Implementation of DMAIC using Machine Learning to Analyze Product Defects in the Welding Consumables Industry

Evan Adriel and Christian Harito
Industrial Engineering Department, BINUS Graduate Program – Master of Industrial Engineering,
Bina Nusantara University,
Jakarta, Indonesia 11840.
evan.adriel@binus.ac.id, christian.harito@binus.ac.id

Abstract—**Ensuring high product quality while minimizing defect rates remains a critical objective in manufacturing, particularly in the welding consumables industry. This study applies Lean Six Sigma (LSS) methodology through the Define, Measure, Analyze, Improve, and Control (DMAIC) framework, integrated with Machine Learning (ML) techniques to accelerate defect analysis. Over a 25-day period, defect and process data were analyzed using Fishbone Diagrams, Fuzzy Failure Mode and Effects Analysis, and a Random Forest-based multi-class classification model. Four key production parameters high and low extrusion pressure, flux temperature, and flux density were identified as major contributors to defects. Improvement strategies, including setting operational pressure limits, enhancing cooling system protection, and introducing durometer-based flux hardness measurements, were implemented. Following these actions, the Rolled Throughput Yield (RTY) improved from 98.5% to 98.7%, with a decrease in high-risk defect codes, as validated by updated FMEA results and more stable control charts, signaling a promising initial result. This improvement represents an early stage, with further gains expected through continued operator adaptation and process refinement. Overall, the integration between field-based observations and machine learning analysis significantly enhances root cause identification and promotes sustainable quality improvement in manufacturing operations.**

I. Introduction

In today's highly competitive market, manufacturing companies are finding it increasingly difficult to maintain their competitiveness [1]. To remain viable, it is necessary to adopt a renewed mindset that prioritizes customer satisfaction as a critical factor for long-term success. Customers are more likely to be satisfied when companies offer high-quality products at reasonable costs [2]. To achieve this, manufacturers must continuously improve their production processes, particularly by reducing various forms of waste.

One critical type of waste in industrial production is product defects, which not only result in financial losses but also damage customer trust and brand reputation. In a case observed at a major welding consumable manufacturer in Indonesia, the average defect rate between January and June 2023 was recorded at 2.29%, exceeding the company's internal quality threshold of 1.5%. These recurring defects not only reflect production inefficiencies but also represent a bottleneck in the company's quality improvement efforts [3]. However, identifying the root causes of such defects remains a complex task, especially in traditional manufacturing environments that still rely heavily on manual inspection and experience-based decision making.

This challenge is particularly evident in the welding consumable industry, which plays a vital role in infrastructure development but has been slow to adopt modern quality management practices. In many cases, outdated technologies and low awareness of quality systems contribute to recurring defects and reduced operational efficiency.

Unlike several previous studies that applied Define, Measure, Analyze, Improve, and Control (DMAIC) methodology to improve product quality such as reducing defects in manufacturing industries [11, 12, 13, 14] and minimizing waste in the food industry [5] by relying solely on conventional Lean Six Sigma (LSS) tools, this paper proposes an evolved DMAIC framework enhanced by ML techniques. While the application of ML has been increasingly widespread in recent years, conventional LSS tools are still incorporated in this study as a comparative analytical approach. In contrast to the approach in [7], which employed ML exclusively, the use of conventional tools remains essential due to the relatively traditional working environment in the welding consumables industry. The significant human involvement in operational processes introduces the potential for unquantifiable "X factors" that may not be easily detected through data-driven methods alone. Therefore, conventional tools are employed to complement the analysis by providing insights from real-world operational perspectives. This integrated approach aims not only to identify the root causes of product defects but also to predict the likelihood of specific failures occurring during production. The ultimate goal is to support continuous improvement and enhance product quality in the electrodes manufacturing industry in Indonesia.

