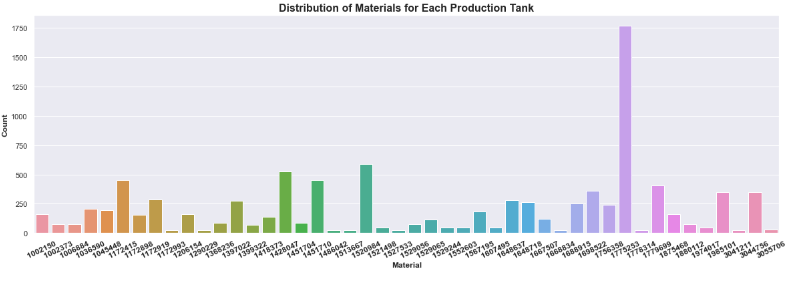


Fig 1 Distribution of Batches for each Production Tank

1. Fig 2 shows the distribution of material data records for each tank and allows you to see which materials were produced the most. This was beverage batch material 1756358 – produced 61 times.



**Figure 2: Distribution of materials overall production tanks**

**5.4.2.2 Multivariate Analysis:**

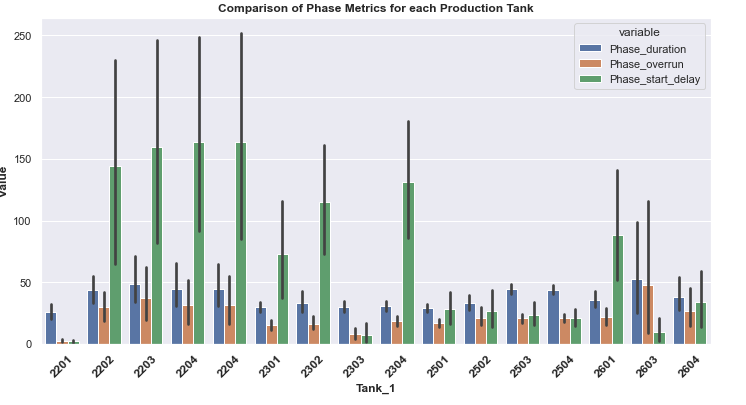
The Bar Chart, in Fig 3, gives visual representation of how phase duration, phase overrun, and phase start delay varies across the different production tanks. Each bar represents a tank, and the height is the phase metric for that tank. All the 22MT and 23MT tanks which have the capacity of 20 tonne, show the highest values for the metric phase start delay, i, e it takes longer for the phase to start. for a production of a batch. In comparison to the lower capacity tanks of 25 and 25 MT tanks.

Phase duration metric is higher for 22MT than it is 23 MT tanks, showing that the batch production is slower in the 22MT tanks than 23MT tanks. There is phase overrun times (downtimes) for every production tank but higher in the 20 tonne tanks in specific 22 MT.

Notable outliers are Tanks 22MT01 and 23MT03 for low phase overrun, however these tanks were destination tanks and only a small number of batches were produced. In line with the information from participant 3, the production operative, he mentioned that tanks 22MT01 and 23 MT05 were the idea tanks to use for production, this is show in the graph where the 3 variables are the lowest.

This chart gives an overview of the tanks, however the number of batches per tank need to be considered, giving from table 2 earlier, more batches were produced in the 25MT tanks, yet the phase duration, overrun and phase start delay metrices were all lower than for those tanks with bigger capacity and less batches produced. This is the reason why the production tanks will be examined in their capacity groups and not individually.

Another important point is the quantity of mucilage gum ingredient material is much higher in batches that are produced in the tanks 22 MT and 23 MT , where there is the high phase start delay times and phase overrun times. This highlights the issues raised by the interview participants where they noticed a longer production times for mucilage containing batches where the higher the ingredient quantity the longer the agitation times and then the longer deaeration times. The other production tank groups have smaller capacity and gum ingredient required is much smaller .



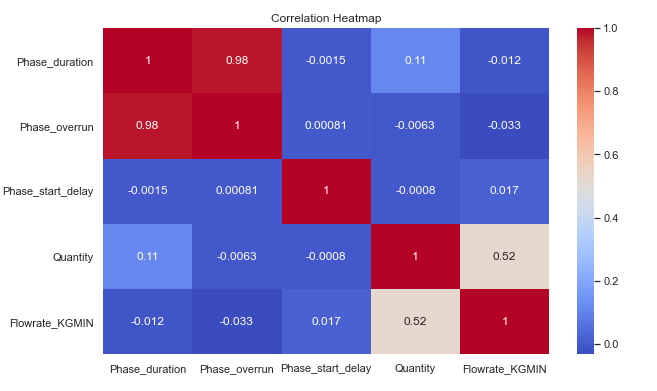
**Figure 3: Comparison of Phase Metrics for each Production Tank**

**5.4.2.3 Correlation Analysis**:

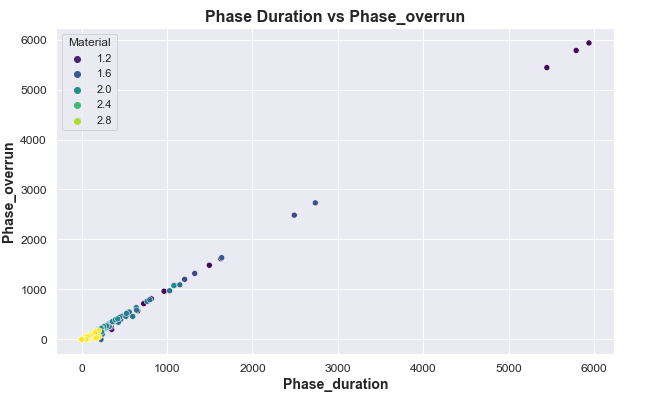
The main objective of this research is to determine the downtime for the production tanks. The variable phase overrun is representative of the downtime that can be experienced by beverage mucilage batch production. Phase start delay is another variable that be measured as downtime.

The correlation heatmap was created to determine the relationships between the variables. The values closest to I have the strongest positive correlation which is seen in the relationship between the phase overrun and phase duration variable, value of 0.98. Another close relationship highlighted is batch quantity and the flowrate, 0.52. There are no other high intercorrelations among the independent variables, therefore ruling out the multicollinearity.

To further examine the relationship between the phase duration and phase overrun, a scatterplot was created, and linear relation was observed, figure 4



**Table 5. Correlation Heatmap**



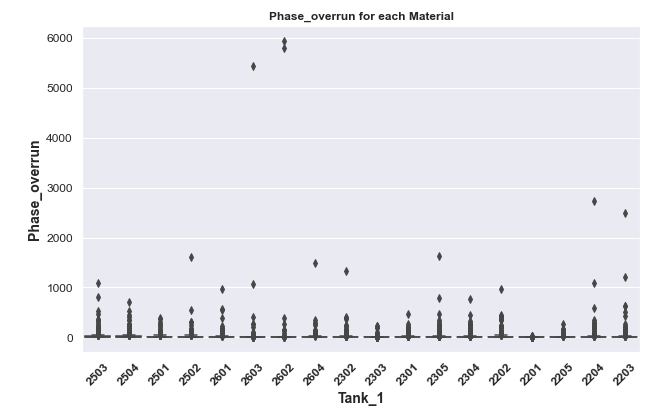
**Figure 4: Phase Duration vs Phase Overrun for all materials produced in all Tanks.**

In conclusion there is a strong relationship between the phase overrun and phase duration. Given the high correlation between phase duration and phase overrun, a linear regression model might be a good starting point. This allows prediction of the phase overrun based on the phase duration.

In comparison there is only a moderate relationship between batch Quantity and flowrate, value of 0.52, it indicates that there are other factors affecting the tank flowrate.

**5.4.2.4 Handling Outliers**

To determine the presence of outliers, boxplots were utilised. Figure 7 displays the distribution of data overall the production tanks for the phase overrun variable. There is the presence of outliers for each tank, with points present outside the range. Under each investigation of the selected phases and production tanks, outliers were assessed and removed prior to machine modelling. Table below shows a summary of the no. of outliers removed from each group production tanks for each batch produced.



**Figure 7: Boxplot of Phase Overrun for each production tank.**



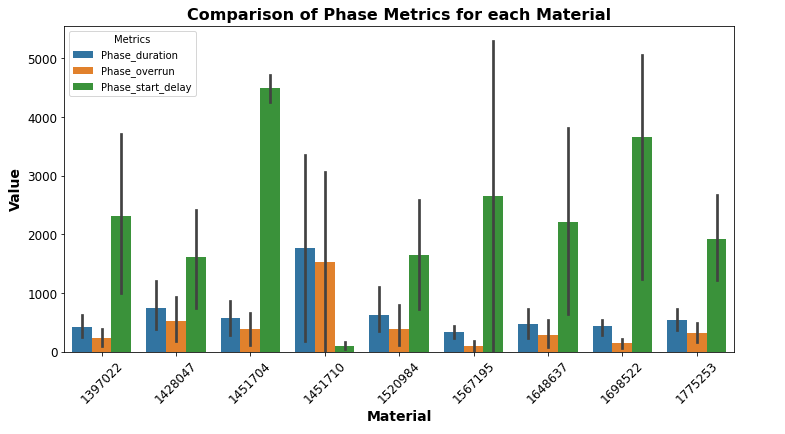
1. **Exploratory Data Analysis for 22 MT. production Tanks**
2. **Model Evaluation for Predicting Phase Overrun for 22 MT. production Tanks for the All phases.**

|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **All phase** |
| **Production Batches** | 73 |
| **Production Batches after Outlier removal** | 59 |
| **Target Variable** | Phase Overrun |
| **Instruction steps/Phases per batch** | 27 |

NB: For each batch produced, there are two production tanks used and thus included in this analysis. One for the actual production and the other for storage as in the destination which has itself a phase step and a phase overrun metric.

**Univariate Analysis: Numerical Features**:

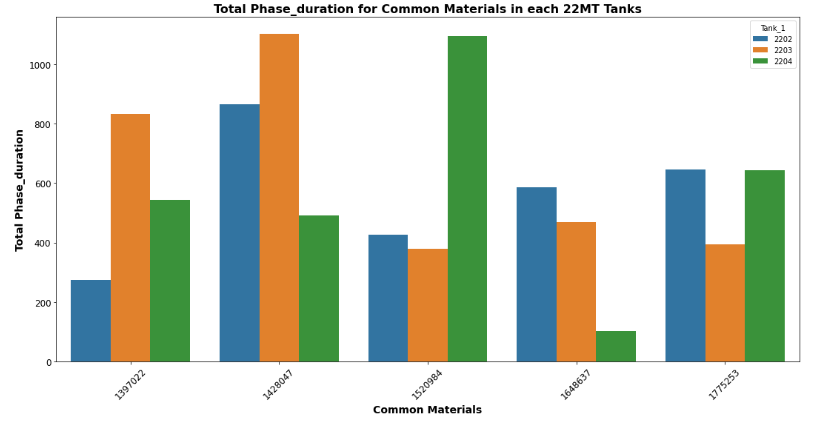
The comparison of Phase Metrics for each Material for the 22MT TanksThis histogram gives a performance overview of each of the 22MT tanks, looking at common materials that were produced. Each material produced in the 22MT tanks experienced phase start delay, phase overrun, and the production time was different between all tanks.



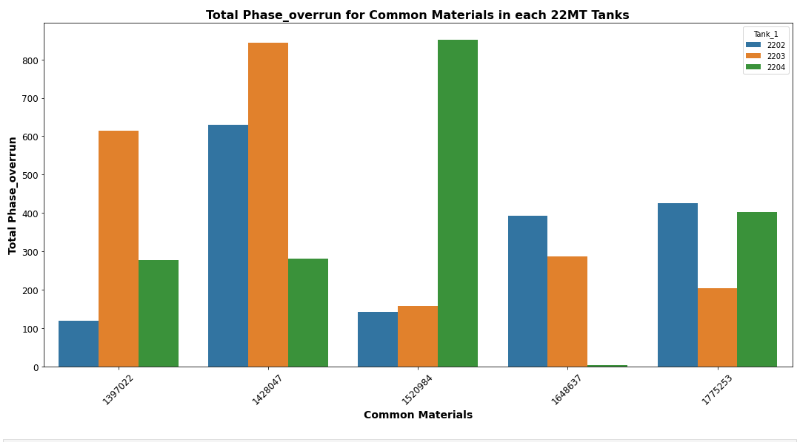
**F****igure 8: Bar chart of 3 phase matrix phases for each material in the 22MT production tanks.**

The phase duration and phase overrun for common materials in each 22 MT tanks are shown in figure 9 and 10 below, its shows that there is difference in production time for materials between all 3 production tanks. For example, for material 1297022, the histogram shows that it would be better to produce in tank 2202 as the phase duration is the lowest and the phaser overrun time is lowest for this tank. Important to note, that for each of the materials there is downtime logged, phase overrun.

In relation to the interview with participant no. 3 the production operative, he mentioned one problematic material, reference no. 1428047. This material is produced in the 22 MT tanks with their 20-tonne capacity, from the graph, the phase duration time and phase overrun times are high. When looking at the instruction steps for these materials, there is two high quantity additions of the mucilage material. requiring two agitation times, but the production tanks 22 MT show differences in the phase duration times for each tank with 2204 tank being the better tanks to use for fast production time and lower overrun times.

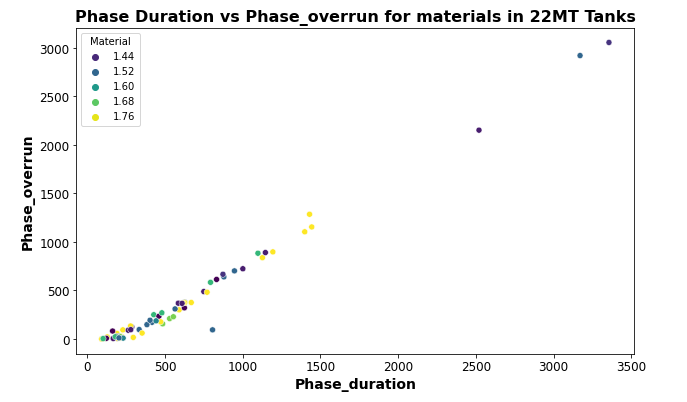


**Figure 9: Bar chart of phase duration phases for each material in the 22MT production tanks.**



**Figure 10: Bar chart of phase overrun phase for each material in the 22MT production tanks.**

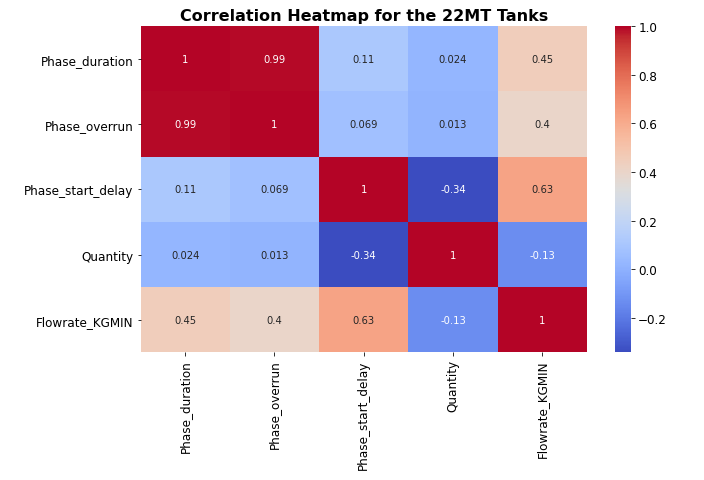
Figure 11 highlights the linear relationship between the target variable phase overrun and phase duration time for all 22 MT production tanks.



**Figure 11: Scatter chart of phase overrun phase vs phase duration 22MT production tanks.**

**Bivariate and Multivariate Analysis:**

**Correlation Analysis**:

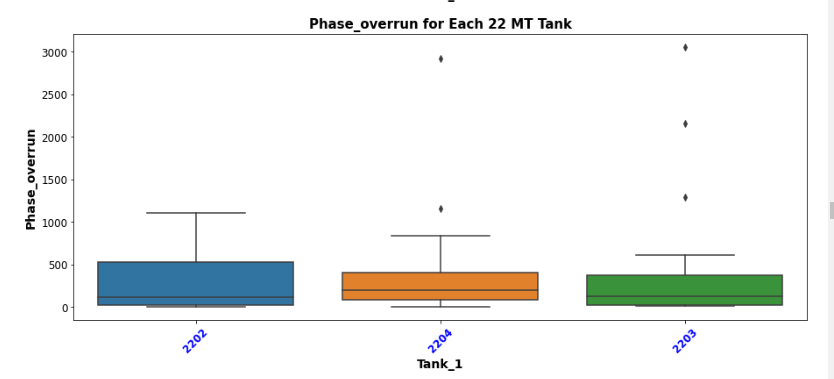


**Figure 12: Correlation analysis heatmap for 22MT production tanks.**

There was high correlation between the Phase overrun and phase duration variables, 0.99.

**Handling outliers**

Figure 13 a boxplot of the phase overrun batch data for each of the 22MT production tanks. There is the presence of points above the highest observations, the horizontal line call whiskers. Also majority of data are distributed above the median line for all three 22 MT tanks



**Figure 13: Boxplot for each 22MT production tanks. For the target variable Phase overrun**

The Interquartile range (IQR) method was used to remove outliers from the production tank data.

There were 14 rows of production data removed from the dataframe as potential outliers.

**Data standardization**



Based on the descriptive statistics for each of the production tank batch data, there was wide variation in the data for phase duration, phase\_start\_delay and the phase overruns., which can affect certain algorithms. Based on this, the data was standardised using the Z-score normalization using StandardScaler from Sklearn.preprocessing library. Another advantage is that its consistent and highly interpretable.

**Applying Machine Learning Models**

Various Machine learning models were applied to the dataframe, reference: Production 22\_df1. Full details of these are in the appendix.

Results for each of the investigations below show the top and poor performing models. Visual results of the top performing models are given in the predicted vs actual graphs and the residual plots.

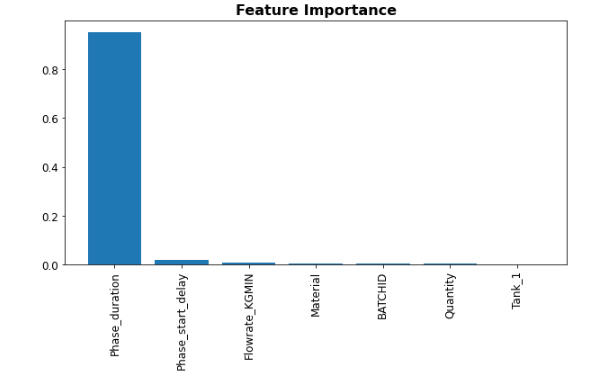
1. **Model Evaluation for Predicting Phase Overrun for 22 MT. production Tanks for the All phases.**

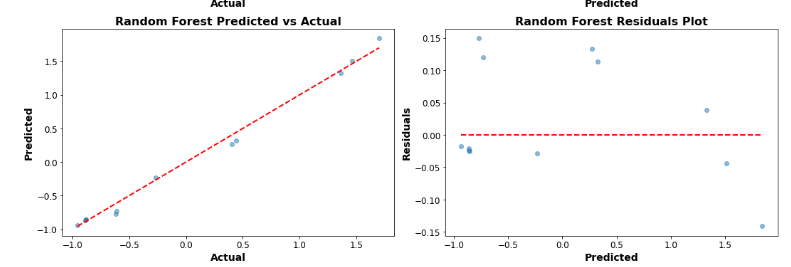
|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **All phase** |
| **Production Batches** | 73 |
| **Production Batches after Outlier removal** | 59 |
| **Target Variable** | Phase Overrun |
| **Instruction steps/Phases per batch** | 27 |

NB: For each batch produced, there are two production tanks used and thus included in this analysis. One for the actual production and the other for storage as in the destination which has itself a phase step and a phase overrun metric.



Random Forest Regressor achieves almost perfect scores on both the training and testing datasets before applying tuning or cross validation. This regressor model provides stability and is less likely to overfit. Another of its advantage is interpretability of feature importance, the graph below shows which feature phase duration as the driver of the predictions. The scatterplot and residual plot show good model performance, with no obvious outlier points.







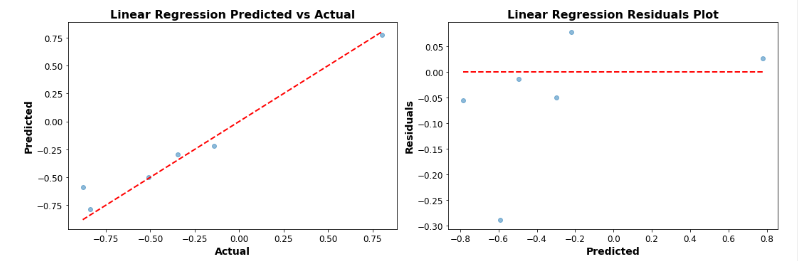
The poorest performing model was the LSTM Neural Network with a poor accuracy score R2 of -0.14 for R^2. This is highly likely due to the target variable not been a time series or text format. LSTM require a lot of data for the training, and this could affect the hyperparameter choices as the top performing models are much simpler models, the complexity of using LSTM is redundant.

1. **Model Evaluation for Predicting Phase Overrun for 22 MT. production Tanks for the deaeration phases.**

|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **deaeration phase** |
| **Production Batches** | 38 |
| **Production Batches after Outlier removal** | 30 |
| **Target Variable** | Phase Overrun |



Given the results, the linear regression model, a perfect accuracy score R2 of 1.00 on training data and a score of 0.94 on test data. This indicates that this model can explain 100% of the variance in the training dataset and 94% in the test dataset. This indicates that these models can explain 100% of the variance in the training dataset and 94% in the test dataset, reflecting an excellent fit. Furthermore, the Mean Squared Error (MSE) for these models is 0.00 for both training and test datasets, showcasing the model's precision in estimating phase overrun in production tanks. The consistent performance in both tuned and untuned metrics shows the strength of linear regression model in capturing the linear relationships in the given dataset for production tank group 22MT, making it a prime choice for predicting phase overrun for the deaeration production phase. The accuracy is shown in the following graphs





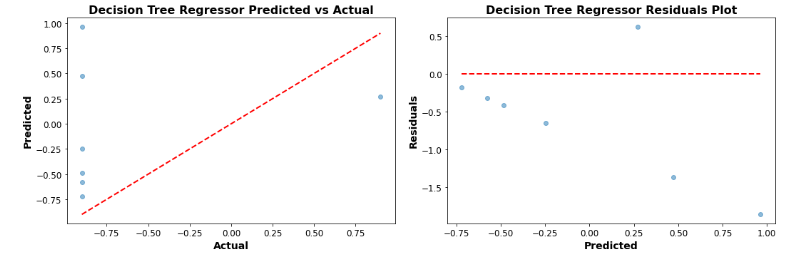
The Long Short-Term Memory (LSTM) Neural Network, a type of Recurrent Neural Network (RNN), presented mixed results when predicting phase overrun in production tanks. The training values R2 were notably negative, standing at -0.91 and -0.79 for the untuned and tuned models respectively. This indicates that the LSTM model failed to capture the underlying patterns in the training data. On the test side, the values further declined to -1.24 and -1.10 for the untuned and tuned models, respectively, suggesting the model's predictions were worse than a basic horizontal line mean prediction. Additionally, the high Mean Squared Error (MSE) values, 1292.83 for training and 424.13 for testing in the untuned model, underscore its lack of precision. While tuning did bring about some improvements, they were marginal. In this context, the LSTM model appears unsuitable for the task, possibly due to its inherent complexity and tendency to model sequential or time-dependent data, which might not align with the nature of the provided dataset.

**Model Evaluation for Predicting Phase Overrun for 22 MT. production Tanks for the Agitation phase phases.**

|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **Agitation phase** |
| **Production Batches** | 47 |
| **Production Batches after Outlier removal** | 34 |
| **Target Variable** | Phase Overrun |



The Decision Tree Regressor top performing model when assessing the Agitation Phases for Production Tanks 22MT. Both the tuned and untuned models presented good results with a Train and Test score R2 of 1.00, indicating that the model perfectly explained the variance in the target variable. The Mean Squared Error (MSE) for both training and testing was impressively low at 0.00 and 0.01 respectively, highlighting the model's precision. These results suggest that the Decision Tree Regressor, without any need for parameter tuning, can capture the inherent patterns and relationships in the dataset with high accuracy.





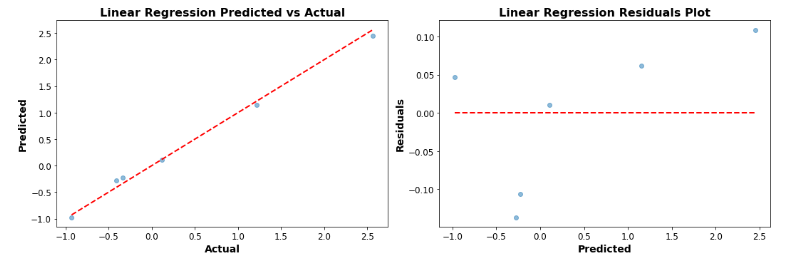
On the other hand, the Dense Neural Network (FCN) displayed the weakest performance among all with a R 2 -of0.72 which declined significantly for the test data, registering a poor R 2 of -2.14. Tuning the model parameters led to some improvements in the training. R 2, but not the test scores. The negative values, R 2, especially for the test data, point towards the model's inability to predict the phase overrun in Production Tanks 22MT reliably. Moreover, the high MSE values reinforce the model's lack of precision in its predictions. Despite the inherent capabilities of neural networks, in this context, the Dense Neural Network failed to generalize well to unseen data.

**Model Evaluation for Predicting Phase Overrun for 22 MT. production Tanks for the Gum addition phase**

|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **Gum addition phase** |
| **Production Batches** | 43 |
| **Production Batches after Outlier removal** | 29 |
| **Target Variable** | Phase Overrun |



For the Gum Addition process in Production Tanks 22 MT, the Linear Regression model demonstrated superior performance. Both the tuned and untuned variants exhibited excellent predictive capabilities, as evident from the Train and Test R 2 scores which consistently hovered around 0.98 and 0.99 respectively. These scores suggest that the model was able to explain almost 99% of the variance in the test data. The MSE values further bolster the model's credibility with figures as low as 0.02 for training and 0.01 for testing, indicating accurate predictions with minimal error. In this context, Linear Regression, without any need for parameter tuning, seems to adeptly capture the underlying relationships in the dataset, offering a reliable model for predicting outcomes in the Gum Addition process. This is visualised in the graphs below.

