International Journal of Engineering Works
https://www.ijew.io/
ISSN-p: 2521-2419
ISSN-e: 2409-2770
Vol. 11, Issue 09, PP. 177-187 September 2024
https://doi.org/10.34259/ijew.24.1109177187

Enhancing Quality and Reducing Wastage in ATN Soap Industry Peshawar Using Six Sigma and Artificial Intelligence (AI)

Syed Muhammad Shakir Bukhari[1], Rehman Akhtar[2]
[1,2]Industrial Engineering Department, University of Engineering and Technology, Peshawar, Pakistan
smshakirbukhari@gmail.com[1], rehman_akhtar@uetpeshawar.edu.pk[2]
Received: 01 August, Revised: 29 August, Accepted: 11 September

Abstract— Peshawar's ATN Soap Industry is confronted with issues pertaining to substandard quality and ineffective waste management, which influence overall operational performance. This study suggests a thorough strategy that combines artificial intelligence (AI) and Six Sigma methodology to address these problems. AI technology will be used to improve predictive maintenance, maximize resource usage, and eliminate defects; Six Sigma techniques will be used to discover and reduce variances in the soap production process. The research will begin with a thorough examination of the way things are now run, highlighting major issues with waste management and quality. The soap production process will be systematically analyzed and improved through the application of Six Sigma techniques, such as the DMAIC (Define, Measure, Analyze, Improve, and Control) methodology. To do this, quantifiable targets must be defined, pertinent data must be gathered, and focused changes must be put into place to get rid of flaws and increase the overall quality of the product. In the context of the ATN Soap Industry, the research seeks to illustrate the synergistic benefits of combining Six Sigma and AI. Significant gains in product quality, a decrease in defects, increased operational effectiveness, and a sustainable waste management strategy are among the anticipated results. This all-encompassing strategy can help improve industrial procedures in the area by acting as a model for other businesses dealing with comparable issues.

Keywords— Artificial Intelligence (AI), Defect Reduction, DMAIC Methodology, Process optimization, , Process variability, Product quality, Process Improvement, Quality defects, Six Sigma, Waste control.

I. INTRODUCTION

1. Background and Motivation

The worldwide soap market is a fiercely competitive one that requires constant process development to guarantee premium goods and effective use of resources. The competitiveness and sustainability of the ATN Soap Industry in Peshawar, Pakistan, are negatively impacted by issues with poor waste control and quality faults. By putting forth a synergistic strategy that blends Artificial Intelligence (AI) technologies with Six Sigma methodology, this research seeks to address these issues. Six Sigma is a data-driven methodology that is commonly used in many different industries to find and fix errors, minimize variances in processes, and boost overall productivity. Its methodical technique, dubbed DMAIC (Define, Measure, Analyze, Improve, Control), offers a structured framework for process improvement and problem-solving. The ATN Soap Industry can improve its quality control procedures and lower production inefficiencies by incorporating Six Sigma methods.

Conversely, artificial intelligence offers enterprises a game-changing chance to transition to smart manufacturing. Artificial intelligence (AI) technologies, such as machine learning and predictive analytics, provide sophisticated instruments for process optimization, predictive maintenance, and real-time monitoring. Predictive maintenance minimizes downtime, proactive defect identification is made possible, and resource optimization promotes sustainable practices when AI is integrated into the production process.

The first step in this research will be a thorough examination of the ATN Soap Industry's present operational status, with a focus on pinpointing problem areas and waste production. The systematic improvement of the soap manufacturing process will be guided by the application of Six Sigma methodology, with an emphasis on key performance indicators linked to waste reduction and quality.

AI technologies will be used in tandem to establish a smart manufacturing environment. Machine learning algorithms will use past data analysis to forecast possible flaws, and real-time monitoring will make sure that any deviations from the norm are quickly found and fixed. Models for predictive maintenance will be created to maximize equipment efficiency and reduce waste and downtime. It is anticipated that the ATN Soap Industry will see notable gains in product quality, a decrease in defects, and increased operational efficiency because of the effective integration of Six Sigma and AI. The research's comprehensive strategy fits well with current global manufacturing trends and lays the groundwork for Peshawar's soap industry to be both competitive and sustainable.

Authors retain all © copyrights 2024 IJEW. This is an open access article distributed under the CC-BY Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The production of soap is a vital part of the consumer goods sector, giving people all around the world access to basic hygiene products. Operating in a competitive market, the ATN Soap Industry in Peshawar, Pakistan, places a premium on reliable product quality and effective waste management. However, due to intrinsic process variability and the dynamic nature of production environments, it confronts difficulties, like many industries, in sustaining these standards. When soap is produced with poor quality, not only do customers become dissatisfied but operating expenses also rise since waste disposal and rework are required. Inadequate management of manufacturing wastes simultaneously threatens the long-term viability of the sector and adds to environmental issues. The creative and all-encompassing solution to these problems lies in the fusion of artificial intelligence (AI) with Six Sigma methodology. Artificial Intelligence (AI) technologies provide real-time monitoring, predictive analytics, and smart manufacturing capabilities, which help to create more sustainable and efficient production processes. Six Sigma's structured methodology makes it possible to identify and eliminate faults.

II. Literature Review

Bill Smith, an engineer at Motorola, invented the Sig Sigma methodology in the 1908s [1], which aims at figuring out and reducing the major causes of defects with the production process [2]. This approach is famous for its capacity of improving the overall operational and financial performance and achieving the customer's satisfaction [3]. DMAIC, a Six Sigma methodology, can be used to control the process variation that also causes product nonconformities [4]. Most people mistake six sigma with manufacturing, although this methodology may be used in any business and for any kind of process [5]. A common use of Six Sigma is the establishment of a management system that identifies errors and offers solutions for their removal. The goal of this methodology is to minimize process variance [6], as it provides an effective methodology when combined with Lean [7]. Moreover, it also helps to address the major causes of CTQS [8]. This has also transformed from metric to methodology, and eventually into a managerial strategy [9]. The DMAIC Six sigma methodology consists of five steps [10]. In first stage, problem is defined. In second stage the extent of problem is measured. The third stage of DMAIC methodology is used to analyze the data [11]. The fourth stage is to improve the process, while in fifth stage the new process is controlled [12]. One popular strategy in operations and business management is lean manufacturing. Since the Lean concept lowers costs by raising production rates and lowering waste percentages to a minimum, it was developed in Japan when the country's manufacturing industry could not afford to make significant expenditures in rebuilding its industries [13], because it was originated in order to maximize the utilization of resources by minimizing waste [14]. This methodology aimed at reducing cost throughout manufacturing process whether it's designing phase, fabricating phase, or finishing using reviews of previous business [15]. Depending on the intended size of the company, further techniques can also be employed to lower losses and increase production. Among these is kaizen, a lean technique that lowers overall production costs by incorporating every worker in the process of improvement and increases profit by minimizing losses [16]. This instrument is utilized in multiple domains, encompassing big, Small, and Medium-Sized Businesses (SMEs). Lean has several advantages, such as shorter lead times, lower operational costs, and higher quality. Economist's measure and monitor the flow of raw materials to finished goods, their transportation, customer delivery, cost holding, and ordering to cut out non-value-added operations and increase profitability while cutting costs [17]. Two methodologies such as Lean and Six Sigma, when combined, developed an effective approach know as Lean Sis Sigma [18]. Lean Six Sigma focuses on the advantages of both Lean Manufacturing and Lean Six Sigma, as lean is not the best option for problems requiring complex data collection and analysis. The integration of these methodologies aimed to overcome the limitation of implementing individually to have a complementary relationship [19].