Evan Adriel is with the Industrial Engineering Department, Faculty of Engineering, Bina Nusantara University, Jakarta, Indonesia 11480 (e-mail: evan21adriel@gmail.com)

Christian Harito is with the Industrial Engineering Department, Faculty of Engineering, Bina Nusantara University, Jakarta, Indonesia 11480 (e-mail: christian.harito@binus.ac.id)

Notations and Preliminaries

Let $\mathfrak{D} = \{(x_i, y_i)\}$ for i = 1 to N represent the dataset, where:

- $x_i \in \mathbb{R}^n$ is the feature vector representing process parameters for the *i-th* extrusion cycle,
- $y_i \in \{0, 1, 2, 3, 4, 5, 6\}$ is the corresponding class label indicating one of seven defect types: *Chipped (Chp)*, *Concentricity > Standard (Conc>Std)*, *Crack (Crk)*, *Porous (Prs)*, *Puffed (Pff)* , *Scratch (Scth)*, and *Others (Oth)*.

The classification function is defined as $\hat{y}_i = f(x_i; \theta)$, where f is a machine learning model parameterized by θ, such as Random Forest, Gradient Boosting, Support Vector Machine, k-Nearest Neighbors, or Naïve Bayes.

Key production parameters used as input features include:

- t_{wet} : Wet Mixing Duration (minutes)
- V_b : Binder Volume (L / 10 kg Flux)
- p_{ext} : Extrusion Pressure (kg/cm^2)
- ρ_{flux} : Flux Density (HA)
- T_{flux} : Flux Temperature (˚C)
- Con_w : Wire Concentricity (mm)
- W_{rf} : Recycled Flux Quantity (kg)

The Rolled Throughput Yield (RTY) is calculated as:

$$RTY = \frac{(Opportunities - Defect)}{Opportunities} \times 100\%$$

Fuzzy FMEA root causes within 80% Pareto threshold:

- D4 : Excess Binder Volume
- D5 : Too few Binder Volume
- D6 : Excess Recycled Flux Quantity
- D7 : Too high Extrusion Pressure
- D8 : Too low Extrusion Pressure
- D9 : Inconsistent Flow inside Container Flux

II. METHODOLOGY

A. Data Collection

The data collection was conducted in July 2024, spanning from July 2nd to July 31st, and was based on Coating Machine 01 at a leading welding consumables manufacturer in Indonesia. This machine produces AWS E6013 welding electrodes with a diameter of Ø3.20 mm and was selected due to its significant contribution to the overall defect rate. Two types of data were collected: primary and secondary. The primary data was obtained through direct observation, focusing on historically applied problem-solving procedures. In contrast, the secondary data was sourced from production records, including process parameters and defect quantities, resulting in approximately 2500 rows of data.

B. DMAIC Implementation

TABLE I. FEATURES AND TARGETS VARIABLE FOR MULTI-CLASS CLASSIFICATION MODELS

Variables for Multi-class Classification	
Features	***Targets***
Dry Mixing Duration	Scratch
Wet Mixing Duration	Concentricity > Standards
New Flux Quantity	Puffed
Binder Volume	Porous
Mixing Speed	Crack
Flux Temperature	Chipped
Ambient Temperature	Others
Flux Density	-
Recycled Flux Quantity	-
Piston Ring Diameter	-
Flux Container Diameter	-
Nozzle to Dies Distance	-
Wire Length	-
Rod Feeding Speed	-
High Extrusion Pressure	-
Middle Extrusion Pressure	-
Low Extrusion Pressure	-

The DMAIC methodology in this study is structured to systematically address quality issues in welding electrode production. In the Define phase, customer requirements are translated into measurable parameters using the Critical to Quality (CTQ) framework, ensuring alignment with the Voice of Customer (VOC) and focusing on key quality attributes [4]. During the Measure phase, production data such as defect counts and process parameters are gathered to establish baseline performance, with tools like RTY and control charts used to quantify overall process efficiency [5]. The Analyze phase combines traditional quality tools, such as Fishbone Diagrams and Fuzzy FMEA, with a data-driven approach using supervised machine learning models to classify and predict defect types. This hybrid approach enhances root cause identification through both qualitative insights and quantitative validation. In the Improve phase, corrective actions are developed based on the analysis results, including proposed enhancements to measurement tools, procedures, and other improvements aimed at increasing product quality. Finally, the Control phase ensures the sustainability of these improvements by updating or establishing standard operating procedures (SOPs), enabling consistent implementation during the welding consumables production process.