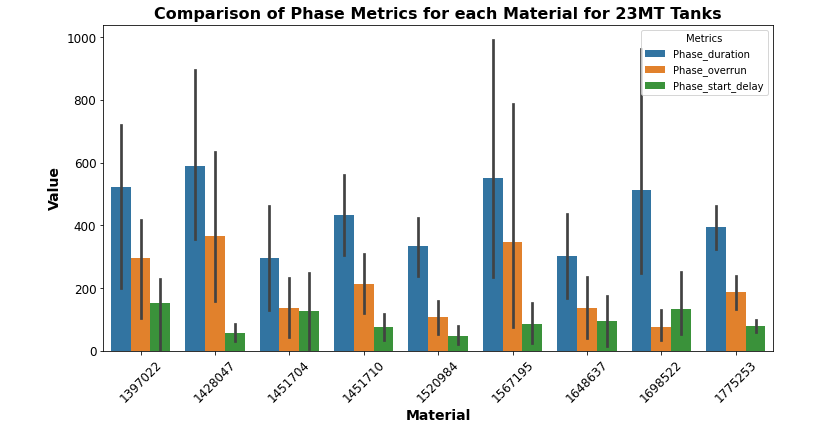




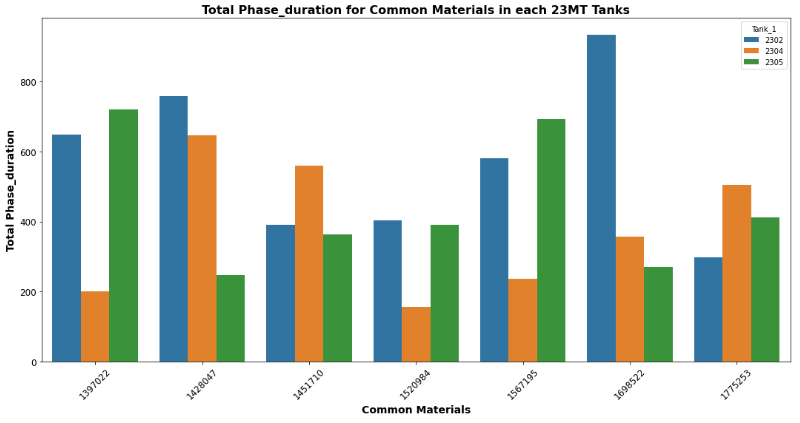
On the opposite end of the spectrum, the Simple Neural Network model proved to be the least efficient in predicting outcomes for the Gum Addition process. While the untuned model presented a discouraging Train R 2of -0.71, the situation worsened after tuning, dropping the score to -0.47. This negative indicates that the model's predictions are worse than simply predicting the mean of the target variable. The test data didn't fare much better with R 2scores of -1.10 and -0.82 for the untuned and tuned models respectively. Furthermore, the high MSE values, especially in the test data, emphasize the model's imprecision. Despite the inherent power of neural networks, in this specific scenario, the Simple Neural Network seems ill-equipped to generalize or make accurate predictions.

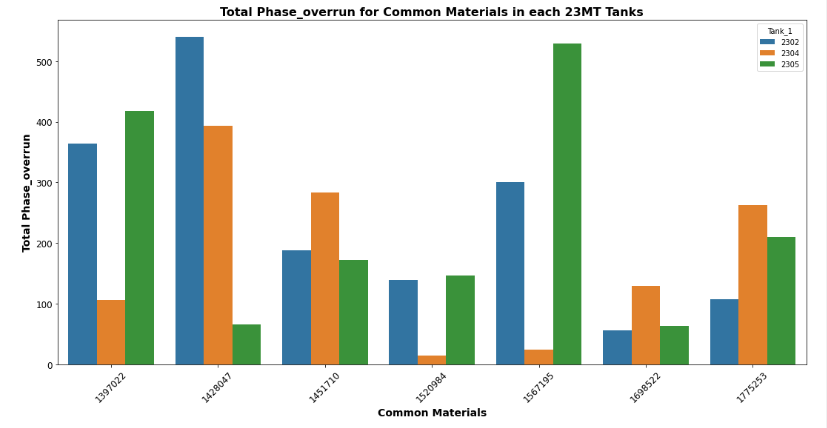
**Exploratory Data Analysis Production Tank 23MT**

**Univariate Analysis: Numerical Features**:



For all the materials produced in the 23 MT production tanks , the bar chart above shows that there was little delay in the starting of the production phases for all tanks , however again there was phase overrun downtimes noted

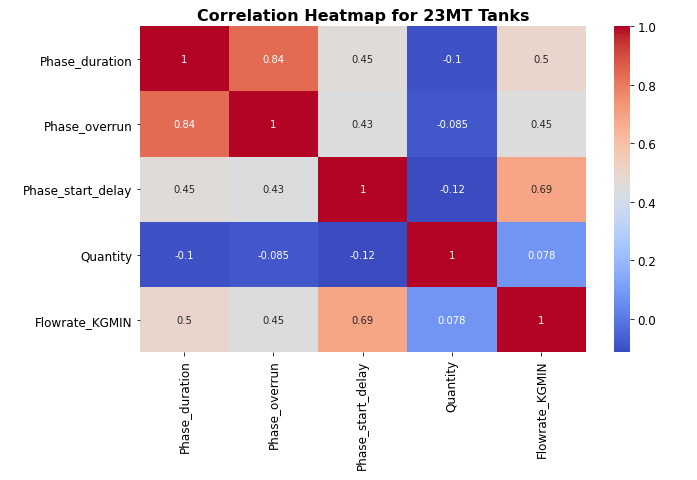




Bar Chart giving visual representation of how phase duration and phase overrun varies for the common materials produced across the 23MT production tanks. Each bar represents a tank, and the height is the phase metric for that tank. There is a lot of variability, and each material would have to reviewed separately, but for material 1567195, it seems that tank 23MT 04 would be the better tank to use, as the phase overrun times were lower in this tank. In general, for tanks that have the same capacity and materials with the same quantity and ingredient addition, there is a lot of variation in the production time and the down time.

**Bivariate and Multivariate Analysis:**

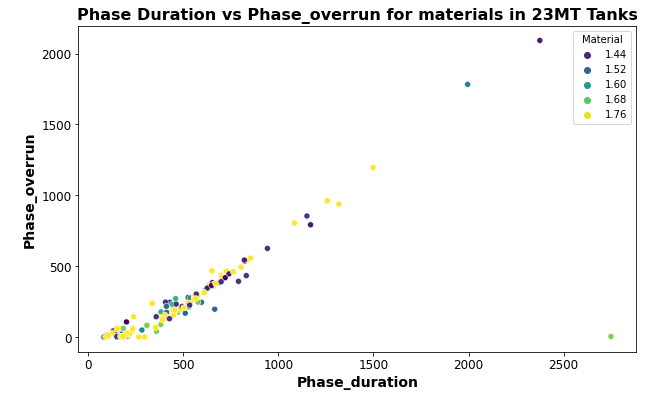
**Correlation Analysis**: A correlation matrices for numerical features to determine if there were any linear relationships between the variables.



The correlation heatmap was created to determine the relationships between the variables. The values closest to I have the strongest positive correlation which is seen in the relationship between the phase overrun and phase duration variable, value of 0.84. Another close relationship highlighted is phase start delay and the flowrate, 0.69.

To further examine the relationship between the phase duration and phase overrun, a scatterplot was created, and linear relation was observed, figure 4

**Pair Plots & Scatter Plots**: For visual inspection of relationships and trends between pairs of numerical features. Looking at the relationship between the phase duration vs phase overrun times , which is linear as per the results of the correlation chart



**Handling outliers#**

**Boxplots**

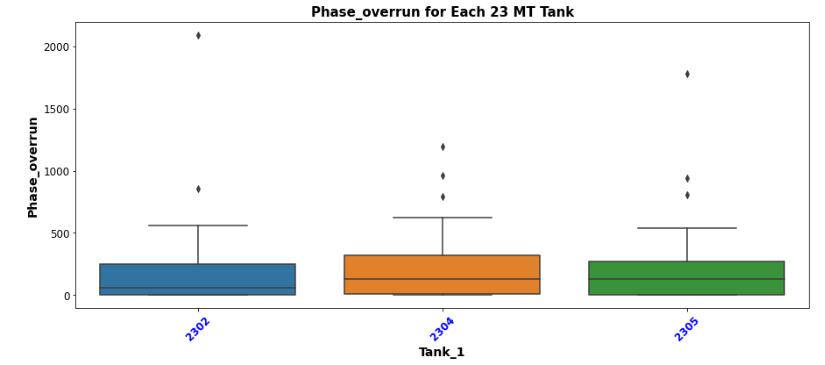


Figure 13 a boxplot of the phase overrun batch data for each of the 23MT production tanks. There is the presence of points above the highest observations, the horizontal line call whiskers. Also, majority of data are distributed above the median line for all three 23 MT tanks. The Interquartile range (IQR) method was used to remove outliers from the production tank data. There were 7 rows of production data removed from the dataframe as potential outliers.

**Data standardization**



Based on the descriptive statistics for each of the production tank batch data, there was wide variation in the data for phase duration, phase\_start\_delay and the flowrates.

These wide variations could impact the performance of certain machine learning models, a single feature disproportionately in scale could influence models that use distance to compute. Based on this, the data was standardised using the Z-score normalization using StandardScaler from Sklearn.preprocessing library. Another advantage is that its consistent and highly interpretable.

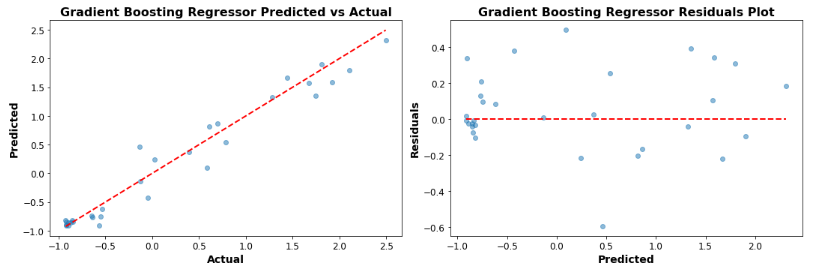
**Model Evaluation for Predicting Phase Overrun for 23 MT. production 23 Tanks for the Agitation phase.**

|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **All phase** |
| **Production Batches** | 162 |
| **Production Batches after Outlier removal** | 39155 |
| **Target Variable** | Phase Overrun |
| **Instruction steps/Phases per batch** | 27 |

NB: For each batch produced, there are two production tanks used and thus included in this analysis. One for the actual production and the other for storage as in the destination which has itself a phase step and a phase overrun metric.



The Gradient Boosting Regressor is the star performer for the Production Tanks 23 MT dataset. Even without tuning, it exhibits a pristine R 2score of 1.00 on the training set, indicating a flawless fit. Upon tuning, this model maintains its high fidelity on the training data and demonstrates robust predictive power on the test set, achieving an R 2score of 0.95. The optimal parameters that accentuate its performance include a learning rate of 0.2, a max depth of 3, and 300 estimators. This model's ability to systematically build trees by adjusting to the errors of the previous ones gives it an edge in capturing intricate data patterns.





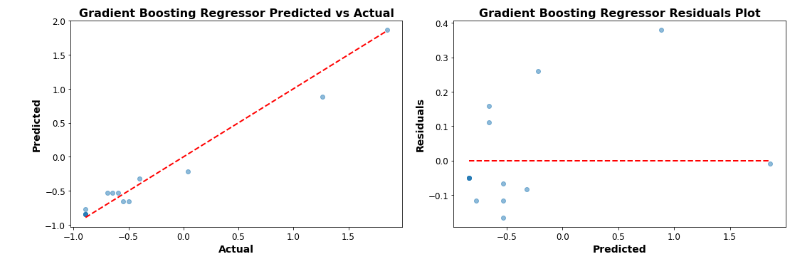
The K-Nearest Neighbors (KNN) model is considered the poorest performer because, despite its perfect training score of 1.0 after tuning, it failed to replicate this high performance on the test data, only achieving a score of 0.70. This large discrepancy between the training and test scores indicates potential overfitting. Overfitting occurs when a model performs exceptionally well on the training data but poorly on new, unseen data. The model becomes too tailored to the specific details and noise of the training set, making it less generalizable to new data. In practical applications, a model's performance on test or unseen data is more crucial than on the training data, hence why the KNN's subpar test score positions it as a weak performer in this context.

**Model Evaluation for Predicting Phase Overrun for 23 MT. production Tanks for the Deaeration phase.**

|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **Deaeration phase** |
| **Production Batches** | 83 |
| **Production Batches after Outlier removal** | 67 |
| **Target Variable** | Phase Overrun |
|  |  |



The Gradient Boosting Regressor stands out as the best performer. It demonstrates near-flawless results with a Train R 2 score of 1.00 and a Test R 2 score of 0.99. After tuning with a learning rate of 0.1, a maximum depth of 3, and 200 estimators, it consistently maintains its top-tier performance, showing its robustness and adaptability.



The Gradient Boosting Regressor stands out as the best performer. It demonstrates near-flawless results with a Train R 2 score of 1.00 and a Test R 2 score of 0.99. After tuning with a learning rate of 0.1, a maximum depth of 3, and 200 estimators, it consistently maintains its top-tier performance, showing its robustness and adaptability.



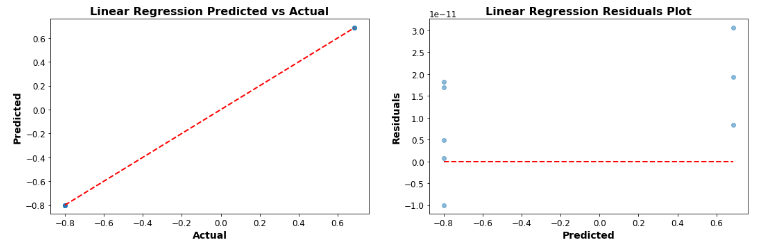
The K-Nearest Neighbors (KNN) model falls short in its test performance. While it achieves a perfect Train R 2 score of 1.00 after tuning, it plummets to a Test R 2score of just 0.4. This indicates that the model might be overfitting the training data. Despite an exhaustive tuning process involving 160 fits to pinpoint the best parameters, like using the 'auto' algorithm, selecting 3 neighbours, and distance-based weights, the model doesn't translate its training success to the test set. This stark difference emphasizes the importance of a model's ability to generalize beyond its training data.

**Model Evaluation for Predicting Phase Overrun for 23 MT. production 23 Tanks for the Agitation phase.**

|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **Agitation phase** |
| **Production Batches** | 82 |
| **Production Batches after Outlier removal** | 39 |
| **Target Variable** | Phase Overrun |
|  |  |



The **Linear Regression model** stands out with perfect results. It achieves a Train R 2score and a Test R 2score both of 1.00, indicating that it can explain 100% of the variance in the target variable for both training and test datasets. This is mirrored in the extremely low Mean Squared Error (MSE) values, which approach zero. No tuning was necessary for this model, suggesting that a simple linear relationship was likely sufficient to capture the patterns in the data.





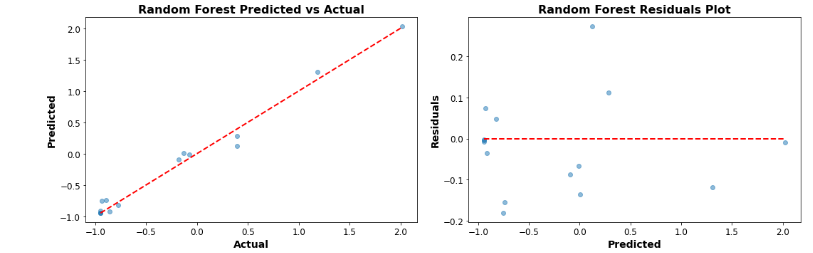
The **Lasso Regression** model has shown the least promising performance before tuning. Its Train R 2 score of 0.17 and Test R 2score of 0.02 suggests that it struggled to capture the variance in the target variable initially. However, after tuning with an alpha value of 0.01, the performance improved drastically to an R 2score of 1.00 for both training and test sets. This indicates a significant improvement, but the stark difference in performance before and after tuning might raise concerns about the model's robustness and its ability to generalize across different datasets.

**Model Evaluation for Predicting Phase Overrun for 23 MT. production Tanks Gum Addition phase**

|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **Gum Addition phase** |
| **Production Batches** | 82 |
| **Production Batches after Outlier removal** | 73 |
| **Target Variable** | Phase Overrun |

:





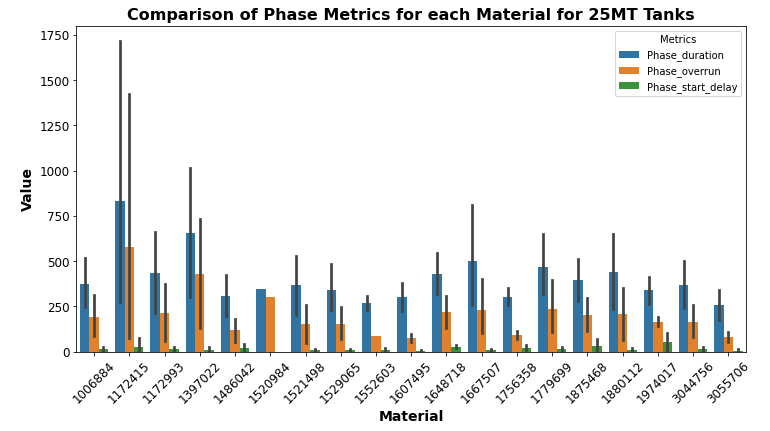
The **Random Forest Regressor** stand out with perfect results. They achieve a Train R2 score and a Test R2 score both of 1.00, which suggests that these models can capture almost all the variance in the target variable for both training and test datasets. The very low MSE values, approaching zero, further confirm this excellence in prediction. The best parameters indicate that for the Random Forest, the optimal number of trees (estimators) is 200 and there's no restriction on depth. These models seem well-suited for predicting phase overrun during the GUM addition phase.



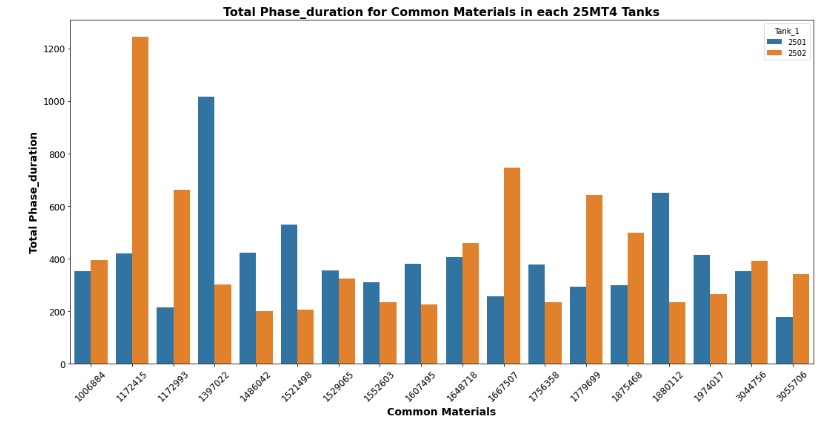
The **Lasso Regression** model initially performed poorly with a Train R2 score of 0.10 and a negative Test R2 score of -0.13, indicating that the model was worse than a horizontal line. However, after tuning with an alpha value of 0.01, the performance dramatically improved to an R2 score of 0.99 for training and 0.97 for testing. Despite this improvement, the initial poor performance might cause concerns about the reliability and robustness of this model for this specific dataset.

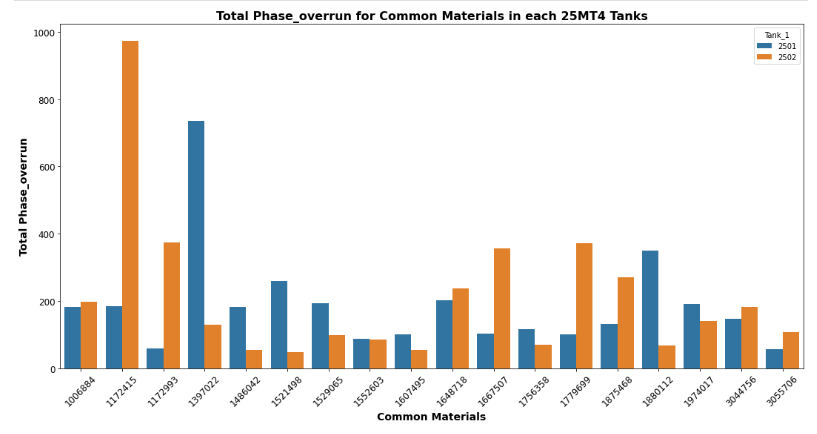
**Exploratory Data Analysis Production Tank 25MT 4:**

**3.4.1 Univariate Analysis:Numerical Features**:



1. The distribution of Material data records for each tank and allows you to see which tanks produced materials in the dataset. There is phase overruns for every material produced in these tanks ,

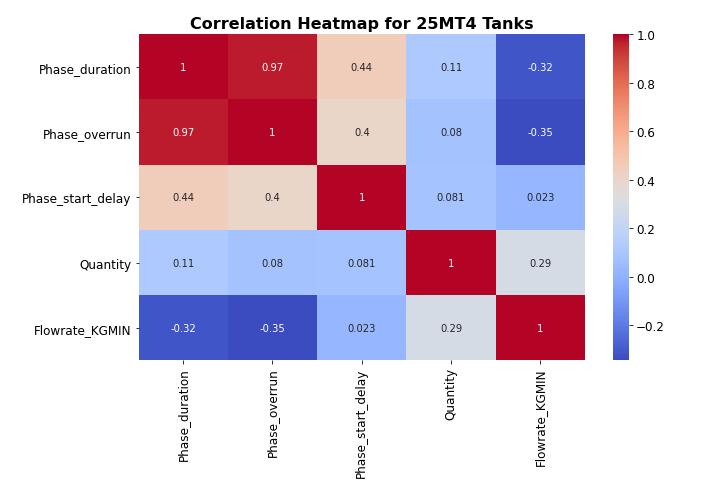




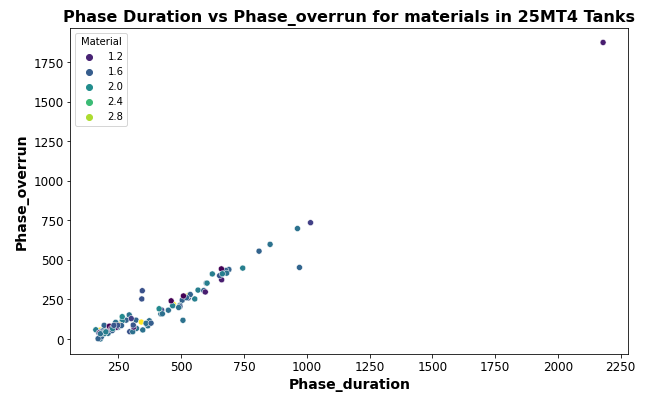
Bar Chart giving visual representation of how phase duration and phase overrun varies across the different production tanks. Each bar represents a tank, and the height is the phase metric for that tank. Apart from 2 high phase overrun values in both tanks , the rest of the materials produced have low phase overrun values

**Bivariate and Multivariate Analysis:**

**Correlation Analysis**: Compute correlation matrices for numerical features to determine if there were any linear relationships between the variables.A high correlation between phase overrun and phase duration variables.

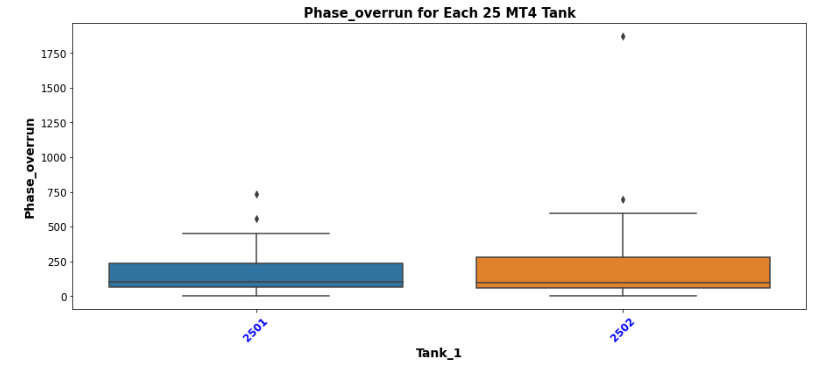


**Pair Plots & Scatter Plots**: For visual inspection of relationships and trends between pairs of numerical features of phase overrun and phase duration



**Handling outliers#**

**Boxplots**



Based on the boxplots in the EDA, outliers were determined to be presence in the data for both tanks of the 25 MT4

**Data standardization**



Based on the descriptive statistics for each of the production tank batch data , there was wide variation in the data for phase\_duration , phase\_start\_delay and the flowrates.

These wide variations could impact the performance of certain machine learning models, a single feature disproportionately in scale could influence models that use distance to compute. Based on this, the data was standardised using the Z-score normalization using StandardScaler from Sklearn.preprocessing library. Another advantage is that its consistent and highly interpretable.

**Model Evaluation for Predicting Phase Overrun for 25 MT. 4 production Tanks All phases**

|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **All phases** |
| **Production Batches** | 98 |
| **Production Batches after Outlier removal** | 81 |
| **Target Variable** | Phase Overrun |
| **Instruction steps/Phases per batch** | 27 |

:

NB: For each batch produced, there are two production tanks used and thus included in this analysis. One for the actual production and the other for storage as in the destination which has itself a phase step and a phase overrun metric.



The Random Forest Regressor stood out as the star performer for predicting phase overrun across all production phases in Production Tanks 25MT. It showcased superior accuracy with an R2score of 1.00 on the training data and an impressive 0.98 on the test data. These metrics reflect the model's excellent capacity to capture the underlying patterns in the data, positioning it as an ideal choice for predicting phase overruns in this context.





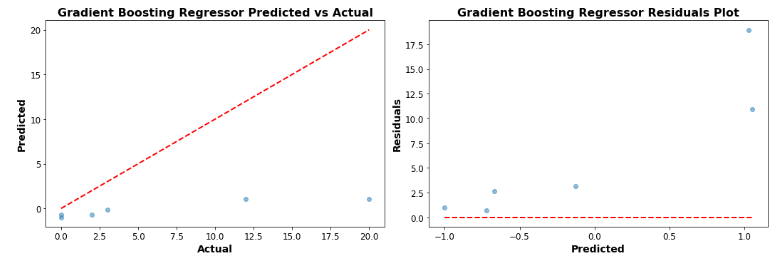
Conversely, the LSTM Neural Network struggled significantly in this predictive task. With an R2score of -0.57 on the training set and an even more concerning -1.26 on the test set, the model displayed a conspicuous inability to make accurate predictions for the given data. This underwhelming performance suggests that the LSTM, in its current configuration, might not be suitable for predicting phase overruns for Production Tanks 25MT.