It has been observed that applying the Lean Six Sigma approach as a management strategy can give businesses a significant competitive advantage in helping them compete with rivals and thrive in the global market. By using this process as a quality assurance tool, they are able to boost their bottom line [20]. This study aims to demonstrate how the Lean Six Sigma technique has been successfully applied in the garment business. Therefore, we investigated how the Lean Six Sigma approach helps identify the main flaws and their causes and use Kaizen to enhance the sewing process, which lowers the percentage of defects and increases profitability, with the use of literature [21]. Like many other manufacturing sectors, the ATN Soap Industry in Peshawar, Pakistan, struggles to maintain high standards of quality and manage waste effectively. In addition to lowering consumer satisfaction, poor quality in soap manufacture results in higher expenses for waste disposal and rework. In addition, insufficient waste management exacerbates environmental issues and jeopardizes the viability of the sector. To address these issues, an innovative and technologically advanced approach is being pursued to improve the caliber of soap manufacturing and streamline waste management [22]. In the industrial sector, artificial intelligence (AI) is emerging as a disruptive force with cutting-edge tools and methods that have the potential to completely overhaul established production procedures. The ATN Soap Industry has a promising opportunity to eliminate quality faults and improve waste control through the incorporation of AI technologies [23]. AI facilitates real-time monitoring of key parameters in the soap manufacturing process. By deploying sensors and data analytics, deviations from the standard operating conditions can be promptly identified [24]. This proactive approach allows for immediate corrective actions, reducing the likelihood of defects and ensuring that the production process operates within optimal parameters.

III. Contribution

The product at the ATN Soap Industry is made from raw materials using a variety of machine techniques. Waste in the production process is the main issue the sector is currently experiencing. Raw materials are wasted during the process while the rejection rate is high. It causes the company's profit to decline. Our goal is to reduce the amount of soap waste and rejected that has quality flaws.

IV. Research Methodology

The research methodology employed to examine and simulate the relationship between pressure control and waste generation in the soap manufacturing process. The methodology integrates the DMAIC framework with exploratory data analysis (EDA) and employs machine learning (ML) techniques, specifically the decision tree algorithm.

1. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is crucial for understanding the data patterns and relationships in the soap production process.

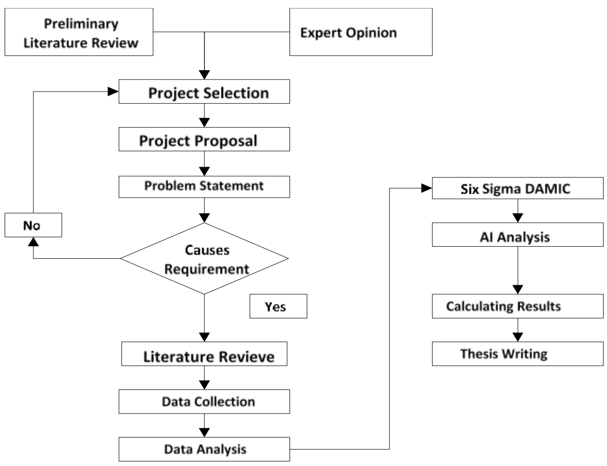

Figure 1. Research Methodology Flow Chart

2. Dmaic on Soup Industry Data

The abbreviation for Define, Measure, Analyze, Improve, and Control is DMAIC. It is an organized technique for enhancing organizational processes. This technique is an organized, data-driven framework intended to enhance, optimize, and stabilize company concepts and processes. Based on Six Sigma concepts, DMAIC serves as the cornerstone of initiatives aimed at improving processes. There are five stages in the process: problem definition, measurement of relevant metrics, data analysis to identify areas for improvement, change implementation, and control mechanism setup to ensure progress. Despite being frequently associated with Six Sigma, DMAIC is flexible and may be used with a variety of process improvement methodologies, including lean concepts. DMAIC improves performance and predictability by identifying and fixing inefficiencies in organizational processes via the use of empirical data and methodical problem-solving techniques shown in fig.1, 2, 3 and 4.

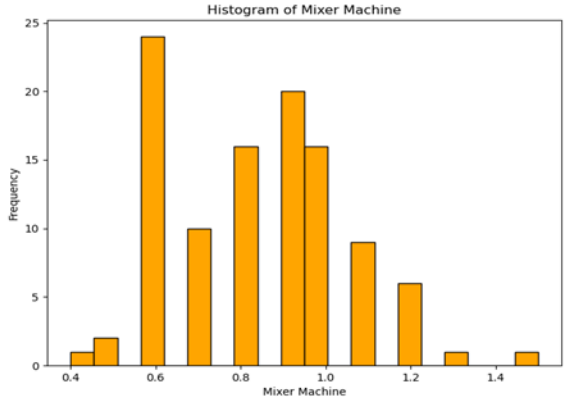

Figure 1. Histogram of Mixer Machine

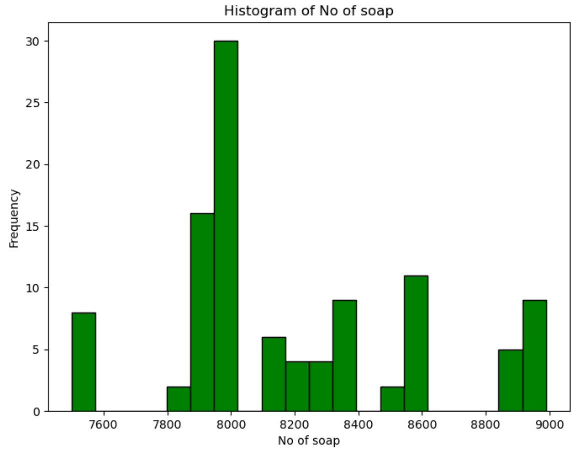

Figure 2. Histogram of No of Soup

Figure 3. Histogram of Plodder Machine

Pronounced "duh-may-ik," the DMAIC technique is based on the PDSA cycle, which statistician Walter A. Shewhart developed in the 1930s. Prominent corporations like as Toyota, Motorola, GE, and Ford Motor Company have influenced the current iteration of this, emphasizing its evolution by practical application and continuous enhancement in many organizational contexts.

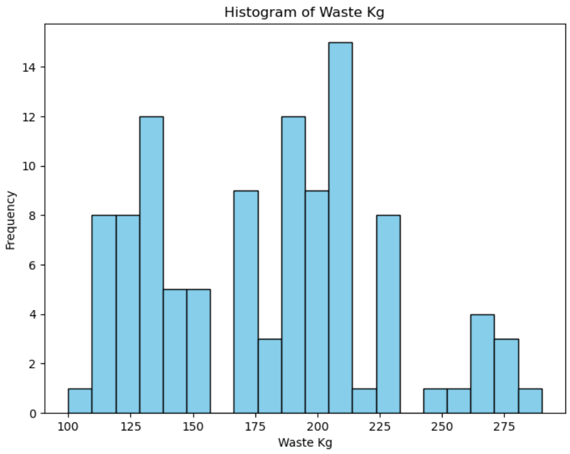

Figure 4. Histogram of Waste Kg

3. Statistical Inference

The statistical significance of observed patterns is then ascertained by using inferential statistics, such as correlation analysis and hypothesis testing, to investigate correlations between variables. Regression analysis, for example, may be used to measure the effect of various operational factors or machine settings on waste creation as shon in fig. 5, 6 7 and 8.