C. Machine Learning

To support defect classification in the Analyze phase, a multiclass-classification ML approach was adopted, as the dataset contained seven distinct defect types. ML, a subset of artificial intelligence capable of predicting outcomes based on historical data without the need for explicit programming rules

[6], was selected for its efficiency in handling complex classification tasks. This setup was modeled using various production parameters as input features, while the defect type served as the target variable. The features and corresponding defect types are summarized in Table 1.

Five algorithms were evaluated: k-Nearest Neighbors (kNN), Random Forest (RF), Gradient Boosting (Gboost), Naïve Bayes, and Support Vector Machine (SVM) using Orange Data Mining. Each was selected based on its suitability for industrial classification tasks and ability to model nonlinear relationships among production factors. kNN was chosen for its straightforward implementation and reliable performance in data classification, where new instances are assigned, labels based on the similarity to their nearest neighboring points [7]. Naïve Bayes because it is easy to implement and requires a small amount of training data to make a prediction [8, 9], which is suitable for this research. RF for its suitability in handling datasets with multiple features and its ability to provide model interpretability through feature importance analysis, which aligns with the objectives of the classification model applied in this study. Gradient Boosting for its high accuracy, even for complex data. Lastly, SVM which has a good ability to model complex linear boundaries [8, 10].

Model performance was assessed using Test & Score; Area Under Curve (AUC), Classification Accuracy (CA), F1-Score, Precision (Prec), Recall, Matthew Correlation Coeffcient (MCC); Confusion Matrix; and ROC Analysis. Due to limited data availability, evaluation was conducted on training data as a preliminary benchmark.

III. Results And Discussion

During the implementation phase of this study, particularly the development of Analyze and Improve Phase were supported through discussions with key personnel, including the Production Manager, Quality Control Manager, and several senior Production Foremans.

A. DMAIC: Define

Through the VOC, three primary drivers were established: welding performance and quality, comfort, and product appearance. Based on these drivers, five Critical to Quality (CTQ) specifications were defined to serve as benchmarks for evaluating current processes and guiding subsequent analysis and improvements. t_{wet} for a minimum of 15 minutes, V_b with a range of 1.45 – 1.60 L / 10 kg of flux, a minimum P_{ext} of 130 kg/cm^2, and respectively, a minimum of 8/100 mm and 40˚C for Con_w and T_{flux}.

B. DMAIC: Measure

In this phase, data were collected over 25-days period to evaluate the current performance of the process. The total production target during this period was 500,000 kg, with the actual output reaching 495,953 kg. Here, 7,333.8 kg were classified as defective, so the total opportunity is 503,268 kg. Using these data, RTY was calculated as defined in (1). The calculation results indicate that the current performance remains slightly below the company's acceptable threshold, which contributes to an estimated 1.5% loss in output.

$$\text{RTY} = \frac{(\text{Opportunities} - \text{Defect})}{\text{Opportunities}} \times 100\% \quad (1)$$

$$\text{RTY} = \frac{(503{,}268 - 7{,}333.8)}{503{,}268} \times 100\%$$

$$\text{RTY} = 98.5\%$$

Process instability was also evidenced by control charts, which highlighted fluctuations in defect rates and deviations from optimal production parameters as seen in Fig. 1.

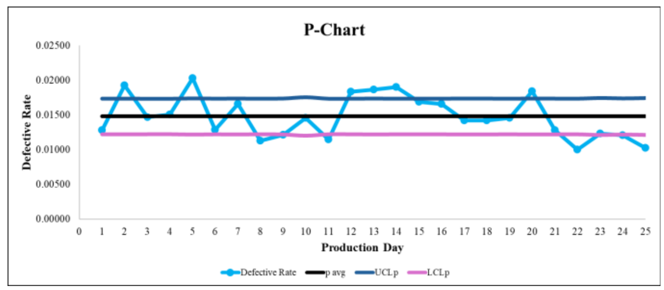

Figure 1. Control chart of current manufacturing performance

C. DMAIC: Analyze

The Analyze phase was conducted to identify the root causes of welding consumables defects through two complementary approaches: field-based investigation and data-driven analysis.

