Opsi p-ISSN 1693-2102
Vol. 18, No. 1, June 2025 e-ISSN 2686-2352

Proposed improvement of product support packaging material defects using the Cross-Industry Standard Process for Data Mining (CRISP-DM) approach

Gina Almas Nabiha [1], Rina Fitriana [1*], Anik Nur Habyba [1], Sannia Mareta [2]
[1]Department of Indutrial Engineering, Faculty of Industrial Technology, Universitas Trisakti, Jln. Letjen S. Parman No. 1 Kampus A, Grogol Petamburan, Jakarta Barat, Indonesia 11440
[2] Faculty of Science and Engineering, University of Nottingham Ningbo, 199 Taikang East Road, Yinzhou District Ningbo, Zhejiang Province, China 315100

*Corresponding Author: rinaf@trisakti.ac.id

Article history:
Received: 16 March 2024
Revised: 23 June 2025
Accepted: 26 June 2025
Published: 30 June 2025

Keywords:
Cross-industry standard process for data mining (CRISP-DM)
Decision tree
Power business intelligence
Statistical process control
Fault tree analysis

ABSTRACT
This research was conducted because the defect rate of packaging materials supporting lithos M products exceeded the Company's tolerance standard of 2%. This research aims to identify the causes and provide suggestions to improve the Quality of product support packaging materials. The methods used in data mining with the CRISP-DM (Cross-Industry Standard Process For Data Mining) approach. The Business Understanding stage determines the problem and research objectives, Power Business Intelligence, SIPOC (Supplier, Input, Process, Output, Customer) Diagrams, Operation Process Chart, QC Action, and CTQ (Critical to Quality). The Data Understanding stage creates a Control P Chart, calculates DPMO and the sigma level obtained by the unscramble machine dented bottle value 762.31 with a Sigma level of 4.66, Sticker 2nd defect Internal 187.47 with a sigma level of 5.06, Cap 2nd defect internal 67.18 with a sigma level of 5.32, and uses Fault Tree Analysis. The Data Preparation stage performs data cleaning, integration, transformation, and preprocessing. The Modelling stage makes classification with C4.5 and the Cart decision tree algorithm. The evaluation stage uses a Confusion Matrix accuracy of 78.8 percent and 89.4 percent, respectively. The Deployment stage produces improvement proposals by creating a Dashboard, Standard Operating Procedure, and Check Sheet.

DOI:
https://doi.org/10.31315/opsi.v18i1.11803

1. INTRODUCTION

The Company can survive in a competitive environment by conducting Quality Control of its products [1]. Quality Control is used to control, select, and assess Quality to produce quality products and reduce defective production [2]. In implementing Quality Control, the ISO (International Organization for Standardization) is needed; this standard is used to achieve quality objectives in implementing quality standards. High-value products can meet customer needs by having good Quality, which increases customer satisfaction and company profits [3].

Quality Control is the activity of maintaining, supervising, and ensuring product quality standards. Defective products occur due to inefficient material handling activities. Material handling involves moving materials from one point to another. An essential aspect of producing a product is the Quality of raw materials, which can affect the final product. Product and service quality is the ability of performance characteristics, features, reliability, serviceability, conformance to standards, durability, perceived Quality, and aesthetics to meet customer demand.

The PT. P is a lubricant industry company that produces lubricants in bottles (lithos), drums, pails, and other industrial needs. The PT. P has two warehouses: the supporting packaging material warehouse and the finished product warehouse. The supporting packaging material warehouse contains supporting packaging materials such as bottles, cartons, label stickers, IBC, and drums received from suppliers, while the finished product warehouse is where the final product is packaged in unit loads with pallets. The PT P's business process consists of receiving, stockpiling, and filling into packaging. Based on the results of interviews with The PT. P, two types of reject material were found: incoming and in-process. The type of reject material is supporting packaging material (bottle, carton, label sticker, capper). Material rejected incoming is material that comes from suppliers, not by specifications, so the material is rejected. In contrast, the material reject process is material that is damaged due to unloading in the material warehouse until the lubricant production process in the production line area, which results in cost losses. Due to the large number of in-process reject materials that occur in the production process, it is necessary to research the PT P to discover the causes of in-process reject materials.

Research was conducted using a Pareto diagram to determine the highest value, which is the problem to be solved first, and using Fault Tree Analysis to find out the cause of the problem [4]. Furthermore, research uses data mining methods with decision tree classification to obtain if-then rule results [5].