Figure 5. Scatter plot for Cutter vs Waste Kg

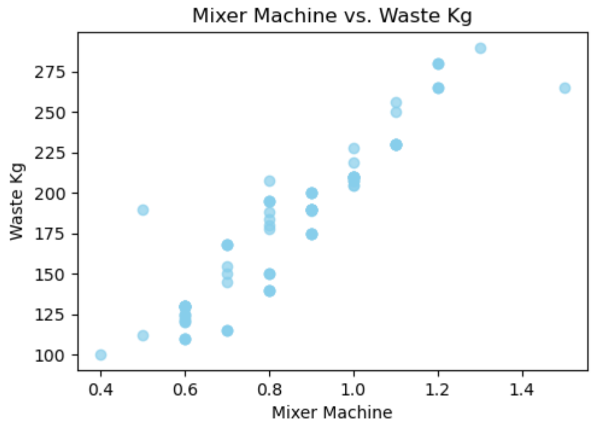

Figure 6. Scatter plot for Mixer Machine vs Waste kg

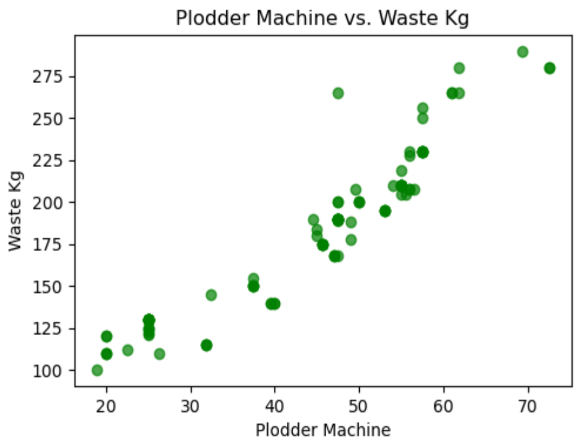

Figure 7. Scatter Plot for Plodder Machine vs Waste Kg

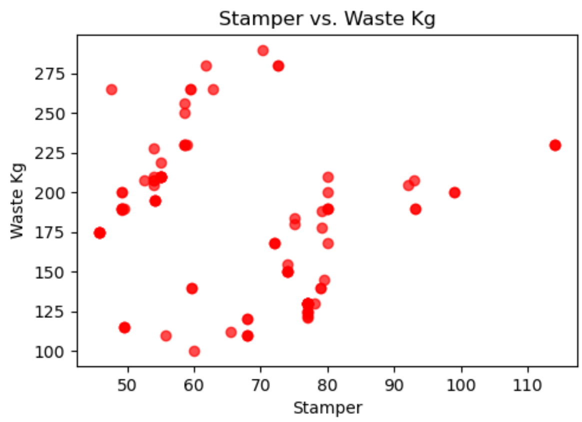

Figure 8. Scatter plot for Stamper vs Waste Kg

An understanding of the correlation between waste generation and machine settings in the soap making process may be gained from the examination of the output. We visually evaluated the relationship between the quantity of waste generated and each machine setting (Mixer Machine, Plodder Machine, Cutter, and Stamper) using scatter plots. Different patterns and connections were shown by the scatter plots. For example, the Mixer Machine setting showed a modest negative correlation with trash creation, suggesting that higher settings might result in less waste being produced. The Plodder Machine setting, on the other hand, displayed a more dispersed distribution of data points, indicating a weaker link. Distinctive distributions were seen in the scatter plots of Cutter and Stamper, signifying varying levels of impact on waste formation. These results provide insightful information for process optimization, pointing out possible areas for development to reduce waste and boost production efficiency in soap.

4. Analysis of Regression

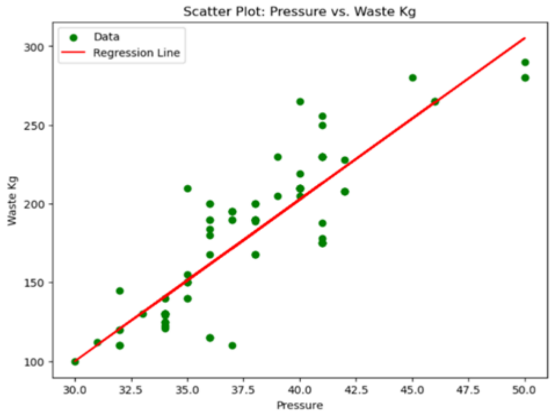

Figure 9. Regression Plot for Pressure vs Waste Kg

The results of the Measure phase's regression analysis showed in fig 9 that pressure levels had a direct impact on the output of waste in the soap-making process. According to the investigation, waste and pressure were positively correlated, meaning that waste production increased as pressure levels rose. This result emphasizes how crucial accurate pressure management is to reducing waste formation and increasing production efficiency. Regression analysis is used to quantify the link between pressure and waste, providing the organization with important information about the underlying causes of waste buildup. This data is essential for creating focused improvement plans that maximize waste reduction and increase overall operational performance through pressure management practice optimization. Furthermore, the regression analysis offers a statistical foundation for comprehending the extent of pressure's influence on waste production, allowing the company to efficiently allocate resources and actions in later stages of the DMAIC process.

5. Exploratory Data Analysis

An essential phase in data analysis is called exploratory data analysis (EDA), which entails analyzing and displaying the traits, connections, and patterns found in a dataset. Gaining understanding of the data's underlying structure, spotting any trends or anomalies, and developing ideas for more research are the main objectives of exploratory data analysis (EDA). In exploratory data analysis (EDA), several statistical methods and visualization instruments are utilized to examine many facets of the dataset, including its distribution, primary tendencies, variability, and relationships across variables, and possible anomalies. EDA facilitates the better understanding of data by analysts and data scientists, enables them to make well-informed decisions on feature engineering, preprocessing, and model selection, and eventually produces insightful findings to aid in decision-making. Before beginning any modeling process, EDA carefully reviews the data to establish the foundation for strong and efficient machine learning models and statistical analysis.

Key elements of exploratory data analysis (EDA) include data summary and visualization, which are essential for comprehending the features and patterns found in a dataset.

6. Histogram

A histogram is a graph that shows how numerical data is distributed. It is made up of a sequence of vertical bars, each of which stands for a bin—a range of values. Each bar's height represents the number or frequency of data points that fall inside that bin.