**Model Evaluation for Predicting Phase Overrun for 25 MT. 4 production Tanks Deaeration phase.**

|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **Deaeration phase** |
| **Production Batches** | 49 |
| **Production Batches after Outlier removal** | 30 |
| **Target Variable** | Phase Overrun |



In the Deaeration Phase of Production Tanks 25MT4, the Gradient Boosting Regressor is the top performer. Displaying minimal error with an almost perfect Train R2of 0.99 and a commendable Test R2of 0.63, it proves its capacity to understand and predict the underlying patterns effectively. Moreover, the model’s performance remains consistent after tuning, with the R2metric still closely hugging the 1.00 mark for training and a stable 0.63 for the test set. The best results were achieved with a learning rate of 0.2, a max depth of 3, and 100 estimators.





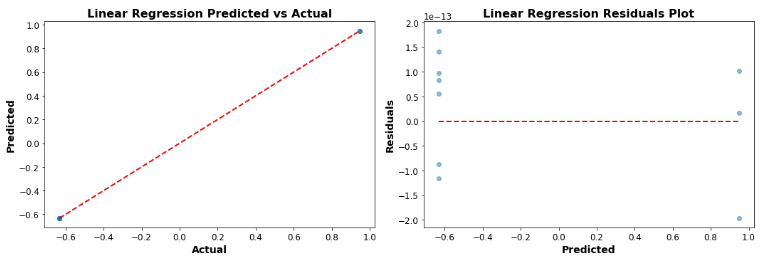
On the other hand, the LSTM Neural Network faced significant challenges during this phase. Originally presenting a negative Train R2 of -0.67 and a Test R2of -0.75, it demonstrates a lack of fit and poor prediction capabilities. While tuning did improve its performance, bringing the Train R2 up to 0.85 and Test R^2 to 0.77, the discrepancy between its original and tuned performances suggests potential overfitting issues or that the model might not be optimally designed for this task.

**Model Evaluation for Predicting Phase Overrun for 25 MT. 4 production Tanks Gum Addition Phase**

|  |  |
| --- | --- |
| **Evaluation Details** | |
| **Instruction Step** | **Agitation Phases** |
| **Production Batches** | 51 |
| **Production Batches after Outlier removal** | 50 |
| **Target Variable** | Phase Overrun |



In the Agitation Phase for Production Tanks 25MT 4, the Linear Regression model demonstrated top performance. Both its Train and Test Mean Squared Error (MSE) are practically at zero, showcasing the model's excellent capability in fitting the training data and generalizing to unseen data. A perfect R2score of 1.00 for both training and test sets, even after tuning, highlighting its predictive power.





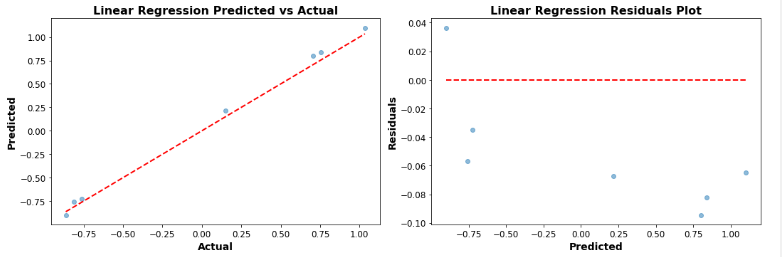
Conversely, the Instance-based K-Nearest Neighbors model underwhelmed in its performance for the Agitation phase. Its Train MSE was 0.43, and it only achieved a low R2score of 0.04 for training and an even more diminished 0.07 for testing. Tuning only marginally improved its R2score, reaching 0.13 for training and remaining at 0.04 for testing. The model's performance suggests that it struggled to find relevant patterns in the dataset for this phase, even after considering 9 neighbors and opting for uniform weighting.

**Model Evaluation for Predicting Phase Overrun for 25 MT. 4 production Tanks Gum Addition Phase**

|  |  |
| --- | --- |
| **Evaluation Details** | |
|  |  |
| **Instruction Step** | **Gum Addition** |
| **Production Batches** | 50 |
| **Production Batches after Outlier removal** | 35 |
| **Target Variable** | Phase Overrun |
|  |  |



In the Gum Addition Phase for Production Tanks 25MT 4, the Linear Regression model emerged as the top performer. The model demonstrated almost perfect fitting with a Train Mean Squared Error (MSE) of 0.00, which held consistently even after tuning. Its testing performance was equally outstanding, with a Test MSE of 0.00 both pre- and post-tuning. The R2score, a measure of the model's predictive power, achieved a flawless 1.00 for both training and testing datasets, highlighting its exceptional efficacy in this phase.



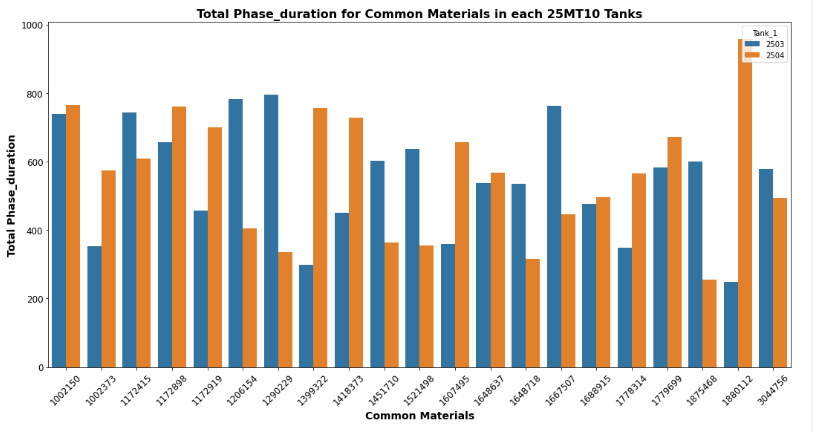


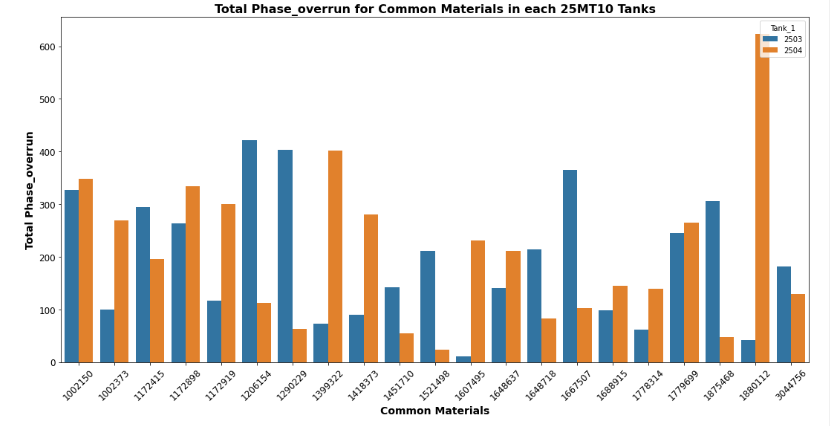
Neural Network (RNN) model using LSTM displayed poor performance. Before tuning, the Train MSE soared to an alarming 3116.96, and the Test MSE reached 2442.13. The R2score, registering at -0.81 for training and an even lower -1.64 for testing, indicated that the model's predictions were substantially worse than simplistic, mean-based predictions. Even after tuning, while there was a marginal improvement in the scores, the results remained unsatisfactory with the Test R2 still lingering at -1.19, underscoring the model's struggle to capture the inherent patterns of the Gum Addition Phase effectively.

**Exploratory Data Analysis : Production Tank 25MT 10**

**Univariate Analysis: Numerical Features**:

The distribution of Material data records for each tank and allows you to see which tanks produced materials in the dataset. The graph below shows the little difference in the total phase duration times between the 2 production tanks of 25 MT10

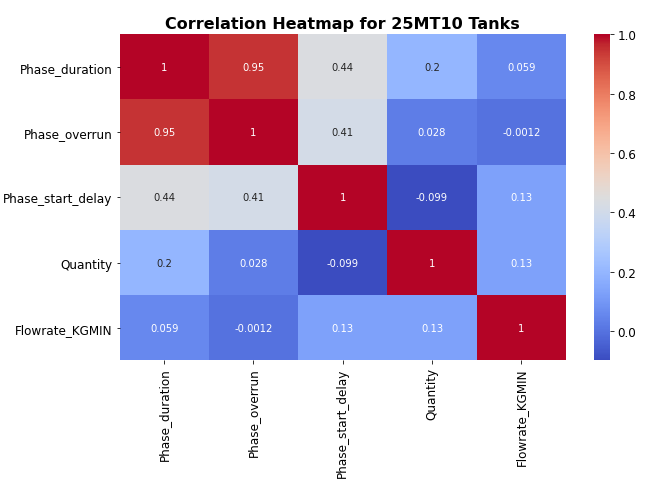




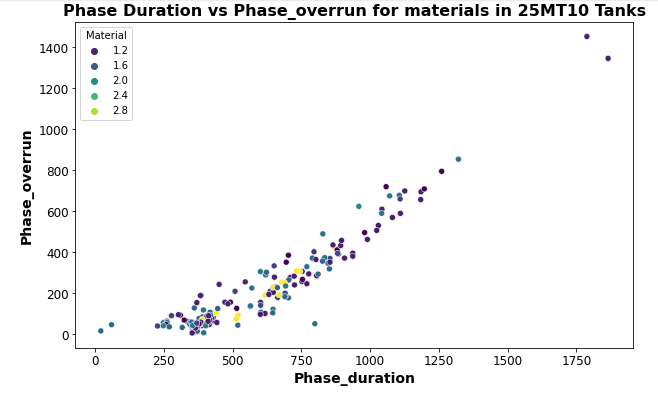
Bar Chart giving visual representation of how phase duration, phase overrun and phase start delay varies across the different production tanks. Each bar represents a tank, and the height is the phase metric for that tank. All materials exhibit Phase overruns in these tanks.

**Bivariate and Multivariate Analysis:**

**Correlation Analysis**: Compute correlation matrices for numerical features to determine if there were any linear relationships between the variables. Good correlation between the phase overrun and phase duration variables.

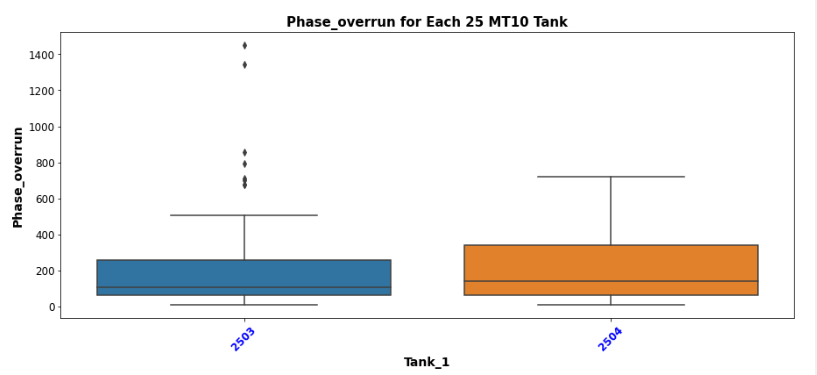


**Pair Plots & Scatter Plots**: For visual inspection of relationships and trends between pairs of numerical features. Good linear relationship observed between the phase duration and phase overrun variables



**Handling outliers**

**Boxplots**



Based on the above boxplots outliers were determined to be presence in the data

**Data standardization**



Based on the descriptive statistics for each of the production tank batch data , there was wide variation in the data for phase\_duration , phase\_start\_delay and the flowrates.

These wide variations could impact the performance of certain machine learning models, a single feature disproportionately in scale could influence models that use distance to compute. Based on this, the data was standardised using the Z-score normalization using StandardScaler from Sklearn.preprocessing library. Another advantage is that its consistent and highly interpretable.

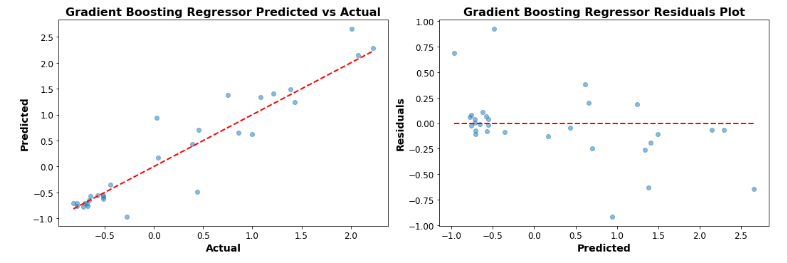
**Model Evaluation for Predicting Phase Overrun for 25 MT. 10 Production Tanks– All Phases**

|  |  |
| --- | --- |
| **Evaluation Details** | |
|  |  |
| **Instruction Step** | **All Phases** |
| **Production Batches** | 194 |
| **Production Batches after Outlier removal** | 150 |
| **Target Variable** | Phase Overrun |
| **Instruction steps/Phases per batch** | 27 |

NB: For each batch produced, there are two production tanks used and thus included in this analysis. One for the actual production and the other for storage as in the destination which has itself a phase step and a phase overrun metric.



The Gradient Boosting Regressor showcased best performance in the modeling for All Phases for Production Tanks 25MT 10. Before tuning, it had a negligible Train Mean Squared Error (MSE) of 0.001 and a Test MSE of 0.088. The model's ability to explain the variance in the data was almost flawless with a training R2score of 0.999 and a testing R2score of 0.905. After tuning, the model managed to achieve a perfect training R2score of 1.000, although with a slight drop in the test R2to 0.887. This was accomplished using a learning rate of 0.2, a maximum depth of 4, and 300 estimators.





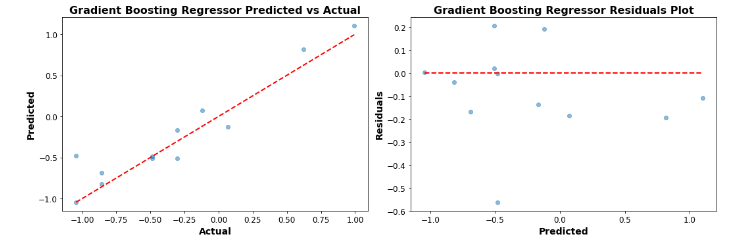
In stark contrast, the Neural Network (RNN) model using LSTM Neural Network displayed significant underperformance. Prior to tuning, it registered a massive Train MSE of 51437.30 and an even larger Test MSE of 63756.90. The R2scores were notably poor, with -0.98 for training data, indicating that the model's predictions were drastically worse than basic mean predictions. The test R2score was at a dismal -1.69. Despite tuning efforts, the results remained subpar with the Test R2improving only slightly to 0.19. The best parameters for the tuned model included 30 neurons for LSTM, 100 epochs, and a batch size of 64.

**Model Evaluation for Predicting Phase Overrun for 25 MT. 10 Production Tanks– Deaeration**

|  |  |
| --- | --- |
| **Evaluation Details** | |
|  |  |
| **Instruction Step** | **Deaeration** |
| **Production Batches** | 92 |
| **Production Batches after Outlier removal** | 58 |
| **Target Variable** | Phase Overrun |
|  |  |



The Gradient Boosting Regressor demonstrates good performance for the deaeration results of the Production Tanks 25MT10. The model, before tuning, delivered a Train MSE of 0.01 and a Test MSE of 0.04. The R2 scores were highly commendable with a score of 0.99 for training data, indicating the model's proficient ability to explain 99% of the variance. Its testing R2 score stood at 0.88. Remarkably, after tuning, the training R2 reached a perfect score of 1.00. The model achieved this exceptional performance with a learning rate of 0.01, a max depth of 4, and 300 estimators. This is also seen in the plots below with a good predictions and residuals.





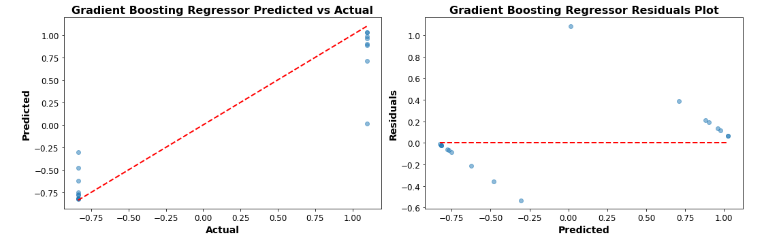
The LSTM Neural Network, a form of RNN, notably struggled in modelling the deaeration results. Pre-tuning, it registered a Train MSE of 54.79 and a much worse Test MSE of 19.09. The R2 values were concerning: -0.67 for training and -0.75 for testing, indicating the model's predictions were significantly worse than a naive mean-based approach. Fortunately, tuning improved the model to some extent, bringing the training R2up to 0.85 and the testing R2to 0.77. This improvement was accomplished with 50 LSTM neurons, 100 epochs, and a batch size of 16.

**Model Evaluation for Predicting Phase Overrun for 25 MT. 10 Production Tanks– Agitation**

|  |  |
| --- | --- |
| **Evaluation Details** | |
|  |  |
| **Instruction Step** | **Agitation** |
| **Production Batches** | 97 |
| **Production Batches after Outlier removal** | 90 |
| **Target Variable** | Phase Overrun |
|  |  |



The Gradient Boosting Regressor stands out as the top-performing model for the agitation phases of the Production Tanks 25MT10. Prior to tuning, it exhibited a Train MSE of 0.01 and a Test MSE of 0.09. Its R2scores are strikingly good, with a score of 0.99 for the training data, which indicates the model's superior capability to account for 99% of the variance. For testing data, it scored 0.90. After optimization, both the training and testing R2scores reached a perfect 1.00. The model achieved this stellar performance using a learning rate of 0.01, a max depth of 4, and 300 estimators.





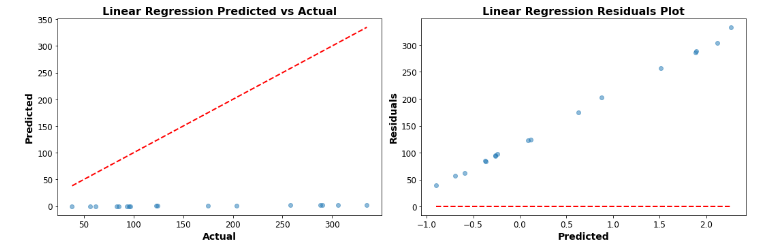
The Lasso Regression model, before tuning, struggled in this task. It reported a Train MSE of 0.90 and an even worse Test MSE of 1.10. The R2scores were not promising: 0.11 for the training set and -0.19 for the testing set. Negative R2 values, especially for the test set, suggest the model's predictions were considerably worse than a simplistic mean-based strategy. While tuning did improve the performance slightly, raising the training R2to 0.32 and the testing R2 to just -0.01, the results were still far from satisfactory. This improvement was possible due to an alpha value of 0.01.

**Model Evaluation for Predicting Phase Overrun for 25 MT. 10 Production Tanks– Gum Addition**

|  |  |
| --- | --- |
| **Evaluation Details** | |
|  |  |
| **Instruction Step** | **Gum Addition** |
| **Production Batches** | 96 |
| **Production Batches after Outlier removal** | 81 |
| **Target Variable** | Phase Overrun |
|  |  |



The top-performing model appears to be the **Linear Regression** under the "Linear" model type. This model exhibits a perfect Train and Test performance with both MSE (Mean Squared Error) and R2 (R-squared or Coefficient of Determination) values. Specifically, it has an MSE of 0.00 for both training and testing phases, and its R2 score stands at 1.00, indicating that the model perfectly predicts the outcomes. Even after tuning, the model retains its superior performance, demonstrating its robustness.

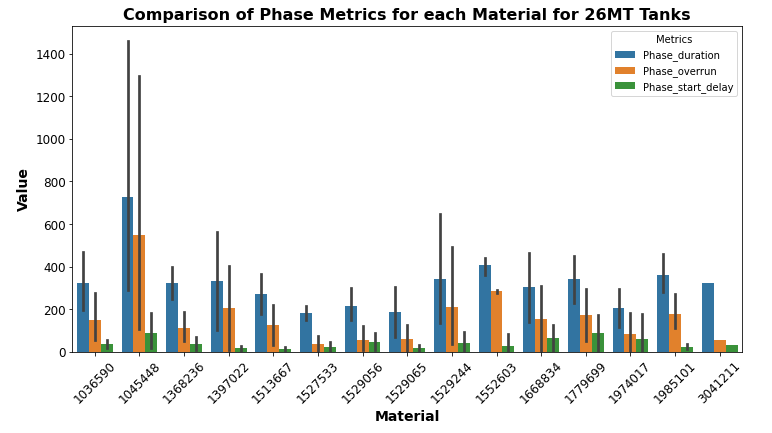




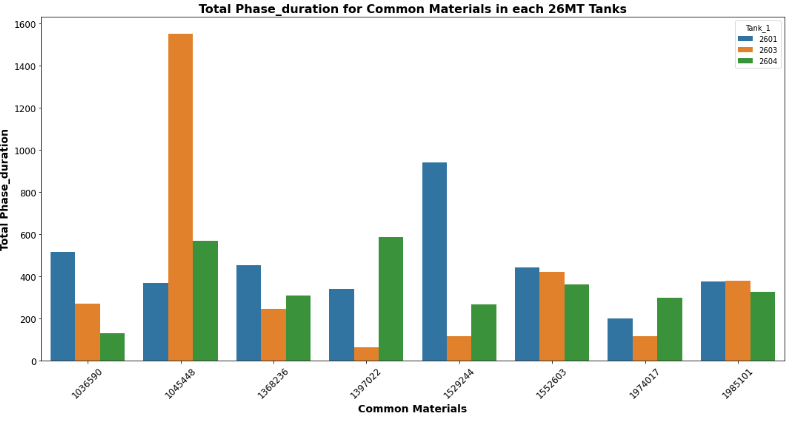
**LSTM Neural Network** under the "Neural Network (RNN)" category showed the poorest performance. Before tuning, it had a Train MSE of 18366.10 and a Test MSE of 34292.40. The R2 values for training and testing were -1.45 and -2.66 respectively. Even after tuning, though there was a reduction in the MSE values (Train MSE tuned to 11195.51 and Test MSE tuned to 20525.29), the R2 scores did not show significant improvement, with -0.50 for training and -1.19 for testing. The negative R2 values indicate that the model is performing worse than a simple horizontal line (mean-based model), signalling that the model might not be the right choice for this particular dataset or task.

**Exploratory Data Analysis** **for** **Production Tank group 26 MT** –

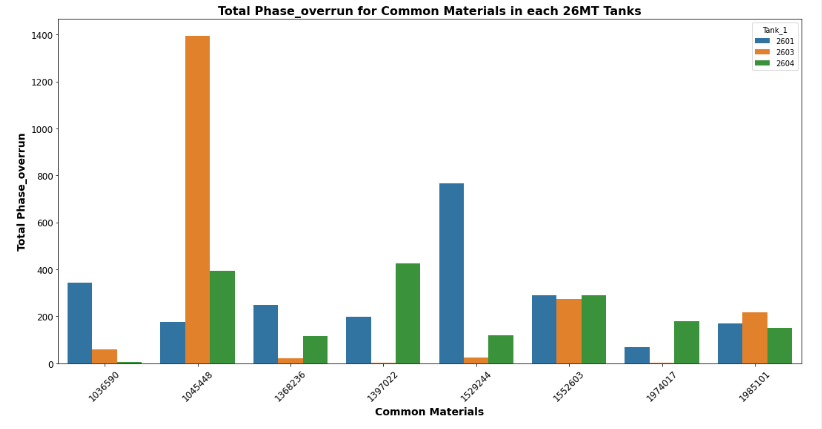
**Univariate Analysis:Numerical Features**:



There are production phase overruns for each material produced in all 26MT tanks. The phase start delay is the lowest metric for all tanks.

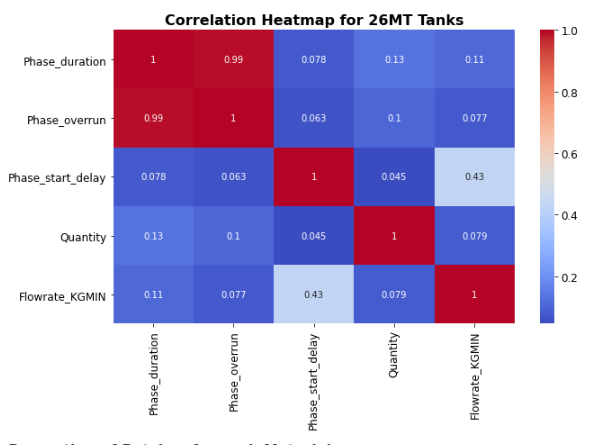


The phase overrun values for common materials between each 26 MT production tanks is seen in the following graph. Bar the unusual overrun value for tank 2603, it’s the tank 2601 that has higher phase overrun times for the common materials.

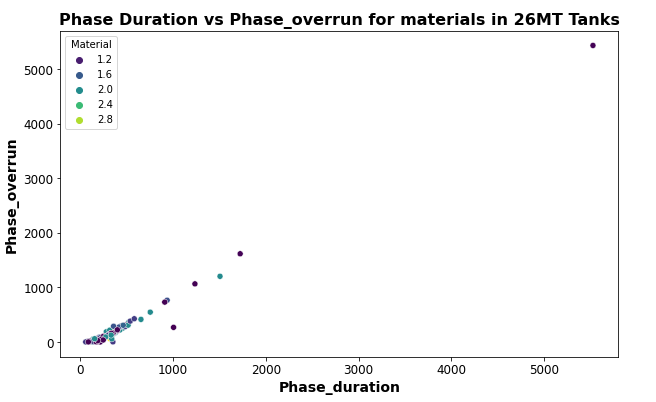


**Bivariate and Multivariate Analysis: Correlation Analysis**:

Compute correlation matrices for numerical features to determine if there were any linear relationship between the variables. The highest correlation is between the phase duration and the phase overrun variables.



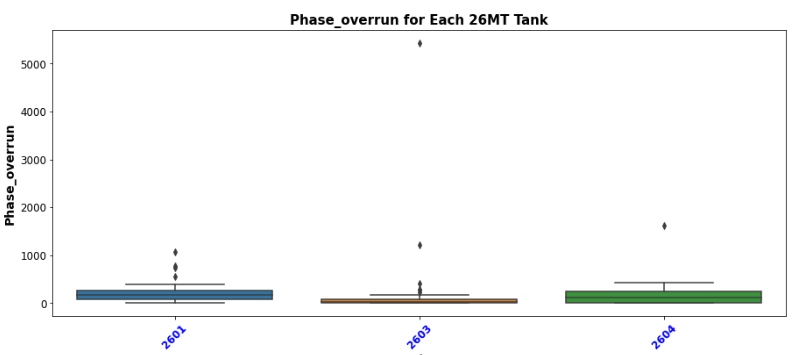
**Pair Plots & Scatter Plots**: For visual inspection of relationships and trends between pairs of numerical features. There is a linear relationship between the phaser overrun and phase duration variables in the dataset.