TABLE II. Current Fuzzy FMEA And RPN Results

Code	*S*	*O*	*D*	*D2*	*RPN*
D4	9	5	3	7	6.05
D5	9	5	7	9	6.83
D6	7	9	8	9	7.20
D7	9	10	9	8	8.28
D8	8	10	8	7	7.97
D9	9	7	10	8	7.71

From the field perspective, root cause identification was performed using the Fishbone Diagram and Fuzzy FMEA. The Fishbone Diagram was structured based on six factors such as man, machine, method, material, environment, and measurement applied to each defect type. The outputs from this diagram were further evaluated using Fuzzy FMEA to quantify the relative importance of each cause. Analysis of FMEA employed four parameters: serverity (S), occurrence (O), detection (D), and dependency (D2). A fuzzy logic system built in MATLAB was used to mitigate ambiguity in rating these factors and to calculate the Risk Priority Number (RPN). Using the Pareto principle, root causes with RPN values exceeding the 80% cutoff were identified as critical contributors to quality degradation. Table 2 presents root causes within the 80% Pareto threshold that are also ranked as important features according to model-derived feature importance metrics.

In parallel, a data-driven analysis was conducted using a multi-class classification approach in Orange Data Mining. This machine learning method was applied to historical production data to identify the most influential features associated with each defect category. Prior to model training, the raw data required label assignment to indicate the dominant defect type for each extrusion cycle. The labeling process was carried out based on field observations, under the assumption that each cycle typically produces one dominant defect.

Prior to evaluating the five machine learning algorithms, a resampling procedure was performed to mitigate class imbalance. Initial confusion matrix analysis revealed a skewed distribution, with minority defect classes underrepresented, potentially biasing model learning. Undersampling was employed to reduce the prevalence of majority classes, resulting in a decrease in the "Crack" and "Others" class proportions from 30.53% and 52.19% to 28.70%, respectively. This adjustment was critical to enhance model generalization and reduce classification bias. Evaluation metrics before and after resampling are presented in Table 3 and 4. As shown in Table 3, some metrics appear to have higher values. However, this increase does not indicate better model performance, but rather reflects bias introduced by the imbalanced dataset.

Among the five algorithms evaluated, RF demonstrated the highest overall performance in terms of classification accuracy and interpretability, and was therefore selected for further analysis, particularly for assessing feature importance across defect types. RF achieved the highest scores across all evaluation metrics, as shown in Table 4. It is important to note that the evaluation primarily focused on the F1-Score, as it provides a balanced measure between Prec and Recall. Moreover, the F1-Score is particularly critical in cases involving imbalanced datasets, as encountered in this study, where it offers a more objective reflection of model performance across both majority and minority classes.

TABLE III. TEST AND SCORE RESULT (BEFORE RESAMPLING)

Model	Defect Type Ranking					
	AUC	***CA***	***F1***	***Prec***	***Recall***	***MCC***
GBoost	0.915	0.771	0.762	0.775	0.771	0.618
RF	0.970	0.846	0.838	0.848	0.846	0.746
kNN	0.920	0.732	0.714	0.725	0.732	0.550
Naïves Bayes	0.725	0.519	0.529	0.544	0.519	0.253
SVM	0.702	0.490	0.484	0.484	0.490	0.178

GBoost = Gradient Boosting; RF = Random Forest; kNN = k-Nearest Neighbor; SVM = Support Vector Machine.

TABLE IV. TEST AND SCORE RESULT (AFTER RESAMPLING)

Model	Defect Type Ranking					
	AUC	***CA***	***F1***	***Prec***	***Recall***	***MCC***
GBoost	0.964	0.813	0.811	0.815	0.813	0.757
RF	0.978	0.832	0.830	0.833	0.832	0.783
kNN	0.894	0.609	0.597	0.614	0.609	0.489
Naïves Bayes	0.753	0.422	0.434	0.460	0.422	0.279
SVM	0.830	0.503	0.478	0.524	0.503	0.346

GBoost = Gradient Boosting; RF = Random Forest; kNN = k-Nearest Neighbor; SVM = Support Vector Machine.

These results confirm the suitability of RF for modeling complex defect classification problems in manufacturing. The key features influencing each type of defect classification are summarized in Table 5. Based on the analysis conducted from two complementary perspectives, Fishbone Diagram and Fuzzy FMEA from the field side, and multi-class classification from the data-driven side several defect causes, and influential production parameters were identified. To strengthen the validity of the findings, the results from both approaches were cross-referenced. As a result, the parameters that consistently emerged as critical in both analyses were selected as the primary focus for improvement initiatives, namely High P_{ext} which represents the D7 and D8 defect code from the FMEA results, ρ_{flux}, W_{rf}, respectively represent by the D9 and D6 code, and lastly, V_b with D4 and D5 defect code.