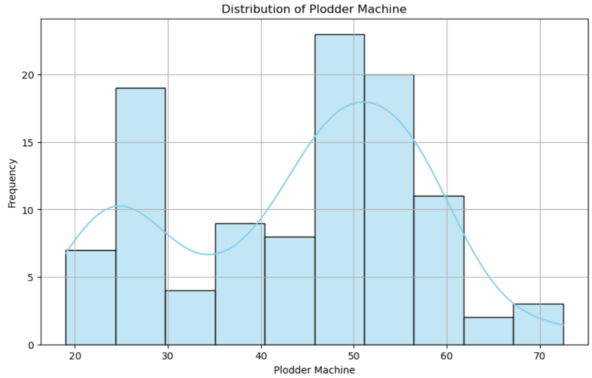

Figure 10. Histogram for Distribution of Plodder Machine

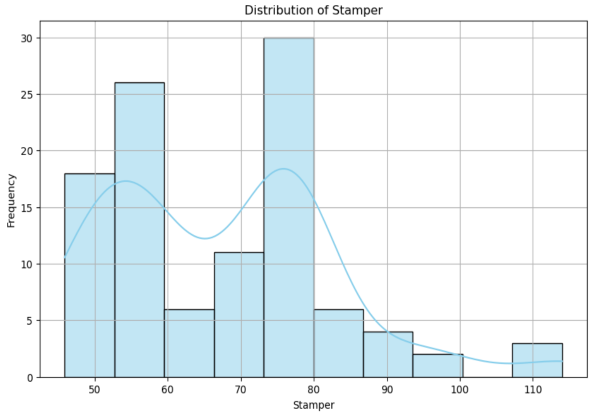

Figure 11. Histogram for Distribution of Stamper

Histograms are useful tools for figuring out a dataset's underlying distribution because they can show characteristics like multimodality, skewness, and symmetry. The visualization's level of detail can be altered by changing the number of bins, enabling a more in-depth examination of the properties of the data. In domains like statistics, data analysis, and machine learning, histograms are frequently employed for conducting exploratory data analysis, detecting anomalies, and evaluating the appropriateness of statistical models.

Here are histograms displaying data for different parameters, including 'Fresh Batch Kg', 'Waste Kg', 'No of soap', 'Mixer Machine', 'Plodder Machine', 'Cutter', 'Stamper', 'Temperature', 'Pressure', and 'Machine Time'. For instance, when analyzing the 'Fresh Batch Kg' parameter, the histogram displays the frequency of various batch sizes utilized in the soap production process.

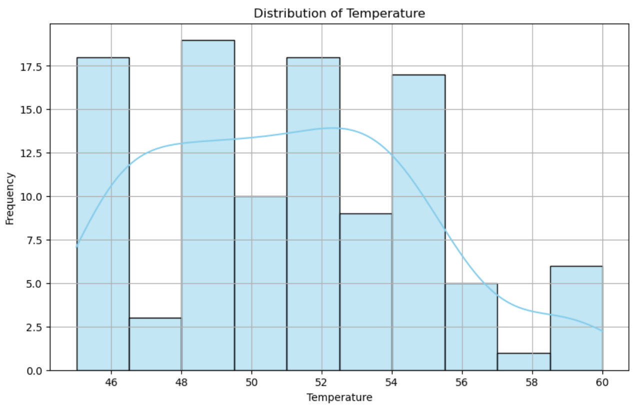

Figure 12. Histogram for Distribution of Temperature

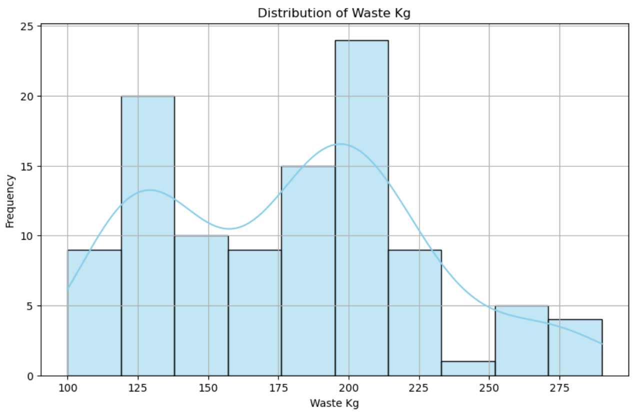

Figure 13. Histogram for Distribution of Waste Kg

Just like an operations manager, the histograms show the distribution of waste generated during production and the frequency of soap bars produced per batch. Every histogram offers valuable insights into the distribution patterns, allowing for a better understanding of the central tendency, variability, and potential outliers in the dataset as shown in figures 10, 11, 12, 13,14, 15, 16 and 17.

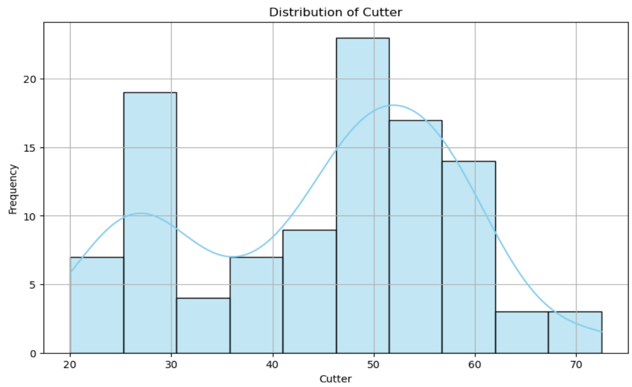

Figure14. Histogram for Distribution of Cutter

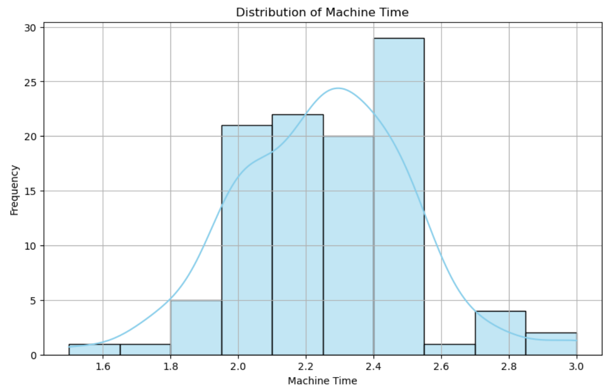

Figure 15. Histogram for Distribution of Machine Time

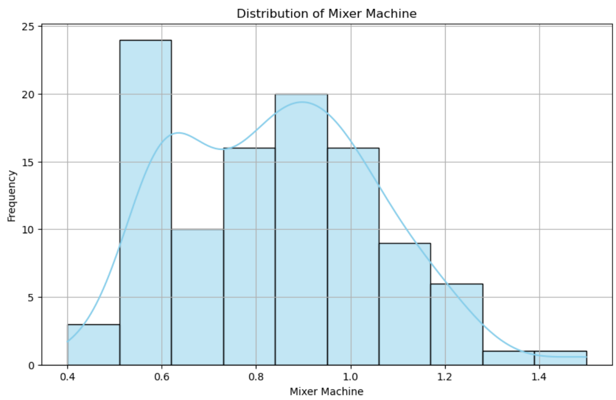

Figure 16. Histogram for Distribution of Mixer Machine

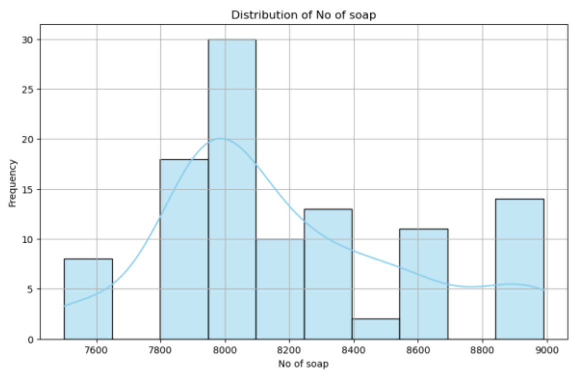

Figure 17. Histogram for Distribution of No of Soap

7. Pie Chart

A pie chart is a statistical visualization as shown in fig.18 that is circular and uses pie slices to represent categorical data. Every slice represents a category, and its size reflects the percentage of the total that it represents. When categorical data is visualized using pie charts, it works especially well when the categories are mutually exclusive and collectively exhaustive. They offer a simple and straightforward method for comparing the proportional sizes of various categories and comprehending how each one contributes to the total. However, because it can be difficult to accurately interpret small differences in slice sizes, pie charts may lose some of their effectiveness when there are too many categories or when the proportions are subtle. Pie charts are frequently used in presentations, reports, and infographics to visually appealingly communicate categorical data, despite this drawback.