**Handling outliers#**

**Boxplots –**

There is the present of outliers in each of the 26 MT production tanks by the appearance of points about the observation line.



The outlier measurement and removal method used was based on the Interquartile Range (IQR). method is robust to extreme values and is preferred over methods like standard deviation-based outlier detection especially if there isn’t a gaussian data distribution.

**Data standardization**



Based on the descriptive statistics for each of the production tank batch data, there was wide variation in the data for phase\_duration , phase\_start\_delay and the flowrates.

These wide variations could impact the performance of certain machine learning models, a single feature disproportionately in scale could influence models that use distance to compute. Based on this, the data was standardised using the Z-score normalization using StandardScaler from Sklearn.preprocessing library.

**Machine Learning Results**

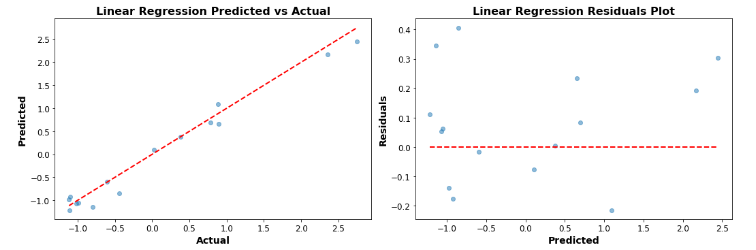
**Model Evaluation for Predicting Phase Overrun for 26 MT Tank for All Phases**

|  |  |
| --- | --- |
| **Evaluation Details** | |
|  |  |
| **Instruction Step** | All Phases |
| **Production Batches** | 46 |
| **Production Batches after Outlier removal** | 27 |
| **Target Variable** | Phase Overrun |
| **Instruction steps/Phases per batch** | 27 |

NB: For each batch produced, there are two production tanks used and thus included in this analysis. One for the actual production and the other for storage as in the destination which has itself a phase step and a phase overrun metric.



In the analysis of Production Tanks 26MT All Phases, the Linear Regression model is the better model. Before tuning, the model yielded a training *R*2of 0.96 and a test *R*2of 0.97, suggesting it could explain approximately 96% and 97% of the variance in the training and test datasets respectively. Post-tuning, the model further improved its test *R*2to an impressive 0.99, indicating an excellent fit to the unseen data and emphasizing its potential in predictive capacities for this dataset. The Mean Squared Error (MSE) metrics had values close to zero for both training and test sets after tuning. Overall, the Linear Regression model stands out as a potent tool for predictions in the context of Production Tanks 26MT All Phases, achieving high predictive accuracy and reliability this is shown in the below plots for the predicted vs actual values and the residual plots





The poorest performer is the LSTM Neural Network from the "Neural Network (RNN)" category. Its initial Train MSE is a 24976.80, and its Test MSE is even higher at 34481.50. Furthermore, the *R*2 values are deeply concerning: -1.35 for training and -0.91 for testing. These negative *R*2 scores imply that the model's predictions are far worse than a simple horizontal line, or a basic mean model. Even after tuning, while there's an improvement in the metrics, they're still quite poor. The best configuration for the LSTM model includes 70 LSTM neurons, 100 epochs, and a batch size of 16.

**Model Evaluation for Predicting Phase Overrun for 26 MT Tank for Deaeration Phase**

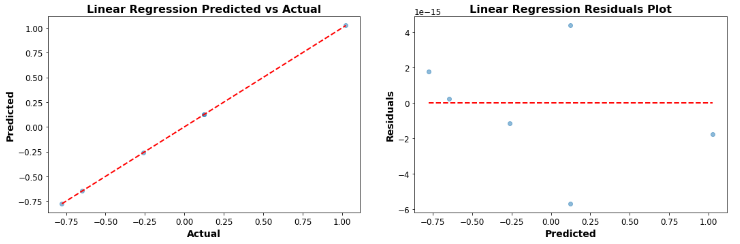
|  |  |
| --- | --- |
| **Evaluation Details** | |
|  |  |
| **Instruction Step** | Deaeration |
| **Production Batches** | 46 |
| **Production Batches after Outlier removal** | 27 |
| **Target Variable** | Phase Overrun |
|  |  |



Linear Regression exhibits good accuracy for this dataset. It achieved a perfect *R*2value of 1 for both the training and test datasets. Additionally, the Mean Squared Error (MSE) for this model is extremely low, nearly zero for both the training and test sets.

The Linear Regression model's performance great for this dataset. To achieve a perfect *R*2 , may indicate a potential for overfitting. Overfitting means the model may not generalize well to unseen data. This perfect score also raises concerns about possible data leakage, where the model might have unintentionally accessed the target variable during training. Other regression methods were looked at , that have regularizations such as lasso and ridge , they also showed good accuracy and precision.

The plots below also highlight how good the predictions made by the linear model area and also low variances in the residuals





The LSTM (Long Short-Term Memory) Neural Network showed a poor performance with the given dataset. The training *R*2value stands at -0.82 and, the test *R*2at -2.38. Such values suggest that the model does a poorer job at predictions than a basic horizontal line. The MSE values further validate this, being relatively high at 129.78 and 73.03 for the training and test datasets, LSTM is not the model for this prediction study

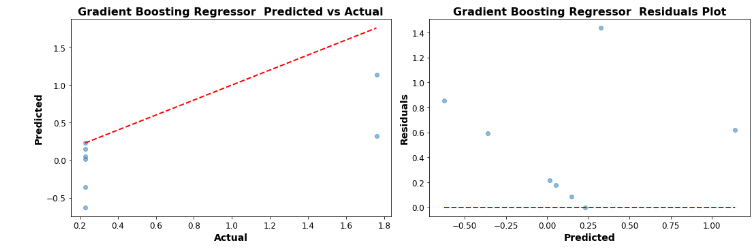
|  |  |
| --- | --- |
| **Evaluation Details** | |
|  |  |
| **Instruction Step** | Agitation |
| **Production Batches** | 46 |
| **Production Batches after Outlier removal** | 40 |
| **Target Variable** | Phase Overrun |
|  |  |

**Model Evaluation for Predicting Phase Overrun for 26 MT Tank for Agitation Phase**



The Gradient Boosting Regressor top performing in terms of model accuracy. With a R2 value of 0.97 for the training set, it indicates that the model captures 97% of the variance in the target variable. The test R2 is lower at 0.28, it's still the highest positive test R2 value among all the models. Its Mean Squared Error (MSE) for both the training and test datasets is low, further indicating its high predictive capability. It provides a good balance between fitting the training data and generalizing to the test data. There is potential for overfitting as there is a difference between the training and test R2 values. Further improvements with hyperparameter tuning, more batch data, or regularization techniques could help this.

The graph below shows that predictions using this model gives errors and the residuals plot shows that there is non linearities in the data





The LSTM Neural Network has a training R2 of -0.42 and a much worse test R2 of -5.51. The negative R2 values, especially one as low as -5.51, imply that the model is doing an exceptionally poor job at predicting the target variable. The high MSE values, 0.62 for the training dataset and 1.22 for the test dataset, further affirm the model's inadequacy.

The LSTM model's poor performance suggests it's not the right choice for this dataset. LSTMs, as a subtype of recurrent neural networks, are primarily designed for sequence prediction problems. The LSTM model might need significant architecture and hyperparameter adjustments.

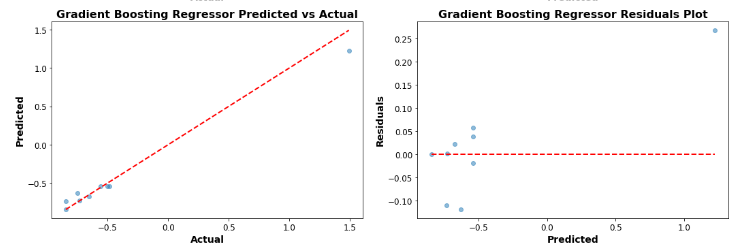
**Model Evaluation for Predicting Phase Overrun for 26 MT Tank for Gum Addition Phase**

|  |  |
| --- | --- |
| **Evaluation Details** | |
|  |  |
| **Instruction Step** | Gum Addition |
| **Production Batches** | 44 |
| **Production Batches after Outlier removal** | 43 |
| **Target Variable** | Phase Overrun |
|  |  |



The Gradient Boosting Regressor, after tuning, displayed good performance. The Train MSE of 1.07513 *E*−13 - 1.07513*E*−13, it essentially got close to a perfect fit for the training data. Moreover, the R2 value of 1.00 for both training and 0.98 for the test dataset indicates a good fit of the model to the data. The test MSE is also considerably low at 0.011. The hyperparameters provided, including a learning rate of 0.2, max depth of 3, and 300 estimators, seem to be optimally tuned for the dataset.

This can be also enhanced by the results from the plots below where there is good alignment of points along the line for the predicted vs actual and good distribution of points around the line in the residual plot





The LSTM Neural Network is ill-suited for this dataset. The high Train MSE of 5461.17 (even after tuning) and the Test MSE of 1773.12 highlight the inadequacy of the model. Although the LSTM model improved its test R2 from -0.34 to 0.01 after tuning, it's still barely better.

These results could be since LSTM models are primarily designed for time-series data where the sequence and temporal dependencies are important.

**Summary of all Machine Learning Models applied for all Production Tank Group under the different phases.**

The primary aim of employing machine learning models in this context is to predict the phase overrun downtime for production batches, specifically those containing mucilage. By accurately predicting this downtime, organizations can gain insights into potential inefficiencies across all production stages. This, in turn, aids in proactive decision-making, optimized resource allocation, enhanced production schedules, and eventually, a reduction in costs due to unanticipated downtimes.

1. **Machine models for each Production Tank group based on the total phase overrun for each batch built.**



Ensemble tree-based methods, particularly the Random Forest and Gradient Boosting Regressors, stood out across different tank capacities and batch sizes. Their superior performance indicates their ability to capture the nuanced relationships between variables, even in the complex setting of mucilage-rich production.

However, there's noticeable model variability among other algorithms. For instance, while Linear Regression proved exemplary with the Tank 26MT dataset, its performance may waver for other tanks. Such variability is indicative of the inherent differences in data characteristics, emphasizing the need to consider the unique nature of the production process.

The consistent underperformance of the LSTM Neural Network possibly results from its need for larger datasets to capture patterns effectively. In contrast, ensemble methods and simpler linear models have shown adaptability, suggesting their suitability for various data sizes and types.

Importantly, the size of the data plays a pivotal role. Large datasets can offer rich insights, favoring some models, while smaller datasets might favor simpler models, as they prevent overfitting.

1. **Machine models for each Production Tank group based on the Phase overrun for the Deaeration Phase for each batch built.**



Analyzing the machine model performance for various production tank groups during the deaeration phase yields valuable insights. Tank groups, each differentiated by their capacity and number of batches, ranged from as compact as 4 tonnes (in the 25 MT 4 tank) to 20 tonnes (in both the 22 MT and 23 MT tanks), and as lightweight as 1400kg in the 26 MT tank. Among them, the number of batches varied between 27 and 67.

A striking observation was the consistent suboptimal performance of the LSTM Neural Network across multiple tanks—22 MT, 25 MT 4, 25 MT 10, and 26 MT. This may suggest that the LSTM, typically better suited for time series data, may not be the optimal choice for predicting phase overrun during the deaeration phase, particularly when the dataset isn't large or sequential enough to leverage the model's strengths.

In contrast, the Gradient Boosting Regressor displayed commendable efficiency, particularly in the 23 MT, 25 MT 4, and 25 MT 10 tanks, implying its robust adaptability across different data sizes and tank capacities. Linear Regression, a simpler model, exhibited top performance in the 22 MT and 26 MT tanks, which had relatively fewer batches (30 and 27 respectively). This could hint at its effectiveness in scenarios where the data volume isn't exceedingly large, or the relationships within the data are more linear.

Interestingly, the K-Nearest Neighbors (KNN) model was the least effective for the 23 MT tank, which had the largest number of batches (67). This deviation from the recurrent LSTM underperformance suggests that KNN may struggle with larger datasets in this specific context.

Ultimately, these findings underscore the imperative of model selection in alignment with dataset characteristics, ensuring optimal predictions, especially in crucial operations like the deaeration phase in production tanks.



For the production tank groups undergoing the agitation phase, we observe a diverse range of machine model performances based on different tank capacities and batch sizes.

Starting with the larger tanks, both 22 MT and 23 MT have a capacity of 20 tonnes but with different batch counts, 34 and 39 respectively. For the 22 MT tank, the Decision Tree Regressor emerged as the top performer, while the Dense Neural Network (often referred to as Fully Connected Network or FCN) lagged. This might suggest that the data's structure is more hierarchical and non-linear, which is a strength of decision trees. On the other hand, the 23 MT tank showed the best results with the simple Linear Regression model, but struggled with Lasso Regression. It implies that for this dataset, the regularization brought in by Lasso may be too strong, leading to an underfit.

The smallest capacity tank in this list, 25 MT 4 with a 4-tonne capacity and 50 batches, exhibited optimal results with the Linear Regression model and faced challenges with the K-Nearest Neighbors (KNN). Given that KNN typically doesn't fare well with a larger number of features due to the curse of dimensionality, it's possible that the dataset for this tank had multiple feature dimensions making KNN inefficient.

The 25 MT 10 tank, having a 10-tonne capacity and a notably larger batch count of 90, demonstrated the best results with the Gradient Boosting Regressor, a sophisticated ensemble technique. This suggests the possibility of complex non-linear relationships in the data, which Gradient Boosting can capture efficiently. Lasso Regression underperformed again in this scenario, reiterating its potential difficulty with this specific context.

Finally, for the 26 MT tank with a weight of 1400kg and 40 batches, the Gradient Boosting Regressor stood out as the top model. The LSTM Neural Network, usually best suited for sequential data, didn't fare well, indicating that the data may not be inherently sequential or the patterns were too complex for the LSTM to capture with the given data size.

In essence, these findings reveal that during the agitation phase, there isn't a one-size-fits-all machine model. The data characteristics of each tank, along with its capacity and batch size, play a significant role in model efficiency, highlighting the importance of bespoke model selection based on specific scenarios.



During the "Gum Addition" phase for the production tank groups, machine model performances exhibited variability contingent on tank capacity and batch size, revealing insightful patterns and possible dataset characteristics.

Both the 22 MT and 23 MT tanks possess a capacity of 20 tonnes but differ in batch counts, with 29 and 73 batches respectively. The 22 MT tank favored the simple linear regression, indicating potential linear relationships in its dataset. Conversely, the Simple Neural Network, which might be better suited for non-linearities, underperformed. This might suggest that the dataset for this tank is predominantly linear. For the 23 MT tank, the more sophisticated Random Forest Regressor excelled. Given that Random Forests are known for handling complex data structures, it can be inferred that the data from this tank might be more intricate or contain non-linear patterns. The underperformance of Lasso Regression might point towards the regularization being too strict, causing a possible underfitting scenario.

The smaller 25 MT 4 tank, with a 4-tonne capacity and 35 batches, best aligned with the linear regression model but fared poorly with the LSTM Neural Network, a model often employed for sequence or time-series data. This suggests that the dataset for this tank, while perhaps linear in nature, may not be inherently sequential, making the LSTM an inappropriate choice.

Similarly, the 25 MT 10 tank, accommodating 10 tonnes and encompassing 81 batches, saw the best outcomes with the linear regression model. The LSTM Neural Network's inferior performance reaffirms the earlier suggestion that the datasets across these tanks may not have significant sequential characteristics to leverage LSTMs efficiently.

Lastly, for the 26 MT tank, weighing 1400kg and with 43 batches, the Gradient Boosting Regressor emerged as the most efficient model. This indicates potential complex non-linear relationships within the dataset. The LSTM's struggle once more underscores the probable non-sequential nature of the data or its inability to capture the patterns given the data size.

In summation, during the "Gum Addition" phase, while linear regression predominated across multiple tanks, the inefficiency of LSTM across tanks was noteworthy. This suggests that while there may be overarching linearity in the processes, sequential data might be scarce or inadequately represented. The results emphasize the need for tailored model selection based on specific data characteristics and operational contexts.

**Interpretability factor :**

When we say you "lose some interpretability" with models like Random Forests or GBMs, it means that while these models might give you better accuracy, they don't provide a straightforward way to understand the reasons behind individual predictions. This trade-off can be significant in industries or applications where explanations are crucial, such as healthcare or finance.

1. **Linear Regression** and its regularized versions (like Ridge and Lasso) are often seen as **highly interpretable**. The reason is that the weight (or coefficient) associated with each feature directly tells us the relationship between that feature and the output. For example, in predicting house prices, if the coefficient for the number of bedrooms is positive, we can directly interpret that having more bedrooms is generally associated with a higher house price.
2. **Decision Trees** are also interpretable to an extent. You can visualize the tree and see the decisions it's making at each node.
3. **Random Forests** and **Gradient Boosting Machines (GBM)** are more complex. While they're based on decision trees, the sheer number of trees and the way they're combined make them harder to interpret directly. A single prediction in a Random Forest, for instance, might be an aggregate result from hundreds or thousands of individual trees.
4. **Deep Learning models**, like neural networks, are often referred to as "black boxes" due to their lack of interpretability. They can have millions of parameters, and understanding the exact reason for a specific prediction is challenging.

**Overall Phase Duration**

**Importance of Phase Duration**: If phase duration is the main feature and your model has a high R^2 value (e.g., 0.98), it indicates that phase duration is a strong predictor of phase overrun. In practical terms, this means changes in phase duration have a substantial and consistent impact on phase overrun, and controlling or monitoring phase duration can help manage phase overrun effectively.

**Interpreting Coefficients**: In a linear regression context, the coefficient of the phase duration would tell you how much phase overrun changes for a unit change in phase duration, holding all else constant. For instance, if the coefficient of phase duration is 2.5, it would mean that for every additional unit (e.g., minute) of phase duration, the phase overrun increases by 2.5 units.

**Linearity Assumption**: Remember that the relationship implied by linear regression is linear. So if the model indicates that for every minute increase in phase duration, phase overrun increases by a set amount, that relationship is assumed to be consistent across all values of phase duration. It's important to validate this assumption with domain knowledge and residual plots.

**Generalization:** While a high R^2 value indicates a good fit to the training data, it doesn't always mean the model will generalize well to new, unseen data. It's essential to ensure that the model isn't overfitting to the training data.

**Practical Implications**: Knowing that phase duration is a primary predictor of phase overrun, operations in the production process can be adjusted to control phase duration. This understanding can lead to more efficient production processes, reduced waste, and better predictability in outcomes.

**Gum Addition Phase Duration**

**Predicting Phase Overrun**: The linear regression model predicts 'phase overrun' (the delay or overshooting of the desired time) using the 'phase duration' (how long the gum addition step takes) as the primary predictor.

**Importance of Gum Addition Time**: A high R^2 value (e.g., 0.98) indicates that the time taken for the gum addition step is a strong predictor of any overrun or delay in the production process. This suggests that the duration of the gum addition phase is crucial for the overall efficiency of the production. If gum addition takes too long or too short, it might lead to an overrun.

**Interpreting Coefficients in Context**: In this context, the coefficient of the phase duration tells you by how much the overrun changes for every unit change in the gum addition time. For instance, if the coefficient for the gum addition time is 3, it means that for every additional minute spent in the gum addition step, the phase overrun increases by 3 minutes.

**Linear Relationship**: This model assumes a consistent, linear relationship between gum addition time and overrun. That is, the overrun's increase or decrease is steady for every unit increase or decrease in the gum addition time. This might be because the gum addition is a critical process, and any deviations in its time cause a domino effect on subsequent phases.

**Practical Implications for Production**: If the gum addition step is known to be a primary predictor of phase overrun, operators can focus on streamlining and closely monitoring this particular step. This could involve ensuring that the gum is always ready and available, machinery used in this step is in optimal condition, or that workers are adequately trained for this task. By optimizing the gum addition step, the entire production process might benefit from reduced overruns.

**Feedback for Improvement**: Knowing this relationship also allows for feedback into the instruction process. If the instructions dictate a gum addition time that consistently results in overruns, perhaps the instructions can be revised or refined.

**Deaeration Phase**

**Predicting Phase Overrun with Deaeration Time**: The linear regression model is used to predict 'phase overrun' using the 'phase duration' of the deaeration phase as the primary predictor.

**Importance of Deaeration Time**: An R^2 value of 0.98 indicates that the time taken for the deaeration phase is a significant predictor of any overrun in the production process. This underscores the importance of the deaeration process in the production workflow. If the deaeration takes longer than expected, or if it's too quick, it can lead to subsequent delays or inefficiencies.

**Interpreting Coefficients for Deaeration**: The coefficient of the phase duration in the deaeration process reveals how much the overrun changes for every unit change in the deaeration time. If the coefficient is, say, 2, then for every extra minute spent in deaeration, the phase overrun increases by 2 minutes.

**Linear Relationship**: This model assumes a consistent, linear relationship between deaeration time and overrun. This suggests that the effects of time changes in the deaeration phase on the overrun are predictable and uniform. In a production setting, this could mean that the deaeration process is sensitive and needs to be consistently timed to prevent subsequent overruns.

**Practical Implications for Production**: The critical nature of the deaeration phase, as indicated by the model, means that production managers should prioritize ensuring that this step is executed efficiently. Strategies could involve regular maintenance of deaeration equipment, training of personnel involved in this phase, and consistent monitoring to ensure the deaeration time remains within the desired limits.

**Feedback and Continuous Improvement**: The clear relationship between deaeration time and phase overrun can serve as feedback for refining the production process. If the standard deaeration time consistently leads to overruns, it may be worth revisiting the process parameters or even the equipment used for deaeration.

In conclusion, employing machine models to predict phase overrun downtime is not just a technical endeavour but a strategic one. Choosing the right model—considering data size and specific challenges like mucilage—can lead to significant operational improvements, cost savings, and enhanced productivity. The insights gleaned can be invaluable in fine-tuning production processes, achieving a competitive edge in the marketplace.



# **Appendix A: All Machine Model Results for all Production Tanks Groups**

* **Tank 22 MT**

1. **All phases**



1. **Deaeration Phase**



1. **Agitation Phase**



1. **Gum Addition Phase**



* **Tank 23 MT**

1. **All phases**



**Deaeration Phase**



1. **Agitation Phase**



1. **Gum Addition Phase**



**Tank 25 MT 4**

1. **All Phases**



1. **Deaeration Phase**



**2- Agitation Phase**



**Gum Addition Phase**



**25MT10**

**All Phases**



**Deaeration Phase**



**Agitation Phase**



**Gum Addition Phase**



**26MT**

**All Phases**



**Deaeration Phase**



**Agitation Phase**



**Gum Addition Phase**



# **Appendix B: Interview Transcripts**

**Participant 1 Interview-**

August 29, 2023, 8:51AM

19m 32s

 **Michelle M. Moran** 0:20  
I'm useless containing bear products, so I just gotta you, would you? You would you be have been involved in setting off the factory talk batch process. Like say the phases and the system.

**Participant 1** 0:36  
Yeah. So basically, so the the system was always there, right? So, but it's at the objective of the system is to execute the batches, right? So there's a whole host of data being collected behind the scenes, but it was basically not in a, it was all there, but it wasn't.

 **Michelle M. Moran** 0:45  
Yes.

**Participant 1** 0:53  
And if I was structured in the in the way a typical way where it was just ready for use for analytics. So I I was involved with. So if you guys, I suppose when they hired when I started three years ago, one of the first things they said to me is, hey, look, we have a black box up in manufacturing. We're like we know the batch starts at for example 8:00 AM in the morning when the guys come in and we know what they finish it at 6:00 PM. But in between is a total black box. I mean we have the recipe, we know exactly how to build it, but we have no tracking of time.  
So I started basically looking in the back end and I built a store procedure that basically looks at every recipe and summarises it. So I look for the key things like.  
I look for the starter batch, I look for all the consumption. I look for all the problems that get answered. I look for like, yeah, the weights that are captured for the materials I look for any agitation time I look for deoration time and I look for homogeneity. I like the processing. So either we have like single modernization, double homogenization, homogenization and pasteurisation and pasteurisation and and basically and then lab, sample and closeout. So if I can get that I can that's basically summarising the batch start to end. And when I did that.  
Face.  
They owe monitors and the manufacturing team were kind of saying Jesus right, you're halfway to building an OE system for us, right? So if you you, you have all the actuals, but we have no. So that's what actually happened. But we have no target, I don't know like if it took like the Dr ration time would say like we have it all, all recipes are set to 480 minutes. So like what's it like 8 hours whatever, but they they don't. There's nothing in between.

 **Michelle M. Moran** 2:14  
Hmm.  
Yeah.

 **Participant 1** 2:32  
And to say there's nothing to say. Well, how so? What's the targets like that? So we have like the plan, but we don't have the target. Likewise for adding powders or drums or hate P process we they didn't basically have like a target. So what I did is I took the data using my stained stored procedure. I just just expanded the time range for the last five years.  
And use that to then get a timing for everything. So I said OK, every time we add powder or every time we add treated water in this system of this size, how long does it take us? So I try to like exclude the outliers. So I took the valley which was the 90th percentile means it was a repeat full time not the fastest time but a repeatable achievable time for us. And then we assess that. So every time now the guys executed batch they have like a target for everything. So then afterwards if we didn't meet our target.

 **Michelle M. Moran** 3:14  
Yeah.

 **Participant 1** 3:24  
He can say why we can say okay there was slow pumping or the froze the the juice that came in from the boat juicer outside it was it was too cold. It was like moving ice not you know. So then we could start having really good meetings saying hey actually.

 **Michelle M. Moran** 3:34  
Yeah.  
Yeah.

 **Participant 1** 3:40  
This is the reason we didn't achieve our target time and then from that over the last three years we've been putting projects in place to say, OK, mobile tanks are a big problem. Okay, whoa, what's going on here, you know, OK. Mobile tanks are a shared resource between the kitchen and manufacturing. So sometimes when the guys go to use mobile tank. Ohh, there's not one, they're available. So how do you solve that? We can look at we didn't analysis then to say okay how many mobile tanks do manufacturing need? How many do the kitchen need okay take these 10 mobile.  
Thanks. And manufacturing there yours only, so you manage them now they're not a shared resource anymore. So straight away that problems resolved. So Long story short, by like, yes. If you're question was did I do that work? Yes. And that led on to so many more projects because we could guess we could get detailed behind what was happening and like that what you had looked inwards basically the dehydration time because.  
We have about 15% of our recipes in the company that have that required the dehydration time, but because it's such a long time it's more like, hey, what are the factors affecting this is it? Does it change winter to somewhere? Is it, you know, does the flow rate impacted? Is it the system like is it like what can change this cause you know we wanted to to reduce it and I know engineering are actually looking into buying a high share mixer which is essentially a blender to blend all the stuff together to get a more consistent.  
I suppose consistent density and forward their product which will put it will ultimately reduce the duration time. So Long story short, bazi involved yes.