TABLE V. FEATURE IMPORTANCE RANKING

Top Features	Defect Type Ranking						
	Chp	***Conc > Std***	***Crk***	***Prs***	***Pff***	***Scth***	***Oth***
High $\boldsymbol{P_{ext}}$	1	1	2	1	4	3	1
$\boldsymbol{V_b}$	5	3	1	6	2	2	2
$\boldsymbol{W_{rf}}$	8	7	5	7	7	7	3
$\boldsymbol{\rho_{flux}}$	4	2	4	2	1	5	4

D. DMAIC: Improve

At this stage, improvement actions are designed for the four production parameters identified in the Analyze phase. The following section presents a detailed discussion of each.

- High Extrusion Pressure (P_{ext})

  The pressure was improved by defining both minimum and maximum allowable limits. This adjustment provides operators with a clearer reference range, enabling more precise control of extrusion pressure during production and reducing the likelihood of defect occurrence. Previously, the production guidelines only specified a minimum pressure, without an upper limit. However, analysis showed that excessively high pressure contributed to a higher rate of defects. As a result, a new operational range of 130–180 kg/cm² has been established and formalized within the "Production Data Sheet".

- Binder Volume (V_b)

The improper usage of binder was identified as being caused by non-standard conditions of the binder tank. Inadequate inspection and cleaning procedures led to sediment accumulation at the bottom of the tank, which is influenced by the binder's inherent property to settle when left unused for extended periods. The dosing method, which relies on surface-level height, becomes biased in the presence of such sedimentation. As a corrective action, an additional activity has been incorporated into the monthly 5S routine to clean and calibrate the binder tank, ensuring dosing accuracy and process consistency.

- Recycled Flux Quantity

In the Improve phase, corrective action was focused on enhancing the handling of recycled flux, which was previously limited due to its hardened and dried condition. The existing method, spraying water to retain moisture proved insufficient, as the flux continued to harden after being left unused for several days. To resolve this, a regeneration process was introduced involving grinding and sieving to restore the physical characteristics of the flux, making it suitable for reuse. This approach ensures a more consistent particle size and prevents performance degradation caused by poor material conditioning, thereby increasing the reusability of recycled flux.

- Flux Density

The hardness of the flux mixture affects the extrusion pressure and overall coating stability. Previously, no practical tool was available to measure flux density, resulting in inconsistent mixing quality. To solve this, a durometer was introduced to measure the hardness of the flux. The measured values were then matched with actual extrusion pressures to create a reference guide. This guide helps operators achieve more consistent flux mixing and pressure control. Durometers are now provided on every production line.

TABLE VI. NEW FUZZY FMEA AND RPN RESULTS

Code	*S*	*O*	*D*	*D2*	*RPN*
D4	9	1	2	7	5.01
D5	9	1	5	9	5.61
D6	7	3	5	9	6.05
D7	9	5	7	8	6.45
D8	8	5	6	7	6.05
D9	9	4	9	8	6.83

The effectiveness of the implemented improvements is demonstrated through the revised RTY, updated FMEA, and control chart. For the new RTY, data were collected over 16 days period to assess the performance of the improved manufacturing process. During this time, the total production target was 342,400 kg, with an actual output of 320,542 kg and 4,330.2 kg classified as defective. This resulted in a total opportunity of 324,872.2 kg. The calculation below results in a 0.2% reduction in the proportion of product defects relative to total output and encouraging improvement.

$$\text{RTY} = \frac{(320{,}542 - 4{,}330.2)}{320{,}542} \times 100\%$$

$$\text{RTY} = 98.7\%$$

Table 7 highlights several defect causes that were previously selected, along with their new RPN values after the improvement actions were applied. Fig. 2 presents the new control chart.