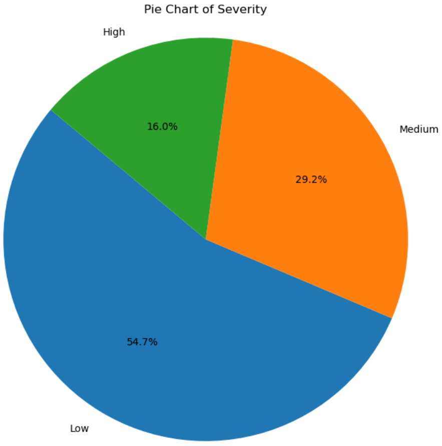

Figure 18. Pie Chart showing Severity percentage

V. Machhine Learning Pipeline and Results

Machine learning is a subset of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to perform tasks without explicit programming. The fundamental purpose of machine learning is to enable systems to learn from data, identify patterns, and make decisions or predictions based on those patterns. Instead of being explicitly programmed to carry out specific tasks, machine learning algorithms learn from the data provided to them and improve their performance over time through experience. The essence of machine learning lies in its ability to extract insights and knowledge from data, enabling systems to perform tasks more accurately and efficiently as they encounter new data. This capability is particularly valuable in fields where traditional rule-based programming approaches are impractical or insufficient due to the complexity and variability of the data.

1. Data Preprocessing

Data preprocessing lays the groundwork for effective machine learning models by transforming raw data into a format that is suitable for analysis and model training. This crucial step involves a series of operations aimed at cleaning, organizing, and enhancing the data to extract meaningful insights and facilitate accurate predictions. At its core, data preprocessing involves several key tasks. Firstly, data cleaning entails identifying and handling missing values, outliers, and inconsistencies within the dataset to ensure its integrity and reliability. This may involve techniques such as imputation, where missing values are replaced with estimated values based on the available data, or outlier detection and removal to mitigate their influence on the model's performance. Below table has the cleaned Raw Data.

Model Traning

2. Decision Tree Algorithm

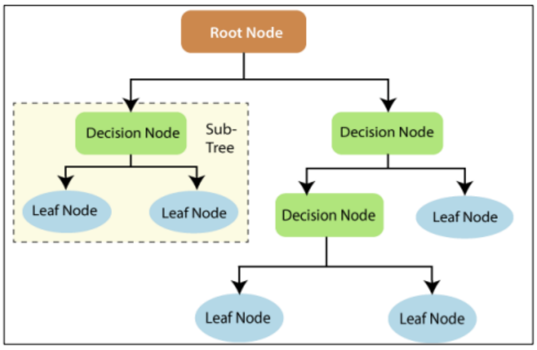

Figure 19. Representation of Decision Tree

Decision trees play a crucial role in data mining and machine learning, providing a clear and effective method for classifying tasks as shown in fig. 19. With a hierarchical structure, decision-making becomes easier as features are evaluated against threshold values, resulting in intuitive rules for classification. Decision trees prioritize conceptual rules over numerical weights, which enhances their interpretability and makes them easier to construct, unlike complex neural networks. The simplicity of these systems, along with their versatility in handling different types of data, has resulted in their widespread use in a range of fields, such as image processing and pattern recognition. Illustrating their usefulness, Figure demonstrates a common decision tree, displaying how nodes represent features and branches indicate decision outcomes. Visual representations are incredibly helpful in simplifying intricate decision boundaries and aiding in the comprehension of class relationships. Decision trees are incredibly versatile tools that strike a balance between simplicity and accuracy. They are indispensable for a wide range of tasks, from predictive modeling to data analysis in real-world scenarios.

The decision tree algorithm was implemented with great attention to detail, utilizing the powerful functionality and user-friendly interface of the MATLAB Classification Learner Toolbox. The model was carefully adjusted with specific hyper parameters to maximize its performance. It is worth mentioning that the maximum number of splits was limited to 100, which strikes a good balance between the complexity and depth of the decision tree. Guided by the Gini's diversity index, a commonly used metric for assessing node impurity, the split criterion was chosen to enhance the model's discriminative power. In order to ensure the accuracy of the decision tree structure and reduce the risk of errors, the option to use surrogate decision splits was turned off.

Through careful model tuning and evaluation, the decision tree model's generalization performance was rigorously assessed using cross-fold validation. Cross-fold validation is a commonly used technique in machine learning to estimate the performance of a predictive model. During this process, the dataset is divided into k subsets, or folds. The model is then trained on k-1 folds and tested on the remaining fold. This procedure is repeated multiple times, with each iteration using a different fold as the test set. The performance metrics are then averaged across all iterations. Through a systematic process of rotating subsets of the data for training and testing, cross-fold validation offers a reliable estimate of the model's performance on unseen data.

When it comes to the decision tree model created using MATLAB Classification Learner Toolbox, cross-fold validation played a crucial role in confirming its accuracy and dependability across various subsets of the dataset. By following a meticulous validation approach, we can ensure that the model's performance is not excessively reliant on a specific subset of the data. This allows us to obtain a more accurate estimate of its ability to generalize. By incorporating cross-fold validation into the model evaluation process, the confidence in the model's ability to accurately classify unseen data is greatly enhanced. As a result, decision-makers can have full trust in the insights produced by the decision tree model, enabling them to make well-informed decisions in critical areas like healthcare and finance, where accuracy and reliability are of utmost importance.

3. Confusion Metrix

A confusion matrix is shown in fig. 20 an essential tool for assessing the performance of a classification model by offering a thorough analysis of its predictions in comparison to the actual class labels in the dataset. The confusion matrix of the model output is shown below and categorizes severity into three classes: "high severity," "medium severity," and "low severity."

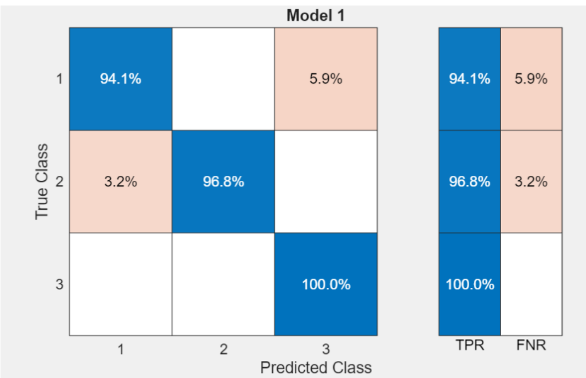

Figure 20. Confusion Matrix for all classes

Each row in the matrix corresponds to the true class labels, whereas each column represents the anticipated classes. The top row shows that 94.1% of incidents categorized as "high severity" were correctly classified, with 0 instances misclassified as "medium severity" and 5.9% mistakenly tagged as "low severity." The second row shows that 3.2% of incidents designated as "medium severity" were misclassified as "high severity," while 96.8% were correctly recognized. The algorithm accurately categorized all occurrences of "low severity," demonstrating its skill in differentiating this class. The confusion matrix provides a detailed summary of the model's performance, helping to pinpoint areas that need improvement and optimization in activities related to severity categorization. The overall accuracy of mode is 98.1%.