 **Michelle M. Moran** 5:14  
Yoga.  
Yeah, you answered a number of questions there for me. Thank you. I so I just a question then. So when I, I took a look at let's say I printed out just a sample of let's say a product and the what the different phases are involved in this particular product. Now you have your star process and your step one cons which is the addition of all the raw materials which are manual and the water treats it. Then you have step 1-2 and three which is the agitations phases.

 **Participant 1** 5:43  
Yeah.  
Yeah.

 **Michelle M. Moran** 5:47  
But there is no specific duration phase.

 **Participant 1** 5:51  
There is, I suppose, yeah, because I suppose it it's.

 **Michelle M. Moran** 5:53  
There it's not. It's it's not, it's not like, let's say, everything else is spelled out. Step one calms step one agitation. There is no duration phase. So why is that?

 **Participant 1** 5:58  
Yeah, yeah, yeah, it's.  
It.  
Well, it did well it I suppose it's the understanding of the recipes. So the deoration will always come.  
And after the agitation before the addition of the first mobile tank. Right. So I kind of knew because I'd worked on the recipes for so long, I knew exactly where it was. So I know if you look at the recipes, you'll see a long phase start delay between the end cause I'm looking only the key phases, which is agitation at consumption, tank status and prompts and things like that. So like necessarily like dehydration is a process it. So I just kind of knew it was in there. I know exactly where it should be and.

 **Michelle M. Moran** 6:19  
Yeah.  
Yeah. Prompts and different prompts and.  
Yeah.  
Yeah.

 **Participant 1** 6:40  
Like that, you know I it it it you know the recipe just calls for. So I knew I didn't necessarily need to call it out as suppose I knew it was there.

 **Michelle M. Moran** 6:47  
Ohh OK that's fine. Um and then the different kind of metrics are things that you're you're quantifying. So let's say the phase duration times, then there's the phase start delay and then there's the flow rates and then you're phase overruns. So would you look at so the face start delay then that was just that's more or less just like how long it took to start it, is it?

 **Participant 1** 7:03  
Yeah.  
Well, it's the time between. So if one phase ends like say that the aeration, right, take that for an example, you might have a phase that ends at 12:00 PM in the like midday. Then the next phase didn't start till 6:00 PM, but that was the duration time they like. There was just we didn't necessarily call it out as as a phase, but that was it. The the time between one phase and the next phase was your duration time because you first whatever reason we could end a phase and there could be a breakdown on the on the homogeniser will say or something like that. There could be some some issue like the guys go offline.

 **Michelle M. Moran** 7:21  
Yeah.  
Yeah.

 **Participant 1** 7:43  
Lift handle for whatever it is. They would end one phase and they don't necessarily stack the next phase and it's just capturing that time between 1:00 and the other so that we can account for it down the line.

 **Michelle M. Moran** 7:52  
Okay so that wouldn't be reflected. Let's say that wouldn't be such a, let's say that wouldn't be a phase overrun. As such, the downtime, the downtime is just between the start and the end of the phase.

 **Participant 1** 7:59  
Yeah, it know.  
Exactly, yeah.

 **Michelle M. Moran** 8:03  
Yeah, alright, OK, now it's perfect and.  
So what? How does what kind? How is he OE calculated from these results?

 **Participant 1** 8:15  
It.  
Yeah. So it's essentially what we do is. So our production week is from 7:00 AM Saturday morning until Thursday evening, 7:00 PM, so that is per tank. That's 132 hours. If you work it out. So what we do is we say okay, let's say we're going to run.

 **Michelle M. Moran** 8:28  
Yeah.

 **Participant 1** 8:33  
50 batches, 20 batches would say 20 batches and.  
We sum up all the batch running time that might come to 80 hours. So then we know we had 132 hours 80 hours divided by 132 hours will come out. What's that 80 divided by?

 **Michelle M. Moran** 8:42  
Yeah.

 **Participant 1** 8:51  
Duck that about 60% awe. So essentially we're all we're really doing what the OE is the target is, I suppose this supposed it's giving you information on how much you're actually using your equipment. So one thing is like, OK, how much you're using it, but just because the machine is on doesn't mean it's necessarily effective. That's what we, that's where we're measuring the OE 2 will say or the OE3. That's why we put a target against every phase to say not it goes to that next level of okay perfect. You said you were adding bulk water, but how long it should have taken you 20 minutes. Why did it take you 30 minutes?

 **Michelle M. Moran** 9:09  
Hmm.

 **Participant 1** 9:21  
And then that's you say. Ohh well, that's just the way it. That's just the race of pumped. I started a machine that just ran to that long. That's when you might get the engineers involved and say hey, look at take a look at their.  
And that, you know, that might say, you know, geez, that, that pull, that bong that you know we we haven't changed that that piece of equipment and you know 10 years let's look at that and then they might get some sort of an uplift or whatever but that's the reason I suppose we will give you an indicator of how much you're using your equipment and how effective you are when you are using the equipment.

 **Michelle M. Moran** 9:32  
Yeah.  
Ohh Okay and that is um.  
I was with for want of a better word, broadcast it on the dashboard.

 **Participant 1** 9:57  
Yeah, it is. Yeah. So well, power BI and we actually built this DPM. So it's said the digital performance management screen. So it's available in all the areas and actually where we're rolling it out to all the sites in the world.

 **Michelle M. Moran** 9:57  
Barbie. Yeah.  
OK.  
Wow.  
Um, actually screen. OK, that's perfect screens.  
So for example, I was talking to Ollie and uh Thomas last night and um, I'm looking the results that I am looking at the majority of the time the downtime or the overrun is to do with the manual edition of the gums.

 **Participant 1** 10:31  
Yeah.

 **Michelle M. Moran** 10:31  
And for all batches. So there is an overrun.  
For every one of them, so it overruns the target.

 **Participant 1** 10:38  
He tells you have such a problem.

 **Michelle M. Moran** 10:38  
And.  
Yeah, there's a problem. So. So like it's not a solution to change the target, but um.

 **Participant 1** 10:47  
Well, we don't. Well, think about it like you wouldn't necessarily want to change the target cause then you, you you're accepting poor performance, I suppose what you want to do is, hey, you get the smart people in the room, you get the pro and you get there, you get the associates, you get the team leads, you get the OE manager, you get the engineers, you get all the people around the room and say, hey, this is a problem and then you say, OK, well, what's our target? OK, it's 60 minutes. And what are we getting now? We're getting 75 minutes. OK, well, let's look at why is that. And then you start looking at, well, is it to do with like?

 **Michelle M. Moran** 10:51  
No.  
Yeah, poor performance.  
Yeah.

 **Participant 1** 11:17  
Don't be material, is it to do with like we're not like there was some reason basically, but we wouldn't necessarily accept we if we bend to low performance then we'll get a high OE. But we're not necessarily solving any problem. We're just we're just making the numbers look a bit better you know.

 **Michelle M. Moran** 11:32  
Yeah, absolutely.  
And.  
OK.  
And yeah, no, that's that's just your you've answered everything a few questions. Few questions for me, but you've expanded which is fantastic. No that's that's brilliant. Ohh and thank you so so much.

 **Participant 1** 11:44  
Yeah.  
No worries and I suppose it's it's great to see that you're getting into it and more people getting into it because the company is changing, you know, like the whole world is changing. Everyone's becoming a bit more data-driven like everyone has a smartphone. Most people now have smart watches like you track your steps, you track everything, you know, like now in work, like, I suppose it's not odd to think that we're going to be looking at how do we take use data to our advantage to make things more effective and efficient. And having spoken to consultancy firms out there and having matched.

 **Michelle M. Moran** 11:48  
Um.  
Yeah.  
Yes.

 **Participant 1** 12:18  
Us against all other peers in the industry, top performers like we are ahead of the curve actually, so that what they're saying well we what we're doing is actually is world class. And while we're doing actually is we're not necessarily going out and buying software packages and spending hundreds of thousands of millions of euro, we built a lot of the stuff in house because we're just reusing our data smartly. So but there's this is not going away. It's only getting bigger and bigger. So it's great to see that actually people like yourself are actually going out, getting the skills upskilling themselves.

 **Michelle M. Moran** 12:47  
Ohh.

 **Participant 1** 12:48  
Preparing themselves for the future, you know, and it goes to everywhere. It's not necessarily manufacturing like the lab have a lot of self required as well like you know.

 **Michelle M. Moran** 12:55  
I know. Yeah, absolutely.

 **Participant 1** 12:57  
So it's everywhere.

 **Michelle M. Moran** 12:58  
It's a it's amazing. It's really is. I I I am, however hard it is and getting my head around these things as it is, it's actually very interesting just to see what it can tell you. You know, data is that data is key, I suppose. Just a quick question like so in terms of data analytics and machine learning and anything like is that currently is that that's what's applied is that?

 **Participant 1** 13:08  
Yeah, yeah, 1%.  
Yeah.  
But we're not necessarily applying machine learning at the moment because we're our fingers are on the pulse like because we've done like so like machine learning.

 **Michelle M. Moran** 13:22  
And what?  
No.  
You you you've skipped that part. Really. You've gone straight to to the.

 **Participant 1** 13:31  
Well, like, yeah, advanced analytics, who because we're like we're we're tracking everything in real time like we don't necessarily like machine learning would be fantastic in one way where like you know like one problem we have in in the on the site in our company is like the schedule, right. So you can imagine like every area, every PO trying to put that schedule together every week is like trying to get a lot of numbers. So like if you actually asked the guys like how do you do it, you actually like find out that it's a very, very manual process.

 **Michelle M. Moran** 13:37  
Yeah.  
Yes.  
Oh, oh gosh. Mental.

 **Participant 1** 14:03  
So like whilst our team was only in existence for three years and we've done a lot of good work, there's a lot more to do, like there's kind of there's kind of an evolution like of data analytics where it's like okay you get like the prescriptive analytics will what has happened and then you're saying well what should have happened and then how do I affect that, what will happen and then like so then it's like there's an evolution and we're kind of getting there at the moment. But I just had to start. Our team is life. So we're trying to expand so, but it's a machine learning is on the way. But I suppose it's not necessarily like you know the.

 **Michelle M. Moran** 14:07  
Uh huh.  
Ohh yes.  
Yeah.  
Yes, exactly.

 **Participant 1** 14:34  
Answer to all problems and some some business stakeholders do think that they're like just throw machine learning at it's like you don't like, it's that will not. What will that tell us? We won't tell us anything you know.

 **Michelle M. Moran** 14:35  
No. Ohh no no no.  
No. You have to quantify it. You have to kind of justify it. And in fairness like is it going to answer what you want at this particular moment in business in the time like so I do get that. So another thing would be the preventative maintenance side of things. So let's say if the tank production tank goes down and it's a mechanical issue.

 **Participant 1** 14:53  
Yeah.  
Yeah. So we are, we're.

 **Michelle M. Moran** 15:03  
Rather than a software issue, so are is the prevent. Is there preventative maintenance being tracked?

 **Participant 1** 15:09  
There is, yeah. So for the Homogenizers will still last year we had an issue where like the homogeniser seals broke. So the the the engineering team was are the main team was saying well that's kind of unusual like we wouldn't have expected to see that they you know. So basically they said well you know what what's the most pressure on those and it's to do with the like trying to get the tanks into terminal balance to basically.

 **Michelle M. Moran** 15:33  
Yeah.

 **Participant 1** 15:33  
To blast it in the pasteurisation process to kill all germs or whatever and but basically the teams were leaving.  
And I suppose it just goes back to the training of the teams. They were leaving us in thermal balance for too long, so that was putting way too much pressure on the seal. So now we're tracking that in real time and maintenance can see that and they can give feedback back to the team. Like, you know, I think it's just more awareness around it. So like analytics helped us to answer the question of, yeah, there was a problem here. And this is the answer. And now the solution as well. It's just more awareness, you know, the kind of way it doesn't need to because we solved the problem upstream. We don't necessarily need machine learning to tell us a problem is about to happen here. You know, it's just more awareness.

 **Michelle M. Moran** 16:06  
Yes.

 **Participant 1** 16:11  
And then it's just not gonna happen again.

 **Michelle M. Moran** 16:14  
Perfect. Yeah. No, no, I just. I just thought of that. Alright. Cause a lot of the research I've I've looked at is more preventative maintenance rather than the predictive. Like like what she has said that while we are currently doing it's not really applied yet because you're getting what you need from the data for the lowly or the OE, whatever measure metrics. So yeah, a lot of it was basically on preventative maintenance and anticipating issues and documenting them on that. So yeah, no, no, that's that's just wanted to know have has that been.

 **Participant 1** 16:22  
Yeah.  
Yeah.

 **Michelle M. Moran** 16:44  
Applies in there, but yeah, no, that's brilliant, Owen. I think that's it actually, to be honest with you. Thank you so, so much for your time and all your help and.  
I.

 **Michelle M. Moran** stopped transcription

**Participant 2 Interview-Meeting Script**

0:0:0.0 --> 0:0:0.390  
Michelle M. Moran  
OK.

0:0:0.-310 --> 0:0:1.910  
Participant 2  
I met Mary Walsh this morning. She was in great form.

0:0:2.540 --> 0:0:3.760  
Michelle M. Moran  
Ohh yeah, very good.

0:0:4.410 --> 0:0:4.660  
Participant 2  
Yeah.

0:0:21.270 --> 0:0:21.470  
Participant 2  
Yeah.

0:0:25.10 --> 0:0:25.460  
Participant 2  
No problem.

0:0:6.560 --> 0:0:33.630  
Michelle M. Moran  
OK. So we will get going and we won't waste your time. So thanks million Ollie for doing the our Miss Oliver for doing this for me. I have sent you on the consent form earlier on. So if you could organise that for yourself and Thomas that would be much appreciated. It's just for a preliminary thing really. So I'm just going to ask you just a few questions based on round my research objectives. So the first was just really the explanation.

0:0:33.850 --> 0:0:42.200  
Michelle M. Moran  
Operation on the quantification of production downtimes is what I'm looking at. So I just want to just your expert knowledge on it.

0:0:43.530 --> 0:0:50.860  
Michelle M. Moran  
So, do you know how the company is currently documenting and quantifying downtimes during the production process at the moment?

0:0:51.780 --> 0:0:54.350  
Participant 2  
Yes. Yeah. Michelle, the company are using.

0:0:55.650 --> 0:1:9.620  
Participant 2  
Data-driven results from operations and it's used to calculate OE. So you have OE figures for every single batch that's built in the plant and those are reviewed.

0:1:9.750 --> 0:1:17.820  
Participant 2  
And every morning at what's called the 9:15 meeting, it's the operations meeting where they go through.

0:1:18.160 --> 0:1:48.910  
Participant 2  
And all the different systems and any downtime which has affected those systems in the last 24 hours. So it's a concise piece of information that the guys didn't take actions from. And as this 9:15 meeting is a cross functional meeting, you have OPS, you have engineers, you have PC present and there may be actions required for each of those areas to be completed and these are reviewed on.

0:2:5.270 --> 0:2:5.720  
Michelle M. Moran  
You OK?

0:1:49.240 --> 0:2:6.130  
Participant 2  
Different lists the guys would take, they would generate a list of actions and would follow up on those actions in due course where they'd be the following morning, whether it be in a week's time or whether it be in two to three weeks time so that they're not lost.

0:2:6.750 --> 0:2:18.380  
Michelle M. Moran  
That's perfect. So in the case of, um, in the case of the tanks, the production area that I'm looking at, um, what you call it? So I'm looking at the 20.

0:2:27.10 --> 0:2:27.810  
Participant 2  
Correct. Yeah.

0:2:19.70 --> 0:2:30.790  
Michelle M. Moran  
22 to 23 systems the 25 and the 26. So they're all governed by owe as well. So for batches there meds with mucilage containing batches.

0:2:31.600 --> 0:2:33.200  
Michelle M. Moran  
They are all governed by OE.

0:2:33.940 --> 0:2:34.830  
Participant 2  
That's correct, yes.

0:2:35.400 --> 0:2:49.10  
Michelle M. Moran  
OK. So then do they look at different phases, you know the activities, so the different stages or does it look and in terms of the the the batch start to the minute the batch ends?

0:2:53.100 --> 0:2:53.860  
Participant 2  
No, you're you're.

0:2:50.10 --> 0:3:7.920  
Michelle M. Moran  
And that's how we calculate your operate your your owe. You don't look at the, you don't look at the individual. Let's say the the agitation. You're the ingredient addition followed by the agitation followed by the duration. Or do you just look at it as a whole as a batch?

0:3:8.590 --> 0:3:9.180  
Michelle M. Moran  
Mate.

0:3:10.170 --> 0:3:20.120  
Participant 2  
No, no there is. There is certain visuals on our GPM, which is our our our local software which calculates.

0:3:35.90 --> 0:3:35.400  
Michelle M. Moran  
Yeah.

0:3:42.430 --> 0:3:42.860  
Michelle M. Moran  
Yes.

0:3:21.300 --> 0:3:51.890  
Participant 2  
How the batch has been built as a whole, but in relation to E it does go down to the phase level. You have a different phase for every single action in a in a batch build, each of those phases is completed or has prerequisite time associated with them, so therefore it is a gauge of how well the associate is or has completed the build as per the particular time associated with that.

0:3:51.960 --> 0:3:53.530  
Participant 2  
Is so there.

0:3:52.960 --> 0:3:54.800  
Michelle M. Moran  
Alright, OK, well, very good.

0:4:0.420 --> 0:4:0.850  
Michelle M. Moran  
Horse.

0:3:54.340 --> 0:4:11.610  
Participant 2  
Therein lies the community there they therein lies the the calculation of our OE. So for example, we have set time and you mentioned come earlier, we have set time put aside for the DA relation of a gum. So it might not necessarily deaerate.

0:4:30.300 --> 0:4:30.750  
Michelle M. Moran  
Hmm.

0:4:12.30 --> 0:4:41.790  
Participant 2  
Uh to that level of acceptance within that time, therefore, we have to extend the time this is all visual. This has taken a sample measuring the density. The associate will then make a decision whether they can progress with the next phase of the match, which is edition of your oil, or they have to wait for another while for it to do a rate further. Again, it's a quality check that is completed that will dictate whether the adoration is fully completed or not.

0:4:49.220 --> 0:4:49.780  
Participant 2  
I can, yeah.

0:4:42.440 --> 0:5:11.990  
Michelle M. Moran  
Okay. So for example I just this is just an example, I know you probably can't see it, but this is just for Kenya. So this is just an example of a batch, A batch from start to finish and the different phases that were involved in each one. It's just a sample and so let's say it doesn't say pacifically the generation stage or face. OK, but these do have it because these are specifically the ones I'm looking at and you.

0:5:12.410 --> 0:5:15.930  
Michelle M. Moran  
I think previously had said to me that anything that contains gum.

0:5:16.660 --> 0:5:18.430  
Michelle M. Moran  
Would require to be settled.

0:5:19.560 --> 0:5:21.450  
Participant 2  
That's correct. Yeah. That's correct. Yeah.

0:5:19.480 --> 0:5:28.660  
Michelle M. Moran  
Isn't that correct? So I just wanna know the after what's from what days to what days would be the duration time.

0:5:50.760 --> 0:5:51.70  
Participant 2  
Yes.

0:5:29.660 --> 0:5:52.10  
Michelle M. Moran  
So I have let's say, so the batch starts and process the S3 batch and process tank status. Then this is step one comes. So all the step one cons is the addition of the raw material which is the water, the sodium benzoate, the citric, the 2 the gums and then further water addition. Then there's step one, Step 2 and Step 3 agitation. Then there's HP.

0:6:9.950 --> 0:6:10.280  
Participant 2  
Yes.

0:5:53.90 --> 0:6:13.760  
Michelle M. Moran  
And then there's the selection of the destination tank. Then the batch complete QA pending. Take a sample and submit for QA sample to the lab results are OK and then HP again and then it's. So I just want to understand the. So once the after the government added and then the water is added so there's a flow, there's a flow rate there.

0:6:14.390 --> 0:6:14.700  
Participant 2  
Yes.

0:6:14.620 --> 0:6:17.420  
Michelle M. Moran  
And then step 212 and three is the agitation.

0:6:18.780 --> 0:6:23.660  
Michelle M. Moran  
This is followed by the HP. So what part is it stopped? And let's settle.

0:6:33.200 --> 0:6:33.510  
Michelle M. Moran  
Yeah.

0:6:34.540 --> 0:6:34.890  
Michelle M. Moran  
Yeah.

0:6:24.430 --> 0:6:35.840  
Participant 2  
So after after your initial agitation, so when we put all our materials into the mobile tank, sorry, into the main tank and we add our final.

0:6:38.630 --> 0:6:38.940  
Michelle M. Moran  
Yeah.

0:6:43.400 --> 0:6:43.760  
Michelle M. Moran  
Yes.

0:6:46.150 --> 0:6:46.520  
Michelle M. Moran  
Right.

0:6:37.160 --> 0:6:47.920  
Participant 2  
Amount of water it there will be an agitation step there, which is a quite an aggressive agitation to blend all of these into place at the end of that one phase.

0:6:50.950 --> 0:6:51.250  
Michelle M. Moran  
Yeah.

0:6:49.370 --> 0:6:59.370  
Participant 2  
Which would be timed. You will have the agitator knocked off that. That then is the DA bration time from the from the moment that the agitator is stopped.

0:7:8.590 --> 0:7:8.890  
Michelle M. Moran  
Yeah.

0:7:11.370 --> 0:7:11.790  
Michelle M. Moran  
Yes.

0:7:15.100 --> 0:7:15.530  
Michelle M. Moran  
Sorry.

0:7:0.450 --> 0:7:18.620  
Participant 2  
It's it's no derating. So because we had to agitate at such high levels to bring this into a a homogeneous solution, we would have introduced so much air into it. And it's at that stage again, I repeat that stage where the educator is knocked off and that it's left to the year 8.

0:7:19.350 --> 0:7:24.150  
Michelle M. Moran  
Okay so in can I just what's the HP phase then?

0:7:24.310 --> 0:7:28.430  
Participant 2  
The HP phase is later on, so the HP phase.

0:7:27.920 --> 0:7:32.860  
Michelle M. Moran  
This. This is dad. So the hphp phase follows the generation.

0:7:33.450 --> 0:7:41.980  
Participant 2  
It it would follow the addition of oil after the aeration. So your, your, your, your mucilage is your base basically.

0:7:42.270 --> 0:7:42.620  
Michelle M. Moran  
Yeah.

0:7:49.200 --> 0:7:49.590  
Michelle M. Moran  
Yeah.

0:8:3.660 --> 0:8:3.970  
Michelle M. Moran  
Yeah.

0:8:11.560 --> 0:8:11.910  
Michelle M. Moran  
Okay.

0:7:42.860 --> 0:8:12.140  
Participant 2  
At the same time as the Mucilages settling, the associate would go and they would build their oil mobile tank or their colour and they they know that they have sufficient time while it's deorating to get their oil mobile built. Once they're happy. So once they have taken their sample of the D rated solution and gotten the desired density, it's only then that they would go on to the next step, which would be oil addition or colour edition.

0:8:12.520 --> 0:8:20.80  
Michelle M. Moran  
So the oil edition of the colour edition, right? So for let's say the oil they would become they would, is that what you, 85 Mt are?

0:8:20.630 --> 0:8:21.100  
Participant 2  
Correct.

0:8:21.800 --> 0:8:22.210  
Participant 2  
Correct.

0:8:21.470 --> 0:8:33.520  
Michelle M. Moran  
These exhaustive cause, so much, so much, I'm looking at some of the batches have colour yes, and some have these additional ingredients from these U85 mtts. So these are the oil additions that are necessary.

0:8:39.30 --> 0:8:39.670  
Michelle M. Moran  
OK.

0:8:40.410 --> 0:8:40.780  
Michelle M. Moran  
Yeah.

0:8:33.140 --> 0:8:43.920  
Participant 2  
The the, the they they are oil additions and also you could have a flavour booster edition towards the end which is also coming in 85 Mt either 60.

0:8:49.570 --> 0:8:50.160  
Participant 2  
Correct.

0:8:51.70 --> 0:8:51.600  
Participant 2  
That's correct.

0:8:43.70 --> 0:9:13.680  
Michelle M. Moran  
Them them actually 586 them once they're all to do with the separate what they're making separately to add on after the duration. OK, so OK, so you know, I just wanted to get into my head because I'm there in, in the batch details that out from the factory floor shop batch and it gives these to the each of the phases, but it just doesn't specifically mention duration phase. So I just wanted to know at what stage what timings am I looking at in terms of so it's before?

0:9:13.780 --> 0:9:15.220  
Michelle M. Moran  
It's the step before the HP.

0:9:15.890 --> 0:9:17.150  
Participant 2  
Absolutely, yeah, yeah.

0:9:18.260 --> 0:9:18.460  
Participant 2  
Yeah.

0:9:20.80 --> 0:9:20.330  
Participant 2  
Your.

0:9:16.990 --> 0:9:21.970  
Michelle M. Moran  
OK, that's fine. And that is fine. So thank you. No, that's that's clear actually.

0:9:26.390 --> 0:9:27.330  
Michelle M. Moran  
Nearly yes, yeah.

0:9:21.20 --> 0:9:28.790  
Participant 2  
Your your HP, your HP is ultimately the the batches nearly there. So all your all your magicians are in place.

0:9:29.330 --> 0:9:29.680  
Michelle M. Moran  
Yeah.

0:9:32.520 --> 0:9:32.810  
Michelle M. Moran  
Yeah.

0:9:37.590 --> 0:9:37.940  
Michelle M. Moran  
Yeah.

0:9:29.570 --> 0:9:40.570  
Participant 2  
And they've it's been completely mixed and now the HP is the point of no return. Once we once we put it through the HP we have, we have a marginalised and pasteurised.

0:9:41.250 --> 0:9:43.860  
Michelle M. Moran  
Ohh. Perfect. OK, so then I am.

0:9:44.660 --> 0:9:50.640  
Michelle M. Moran  
The S4 is batch completes QA pending. What's that about what? What you what are you looking for there?

0:10:5.290 --> 0:10:7.10  
Michelle M. Moran  
I watched that for you.

0:10:10.10 --> 0:10:11.590  
Michelle M. Moran  
Depending on the the mixture.

0:10:12.610 --> 0:10:13.450  
Michelle M. Moran  
And the material?

0:10:18.740 --> 0:10:19.110  
Michelle M. Moran  
Works.

0:10:19.800 --> 0:10:20.200  
Michelle M. Moran  
Yep.

0:9:51.670 --> 0:10:22.530  
Participant 2  
So once the guy is completed, their HP or any type of a blend or any any batch build upstairs, there will be a quality check that must be completed before they can closeout the batch. It's either, it's can either be a density check or a bricks check. OK, so dependent upon the components in the batch. So obviously juice based materials will be closed out on bricks and other batches will be closed out on density.