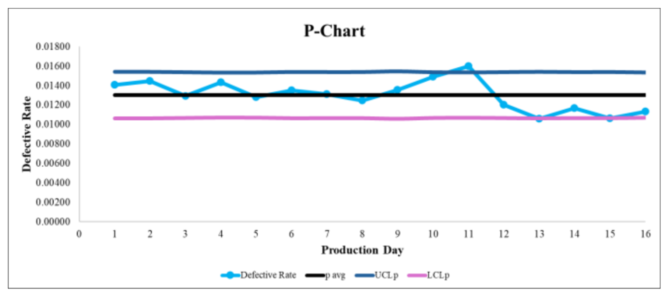

Figure 2. Control chart of improved manufacturing performance

E. DMAIC: Control

To sustain the improvements implemented, a structured control mechanism was introduced under the oversight of the Production Foreman. Daily monitoring activities include recording extrusion pressure values in the "Production Process Report" and checking real-time pressure readings directly from the extruder gauge during operation. Binder tank conditions are inspected at the start of each production day, alongside periodic measurements of its volume to verify the dosing system accuracy. A visual inspection is performed on the physical condition of the recycled flux prior to its use in the Wet Mixing process to ensure it meets safety and quality standards. The flux mixture's hardness is assessed using a durometer after each wet mixing cycle, with all control data systematically recorded and reviewed for consistency. These measures are intended to maintain process stability, preserve product quality, and reinforce the sustainable application of improvements.

IV. CONCLUSION

The implementation of the DMAIC methodology to improve welding electrode quality yielded several significant outcomes. In the Analyze phase, four production parameters were identified as having the most substantial impact on product quality: High P_{ext}, V_b, W_{rf}, and ρ_{flux}. These parameters were determined by aligning insights from both field-based analysis (Fuzzy FMEA) and data-driven analysis (Random Forest feature importance), which showed strong correlation between the two perspectives.

Several targeted improvements were introduced: a defined extrusion pressure range (130–180 kg/cm²) was added to the Production Data Sheet; binder tank inspection and cleaning were integrated into the monthly 5S activities; a revised handling method for recycled flux was proposed; and durometers were deployed to standardize flux hardness evaluation. ML, implemented through a multi-class classification model, proved highly effective, objective, and consistent in accelerating root cause analysis. Among the algorithms evaluated, RF delivered the best performance, with an AUC of 0.978, CA of 0.832, and F1 of 0.830. Its ROC curves also confirmed its robustness across most defect types. Despite challenges with class imbalance in the dataset arising from naturally skewed defect distributions, undersampling was successfully applied to reduce training bias.

Overall, the improvements led to measurable success. The RTY increased, with the defective contribution to production output reduced by 0.2% (from 1.5% to 1.3%). FMEA updated showed a reduction in high-risk defect codes, and control charts more stable defect rates, with data points predominantly remaining within control limits. The results confirm that integrating DMAIC with machine learning effectively improves quality in welding consumables production. Future work may extend this method to other stick electrode types and address additional root causes to further enhance results.

Acknowledgment

The author would like to express sincere gratitude to the Production Manager, Quality Control Manager, senior Production Foremans, at a leading welding consumables manufacturer in Indonesia for their valuable support, discussions, and contributions during the research and implementation process.

Data Availability

The data used for this research can be accessed through the following link: DATASET.xlsx.

Author Contribution

The following statements were used: Conceptualization, E.A., C.H.; methodology, E.A., data collection, E.A.; experiments, E.A.; analysis, E.A.; writing, E.A.; review and supervision, C.H. All authors have read and agreed to the published version of the manuscript.

References

[1] E. Lodgaard and D. Powell, "The changing role of shop-floor operators in zero defect manufacturing," *Procedia CIRP*, vol. 104, pp. 594–599, 2021. [Online]. Available: https://doi.org/10.1016/j.procir.2021.11.100.

[2] A. Palange and P. Dhatrak, "Lean manufacturing: a vital tool to enhance productivity in manufacturing," *Materials Today: Proceedings*, vol. 46, pp. 729–736, 2021. [Online]. Available: https://doi.org/10.1016/j.matpr.2020.12.193.

[3] H. Zhang, R. Pan, F. Chang, L. He, Z. Dong, and J. Yang, "Zero-DD: Zero-sample defect detection for industrial products," Computers and Electrical Engineering, vol. 105, p. 108516, 2023. [Online]. Available: https://doi.org/10.1016/j.compeleceng.2022.108516.