4. ROC Curve

The ROC curve is a visual tool used to assess the effectiveness of a binary classification model at various threshold levels.

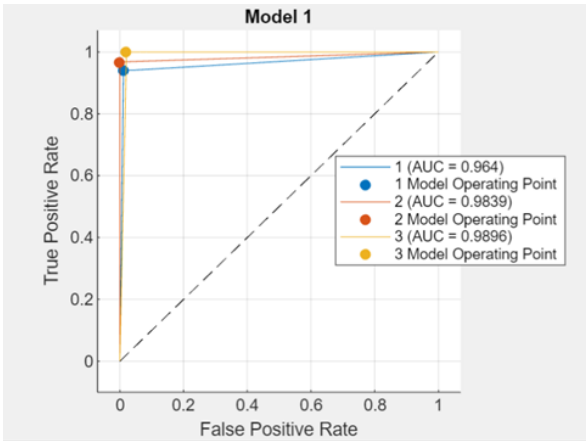

Figure 21. ROC Curve for all classes

The graph displays as shown in fig. 21 the relationship between the true positive rate (TPR) and the false positive rate (FPR) across different threshold settings. The AUC (Area under the Curve) is a single statistic that measures the overall effectiveness of the ROC curve.

The ROC curve for class 1 has an AUC value of 0.964. The model has a strong capacity to differentiate class 1 examples from instances of other classes, with a 96.4% probability of ranking a randomly selected positive instance higher than a randomly selected negative one. A high AUC score indicates that the model is proficient at accurately classifying class 1 cases with minimal misclassifications.

Class 2 achieved an AUC value of 0.9839, which is greater than that of class 1. The model excels at identifying class 2 occurrences from instances of other classes, achieving a 98.39% accuracy in classification. A higher AUC value suggests that the model has excellent discriminating ability for class 2, making it very dependable in finding occurrences of this class.

Class 3 achieved the greatest AUC value of 0.9896 among the three classes based on the ROC curve. This indicates outstanding performance in accurately identifying class 3 examples compared to cases from other classes, with a 98.96% probability of right classification. The model's excellent AUC score indicates its high accuracy in recognizing occurrences of class 3, demonstrating its effectiveness in categorizing this specific class as given in table 1, 2.

Table 1. Soap Waste Data

Sr. #	Waste Kg	No of soap	Mixer Machine	Plodder Machine	Cutter	Stamper	Temperature	Pressure	Machine Time	Severity
0	265.0	8120.0	1.5	47.5	48.5	47.5	50.0	40.0	2.0	High
1	210.0	7907.0	1.0	54.0	54.0	54.0	48.0	35.0	2.5	Medium
2	190.0	7890.0	0.5	44.5	49.5	49.5	52.0	38.0	2.2	Low
3	208.0	8567.0	0.8	49.6	52.6	52.6	55.0	42.0	2.3	Medium
4	110.0	8964.0	0.6	26.3	28.3	55.8	47.0	37.0	2.1	Low
---	---	---	---	---	---	---	---	---	---	---
101	130.0	8567.0	0.6	25.0	27.0	77.0	46.0	34.0	2.5	Low
102	210.0	8324.0	1.0	55.0	55.0	55.0	54.0	40.0	2.0	Medium
103	140.0	7950.0	0.8	40.0	41.0	79.0	50.0	34.0	2.4	Low
104	168.0	7994.0	0.7	47.0	48.0	72.0	51.0	38.0	2.1	Low
105	190.0	8000.0	0.9	47.5	48.5	80.0	49.0	37.0	2.2	Low

Table 2. Raw Data

Sr. #	Waste Kg	No of soap	Mixer Machine	Plodder Machine	Cutter	Stamper	Temperature	Pressure	Machine Time	Severity
0	265.0	8120.0	1.5	47.5	48.5	47.5	50.0	40.0	2.0	High
1	210.0	7907.0	1.0	54.0	54.0	54.0	48.0	35.0	2.5	Medium
2	190.0	7890.0	0.5	44.5	49.5	49.5	52.0	38.0	2.2	Low
3	208.0	8567.0	0.8	49.6	52.6	52.6	55.0	42.0	2.3	Medium
4	110.0	8964.0	0.6	26.3	28.3	55.8	47.0	37.0	2.1	Low
---	---	---	---	---	---	---	---	---	---	---
101	130.0	8567.0	0.6	25.0	27.0	77.0	46.0	34.0	2.5	Low
102	210.0	8324.0	1.0	55.0	55.0	55.0	54.0	40.0	2.0	Medium
103	140.0	7950.0	0.8	40.0	41.0	79.0	50.0	34.0	2.4	Low
104	168.0	7994.0	0.7	47.0	48.0	72.0	51.0	38.0	2.1	Low
105	190.0	8000.0	0.9	47.5	48.5	80.0	49.0	37.0	2.2	Low

Data preprocessing lays the groundwork for effective machine learning models by transforming raw data into a format that is suitable for analysis and model training. This crucial step involves a series of operations aimed at cleaning, organizing, and enhancing the data to extract meaningful insights and facilitate accurate predictions. At its core, data preprocessing involves several key tasks. Firstly, data cleaning entails identifying and handling missing values, outliers, and inconsistencies within the dataset to ensure its integrity and reliability.

Table 3. Standardized Data

Sr. #	Waste Kg	No of soap	Mixer Machine	Plodder Machine	Cutter	Stamper	Temperature	Pressure	Machine Time	Severity
0	1.827978	-0.176167	3.167578	0.281608	0.282613	-1.299234	-0.274161	0.562334	-0.995852	High
1	0.651956	-0.727831	0.752015	0.761401	0.699103	-0.877778	-0.812329	-0.684683	0.966236	Medium
2	0.224312	-0.771861	-1.663548	0.060166	0.358338	-1.169555	0.264007	0.063527	-0.211017	Low
3	0.609191	0.981551	-0.214210	0.436618	0.593087	-0.968553	1.071259	1.061141	0.181401	Medium
4	-1.48626	2.009771	-1.180435	-1.283253	-1.24704	-0.761067	-1.081413	-0.185876	-0.603435	Low
---	---	---	---	---	---	---	---	---	---	---
101	-1.05862	0.981551	-1.180435	-1.379212	-1.34548	0.613527	-1.350497	-0.934087	0.966236	Low
102	0.651956	0.352188	0.752015	0.835215	0.774828	-0.812939	0.802175	0.562334	-0.995852	Medium
103	-0.84479	-0.616463	-0.214210	-0.271998	-0.28532	0.743206	-0.274161	-0.934087	0.573818	Low
104	-0.24609	-0.502504	-0.697323	0.244701	0.244750	0.289330	-0.005077	0.063527	-0.603435	Low
105	0.224312	-0.486964	0.268902	0.281608	0.282613	0.808045	-0.543245	-0.185876	-0.211017	Low

This may involve techniques such as imputation, where missing values are replaced with estimated values based on the available data, or outlier detection and removal to mitigate their influence on the model's performance. Above table has the cleaned Raw Data.