0:10:23.10 --> 0:10:23.620  
Michelle M. Moran  
Ours.

0:10:32.750 --> 0:10:33.80  
Michelle M. Moran  
OK.

0:10:35.670 --> 0:10:36.120  
Michelle M. Moran  
Hmm.

0:10:23.130 --> 0:10:39.590  
Participant 2  
As for the point to note, if we have a brand new batch, no matter what it is, it would always look to close on density in its very first manufacturer and it's from that then that we get a bricks reading which will be used in in batches 234 and so on after that.

0:10:40.130 --> 0:10:45.490  
Michelle M. Moran  
Yours and where they they're logged in the the BMR. Or is it on system?

0:10:44.810 --> 0:11:10.560  
Participant 2  
They there will be, there will be initially lodged in open batch so they guys would key in the result that they got and that then subsequently would be relayed onto the paperwork. So we have gone paperless to a certain extent in, in beverage base. But at the very end we print out all the results onto 2, maybe three pieces of paperwork, which is now our new BMR.

0:11:11.670 --> 0:11:12.50  
Michelle M. Moran  
OK.

0:11:11.290 --> 0:11:26.330  
Participant 2  
And it used to be 10 pages long. Now we have it. We have it, we have it compacted down to two. What, what with all the data printed, not handwritten in it's all printed. And they are for to to keep, for, for, for the next few years.

0:11:27.370 --> 0:11:29.20  
Michelle M. Moran  
Very good. Um.

0:11:29.750 --> 0:11:32.440  
Michelle M. Moran  
I says sample to the lab results OK.

0:11:33.570 --> 0:11:33.770  
Participant 2  
Ohh.

0:11:33.170 --> 0:11:33.960  
Michelle M. Moran  
It's it's.

0:11:44.630 --> 0:11:45.130  
Michelle M. Moran  
I think that.

0:11:46.660 --> 0:11:46.980  
Michelle M. Moran  
Here.

0:11:34.770 --> 0:12:4.550  
Participant 2  
So. So in certain circumstances, like in flexible manufacturing, I suppose a better example, you may have a clarity sample to send to the lab in order to get the green light to proceed to closeout or top up or whatever, whatever is needed to close it out in relation to our beverage base, main systems and the only sample that we will be sending to the lab would be retained.

0:12:13.690 --> 0:12:14.80  
Michelle M. Moran  
Hmm.

0:12:4.700 --> 0:12:16.950  
Participant 2  
And the retention samples for the bigger batches, um, there would be no batch that would require intermediate testing or in process testing prior to closeout.

0:12:17.920 --> 0:12:19.210  
Michelle M. Moran  
OK, that's perfect.

0:12:18.490 --> 0:12:30.620  
Participant 2  
But again, as I said already, that's not the same in flexible and for some of the coke or Sprite matches that we have where in process tests are required to to give the go ahead to closeout.

0:12:31.230 --> 0:12:41.350  
Michelle M. Moran  
Okay. That's perfect. Thanks, Olly. So then um, just just random questions are really, but are there any initiatives and strategies in place to address and reduce potential downtimes?

0:12:42.360 --> 0:12:43.170  
Participant 2  
Yes there is.

0:12:44.630 --> 0:12:54.370  
Participant 2  
I suppose we have a lot or we had a lot of different types of additions into main tanks, so anything from liquids?

0:12:55.150 --> 0:13:3.460  
Participant 2  
From drums, we have bulk charges. We have mobile bills that are then put into them into the the main tank.

0:13:4.60 --> 0:13:22.200  
Participant 2  
But what we did a project that was completed over the last 18 months to two years was trying to pre weigh bulk powders into tots that they can be, you know put into tots in a very safe manner using a gantry and a hoist.

0:13:22.900 --> 0:13:52.250  
Participant 2  
At with with calibration scales underneath, what we would do then is we'd take those tots, we'd say for example 726 kg of sodium gluconate and transfer that directly into the main tank. You compare that then to what we used to do prior to this, which was to handle this ingredient in 25 kg bag. So you're talking about an awful lot of bags to be handled.

0:13:52.320 --> 0:14:11.180  
Participant 2  
Manually to be pushed straight through the memory of a tank compared to the initiative that the guys came up with where we'd pre weigh super sacks of this ingredient into a tote and that tote then would be fork tripped over as far as the area it would be.

0:14:12.220 --> 0:14:31.880  
Participant 2  
Moved over the Manway and the gate at the bottom of the tolls will be opened and we would dispense in exactly what we were looking for, which has already been pre weighed. So that's one initiative that has really helped two of our main runners in there.

0:14:36.860 --> 0:14:37.200  
Michelle M. Moran  
Yeah.

0:14:33.500 --> 0:15:3.10  
Participant 2  
And it's it has. It has saved us some time. I suppose you might ask the question well, does it not take time to build these thoughts? When we looked at the schedule, we found that there was opportunities earlier on in the week where matches were processing and there were long processes where really all they need is 1 associate keeping an eye on a screen. The other associate then could go and build a tots in advance so that that's where the saving was. So that we're utilising.

0:15:3.410 --> 0:15:5.330  
Participant 2  
The associates time to the best we could.

0:15:7.130 --> 0:15:7.310  
Participant 2  
Yep.

0:15:6.170 --> 0:15:10.950  
Michelle M. Moran  
Absolutely sounds good and that would be, let's say for the likes of the addition of the gums with it.

0:15:36.360 --> 0:15:36.680  
Michelle M. Moran  
Hmm.

0:15:11.370 --> 0:15:41.560  
Participant 2  
Uh, the gums, I suppose, could be used. I know for the system 24 and 25, which are medium sized tanks, and I know there's a a batch out there which requires I think it's 650 odd kg of gum. So our gum comes in either 25 kg bags or 500KG super sacks. So there is going to be a certain level of manual intervention where an associate.

0:15:47.670 --> 0:15:47.940  
Michelle M. Moran  
Yeah.

0:15:41.630 --> 0:15:51.20  
Participant 2  
You'll have to physically lift A25KG bag over the manway and dispense it to make up to the 600 odd kg, but.

0:15:50.710 --> 0:15:51.70  
Michelle M. Moran  
Yes.

0:15:56.500 --> 0:15:56.950  
Michelle M. Moran  
Yeah.

0:15:51.640 --> 0:15:59.820  
Participant 2  
That the density of gum is quite large and we don't have torts big enough at the moment to actually.

0:16:0.820 --> 0:16:8.250  
Participant 2  
Fill out our pre way such a large amount of gum so the guys do have a another.

0:16:25.30 --> 0:16:25.640  
Michelle M. Moran  
Much better.

0:16:8.340 --> 0:16:30.320  
Participant 2  
And in in engineered Auger in place where they can just spend the ingredient as slowly as they can into their tanks. Therefore, it might help the agitation and blending it in much better. But again that that's project work that, that, that still has a lot of work to be completed on.

0:16:31.700 --> 0:16:32.180  
Participant 2  
What's that?

0:16:30.860 --> 0:16:45.700  
Michelle M. Moran  
Yeah. So thank you. So let's say for example, I'm just look at there are looking at the results of what I've been doing so far of all the ingredients. That's the gums that takes the longest cause and it is manual. It's a manual process.

0:16:46.160 --> 0:16:58.970  
Michelle M. Moran  
And I just wanna know that the difference. So I'm in the research results that I got downloaded and it has the phase duration. So how long it takes then it has the start phase.

0:16:59.780 --> 0:17:8.520  
Michelle M. Moran  
Delay measure and then it has the and then it calculates the phase overrun, which is the downtime which you know what makes it extra but.

0:17:8.920 --> 0:17:27.390  
Michelle M. Moran  
And the face start to delay. So and I and and then there's the there's the target start delay and then there's the target duration who sets the targets, who who, who sets them targets? Let's say there's also a target for the flow rate as well. So are these all just preset?

0:17:45.30 --> 0:17:45.420  
Michelle M. Moran  
Yeah.

0:17:27.970 --> 0:17:46.540  
Participant 2  
Yeah, the the would have been investigations completed and timings completed for a lot of these. So for instance you you mentioned the flow rate, the particular flow rate, A-Team looked at the rate of flow, we'll say for bulk orange juice if was.

0:18:3.910 --> 0:18:4.230  
Michelle M. Moran  
To.

0:17:48.100 --> 0:18:13.310  
Participant 2  
It was pumping from the bulk storage area out of a container out there 22 odd tone which was temperature set at -5 degrees. So obviously the viscosity at that temperature was quite quite high, and the speed that the liquid was able to come up to the beverage base area was was much slower. So it team of engineers and.

0:18:13.930 --> 0:18:43.430  
Participant 2  
Um associates and beverage base, along with process quality assurance, completed a project whereby they brought the temperature of the juice to 0 degrees OK and they did some cheques and trials at having the pump speed higher, which eventually showed us that we could get the juice in much quicker. Chests were done on the stability of the juice at that temperature and results came back to say that everything was going to be okay. So the guy who's.

0:18:43.520 --> 0:18:56.720  
Participant 2  
Put that in place as. Now when you set point that our juice will now be stored at 0 degrees Celsius and our pump speed now can go to. I believe it was 180 litres per minute.

0:18:58.580 --> 0:18:58.920  
Michelle M. Moran  
OK.

0:18:58.70 --> 0:19:8.510  
Participant 2  
So that was one collaboration between many different groups in order to help the efficiency of a batch Bild in relation to the pumping of bulk juice.

0:19:9.860 --> 0:19:10.450  
Michelle M. Moran  
OK.

0:19:12.550 --> 0:19:12.870  
Michelle M. Moran  
Now.

0:19:14.130 --> 0:19:14.820  
Michelle M. Moran  
Yeah, yeah.

0:19:11.20 --> 0:19:16.380  
Participant 2  
So that was that. That was a very good initiative and I supposed to further on through your question then.

0:19:16.840 --> 0:19:20.650  
Participant 2  
And associates in the area would have done timings on.

0:19:21.670 --> 0:19:30.890  
Participant 2  
The whole process of taking a barrel, sorry, a pallet of barrels into the room removing their lids.

0:19:37.800 --> 0:19:38.110  
Michelle M. Moran  
Here.

0:19:31.500 --> 0:20:0.540  
Participant 2  
Uh, pulling out the inner liners? I'm folding them over the sides of drums, physically taking those drums, tipping them into a hopper, A squeezing the bag and returning the pallet away so they can bring a fresh palette in. They would have completed timings on X number of drums per hour, and those figures then would have filtered into all the phases for all the batches that the guys doing beverage base they're on.

0:20:1.230 --> 0:20:4.250  
Michelle M. Moran  
Nor is OK. Um, OK.

0:20:6.150 --> 0:20:10.720  
Michelle M. Moran  
Um, where would lost my three and thought now? But I mean it's OK. No, that's great. Thank you.

0:20:10.990 --> 0:20:11.460  
Participant 2  
No home.

0:20:12.220 --> 0:20:16.920  
Michelle M. Moran  
What factors are pivotal when setting up who sets up the the production schedule?

0:20:24.820 --> 0:20:25.220  
Michelle M. Moran  
Hmm.

0:20:37.710 --> 0:20:38.20  
Michelle M. Moran  
Yeah.

0:20:17.800 --> 0:20:38.990  
Participant 2  
OK, so I suppose ultimately the production schedule is dependent upon our customer, #1 and #2. Then the planners would look at what trends are out there. You know is there, you know is there what kind of stock is in house and relation to we'll just call it the German market. So we have a German Fanta.

0:20:40.150 --> 0:20:48.340  
Participant 2  
They would then look to see what the forecast is in relation to the next week, 2 weeks, 3 weeks, and they would plan accordingly then.

0:20:49.440 --> 0:20:54.550  
Participant 2  
In relation to their intention to build German Fanta so.

0:21:8.250 --> 0:21:8.720  
Michelle M. Moran  
Hmm.

0:20:55.320 --> 0:21:14.330  
Participant 2  
That's only the the start of it. What they have to do in the background then is they have to get juice. They have to make sure they have sufficient ingredients in place for operations actually to be able to follow through and build said batches. So they do have a lot of work to do, both before and after the forecast is is set.

0:21:14.950 --> 0:21:38.190  
Participant 2  
And so yeah, they would, they would ensure that we have the correct amounts of material, which is in unrestricted available to us. So that would say in week 35 the intention is to build 4 tanks of German Fanta that we would have everything available to us for that week in that particular day that was meant to be built.

0:21:46.160 --> 0:21:46.490  
Participant 2  
Yeah.

0:21:39.70 --> 0:21:48.450  
Michelle M. Moran  
So for for just using this as an example, then just to keep it, just call it orange and watch got it so.

0:21:48.530 --> 0:21:57.80  
Michelle M. Moran  
It, like so in the production area with all these various tanks, depending on the schedule, depending on the quantity this required.

0:21:57.770 --> 0:22:1.860  
Michelle M. Moran  
On the coast of water, the customer wants will depend on what tank it's used. It's built in.

0:22:2.820 --> 0:22:3.800  
Participant 2  
That that is correct.

0:22:3.280 --> 0:22:5.880  
Michelle M. Moran  
Are as it is. The tank is a tank assigned.

0:22:7.770 --> 0:22:8.30  
Michelle M. Moran  
You.

0:22:6.760 --> 0:22:12.480  
Participant 2  
No, no. A system would be assigned for the larger runners, so we have we have many different.

0:22:12.570 --> 0:22:18.190  
Participant 2  
Have a fantastic out there that go to very large markets so.

0:22:18.200 --> 0:22:18.570  
Michelle M. Moran  
Yeah.

0:22:24.290 --> 0:22:24.600  
Michelle M. Moran  
Hmm.

0:22:28.130 --> 0:22:28.380  
Michelle M. Moran  
Yeah.

0:22:18.770 --> 0:22:43.90  
Participant 2  
And we have big systems which are 20 tonne and we have medium to small systems which are tin and four tonne. So dependent upon the volume required, obviously the the bigger runners will go as multiple chains on the larger systems and our smaller runners then would be fitted into the medium to small beverage base tanks.

0:22:43.930 --> 0:22:45.420  
Michelle M. Moran  
Ohh OK very good.

0:22:46.20 --> 0:22:47.490  
Michelle M. Moran  
Um, OK.

0:22:49.20 --> 0:22:49.750  
Michelle M. Moran  
And see.

0:22:50.560 --> 0:22:55.530  
Michelle M. Moran  
So do you foresee any any major challenges so?

0:22:56.850 --> 0:23:3.960  
Michelle M. Moran  
So I didn't realise that you have E actually have OE on every tank on every batch Med.

0:23:4.460 --> 0:23:4.780  
Participant 2  
Yes.

0:23:5.430 --> 0:23:12.640  
Michelle M. Moran  
And that is as a result of the entire phases, the timing of each part part of the batch makeup.

0:23:13.500 --> 0:23:14.590  
Participant 2  
That is correct, yeah.

0:23:14.500 --> 0:23:21.280  
Michelle M. Moran  
And you just go start it. So we have certain targets to meet. And if they don't meet, if you discuss it at the 9:15 meeting.

0:23:21.900 --> 0:23:22.380  
Participant 2  
Correct.

0:23:22.860 --> 0:23:33.830  
Michelle M. Moran  
So let's say you have your German Fanta. It was met in, let's say one of the 20 tonne ones, the 22 to empty ones and it was delayed.

0:23:34.980 --> 0:23:37.270  
Michelle M. Moran  
Or goes downtime.

0:23:38.50 --> 0:23:39.690  
Michelle M. Moran  
And so that's brought up at the meeting.

0:23:40.720 --> 0:23:41.110  
Participant 2  
Yes.

0:23:40.850 --> 0:23:42.510  
Michelle M. Moran  
And you have to have a reason why.

0:24:4.470 --> 0:24:4.910  
Michelle M. Moran  
Or.

0:24:5.630 --> 0:24:6.120  
Michelle M. Moran  
Right.

0:23:43.90 --> 0:24:13.110  
Participant 2  
Yes, we would strive to try and come to the OR get to the bottom of what? What's the what caused the downtime. So I suppose 11 particular example might be that there may have been a delay further down the chain. There may have been a delay in liquid filling which may have held up a tank that we needed as a destination tank for a HP of a particular chain. But that's look at that is an extreme measure that we can't really put an action in place.

0:24:14.740 --> 0:24:15.150  
Michelle M. Moran  
No.

0:24:13.180 --> 0:24:25.130  
Participant 2  
Against Porsche, if there was, if there was would say a breakdown, or if there was a an issue with staging or you know if there was an issue with.

0:24:25.250 --> 0:24:25.730  
Participant 2  
Um.

0:24:26.620 --> 0:24:33.400  
Participant 2  
Uh, the correct amount of juice not being present for some reason or other that we had to go and call up another.

0:24:34.50 --> 0:24:44.740  
Participant 2  
Batch of that same ingredient and it was an hour away from being sent up to the area. There are other types of downtime that wedding car.

0:24:43.940 --> 0:24:49.380  
Michelle M. Moran  
Downtimes, yeah, 100% workers, that's all. Is that all recorded in?

0:24:56.200 --> 0:24:56.480  
Michelle M. Moran  
Yeah.

0:25:5.640 --> 0:25:6.0  
Michelle M. Moran  
Yeah.

0:24:50.40 --> 0:25:16.990  
Participant 2  
That would be recorded in the phase. So obviously if we didn't have enough of a certain juice X that we needed, that phase would keep running until we've all the required quantity in. So there was a set time against the phase and now it's been extended by an hour, an hour and a half. We have to account for that hour and a half and we have we have to put in the explanation as to why, why this downtime occurred.

0:25:17.810 --> 0:25:18.320  
Michelle M. Moran  
OK.

0:25:17.760 --> 0:25:33.900  
Participant 2  
So we'll say for example, if it was juice not available, we'd have to put that in place and then that would philtre into our our OE data. And when we go to review the particular build of that batch, we would find that.

0:25:34.710 --> 0:25:44.190  
Participant 2  
It had a this issue had a serious contributing factor towards the the extended time it took to get it completed in that phase.

0:25:45.280 --> 0:25:48.640  
Michelle M. Moran  
OK. Thank you. So prior to OE.

0:25:49.620 --> 0:25:50.710  
Michelle M. Moran  
How is this managed?

0:25:55.720 --> 0:25:56.260  
Michelle M. Moran  
That that.

0:25:51.720 --> 0:25:56.510  
Participant 2  
This was managed through the start and the finish of a batch, so we would have had.

0:26:0.370 --> 0:26:0.680  
Participant 2  
No.

0:25:57.10 --> 0:26:3.520  
Michelle M. Moran  
Okay so you wouldn't be looking individually at the phases you would have been just looking at the start to the batch to the end of the match and?

0:26:4.390 --> 0:26:5.870  
Michelle M. Moran  
Yeah. OK.

0:26:4.380 --> 0:26:12.980  
Participant 2  
Correct and and and figuring out then. Well if if that batch took 8 hours to build, we'd have to find out the reason why it actually took 9 1/2.

0:26:13.860 --> 0:26:18.30  
Participant 2  
Just from when it started to when it got closed out and we.

0:26:17.490 --> 0:26:28.300  
Michelle M. Moran  
And yeah, and hot water. So what? What has happened between then and now? So is it just a more and more more software, more capability?

0:26:27.500 --> 0:26:29.470  
Participant 2  
Ohh it's it's it's it's data.

0:26:29.70 --> 0:26:31.830  
Michelle M. Moran  
Did that the data analysts excited things? Is it?

0:26:36.200 --> 0:26:36.580  
Michelle M. Moran  
OK.

0:26:31.370 --> 0:26:40.830  
Participant 2  
The Dash Analytics site has been the the main main force or here. It has opened, basically opened her eyes and relation to you know.

0:26:55.50 --> 0:26:55.370  
Michelle M. Moran  
Yeah.

0:27:1.140 --> 0:27:1.520  
Michelle M. Moran  
Hmm.

0:26:41.490 --> 0:27:2.690  
Participant 2  
What's involved in the phase the impact of a phase running over the impact of a phase starting too early? That's another one that can give us a negative OE even though we're trying to get a step ahead. So if I bring you back to the example we discussed earlier where the mucilage is agitating and the ages have been turned off.

0:27:3.670 --> 0:27:11.360  
Participant 2  
If the associate started the mobile prep phase a little bit early because you know they were available to do that.

0:27:11.780 --> 0:27:12.90  
Michelle M. Moran  
Here.

0:27:36.690 --> 0:27:37.100  
Michelle M. Moran  
OK.

0:27:12.610 --> 0:27:43.20  
Participant 2  
That that phase keeps running until the mobile is actually dispensed into the tank. So you're looking at data timers running there. You know that, you know are going to impact the OE, but you know it hasn't impacted the overall build of the batch even though your owe might be down slightly, does a lot, there's still a lot of work to do in the background there to get the most accurate data.

0:27:43.260 --> 0:27:46.890  
Participant 2  
To represent what the associates are doing to build a batch.

0:27:47.390 --> 0:27:56.990  
Michelle M. Moran  
OK, OK, I understand. So do you understand, do you understand the background that's going on? So when you like so?

0:27:58.220 --> 0:28:7.380  
Michelle M. Moran  
And are these results? Are these are automatically calculated in the background and it gives you onto a dashboard like say the power BI is that it?

0:28:8.330 --> 0:28:8.660  
Participant 2  
Yes.

0:28:8.50 --> 0:28:11.640  
Michelle M. Moran  
So what you it's already it's once the match starts.

0:28:12.290 --> 0:28:22.20  
Michelle M. Moran  
Work. And when she start the batch, everything is logged. Everything's documented in the background and it automatically calculates the OE and is displayed on a dashboard.

0:28:28.170 --> 0:28:28.400  
Participant 2  
Yep.

0:28:40.240 --> 0:28:40.450  
Participant 2  
You.

0:28:22.710 --> 0:28:42.420  
Michelle M. Moran  
So we continue on you, you do the prompts on your screen and all of that, but that's as far as you go in terms of you won't unless something happens in on, on site. Are you there are no downtime, you'll see it, but if not, you won't see anything until you get the OE for that batch.

0:28:45.580 --> 0:28:45.690  
Michelle M. Moran  
Is.

0:28:55.270 --> 0:28:55.640  
Michelle M. Moran  
Yeah.

0:28:42.550 --> 0:28:58.660  
Participant 2  
Yeah, correct. Correct. You're you don't really know what the fifth. You don't have your finger on the pulse at the exact time that something is happening. So we'll say if it's an addition phase or an oil mobile build phase, you won't know until after the fact how that is going.

0:29:0.310 --> 0:29:1.180  
Participant 2  
That that does that.

0:28:59.700 --> 0:29:9.430  
Michelle M. Moran  
Okay so like you can't you can't dip in and say let's say at the duration stage you can't dip in to say that all everything else is running according to time.

0:29:10.140 --> 0:29:10.460  
Michelle M. Moran  
No.

0:29:23.210 --> 0:29:24.120  
Michelle M. Moran  
No, no, just.

0:29:25.570 --> 0:29:26.440  
Michelle M. Moran  
Yeah, yeah, yeah.

0:29:11.260 --> 0:29:38.780  
Participant 2  
Well, I suppose you could if you went looking, but it you know, the guys wouldn't have had time really to go looking at that to that level of detail. And myself personally I wouldn't have been looking at it. I'd have been interested in what the data was after. But at the time at the time, you know you, you you get a feel for a phase if there's a problem and you go into the downtime tool that the guys have, you would say you would certainly see if something was running over.

0:29:39.240 --> 0:29:42.860  
Michelle M. Moran  
Yeah. Yeah, yeah, yeah. And with the. So would it be um.

0:30:1.440 --> 0:30:1.780  
Participant 2  
Yeah.

0:29:44.490 --> 0:30:4.100  
Michelle M. Moran  
Let's say for example, for the likes of when I'm talking to Thomas. Now, after you like, he wouldn't even more or less concerned about this. He'll be very concerned about getting the batch done, getting it out basically. And it'll be management that looks after the newest there's anything wrong with the bachelor that late starts or the the downtimes isn't that is it that would be very it.

0:30:8.530 --> 0:30:9.250  
Michelle M. Moran  
Ohh.

0:30:10.200 --> 0:30:11.330  
Michelle M. Moran  
Yeah, absolutely.

0:30:19.920 --> 0:30:20.260  
Michelle M. Moran  
Yeah.

0:30:24.170 --> 0:30:24.660  
Michelle M. Moran  
Ohh.

0:30:9.300 --> 0:30:24.880  
Participant 2  
An issue that he knew of and that he was trying to get sorted but transfer your question. Yeah, there's certain things that he won't know until he closes it out and he has to go and enter his downtime for the different phases at the end and he going, what happened here? He'd have to look back and see.

0:30:27.510 --> 0:30:27.750  
Participant 2  
Yeah.

0:30:33.710 --> 0:30:34.170  
Participant 2  
You're fine.

0:30:25.350 --> 0:30:34.620  
Michelle M. Moran  
Okay okay. Well, I can ask him that. Then when I when I'm talking to him anyways. But let me see anything I need. Anything else? Sorry I've kept you long time first.

0:30:37.680 --> 0:30:42.340  
Michelle M. Moran  
Let me see. I think I've asked you most everything really. You've really very good explaining.

0:30:42.930 --> 0:30:43.640  
Michelle M. Moran  
And.

0:30:52.300 --> 0:30:57.510  
Michelle M. Moran  
Okay so the there isn't, is there? So it's really a dense and research to check the QA check.

0:30:58.680 --> 0:30:58.920  
Participant 2  
Yeah.

0:30:58.330 --> 0:31:1.580  
Michelle M. Moran  
Depending on the March, you could be a prick, the bricks or the density.

0:31:2.710 --> 0:31:3.780  
Michelle M. Moran  
Is there flow metres?

0:31:11.790 --> 0:31:12.580  
Michelle M. Moran  
For water.

0:31:4.500 --> 0:31:19.380  
Participant 2  
The flow metres in line, so flow metres for juice. Bull juice. Sorry. And there will be a flow metre for water also flow metre during HP. Very important that all these criteria are met and that we have those.

0:31:19.470 --> 0:31:23.210  
Participant 2  
Um. Uh, pieces of equipment in place to help us.

0:31:24.0 --> 0:31:26.390  
Michelle M. Moran  
Why is there flow metre for the HP? Sorry.

0:31:28.380 --> 0:31:28.780  
Michelle M. Moran  
Hurry.

0:31:26.710 --> 0:31:30.290  
Participant 2  
Well, when you're when you're moving your your liquid through.

0:31:29.990 --> 0:31:30.470  
Michelle M. Moran  
Ohh.

0:31:31.80 --> 0:31:39.660  
Participant 2  
It has to. It has to go through the plates the the HP. Sorry, the the the pasteurisation plates at a certain.