[4] A. A. Hidayat, M. Kholil, J. Haekal, N. A. Ayuni, and T. Widodo, "Lean manufacturing integration in reducing the number of defects in the finish grinding disk brake with DMAIC and FMEA methods in the automotive sub-industry company," *International Journal of Scientific Advances*, vol. 2, no. 5, pp. 713–718, 2021. [Online]. Available: https://doi.org/10.51542/ijscia.v2i5.7.

[5] I. T. B. Widiwati, S. D. Liman, and F. Nurprihatin, "The implementation of Lean Six Sigma approach to minimize waste at a food manufacturing industry," *Journal of Engineering Research*, 2024. [Online]. Available: https://doi.org/10.1016/j.jer.2024.01.022.

[6] M. M. Hassan, Y. Xu, J. Sayada, M. Zareef, M. Shoaib, X. Chen, H. Li, and Q. Chen, "Progress of machine learning-based biosensors for the monitoring of food safety: A review," *Biosensors and Bioelectronics*, vol. 267, p. 116782, 2025. [Online]. Available: https://doi.org/10.1016/j.bios.2024.116782.

[7] V. Görüş, M. M. Bahşı, and M. Çevik, "Machine learning for the prediction of problems in steel tube bending process," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108584, 2024. [Online]. Available: https://doi.org/10.1016/j.engappai.2024.108584.

[8] A. Subasi, *Practical Machine Learning for Data Analysis Using Python*. Cambridge, MA, USA: Academic Press, 2020.

[9] A. Tella, A.-L. Balogun, N. Adebisi, and S. Abdullah, "Spatial assessment of PM10 hotspots using random forest, K-nearest neighbour and Naïve Bayes," *Atmospheric Pollution Research*, vol. 12, p. 101202, 2021. [Online]. Available: https://doi.org/10.1016/j.apr.2021.101202.

[10] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short-term memory algorithms in machine learning," *Decision Analytics Journal*, vol. 3, p. 100071, 2022. [Online]. Available: https://doi.org/10.1016/j.dajour.2022.100071.

[11] S. Kumar, G. Brahma, P. Kaur, and B. Kumar Gupta, "A case study on quality improvement of connecting rod assembly line using Six Sigma DMAIC cycle," *International Journal of Six Sigma and Competitive Advantage*, vol. 12, no. 3, 2020.

[12] G. C. P. Condé, P. C. Oprime, M. L. Pimenta, J. E. Sordan, and C. R. Bueno, "Defect reduction using DMAIC and Lean Six Sigma: a case study in a manufacturing car parts supplier," *International Journal of Quality and Reliability Management*, vol. 40, no. 9, pp. 2184–2204, 2023. [Online]. Available: https://doi.org/10.1108/IJQRM-05-2022-0157.

[13] A. Mittal, P. Gupta, V. Kumar, A. Al Owad, S. Mahlawat, and S. Singh, "The performance improvement analysis using Six Sigma DMAIC methodology: A case study on Indian manufacturing company," *Heliyon*, vol. 9, no. 3, p. e14625, 2023. [Online]. Available: https://doi.org/10.1016/j.heliyon.2023.e14625.

[14] P. H. Fan, "Applying six sigma to improve the defect rate of electronic components: a six-sigma case study," *South African Journal of Industrial Engineering*, vol. 35, no. 1, pp. 41–56, 2024.

[15] R. Yahmadi, K. Brik, and F. Ben Ammar, "Fuzzy risk priority number assessment for solar gel battery manufacturing defects," *Engineering Failure Analysis*, vol. 124, p. 105327, 2021. [Online]. Available: https://doi.org/10.1016/j.engfailanal.2021.105327.

[16] S. J. Basha, S. R. Madala, K. Vivek, E. S. Kumar, and T. Ammannamma, "A review on imbalanced data classification techniques," in *Proc. 2022 Int. Conf. Advanced Computing Technologies and Applications (ICACTA)*, 2022, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ICACTA54488.2022.9753392.

[17] K. Guleria, S. Sharma, S. Kumar, and S. Tiwari, "Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning," *Measurement: Sensors*, vol. 24, p. 100482, 2022. [Online]. Available: https://doi.org/10.1016/j.measen.2022.100482.