Z-score standardization, also known as standard scaling, is a normalization technique that transforms the features of a dataset to have a mean of 0 and a standard deviation of 1. This method centers the data on its mean and adjusts for differences in scale, making it robust to outliers. The transformation involves subtracting the mean of each feature from its original value and then dividing by the standard deviation. Z-score standardization results in a distribution with a mean of 0 and a standard deviation of 1, allowing for easier interpretation and comparison across features. It is commonly used in statistical analysis, machine learning, and data preprocessing pipelines to standardize the scale of features before applying algorithms that are sensitive to differences in scale, such as linear regression and support vector machines. The data given in table 4.

Table 4. Normalized Data

Sr. #	Waste Kg	No of soap	Mixer Machine	Plodder Machine	Cutter	Stamper	Temperature	Pressure	Machine Time	Severity
0	0.868421	0.416107	1.000000	0.532710	0.542857	0.023495	0.333333	0.50	0.333333	High
1	0.578947	0.273154	0.545455	0.654206	0.647619	0.118943	0.200000	0.25	0.666667	Medium
2	0.473684	0.261745	0.090909	0.476636	0.561905	0.052863	0.466667	0.40	0.466667	Low
3	0.568421	0.716107	0.363636	0.571963	0.620952	0.098385	0.666667	0.60	0.533333	Medium
4	0.052632	0.982550	0.181818	0.136449	0.158095	0.145374	0.133333	0.35	0.400000	Low
---	---	---	---	---	---	---	---	---	---	---
101	0.157895	0.716107	0.181818	0.112150	0.133333	0.456681	0.066667	0.20	0.666667	Low
102	0.578947	0.553020	0.545455	0.672897	0.666667	0.133627	0.600000	0.50	0.333333	Medium
103	0.210526	0.302013	0.363636	0.392523	0.400000	0.486050	0.333333	0.20	0.600000	Low
104	0.357895	0.331544	0.272727	0.523364	0.533333	0.383260	0.400000	0.40	0.400000	Low
105	0.473684	0.335570	0.454545	0.532710	0.542857	0.500734	0.266667	0.35	0.466667	Low

Min-max scaling is a normalization technique that transforms the features of a dataset to a fixed range, typically between 0 and 1. This method ensures that all features have the same scale, making them directly comparable. The transformation involves subtracting the minimum value of each feature from its original value and then dividing by the range, which is the difference between the maximum and minimum values. Min-max scaling preserves the shape of the original distribution but may be sensitive to outliers, as extreme values can disproportionately affect the scaled values. Despite this limitation, min-max scaling is widely used in various applications, including image processing, neural networks, and data mining. The data given in table 5.

Table 5. Encoded Data

Sr. #	Waste Kg	No of soap	Mixer Machine	Plodder Machine	Cutter	Stamper	Temperature	Pressure	Machine Time	Severity
0	1.827978	-0.176167	3.167578	0.281608	0.282613	-1.299234	-0.274161	0.562334	-0.995852	1
1	0.651956	-0.727831	0.752015	0.761401	0.699103	-0.877778	-0.812329	-0.684683	0.966236	2
2	0.224312	-0.771861	-1.663548	0.060166	0.358338	-1.169555	0.264007	0.063527	-0.211017	3
3	0.609191	0.981551	-0.214210	0.436618	0.593087	-0.968553	1.071259	1.061141	0.181401	2
4	-1.486266	2.009771	-1.180435	-1.283253	-1.247040	-0.761067	-1.081413	-0.185876	-0.603435	3
---	---	---	---	---	---	---	---	---	---	---
101	-1.058621	0.981551	-1.180435	-1.379212	-1.345483	0.613527	-1.350497	-0.934087	0.966236	3
102	0.651956	0.352188	0.752015	0.835215	0.774828	-0.812939	0.802175	0.562334	-0.995852	2
103	-0.844799	-0.616463	-0.214210	-0.271998	-0.285328	0.743206	-0.274161	-0.934087	0.573818	3
104	-0.246097	-0.502504	-0.697323	0.244701	0.244750	0.289330	-0.005077	0.063527	-0.603435	3
105	0.224312	-0.486964	0.268902	0.281608	0.282613	0.808045	-0.543245	-0.185876	-0.211017	3

Encoding categorical variables is a fundamental preprocessing step in machine learning, as many algorithms require numerical input and cannot directly handle categorical data. To convert categorical variables into a numerical format, practitioners often utilize techniques such as one-hot encoding or label encoding. By effectively encoding categorical variables into a numerical format, practitioners can facilitate the integration of categorical data into machine learning models, enabling algorithms to process and derive insights from a broader range of datasets.

A positive correlation between pressure and waste indicates that higher pressure leads to more waste. Most of the waste is classified as low severity (54.7%), with significant portions in medium (29.2%) and high (16%) severity levels. The decision tree provides a clear framework for making structured, data-driven decisions. The confusion matrix demonstrates high accuracy in classifying waste severity, especially for high severity waste.

DISCUSSION

The challenge of excessive waste generation within the soap production process presents a multifaceted problem in the ATN Soup Industry. This problem has significant implications for operational efficiency, financial sustainability, and environmental responsibility at the same time. The organization has adopted a dual approach, integrating the DMAIC methodology and machine learning techniques, to address this challenge in a comprehensive manner. A meticulous definition of the waste generation problem and the establishment of clear objectives for improvement are the first steps in the DMAIC framework, which offers a structured roadmap for problem-solving. During the Measure phase, a comprehensive dataset that encompasses a variety of production parameters is gathered and analyzed. This reveals crucial insights into the extent of waste generation and the detrimental impacts that it has on the overall efficiency of production. The next phase, which is called "Analyze," delves even further into the data, thereby revealing correlations between the pressure levels and the amount of waste that is produced. This lays the groundwork for initiatives that are specifically aimed at improvement.

In the process of modeling the connection between pressure levels and waste generation, machine learning, and more specifically the decision tree algorithm, has emerged as an essential tool. With an impressive accuracy rate of 98%, the decision tree algorithm can effectively classify the severity of waste, which enables the identification of key areas within the production process that can be optimized. To reducing the amount of waste generated and improving operational efficiency, this predictive capability gives the organization the ability to make decisions based on accurate information and to drive targeted interventions.

In the future, work within the ATN Soup Industry presents exciting opportunities for further advancement and innovation in waste management and production efficiency. These opportunities are presented in the context of the prospects that lie ahead. An avenue that could be investigated is the possibility of enhancing and broadening the scope of predictive modeling techniques through the application of machine learning. It has been demonstrated that the decision tree algorithm is quite accurate in classifying the severity of waste. However, the incorporation of more advanced algorithms, such as neural networks or ensemble methods, could result in even greater predictive power and more nuanced insights into the patterns of waste generation. The organization can refine its predictive models by utilizing the capabilities of these advanced techniques, which allows it to accurately forecast waste levels and identify potential risk factors or opportunities for improvement with a higher degree of precision. In addition, future efforts might concentrate on improving real-time monitoring and control systems to make proactive decision-making and intervention easier to accomplish. The organization can implement robust monitoring systems that are able to capture real-time data on pressure levels, production parameters, and waste generation rates by leveraging technologies that are connected to the Internet of Things (IoT) and sensor networks. This continuous stream of data can be analyzed in conjunction with historical datasets to identify anomalies, anticipate potential problems, and initiate timely interventions in order to reduce the accumulation of waste and maximize the efficiency of production.