0:31:40.370 --> 0:31:53.160  
Participant 2  
Speed or flow so that we don't burn the product or that we actually comply with what core is saying in relation to the contact time that our beverage based has?

0:32:3.180 --> 0:32:3.560  
Michelle M. Moran  
Yeah.

0:32:11.640 --> 0:32:11.870  
Michelle M. Moran  
Yeah.

0:32:12.980 --> 0:32:13.430  
Michelle M. Moran  
On his.

0:31:54.0 --> 0:32:14.410  
Participant 2  
With the extreme temperatures of 85 plus degrees and also that we bring it back down to the below 20 degrees, then on the on the return back the way so flow is very important there and the rate has to be at a certain level which is pre validated for each match. So that's why it's important.

0:32:14.840 --> 0:32:19.230  
Michelle M. Moran  
And is there, is it he is HP done for all matches that containing me silage?

0:32:20.380 --> 0:32:21.130  
Participant 2  
Um.

0:32:20.110 --> 0:32:22.610  
Michelle M. Moran  
Are just juice containing or.

0:32:23.750 --> 0:32:24.440  
Michelle M. Moran  
Are this?

0:32:27.550 --> 0:32:27.920  
Michelle M. Moran  
Yes.

0:32:38.340 --> 0:32:38.770  
Michelle M. Moran  
Yeah.

0:32:22.790 --> 0:32:46.160  
Participant 2  
It is what the rule of Tom would be. Anything containing a juice more than likely. Now we do have blends out there with juice and they're aseptically filled or they're blends that get filled and get sent straight to the freezer. The rule at home would be more than likely if the seduce present HP will occur, but there is a good few exceptions.

0:32:47.210 --> 0:32:54.890  
Michelle M. Moran  
OK. Yeah, cause I I see a lot of a lot of the batches that I'm looking at are examples of materials. Have the HP space.

0:32:55.370 --> 0:32:55.630  
Participant 2  
Yeah.

0:32:55.780 --> 0:32:58.350  
Michelle M. Moran  
Honest and there doesn't seem to be any juice.

0:33:3.810 --> 0:33:4.280  
Michelle M. Moran  
OK.

0:33:0.290 --> 0:33:11.870  
Participant 2  
What? Ohh, that that could be the case as well. This it could be there could be a highly sensitive material present in that build that we do need to pasteurise before we we fill it out.

0:33:10.660 --> 0:33:13.940  
Michelle M. Moran  
OK, if not necessary, not necessarily juice though.

0:33:14.290 --> 0:33:15.380  
Participant 2  
Not necessarily juice.

0:33:14.890 --> 0:33:19.480  
Michelle M. Moran  
No. OK, that's great and that's good. I didn't realise that and they.

0:33:20.540 --> 0:33:29.130  
Michelle M. Moran  
And have we? Ohh that's. I think that's yeah. No, no, I think I think that's age. Ollie. Thank you so so much.

0:33:29.410 --> 0:33:29.990  
Participant 2  
No problem.

0:33:40.580 --> 0:33:40.820  
Participant 2  
Yeah.

0:33:30.810 --> 0:33:49.870  
Michelle M. Moran  
I'll probably I know you've gone through before with me and but it just didn't I and the duration part was a bit confusing for me is in terms of when it starts and when it didn't like everything else, everything else that everything else is explained in the phases, they're very well explained and times, but that's the only step that's not.

0:33:51.450 --> 0:33:52.380  
Michelle M. Moran  
Documented.

0:33:52.860 --> 0:33:53.150  
Participant 2  
Yeah.

0:33:53.150 --> 0:33:58.560  
Michelle M. Moran  
It's not. It's not a phase on its own, it's not. Duration starts, duration stops.

0:33:59.580 --> 0:34:0.530  
Participant 2  
It it it?

0:33:59.610 --> 0:34:2.810  
Michelle M. Moran  
It's just it's, you know, I mean, I wasn't quite sure.

0:34:12.880 --> 0:34:13.140  
Participant 2  
Yeah.

0:34:3.780 --> 0:34:14.30  
Michelle M. Moran  
Um, you know, so like I said, I had an idea was after step three and HP Step 3 agitation and and HP, but I just wasn't quite sure.

0:34:33.350 --> 0:34:33.580  
Participant 2  
Yeah.

0:34:15.60 --> 0:34:35.390  
Michelle M. Moran  
Where where? It was like it's a for this particular example that I earlier on I was talking to, there's it can be so HP and then select a destination tank and then QA pending which which she said was density or bricks. Take a sample to the lab if necessary and then it says HP again.

0:34:41.210 --> 0:34:41.920  
Michelle M. Moran  
Pasteurising.

0:34:36.150 --> 0:34:45.830  
Participant 2  
Now, if you had certain batches, we have double, we have double pass to complete, so some are double homogenization.

0:34:46.630 --> 0:34:46.960  
Michelle M. Moran  
Yeah.

0:34:47.190 --> 0:34:49.700  
Participant 2  
So that's just goes through the homogeniser.

0:34:54.200 --> 0:34:55.280  
Michelle M. Moran  
Yeah, yeah.

0:34:50.610 --> 0:34:56.200  
Participant 2  
And his return back through the Homogeniser. So there's two homogenization steps, but we do have one.

0:34:57.660 --> 0:35:2.230  
Participant 2  
A product where we complete HP phase.

0:35:3.430 --> 0:35:3.830  
Michelle M. Moran  
Yes.

0:35:8.520 --> 0:35:9.10  
Michelle M. Moran  
Okay.

0:35:3.410 --> 0:35:12.570  
Participant 2  
And then we we complete a homogenization step afterwards. So it's it's it's pasturized once and homogenised twice.

0:35:13.490 --> 0:35:14.560  
Participant 2  
And that.

0:35:16.10 --> 0:35:17.410  
Participant 2  
That explains to his feet.

0:35:13.120 --> 0:35:18.930  
Michelle M. Moran  
All right, okay. So that explains the 2HP because it's not, it's not, yeah, it's not.

0:35:20.660 --> 0:35:24.390  
Michelle M. Moran  
It's not differentiators in the information, it's just says HP twice.

0:35:25.480 --> 0:35:25.870  
Michelle M. Moran  
So.

0:35:34.760 --> 0:35:35.270  
Michelle M. Moran  
Okay.

0:35:39.630 --> 0:35:40.260  
Michelle M. Moran  
Ohh yeah.

0:35:50.650 --> 0:35:51.20  
Michelle M. Moran  
Here.

0:35:24.720 --> 0:35:54.540  
Participant 2  
Yeah. Now if you will find on the medium beverage based systems, medium and small beverage based systems, you will see quite a lot of double homogenization batches. And basically this is in relation to the stability of the beverage. So we would have found at our indeed that we were getting Nick Ring, we're getting separation and after we say day 10 or day 15 of of analysis. So the action then with.

0:35:55.130 --> 0:35:56.180  
Participant 2  
Or indeed was.

0:36:9.930 --> 0:36:10.320  
Michelle M. Moran  
Ohh OK.

0:36:11.870 --> 0:36:13.80  
Michelle M. Moran  
Just forcing it through.

0:35:57.210 --> 0:36:18.690  
Participant 2  
It's not quite stable, so we just we need to put it through the homogeniser again, one more time. And it so happened then that our beverage became more stable afterwards. Homogenised twice, marginalisation is you're just forcing, you're forcing the, forcing it through what really extreme pressure, which will will.

0:36:19.550 --> 0:36:31.130  
Participant 2  
Yeah, I suppose for one to a better word, bind up the different molecules that are in the beverage that bind them together better so that they'll stay stable and present for longer.

0:36:33.160 --> 0:36:33.310  
Participant 2  
Yeah.

0:36:32.100 --> 0:36:33.340  
Michelle M. Moran  
Ohh okay that's perfect.

0:36:37.740 --> 0:36:37.980  
Participant 2  
Yeah.

0:36:33.940 --> 0:36:41.360  
Michelle M. Moran  
And so some, we'll just one more question. So just for this particular example, there is 2 tanks.

0:36:42.400 --> 0:36:44.850  
Michelle M. Moran  
So it starts off in 25 M 204.

0:36:45.620 --> 0:36:45.910  
Participant 2  
Yeah.

0:36:45.590 --> 0:37:1.940  
Michelle M. Moran  
And then after the second homogenising homogenisation, then it goes outside and then it goes to select destination tank which changes to 25 Mt 03 and that's where it's homed. So when it passes, so when it goes through the marginalisation, it goes into another tank.

0:37:2.400 --> 0:37:2.840  
Participant 2  
Correct.

0:37:5.80 --> 0:37:5.370  
Participant 2  
Yeah.

0:37:2.660 --> 0:37:8.470  
Michelle M. Moran  
Is that it? And then another storage tank, so you'll always have two tanks available when you're making a batch, is it?

0:37:9.810 --> 0:37:10.170  
Michelle M. Moran  
No.

0:37:8.890 --> 0:37:12.810  
Participant 2  
No, in this particular circumstance, your system 24.

0:37:13.700 --> 0:37:16.530  
Participant 2  
So that's 25 Mt or one and or two.

0:37:15.220 --> 0:37:17.180  
Michelle M. Moran  
25 yeah.

0:37:18.850 --> 0:37:19.210  
Michelle M. Moran  
Yes.

0:37:20.200 --> 0:37:20.610  
Michelle M. Moran  
Yes.

0:37:22.970 --> 0:37:24.120  
Michelle M. Moran  
Tent on, yeah.

0:37:17.410 --> 0:37:28.370  
Participant 2  
Therefore tone each OK 25 Mt, 03 and 04 are 10 tonne each, so so in this case and this particular example you only have one tank available.

0:37:32.200 --> 0:37:32.540  
Michelle M. Moran  
Yeah.

0:37:29.100 --> 0:37:35.50  
Participant 2  
Okay for the 1st and marginalisation step. OK, the tank you came from.

0:37:44.280 --> 0:37:44.830  
Michelle M. Moran  
Okay.

0:37:50.260 --> 0:37:50.850  
Michelle M. Moran  
Sorry.

0:37:36.790 --> 0:37:56.560  
Participant 2  
Is now then going to be the new destination tank, so there's going to be a delay in CIP ING that tank and to go from three back to four. So I'll just be, I'll just be clear as I can on it. On System 25, OK, if you have a double HP, sorry, a double homogenization batch.

0:37:57.760 --> 0:38:0.160  
Participant 2  
You're building it in 25 into your 4.

0:38:0.580 --> 0:38:0.890  
Michelle M. Moran  
Yeah.

0:38:1.260 --> 0:38:10.770  
Participant 2  
You have everything completed. You're ready to hit start on homogenization, you have to the if that won't happen unless 25 and two or three is clean, ready and available.

0:38:11.560 --> 0:38:14.520  
Michelle M. Moran  
Tick till to keep her to Cockfosters to go. Yeah.

0:38:14.70 --> 0:38:19.190  
Participant 2  
Told exactly so once for the contents of four go into 3.

0:38:20.600 --> 0:38:21.390  
Participant 2  
There's a pause.

0:38:22.190 --> 0:38:29.570  
Participant 2  
4 Now has to be washed down CIPD and it has to be clean and available so that it will take from three.

0:38:30.690 --> 0:38:41.720  
Participant 2  
That's the second pass, so 3 back and four again, that's your double pass completed time. Time is just excruciating. A lot of them batches have 20 hours route minimum.

0:38:42.600 --> 0:38:43.720  
Participant 2  
And it's all down to.

0:38:44.560 --> 0:38:45.160  
Participant 2  
That kind of.

0:38:44.590 --> 0:38:46.30  
Michelle M. Moran  
Please and availability.

0:38:55.430 --> 0:38:55.800  
Michelle M. Moran  
Yeah.

0:38:46.170 --> 0:38:58.990  
Participant 2  
Availability because we we can go from System 24 to 25 obviously because you're going from a smaller tank to a bigger tank, but you can't, you can't go from a bigger tank to a smaller tank, obviously.

0:38:57.490 --> 0:39:0.160  
Michelle M. Moran  
No. If there's going to a smaller tank, obviously. Yeah, yeah, yeah.

0:38:59.680 --> 0:39:11.890  
Participant 2  
So it's it's it's it's you know that dash in itself is a project later on at some stage maybe we need to upgrade the tanks, have them all bigger. It would give us more scope and relation to what we can or can't build.

0:39:12.680 --> 0:39:13.160  
Michelle M. Moran  
Yeah.

0:39:12.680 --> 0:39:13.300  
Participant 2  
You know there.

0:39:16.990 --> 0:39:17.240  
Participant 2  
Yeah.

0:39:13.910 --> 0:39:23.580  
Michelle M. Moran  
Yeah. Okay talk. Jeez, that's mad, mad. And I so sorry. Just a quick question then the 26 Mt, what was quite what capacity are they?

0:39:23.560 --> 0:39:26.950  
Participant 2  
You're talking about upwards up to 1400 kg.

0:39:30.590 --> 0:39:31.890  
Michelle M. Moran  
Very small batches, yeah.

0:39:54.150 --> 0:39:54.430  
Michelle M. Moran  
Yeah.

0:39:27.690 --> 0:39:58.100  
Participant 2  
So they're there for small batches, so I suppose something, something that the area improved on in the last three or four years is, I suppose sharing the load with the kitchen. So we have capacity to marginalised and pasteurise and system 26 and we can obviously it's for smaller batches, the kitchen have what's called a skid, which is a HP, but on a very small scale and they might homogenise and pasteurise from one mobile tank to another.

0:39:58.690 --> 0:40:8.780  
Participant 2  
So there was a little bit of, you know, I suppose trading done in relation to well, if you took these few oil blends and put them into mobiles and build them in the kitchen.

0:40:9.840 --> 0:40:15.110  
Participant 2  
Beverage based guys and Sister 26 can actually take your big runners, make them a little bit bigger.

0:40:15.530 --> 0:40:15.900  
Michelle M. Moran  
Yeah.

0:40:22.880 --> 0:40:23.390  
Michelle M. Moran  
You okay.

0:40:15.870 --> 0:40:26.610  
Participant 2  
And run them on System 26 and utilise it a bit more and that has been a very good friendship between the two areas for quite a while now. There is some very.

0:40:43.470 --> 0:40:43.740  
Michelle M. Moran  
There.

0:40:26.950 --> 0:40:53.140  
Participant 2  
And long winded mobile bills that when we go build them in beverage base, we really don't have the facility on the gantries that they have in the kitchen. So the trade off was if you looked after those mobiles which might have 20 ads in them at a time, we'll look after these three big runners on our bigger system. So we're we're building it once, but we're building it less than you used to build it because our quantities are much higher.

0:40:53.810 --> 0:40:55.410  
Michelle M. Moran  
Ignore. It's OK. Very good.

0:40:55.310 --> 0:40:55.600  
Participant 2  
So.

0:40:56.180 --> 0:40:56.610  
Michelle M. Moran  
Cool.

0:40:56.380 --> 0:40:59.50  
Participant 2  
That's working. Working as a team cross functionally.

0:41:1.870 --> 0:41:2.80  
Participant 2  
Yep.

0:41:8.560 --> 0:41:9.740  
Participant 2  
You OK? You're OK.

0:41:0.440 --> 0:41:16.310  
Michelle M. Moran  
Excellent. Still good to see us. Good to see us and allied. That's it. My God, your head is probably fried. 40 minutes. Thank you so much for your time and patience and your participation and was very in depth and lot. Lot learned definitely.

0:41:16.630 --> 0:41:21.790  
Michelle M. Moran  
And I'm just gonna stop the recording now. How do I do that?

0:41:23.360 --> 0:41:24.610  
Participant 2  
A.

0:41:28.410 --> 0:41:28.870  
Michelle M. Moran  
Yeah.

0:41:26.190 --> 0:41:29.850  
Participant 2  
If you're going to record and transcribe, I'd say probably in there.

**Participant 3 Interview-**

0:0:0.0 --> 0:0:1.240  
Participant 3  
No, you're grand grand.

0:0:22.820 --> 0:0:23.290  
Participant 3  
Yeah.

0:0:2.340 --> 0:0:23.370  
Michelle M. Moran  
Which goes I'm so I just have a few questions basically and for the for my project I have to have a kind of a recorded interview with with the participant, the expert and I have to transcribe the conversation as well. So I have to use it as well if that makes sense so.

0:0:28.250 --> 0:0:28.620  
Participant 3  
OK.

0:0:23.470 --> 0:0:39.720  
Michelle M. Moran  
Um, just a few quick questions. Basically, that's all about the process. So as to start off with the first thing was I'm just looking at production downtimes. So in that area where the tanks of the 25 empty 2223 and 26.

0:0:41.20 --> 0:0:54.490  
Michelle M. Moran  
Which you're working with. So I just wanted to say just a few questions basically on the production downtimes. So how often would you notice stops or interruptions during the, let's say an average production?

0:0:55.580 --> 0:0:56.870  
Michelle M. Moran  
Would would there be many?

0:1:1.790 --> 0:1:2.220  
Michelle M. Moran  
Yes.

0:0:58.130 --> 0:1:3.690  
Participant 3  
And it it is this based on the muscles batches now or is this based on all round?

0:1:2.820 --> 0:1:9.340  
Michelle M. Moran  
Just, just, no, no. Just amuse silage containing batches, ones that isn't that they're gum containing batches.

0:1:9.710 --> 0:1:10.0  
Participant 3  
Yeah.

0:1:11.330 --> 0:1:21.800  
Participant 3  
So I suppose a lot of our dying time will come from the mixing in of our gum, so we'll get a we'll get an allocated time for how long it takes for gum to mix in.

0:1:22.570 --> 0:1:32.60  
Participant 3  
And there was just times it goes way over that time because the gum isn't being pulled in properly because we don't have sufficient agitation. So that may take longer than normal.

0:1:32.510 --> 0:1:33.870  
Michelle M. Moran  
Yeah. OK.

0:1:33.230 --> 0:1:36.640  
Participant 3  
So we'll have to record extra downtime for that and explain.

0:1:37.370 --> 0:1:39.510  
Participant 3  
Why we went over the time allocated?

0:1:40.630 --> 0:1:41.720  
Participant 3  
Over the mixing time.

0:1:40.910 --> 0:1:49.440  
Michelle M. Moran  
OK. And so who who originally do you know who originally allocated that time? Where does that time come from that target time?

0:1:49.910 --> 0:1:51.710  
Participant 3  
Well, when they said up to you OE they.

0:1:52.590 --> 0:1:55.580  
Participant 3  
When they set up the OE process, they looked at the.

0:1:56.370 --> 0:2:3.380  
Participant 3  
But the 20 last batches that were made and they kind of picked a they kind of picked the best times out of that.

0:2:3.940 --> 0:2:7.960  
Michelle M. Moran  
All right, OK, OK. Ohh right. Um, yeah.

0:2:6.460 --> 0:2:9.670  
Participant 3  
So they obviously would have had a batch, maybe dash.

0:2:18.420 --> 0:2:19.140  
Michelle M. Moran  
Ohh right, OK.

0:2:10.500 --> 0:2:19.560  
Participant 3  
Had little gone managed and it didn't need big agitation time and they based it off that particular time then so that meant yeah.

0:2:19.740 --> 0:2:27.770  
Michelle M. Moran  
So the realistic really realistically like the target times that are there for that phases for the commentation phases is not realistic really.

0:2:28.620 --> 0:2:32.180  
Participant 3  
In certain in certain times, no, it's not. Yeah. Yeah.

0:2:29.920 --> 0:2:43.130  
Michelle M. Moran  
It's certain certain batches and would you would you know offhand, you would know if you seen on the schedule a batch coming in, you'd know, would it be dependent on the amount of gum that's going to be in that batch or the size of the batch?

0:2:43.900 --> 0:2:44.410  
Michelle M. Moran  
Or.

0:2:43.980 --> 0:2:56.860  
Participant 3  
It depend, it depends on the amount of gold miners and what tank you're actually building it in and which particular batch it is. There is a batch that we know we see it on our on our plan every week that all get. That's gonna take an extra four or five hours to mix in.

0:2:57.580 --> 0:2:58.560  
Michelle M. Moran  
Yeah.

0:2:57.720 --> 0:3:4.350  
Participant 3  
And as A and as other batches we know then if it's a particular tank, we know that tank has poor agitation, so it's gonna.

0:3:5.10 --> 0:3:8.400  
Participant 3  
We're gonna be me mixing for an extra two or three hours on this tank.

0:3:8.960 --> 0:3:9.630  
Michelle M. Moran  
Ohh okay.

0:3:17.160 --> 0:3:17.570  
Michelle M. Moran  
Yeah.

0:3:9.220 --> 0:3:20.380  
Participant 3  
And then there's other tanks were good agitation. If we build it in those tanks, we know we'll have very little downtime. So it all depends on what tanks are available to us and which batches they are and how much actually going is in the batch.

0:3:20.870 --> 0:3:34.330  
Michelle M. Moran  
OK, OK. OK. So obviously the higher the, my, the bigger the quantity going, the more issue along with the picture and do you get do you get to choose do you choose which tank you're going to make or is that scheduled or?

0:3:34.740 --> 0:3:41.750  
Participant 3  
No, we can, if we have tanks available, we'll pick our better tanks. Our tanks are better vegetation, but that's not always visible because we'll have.

0:3:42.950 --> 0:3:46.540  
Participant 3  
All thanks are in use during the week, so it's kind of hard to.

0:3:47.70 --> 0:3:47.810  
Michelle M. Moran  
Pick and choose.

0:3:50.430 --> 0:3:50.900  
Michelle M. Moran  
Okay.

0:3:47.440 --> 0:3:52.710  
Participant 3  
Take pick and choose the right tank you want. So a lot of times I thought of our controls so.

0:3:53.140 --> 0:3:58.470  
Michelle M. Moran  
Yeah. So have you. Do you know I've found it a particular material or anything that would be?

0:4:0.130 --> 0:4:0.950  
Michelle M. Moran  
Difficult.

0:4:1.820 --> 0:4:7.610  
Michelle M. Moran  
But you'd know would take a long time. That would take longer than it's a, you know, that's the target time.

0:4:8.560 --> 0:4:15.170  
Participant 3  
I don't know. I I can. I don't know the material number off hand, but I know the batch it's a it's a PF55.

0:4:16.100 --> 0:4:16.770  
Michelle M. Moran  
OK.

0:4:18.460 --> 0:4:18.830  
Michelle M. Moran  
And.

0:4:17.70 --> 0:4:19.490  
Participant 3  
And it's got a lot of.

0:4:20.740 --> 0:4:21.70  
Michelle M. Moran  
Go.

0:4:22.470 --> 0:4:23.40  
Michelle M. Moran  
Our.

0:4:20.440 --> 0:4:23.370  
Participant 3  
Colour in it as well. And when you put a collar.

0:4:24.210 --> 0:4:30.300  
Participant 3  
And when you put a colour into a tank as well, it's it's harder for the gum to mix in.

0:4:31.710 --> 0:4:32.260  
Michelle M. Moran  
OK.

0:4:33.40 --> 0:4:34.740  
Michelle M. Moran  
Um so.

0:4:35.140 --> 0:4:36.920  
Michelle M. Moran  
And let's say.

0:4:38.720 --> 0:4:42.290  
Michelle M. Moran  
The come the colour is in Dallas, like when is the colour advice.

0:4:43.400 --> 0:4:44.310  
Participant 3  
On the PF.

0:4:44.950 --> 0:4:46.590  
Michelle M. Moran  
For example, the PX55.

0:4:46.400 --> 0:4:49.550  
Participant 3  
Yeah. So it's added before your gum.

0:4:50.670 --> 0:4:51.300  
Michelle M. Moran  
Okay.

0:4:50.720 --> 0:4:51.350  
Participant 3  
Is put in.

0:4:52.300 --> 0:4:56.860  
Participant 3  
So your colour is added and then there is a small timer after that then.

0:4:57.430 --> 0:4:57.900  
Michelle M. Moran  
Yeah.

0:4:58.320 --> 0:4:59.550  
Participant 3  
And then you put in.

0:5:0.300 --> 0:5:0.710  
Michelle M. Moran  
The gum.

0:5:1.370 --> 0:5:2.640  
Participant 3  
You're going after that.

0:5:3.70 --> 0:5:5.250  
Michelle M. Moran  
Okay so yeah.

0:5:3.940 --> 0:5:6.120  
Participant 3  
And you'll get about and you'll get about.

0:5:7.20 --> 0:5:13.890  
Participant 3  
You get about 2000 kg in that will mix relatively OK, but it's still know of your gum.

0:5:11.570 --> 0:5:13.950  
Michelle M. Moran  
Of the colour are the gum.

0:5:14.810 --> 0:5:15.540  
Michelle M. Moran  
After that.

0:5:14.680 --> 0:5:16.840  
Participant 3  
But it's the last. It's the last.

0:5:20.300 --> 0:5:20.610  
Michelle M. Moran  
Yeah.

0:5:17.800 --> 0:5:25.200  
Participant 3  
Say say 1500 Kg's, we'll just we'll just lie on top of the the mix. It just very hard to pull it in.

0:5:24.930 --> 0:5:29.200  
Michelle M. Moran  
Her to make hard to member. I've seen that I seen as Jessie showed me. It's stuffed.

0:5:34.600 --> 0:5:36.130  
Participant 3  
It will make it more different yet.

0:5:38.50 --> 0:5:38.400  
Participant 3  
To.

0:5:30.460 --> 0:5:39.950  
Michelle M. Moran  
Alright, OK. And you're saying that colour address can make it more difficult for the the gum to move for to, to, to, to, to disperse or whatever?

0:5:39.30 --> 0:5:40.560  
Participant 3  
Yes, this person. The tank. Yeah.

0:5:40.960 --> 0:5:50.690  
Michelle M. Moran  
Okay and um so that just results in just a longer time. That's all that really doesn't. It just takes. It takes a bit longer if to leave it a bit longer to mix.

0:5:50.760 --> 0:5:54.610  
Participant 3  
It it takes a longer mixing time, but the problem you'll run into then is that.

0:6:9.800 --> 0:6:10.210  
Michelle M. Moran  
Yeah.

0:5:56.370 --> 0:6:12.60  
Participant 3  
Would say if we were making a batch on the smaller side, say during the week, and we have to wait an extra 5 hours for it to mix. We're pushing everything out five hours, so it's putting we could pressure on us on a Thursday then to try and get everything say completed.

0:6:12.890 --> 0:6:15.780  
Michelle M. Moran  
Yet these scheduled completes. Uhh, OK.

0:6:14.660 --> 0:6:23.340  
Participant 3  
Yeah. And that's and that's happened on a few occasions now we run into problems on Thursday as the moving patches around from different areas to try and get everything built and.

0:6:25.490 --> 0:6:26.170  
Participant 3  
It does.

0:6:24.680 --> 0:6:28.770  
Michelle M. Moran  
Yeah. So would you, what would you think would be a solution to this?

0:6:30.570 --> 0:6:37.780  
Participant 3  
Well, solution was supposed to. Basic solution is better is better. Agitation in the tanks. That's number one solution.

0:6:36.840 --> 0:6:42.720  
Michelle M. Moran  
Okay that. Yeah. Absolutely. Yeah. Yeah, that's a given. But let's say if.

0:6:46.110 --> 0:6:57.360  
Michelle M. Moran  
You know, when they when they schedule, it depends on the customer and what they want and then they schedule it and then they put you know. But what I'm saying is is if the targets were a bit more realistic.

0:6:58.0 --> 0:6:58.830  
Michelle M. Moran  
To the batch.

0:7:0.480 --> 0:7:2.100  
Michelle M. Moran  
Would that reflect on the schedule?

0:7:5.330 --> 0:7:6.30  
Participant 3  
Yeah.

0:7:3.700 --> 0:7:11.630  
Michelle M. Moran  
With people you know what I mean? If they like, say, would like when they schedule something, they would, they know particularly with material how long it should take.

0:7:12.910 --> 0:7:14.650  
Michelle M. Moran  
Does that make dinner? What I'm trying to say?

0:7:14.160 --> 0:7:20.520  
Participant 3  
Yeah, yeah. They're basing their schedule on the OE and what the OE along was in time wise.

0:7:31.890 --> 0:7:32.940  
Participant 3  
Yeah, yeah.

0:7:33.810 --> 0:7:34.670  
Participant 3  
It's lower yet.

0:7:21.20 --> 0:7:37.670  
Michelle M. Moran  
OK, so let's say your PF55 go runs over so that it gives you a higher a lower OE is it, is it lower OE operational, you know it's it's lower. Yeah. Sorry, it's it's it's lower. So would P would the PF?

0:7:38.670 --> 0:7:40.330  
Michelle M. Moran  
Give you a lower oil all the time.

0:7:41.740 --> 0:7:44.410  
Participant 3  
Ohh always yeah, the PF is always the troublesome batch.

0:7:46.80 --> 0:7:46.230  
Participant 3  
Yeah.

0:7:44.570 --> 0:7:57.60  
Michelle M. Moran  
Through this one batch. So therefore let's say the and the the that that overrun phase overrun or whatever, that would indicate that would reflect on the OE on the lower OE.

0:7:57.460 --> 0:7:57.930  
Participant 3  
Yes.

0:8:1.340 --> 0:8:1.830  
Participant 3  
Yes.

0:7:57.750 --> 0:8:13.740  
Michelle M. Moran  
And that overrun is reflective of the target that's set. So OK, so that's if the target was updated to reflect what really happens with the PF to allow the manual edition of all those tank, all those gums, therefore it wouldn't affect the OE.

0:8:14.650 --> 0:8:14.930  
Michelle M. Moran  
He.

0:8:14.250 --> 0:8:15.740  
Participant 3  
It wouldn't affect your weed then, no.

0:8:16.80 --> 0:8:23.0  
Michelle M. Moran  
No, but I mean if it if it starts, it's all the time affecting the we would they not just look at the schedule and and or try and.

0:8:24.490 --> 0:8:26.620  
Michelle M. Moran  
Do you know when trying to say I mean?

0:8:26.410 --> 0:8:40.940  
Participant 3  
I know what you're trying to say and and now what they have done with that patch is they've moved out to a towards the evening. Well, I towards the afternoon. So if it has to overrun now you're saying we can leave a mixing overnight and they turn off the edge later on the Friday.

0:8:41.380 --> 0:8:41.850  
Michelle M. Moran  
Yeah.

0:8:42.300 --> 0:8:43.510  
Participant 3  
So that was one.

0:8:53.870 --> 0:8:54.630  
Participant 3  
It's that mix up.

0:9:3.200 --> 0:9:3.950  
Participant 3  
Yeah, yeah.

0:9:8.710 --> 0:9:9.220  
Participant 3  
It is.

0:8:44.640 --> 0:9:9.500  
Michelle M. Moran  
Changed the day of production to Thursday so you know for them for that particular batch then. So it's just left left mix on the Friday. So it could it and it would love mix until it somebody on the weekend shift would check it and then if it's mixed they'd letter turn it off and let it settle is that it so then would that so that's still increasing the OE time for that batch or lowering the.

0:9:9.910 --> 0:9:14.750  
Participant 3  
Ohh what is? Yeah, because it's it's gone way over this. It's gone hours and hours over the phase then so.

0:9:15.350 --> 0:9:15.910  
Michelle M. Moran  
OK.

0:9:17.320 --> 0:9:21.920  
Michelle M. Moran  
So like, I don't think I like unless like it's bad. Like unless you get a better agitation.

0:9:23.160 --> 0:9:39.450  
Michelle M. Moran  
And it's it's it that's the only really option that you can with this particular batch cause you still have to add in all that gum still has to be mixed. It still have to be left there. So, but what I'm I'm I'm just kind of trying to figure out if that's the only way we could he could solve that.

0:9:38.830 --> 0:9:47.400  
Participant 3  
No, I think another option that they could do is that if we were to make a highly concentrated batch of gum.

0:9:58.860 --> 0:9:59.340  
Michelle M. Moran  
And then.

0:9:59.440 --> 0:10:2.150  
Participant 3  
I we could use the IBC on the day of production then.

0:10:2.980 --> 0:10:5.170  
Michelle M. Moran  
Ohh yeah, pretty mixers premixes.

0:10:3.470 --> 0:10:6.480  
Participant 3  
Because you're you're premixes. And if you get a.

0:10:7.680 --> 0:10:15.10  
Participant 3  
A bigger quantity, a highly high, highly concentrated one that you could have maybe 10 bags of gum.

0:10:16.900 --> 0:10:18.790  
Participant 3  
And and mix it all in.

0:10:19.520 --> 0:10:26.110  
Participant 3  
Once a month or something like that, so that on the day of production then we we're only putting in BC rather than powdered gum.

0:10:26.690 --> 0:10:27.300  
Michelle M. Moran  
Ohh OK.

0:10:27.90 --> 0:10:29.410  
Participant 3  
That's that would be another alternative.

0:10:28.90 --> 0:10:32.0  
Michelle M. Moran  
That's that's that would be. Absolutely. Yeah. That makes sense. Alright.

0:10:32.460 --> 0:10:43.680  
Michelle M. Moran  
And is there, would you said different times different vegetation rates? So would you know offhand which tanks will, let's say for example the between the 25 M to one and Mt 02?

0:10:44.370 --> 0:10:45.290  
Michelle M. Moran  
Which would be better?

0:10:47.900 --> 0:10:51.530  
Participant 3  
Them ones aren't too bad the yeah.

0:10:49.980 --> 0:10:53.770  
Michelle M. Moran  
They're OK. They're they're only small, small batches, 4 tonne ones, isn't it?

0:10:52.930 --> 0:10:55.700  
Participant 3  
Yeah. Yes, that's all. So they're not too bad.

0:10:55.30 --> 0:10:58.850  
Michelle M. Moran  
So they're all they're pretty, OK, the 25, the old tree and the 04.

0:10:59.470 --> 0:11:0.290  
Participant 3  
Or three.

0:11:1.30 --> 0:11:2.800  
Participant 3  
Two or three is a bad tank.

0:11:4.770 --> 0:11:5.170  
Participant 3  
Yeah.

0:11:3.330 --> 0:11:6.270  
Michelle M. Moran  
That's a bad tank, so that is in slower agitation.

0:11:7.170 --> 0:11:9.880  
Participant 3  
Yeah, the agitation isn't as strong in that tank.

0:11:10.770 --> 0:11:11.140  
Michelle M. Moran  
OK.

0:11:10.780 --> 0:11:21.380  
Participant 3  
And and even though pills only a small batch, it's a like it's an 8, it's a 9 tonne batch, but you've got 1800 kg SA gum and that and that's.

0:11:22.690 --> 0:11:23.990  
Participant 3  
And that just adds.

0:11:26.70 --> 0:11:27.50  
Participant 3  
That had such.

0:11:31.70 --> 0:11:31.500  
Participant 3  
Yeah.

0:11:32.500 --> 0:11:32.710  
Participant 3  
Yeah.

0:11:22.910 --> 0:11:41.300  
Michelle M. Moran  
Yeah, that's that. That adds to it as well. In fairness. So, so like he has a smaller the tank, the more gum that's in the worse it is, is that it? Yeah. Okay. So then for the 22 Mt tanks, there are 20, the 22 and the 2320, they're 2020 times, is it 20?

0:11:40.660 --> 0:11:43.890  
Participant 3  
There. Yeah, there. Yeah. There are 20 tonne tanks, so.

0:11:43.200 --> 0:11:43.960  
Michelle M. Moran  
20 tanks.

0:11:44.580 --> 0:11:48.80  
Participant 3  
The best tanks to build there would be 01 and 05:00.

0:11:48.820 --> 0:11:50.750  
Michelle M. Moran  
Ohh, the of the 22.

0:11:51.90 --> 0:11:56.240  
Participant 3  
Yeah. Ohh. Wanna know five and ohh 105 and.

0:11:59.520 --> 0:12:0.480  
Participant 3  
1/2.

0:12:0.920 --> 0:12:1.430  
Michelle M. Moran  
He.

0:12:2.800 --> 0:12:3.630  
Participant 3  
And our tree.

0:12:5.560 --> 0:12:7.290  
Participant 3  
No, of the and.

0:12:8.390 --> 0:12:8.670  
Participant 3  
Yeah.

0:12:9.290 --> 0:12:11.340  
Participant 3  
No, they are good. Thanks. Yeah.

0:12:4.160 --> 0:12:13.20  
Michelle M. Moran  
Of the 23, no. Of all three as well. They're bad. They're they're, they're good. And then for the 23 tanks then?

0:12:13.530 --> 0:12:17.50  
Participant 3  
It's all 1:00 and 05:00 or two best tanks and that system.

0:12:17.530 --> 0:12:18.50  
Michelle M. Moran  
Ohh okay.

0:12:18.800 --> 0:12:22.410  
Michelle M. Moran  
And so that like and then. OK then the 26.

0:12:29.830 --> 0:12:31.80  
Participant 3  
Yeah. So.

0:12:23.470 --> 0:12:31.470  
Michelle M. Moran  
Um it 20? I'm lonely looking at the 26 old Mt 01. They're the only ones that have batches for that I'm looking at, yeah.

0:12:33.180 --> 0:12:33.550  
Participant 3  
It.

0:12:34.290 --> 0:12:38.60  
Participant 3  
It's not too bad on that because it's not. It's a small amount of gums, so you can.

0:12:43.650 --> 0:12:43.970  
Michelle M. Moran  
It.

0:12:39.60 --> 0:12:45.450  
Participant 3  
It's going in a small bags as well, so you can kind of you can control how it's you allow it to mix in as you're putting it in, so.

0:12:45.770 --> 0:12:46.230  
Michelle M. Moran  
Yeah.

0:12:47.420 --> 0:12:47.930  
Participant 3  
Push.

0:12:47.740 --> 0:12:52.150  
Michelle M. Moran  
OK. Can I ask you then just sorry, you know, thank you.

0:12:53.20 --> 0:13:0.300  
Michelle M. Moran  
So have an example of the the the different phases, let's say for a particular batch and I was just asking Ally so basic.

0:13:0.580 --> 0:13:8.280  
Michelle M. Moran  
And so you have your start to the process, then your step one cons there. The addition of all the ingredients.

0:13:8.730 --> 0:13:9.40  
Participant 3  
Yep.

0:13:9.260 --> 0:13:16.610  
Michelle M. Moran  
That's that prompt is that that's the prompt. And then you have, um, the agitation step 1-2 and three.

0:13:17.820 --> 0:13:18.210  
Michelle M. Moran  
Yeah.

0:13:18.940 --> 0:13:26.0  
Michelle M. Moran  
Um. And then between that last agitation and the HP, that's when you just switch everything off, is that it?

0:13:26.270 --> 0:13:28.800  
Participant 3  
Yes, once you've added all your gourmet in.

0:13:32.630 --> 0:13:32.990  
Michelle M. Moran  
Yeah.

0:13:29.620 --> 0:13:35.890  
Participant 3  
And and you've answered your prompt on your gum. It'll ask you, you get a prompt. Insure gum was mixed in.

0:13:36.180 --> 0:13:36.940  
Michelle M. Moran  
Yes. Yeah.

0:13:37.20 --> 0:13:40.350  
Participant 3  
And once you answer that, then it.

0:13:41.20 --> 0:13:43.290  
Participant 3  
It goes to a water and then after that.

0:13:43.570 --> 0:13:43.940  
Michelle M. Moran  
Yeah.

0:13:44.40 --> 0:13:46.540  
Participant 3  
And then once you answer your water at prompt.

0:13:48.290 --> 0:13:50.340  
Participant 3  
There's a a one hour timer.

0:13:50.950 --> 0:13:51.320  
Michelle M. Moran  
Yes.

0:13:51.290 --> 0:13:55.580  
Participant 3  
I want I want start timer is up then you're always starts running from there.

0:13:57.460 --> 0:14:1.450  
Michelle M. Moran  
And so water arm. So does one hour after the water is added.

0:14:2.420 --> 0:14:2.860  
Michelle M. Moran  
Okay.

0:14:1.790 --> 0:14:6.460  
Participant 3  
Yeah. And then you turn off your and then you turn off your agitator and let it deteriorate, then after that.

0:14:6.910 --> 0:14:10.110  
Michelle M. Moran  
OK, turn off agitator into your ears.

0:14:11.180 --> 0:14:11.790  
Michelle M. Moran  
And.

0:14:11.220 --> 0:14:22.210  
Participant 3  
But what's happening? But what's happening with us is we can't answer that prompt on the insure gum is mixed in because it's still not mixed in, so it can run for hours.

0:14:22.680 --> 0:14:24.870  
Michelle M. Moran  
Yeah. Yeah. OK.

0:14:25.220 --> 0:14:36.410  
Michelle M. Moran  
Um, that's the time. That's the catch. So and for the OE metrics each each phase addition, each phase is measured?

0:14:37.120 --> 0:14:37.560  
Participant 3  
Yes.

0:14:40.780 --> 0:14:42.190  
Participant 3  
No. Every.

0:14:43.290 --> 0:14:43.520  
Participant 3  
Yeah.

0:14:44.790 --> 0:14:45.80  
Participant 3  
Yeah.

0:14:38.120 --> 0:14:47.720  
Michelle M. Moran  
It's not just the start of the batch and the end of the match, it's each individual fit. Everything is accumulated or OK or it's very good.

0:14:48.780 --> 0:14:51.710  
Michelle M. Moran  
OK, let's see what else is there.

0:14:53.560 --> 0:15:4.390  
Michelle M. Moran  
Yeah, that's that one. That's perfect. How do you record or note down any of the issues, is that um on on online or is that a work written or?

0:15:4.780 --> 0:15:8.230  
Participant 3  
No, it's it's all online. So we are we do it all.

0:15:9.700 --> 0:15:13.830  
Participant 3  
Online on the computer's upstairs, that's all collected in the data then so.

0:15:14.260 --> 0:15:14.420  
Michelle M. Moran  
So.

0:15:14.490 --> 0:15:18.760  
Michelle M. Moran  
Ohh, as you were going through the phases, if you run into problems you'll update it as you go along.

0:15:19.490 --> 0:15:22.880  
Participant 3  
Yeah, we tried to update that as we go along. Ready. Yeah.

0:15:21.520 --> 0:15:23.510  
Michelle M. Moran  
As you can as you can, yeah.

0:15:24.710 --> 0:15:25.520  
Michelle M. Moran  
As ebola.

0:15:26.990 --> 0:15:28.410  
Michelle M. Moran  
Hello. OK.

0:15:29.520 --> 0:15:34.620  
Michelle M. Moran  
So, are there certain indicators of signals that alert you before potential downtime?

0:15:38.520 --> 0:15:39.100  
Participant 3  
Ohh.

0:15:42.340 --> 0:15:43.950  
Participant 3  
Say that again. Now is there.

0:15:44.280 --> 0:15:50.100  
Michelle M. Moran  
Yeah. Sorry. Is there any certain indicators or signals that alerts you that there could be a potential downtime?

0:15:51.840 --> 0:15:54.240  
Participant 3  
Just your commendation. You'll always, yeah.

0:15:53.280 --> 0:15:56.40  
Michelle M. Moran  
Just the gym is the gum is the major problem.

0:15:56.110 --> 0:15:58.230  
Participant 3  
That's the major problem with these batches yet.

0:15:57.460 --> 0:15:58.960  
Michelle M. Moran  
Yes, the major problem, yeah.

0:16:0.30 --> 0:16:0.620  
Michelle M. Moran  
Um.

0:16:1.700 --> 0:16:11.190  
Michelle M. Moran  
And like I've we talked about this, he would adjust your workflow you you know, I mean you you you adjust it so if you know with particular batch has more gum in it and there all that so you've said that already to me.

0:16:11.850 --> 0:16:12.520  
Michelle M. Moran  
Um.

0:16:13.530 --> 0:16:26.630  
Michelle M. Moran  
What tools and resources would help you preemptively address it? But you've already said that, so if the potential of premixing it the golem and then it adding it to the batch, have you said that to people or has that ago or anybody?

0:16:26.960 --> 0:16:32.770  
Participant 3  
No, we it has been brought up but hasn't been brought to the forefront yet.

0:16:33.270 --> 0:16:33.690  
Michelle M. Moran  
Yeah.

0:16:35.150 --> 0:16:38.310  
Michelle M. Moran  
That's how it's a pretty good idea, though. You know, kind of.

0:16:37.700 --> 0:16:39.760  
Participant 3  
I think it would be a good idea because it would.

0:16:46.0 --> 0:16:46.560  
Michelle M. Moran  
Yeah.

0:16:41.970 --> 0:16:46.860  
Participant 3  
Suppose it adds a bit more cost to it because they're using BC and maybe I don't know.

0:16:47.660 --> 0:17:0.120  
Michelle M. Moran  
You'd have to weigh it, weigh it up, like quite good. So you know, for the some some batches require additional essay flavour additions or oil editions.

0:17:1.130 --> 0:17:1.730  
Participant 3  
Yes.

0:17:1.180 --> 0:17:3.680  
Michelle M. Moran  
I got seems to be after.

0:17:4.980 --> 0:17:6.550  
Michelle M. Moran  
HP is it or?

0:17:6.390 --> 0:17:7.710  
Participant 3  
No, that's.

0:17:7.310 --> 0:17:8.960  
Michelle M. Moran  
Where's that when when's happy?

0:17:9.490 --> 0:17:16.940  
Participant 3  
That's after your after your batch is derated and you're densities are raining, you've taken your tested and you're density is in in.

0:17:17.640 --> 0:17:18.190  
Participant 3  
Spec.

0:17:18.660 --> 0:17:19.50  
Michelle M. Moran  
Yeah.

0:17:19.460 --> 0:17:21.950  
Participant 3  
10 year olds are added to your gum then.

0:17:22.960 --> 0:17:26.780  
Michelle M. Moran  
Ohh it's OK and that I do they pose any problems or anything?

0:17:27.150 --> 0:17:28.920  
Participant 3  
No, they don't pose any problems.

0:17:28.570 --> 0:17:31.720  
Michelle M. Moran  
At that stage everything is good, you know, so it's just.

0:17:35.270 --> 0:17:35.630  
Michelle M. Moran  
Yeah.

0:17:30.700 --> 0:17:38.900  
Participant 3  
At that stage everything is good. Yeah, it's after the rating. Everything is fine. You add your oils and there's a 30 minute timer after you add your oils.

0:17:39.970 --> 0:17:44.810  
Participant 3  
For the allowed them to mix in properly and then and then you do the homogenization.

0:17:45.930 --> 0:17:46.410  
Michelle M. Moran  
Ohh.

0:17:45.710 --> 0:17:46.920  
Participant 3  
Process after dash.

0:17:47.790 --> 0:17:50.360  
Michelle M. Moran  
Ohh it's OK and.

0:17:51.620 --> 0:17:52.950  
Michelle M. Moran  
Was going to say to you.

0:18:1.960 --> 0:18:2.470  
Participant 3  
No.

0:17:53.850 --> 0:18:5.120  
Michelle M. Moran  
You know how you said certain certain times you you don't you don't know why certain tanks are, you know, the agitations are different. There's no like they were all. They're all seem to be the same make same.

0:18:4.690 --> 0:18:10.30  
Participant 3  
They're all. Yeah, they're all. They're all the same. Some tanks just seem to have a better Poland than than others.

0:18:10.570 --> 0:18:11.440  
Michelle M. Moran  
Yeah.

0:18:10.900 --> 0:18:13.90  
Participant 3  
Digitation it just we can't.

0:18:12.670 --> 0:18:14.100  
Michelle M. Moran  
It's just seems to be, yeah.

0:18:14.930 --> 0:18:15.500  
Michelle M. Moran  
Um.

0:18:15.110 --> 0:18:16.620  
Participant 3  
I suppose you could.

0:18:17.350 --> 0:18:18.960  
Participant 3  
You could possibly say that.

0:18:20.450 --> 0:18:26.210  
Participant 3  
That the good tanks are the ones that we usually process into to kind of process tanks to destination tanks.

0:18:27.90 --> 0:18:29.540  
Participant 3  
They're not majority. The time building tanks.

0:18:33.820 --> 0:18:35.190  
Participant 3  
Yeah, yeah.

0:18:30.50 --> 0:18:36.940  
Michelle M. Moran  
Okay so the good tanks are the destination tanks that I see here. Uh, yes. Select destination tank.

0:18:37.380 --> 0:18:37.720  
Participant 3  
Yeah.

0:18:37.950 --> 0:18:38.420  
Michelle M. Moran  
Yeah.

0:18:39.330 --> 0:18:42.980  
Participant 3  
So so normally they don't do lot of vegetation.

0:18:43.820 --> 0:18:45.290  
Participant 3  
There's not a lot of product being.

0:18:45.960 --> 0:18:46.900  
Michelle M. Moran  
Introduced.

0:18:45.910 --> 0:18:49.120  
Participant 3  
Being poured into them. Yeah, everyday to them. So unless.

0:18:48.950 --> 0:18:49.670  
Michelle M. Moran  
Ours.

0:18:49.990 --> 0:18:52.520  
Participant 3  
That probably has a big bearing on that.

0:18:53.270 --> 0:18:59.580  
Michelle M. Moran  
You'll probably okay no problem in HN, no problem with it. There's no need for agitation for them.

0:19:0.10 --> 0:19:0.400  
Participant 3  
Yeah.

0:19:0.860 --> 0:19:1.220  
Michelle M. Moran  
Yeah.

0:19:2.550 --> 0:19:3.320  
Michelle M. Moran  
OK.

0:19:4.350 --> 0:19:4.970  
Michelle M. Moran  
And.

0:19:5.880 --> 0:19:12.370  
Michelle M. Moran  
That's it, Thomas. I think that's all. I thank you so, so much for your time. I I know.

0:19:11.580 --> 0:19:13.30  
Participant 3  
Ohh no, it's fine, it's grand.

0:19:13.60 --> 0:19:21.770  
Michelle M. Moran  
No, no, no. I do appreciate it. And it's it's just the process I have to have it. I have to have evidence that I talk to you, you know.

0:19:21.690 --> 0:19:26.60  
Participant 3  
OK. So they'll have you think too this interview when you're thesis are?

0:19:26.350 --> 0:19:26.690  
Michelle M. Moran  
Yeah.

0:19:27.20 --> 0:19:27.460  
Participant 3  
OK.

0:19:42.100 --> 0:19:42.840  
Participant 3  
Ohh OK.

0:19:52.140 --> 0:19:52.730  
Participant 3  
Yeah.

0:19:27.930 --> 0:19:58.540  
Michelle M. Moran  
Yeah. And I've been doing figures here and that on its 30 years of batch from muslish containing batches. And So what you have said, what Ollie has said corresponds to the results that I'm getting. So I know Joe, as in you've I just asked you which tanks are issues. So I can see straight away I knew that PF55 was one of them. One of them issue ones because it has the higher your gum. So that's it's just it's just kind of tying up what I'm seeing from the data.

0:19:58.960 --> 0:20:1.690  
Michelle M. Moran  
And what you're seeing on the floor and experiencing?

0:20:1.750 --> 0:20:4.840  
Participant 3  
OK, yes, that there it's yeah.

0:20:9.210 --> 0:20:9.760  
Participant 3  
Yes.

0:20:3.550 --> 0:20:13.370  
Michelle M. Moran  
That there you know, you, you you're physically seeing the problem, whatever. But from the data I pull it you can see it but you see you already see that through the OE or the OE figure.

0:20:13.730 --> 0:20:14.240  
Participant 3  
Yes.

0:20:30.850 --> 0:20:31.800  
Participant 3  
OKOK.

0:20:34.470 --> 0:20:35.50  
Participant 3  
Brilliant.

0:20:40.560 --> 0:20:42.780  
Participant 3  
Yeah. Ollie. Yeah. Ollie. Yeah, we'll.

0:20:43.910 --> 0:20:44.230  
Participant 3  
Yeah.

0:20:14.520 --> 0:20:44.670  
Michelle M. Moran  
So this is the dot. This is one way of doing it that can connect to the wee. But in in work and they use a different software, different project, different things. So but it's it's just tying in what you're saying, what is actually happening. You can see it from the the the numbers you know so and that's that's why I'm using you you know I know I don't know about that. So what you got Olly has the consent form, so it's just basically to protect your rights to make sure I'm not.

0:20:45.100 --> 0:20:49.400  
Michelle M. Moran  
Gdpr and all of that kind of thing. But just to be sure you're good and.

0:20:47.210 --> 0:20:50.180  
Participant 3  
Ohh yeah, yeah, we get them, we get them back to you.

0:20:57.950 --> 0:20:58.390  
Participant 3  
OK.

0:21:1.230 --> 0:21:1.870  
Participant 3  
No problem.

0:20:50.530 --> 0:21:3.850  
Michelle M. Moran  
Yeah, there's no problem. Um Ali, our nationally just come from so tired and confused. But come here. Thomas. Thank you so much. I'll see you when I get back to work anyways. Which would be next week. But I will definitely catch up with you then in person. But I totally, totally appreciate your time.

0:21:3.770 --> 0:21:7.0  
Participant 3  
Ohh God, you're very welcome. No problem. OK then.

0:21:5.800 --> 0:21:10.270  
Michelle M. Moran  
I don't. Alright. Thank you so much. Take care. Bye bye. Bye. Thank you.

0:21:8.480 --> 0:21:10.340  
Participant 3  
Right, right, right, bye, bye.

# Appendix C: Data Permissions

# Appendix D: Consent Forms

# Reference List