CONCLUSION

As a conclusion, the ATN Soup Industry has been able to effectively address the challenge of excessive waste generation in its production process because of the integration of the DMAIC methodology and techniques for machine learning. The organization has gained valuable insights into the underlying causes of waste as well as opportunities for improvement because of the systematic definition of the problem, the analysis of production data, and the modeling of the relationship between pressure levels and waste generation. A powerful instrument for guiding decision-making and driving targeted interventions, the decision tree algorithm has emerged because of its high accuracy in classifying the severity of waste.

Moving forward, the organization has the potential to build upon these accomplishments by further refining techniques for predictive modeling, improving real-time monitoring systems, and fostering collaboration with industry partners and regulatory bodies. The ATN Soup Industry can continue to optimize its production processes, reduce the amount of waste it generates, and uphold its commitment to operational efficiency and environmental sustainability if it embraces innovation, collaborates with others, and strives for continuous improvement. The industry could pave the way for a manufacturing ecosystem that is more efficient, resilient, and environmentally responsible if it takes proactive measures and forms strategic partnerships.

REFERENCES

[1] Harry, M., & Schroeder, R. (2000). Six Sigma: The Breakthrough Management Strategy Revolutionizing the World's Top Corporations. Doubleday.

[2] Ross, P. (2019). Artificial Intelligence: A Guide for Thinking Humans. Little, Brown Spark.

[3] Pyzdek, T., & Keller, P. A. (2014). The Six Sigma Handbook. McGraw-Hill Education.

[4] Davenport, T. H., & Harris, J. (2007). Competing on Analytics: The New Science of Winning. Harvard Business Review Press.

[5] M. Mohan Prasad, J. M. Dhiyaneswari, J. Ridzwanul Jamaan, S. Mythreyan, and S. M. Sutharsan, “A framework for lean manufacturing implementation in Indian textile industry,” Materials Today: Proceedings, vol. 33, pp. 2986–2995, Jan. 2020, doi: 10.1016/j.matpr.2020.02.979.

[6] N. Kumar, S. Shahzeb Hasan, K. Srivastava, R. Akhtar, R. Kumar Yadav, and V. K. Choubey, “Lean manufacturing techniques and its implementation: A review,” Materials Today: Proceedings, vol. 64, pp. 1188–1192, Jan. 2022, doi: 10.1016/j.matpr.2022.03.481.

[7] S. Vinodh, S. V. Kumar, and K. E. K. Vimal, “Implementing lean sigma in an Indian rotary switches manufacturing organisation,” Production

Planning & Control, vol. 25, no. 4, pp. 288–302, Mar. 2014, doi: 10.1080/09537287.2012.684726.

[8] J. Bhamu and K. Singh Sangwan, "Lean manufacturing: literature review and research issues," International Journal of Operations & Production Management, vol. 34, no. 7, pp. 876–940, Jan. 2014, doi: 10.1108/IJOPM-08-2012-0315.

[9] "Possibilities of Maintenance Service Process Analyses and Improvement Through Six Sigma, Lean and Industry 4.0 Implementation | SpringerLink." Accessed: Sep. 24, 2023. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-01614- 2_43

[10] EV Gijo, J Scaria, J Antony, "Application of six sigma methodology to reduce defects of a grinding process". Quality and Reliability Engineering International, Vol. 27, pp. 1221-1234, 4 May, 2011, doi: doi.org/10.1002/qre.1212

[11] Suriya Lertwattanapongchai, Fredric Willain Swierczek "Assessing the change process of Lean Six Sigma: a case analysis". International journal of Lean Six Sigma, Vol. 5, 28 Oct. 2014

[12] DIEgo Pacheco, Isaac Pergher, Carla ten Caten, "18 comparative aspects between Lean and Six Sigma: Complementarity and implications". International journal of Lean Six Sigma, Vol. 5, 01 June. 2015

[13] "Cultural Impact on Lean Six Sigma and Corporate Success: Causal Analyses ... - Miriam Jacobs - Google Books." Accessed: Jul. 28, 2023. [Online]. Available: https://books.google.com.pk/books

[14] Ruth Dede Adikorley, Lori Rothenberg, Aron Guillory, "Lean Six Sigma applications in the textil". International journal of Lean Six Sigma, Vol. 8, 5 June, 2017 doi: 10.1108/IJLSS-03-2016-0014

[15] M. S. Kaswan et al., "Integrated Green Lean Six Sigma-Industry 4.0 approach to combat COVID-19: from literature review to framework development," International Journal of Lean Six Sigma, vol. ahead-of-print, no. ahead-of-print, Jan. 2023, doi: 10.1108/IJLSS-11-2022-0227.

[16] R. Rathi, M. S. Kaswan, J. Antony, J. Cross, J. A. Garza-Reyes, and S. L. Furterer, "Success factors for the adoption of green lean six sigma in healthcare facility: an ISMMICMAC study," International Journal of Lean Six Sigma, vol. 14, no. 4, pp. 864–897, Jan. 2022, doi: 10.1108/IJLSS-02-2022-0042.

[17] B. Chaurasia, D. Garg, and A. Agarwal, "Framework to improve performance through implementing Lean Six Sigma strategies to oil exporting countries during recession or depression," International Journal of Productivity and Performance Management, vol. 65, no. 3, pp. 422–432, Jan. 2016, doi: 10.1108/IJPPM-01-2015-0011.

[18] V. S. M., "Corporate perspectives: commonalities and differences between Six Sigma and Lean," International Journal of Lean Six Sigma, vol. 6, no. 3, pp. 281–288, Jan. 2015, doi: 10.1108/IJLSS-12-2014-0038.

[19] R. Jamil, I. Jamil, Ming Li, Zhao Jinquan, Development Tendency of Energy: A Short Review, International Scholarly and Scientific Research & Innovation 7(9) 2013.

[20] S. B. Rane, S. Wavhal, and P. R. Potdar, "Integration of Lean Six Sigma with Internet of Things (IoT) for productivity improvement: a case study of contactor manufacturing industry," Int J Syst Assur Eng Manag, vol. 14, no. 5, pp. 1990–2018, Oct. 2023, doi: 10.1007/s13198-023-01980-7.

[21] Lee, J., Bagheri, B., & Kao, H. A. (2015). A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters, 3, 18-23.

[22] Lee, J., Lapira, E., Bagheri, B., & Kao, H. A. (2013). Recent Advances and Trends in Predictive Manufacturing Systems in Big Data Environment. Manufacturing Letters, 1(1), 38-41.

[23] Schuh, G., Anderl, R., Gausemeier, J., ten Hompel, M., Wahlster, W., & Westkämper, E. (Eds.). (2017). Industrie 4.0 Maturity Index: Managing the Digital Transformation of Companies. Springer.

[24] Carbone, J. (2018). Artificial Intelligence for Predictive Maintenance. CRC Press

Syed Muhammad Shakir Bukhari has completed his MS degree in Engineering Management at the Industrial Engineering Department, University of Engineering and Technology, Peshawar, Pakistan, in 2024.

Dr. Rehman Akhter (PhD USA) is an Associate Professor in the Department of Industrial Engineering, University of Engineering and Technology ,Peshawar, Pakistan.

How to cite this article:

Syed Muhammad Shakir Bukhari, Rehman Akhtar "Enhancing Quality and Reducing Wastage in ATN Soap Industry Peshawar Using Six Sigma and Artificial Intelligence (AI)" International Journal of Engineering Works, Vol. 11, Issue 09, PP. 177-187, September 2024. https://doi.org/10.34259/ijew.24.1109177187.