There was much research about quality and data mining. An essential component of groundwater resource management is the use of data mining algorithms (DMAs) to model groundwater quality in coastal areas [6]. The decision tree mining technique was applied to construct the water conservation informatization data sharing system since the conventional sharing system's low-quality mining dataset hinders real-time sharing and system scalability [7]. Through a systematic review and subsequent categorization of each of the CRISP-DM steps, this research presented a didactic and utilitarian model based on the most recent developments in the literature, through the lens of production engineering, and suggested guidelines from the fields of quality management and risk management [8]. Understanding the intricate interdependencies in battery cell manufacture and identifying areas for improvement are essential for ensuring optimal cell quality and enabling quality control and established data mining techniques in the context of intricate operations and data structures [9]. The findings of the study offer a fast and precise way to classify quality reports, which aids in the development of the engineering quality knowledge system [10]. The study on the best kernel function forecasting model was conducted by predicting ground rod sales using the Support Vector Regression (SVR) method, which was displayed in an ideal data visualization [11]. The study used the Support Vector Regression (SVR) method to model the sales experience, analyze the accuracy of predictions, and create a dashboard of prediction results using Power Business Intelligence (Power BI) software [12]. The SERVQUAL method at Soekarno-Hatta International Airport was the subject of a study that failed to achieve passenger satisfaction standards and the findings of the Binary Logistic Regression analysis show that passenger satisfaction is influenced by parking price and ambiance [13], The results of this study, which used the DMAIC (Define–Measure–Analyze–Improve–Control) Six Sigma approach, demonstrated that the jacket product's quality was higher than the industry average in Indonesia and was categorized as the industry average in the USA [14], It is essential to perform quality control on raw material inventory loss using the control chart, a fishbone diagram and Root Cause Analysis (RCA) to enhance quality control in the raw material unloading process [15]. The findings of the study on six sigma DMAIC are the sigma level for the variable data and the attribute data. The tin ingot test failure type has the highest RPN score [16]. It is essential to perform quality control on raw material inventory loss using the control chart as well as to analyze with a fishbone diagram and Root Cause Analysis (RCA) [17].

This research uses the CRISP-DM approach, decision tree data mining, Statistical Quality Control and Power Business Intelligence. CRISP-DM (Cross Industry Standard Process For Data Mining) is a method that supports data processing in modeling cases with diverse structures [18], The Cross Industry Standard Process for Data Mining (CRISP-DM) has grown to be one of the most popular in the industry [8]. CRISP-DM consists of six stages: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, Testing, and Deployment [19], [20]. Statistical Process Control is a statistical technique for monitoring measurement standards and taking action for product improvement. Statistical Process Control has seven tools: Check Sheet, Scatter diagram, parts diagram, fishbone diagram, flow chart, histogram, and Control Chart. In addition, this research uses Power Business Intelligence for data visualization. Power Business intelligence is a tool for analyzing data visual [21]. In data mining for modeling, the Decision tree classification method is used. Decision tree is a prediction model used for classification and prediction tasks. A decision tree is a diagram with a tree structure [22].

2. MATERIALS AND METHODS

The method in this research is to collect data from research problems. The data collected are primary data and secondary data. After all the data is collected, the next step is to process the data from the problem using the CRISP-DM approach. Figure 1 shows the stages of data processing with the CRISP-DM approach.

The research methodology uses the Data Mining framework with CRISP-DM, which consists of Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. The approach aims to find out the cause of the problem and provide suggestions for packaging quality. At the Business Understanding stage, research problems were identified using SIPOC diagrams, operation process charts, QC Action in each production process, and Critical Quality to find problems with Pareto diagrams. The data Understanding stage was used to see the stability of the production process with a control p chart, calculate the sigma level value, and find the cause of packaging defects using fault tree analysis. The Data Preparation stage prepares data before the modeling stage with data integration, transformation, reduction, and cleaning. The modeling stage was used to create a decision tree with the C4.5 algorithm with R studio and the Cart algorithm with Minitab, which produces an if-then rule for the number of defects, low, high, and medium. The evaluation stage was used to see the Accuracy of the results in forming the decision tree model. Deployment stage to provide improvement proposals by creating a Dashboard, Standard Operating Procedure, and Check Sheet.

3. RESULTS AND DISCUSSION

3.1. Business Understanding

Business Understanding is the stage used to determine the objectives to be achieved in this research [23]. The business aims to analyze the causes of defects in packaging materials supporting lithos M products and provide suggestions for improving the Quality of packaging materials supporting lithos M products. Data Mining aims to explore the low, high, or medium defects pattern and find the if-then rule.

This stage is to describe the resources and limitations that exist at PT P .The following are the resources and limitations that PT P has:

Resources from Research:

- The data collected is Lithos M products' in-process reject material data in June, July, August, September, and October 2023.
- Data resources are stored in a database owned at PT P.

Limitation of Research:

- The result of the decision tree is knowing the if-then rule of the number of low, high, or medium defects.
- If there is incomplete data, the data will be processed through data preparation.
- Limitations in the deployment stage are only limited to suggestions because the decision is in the Company determined by PT P.

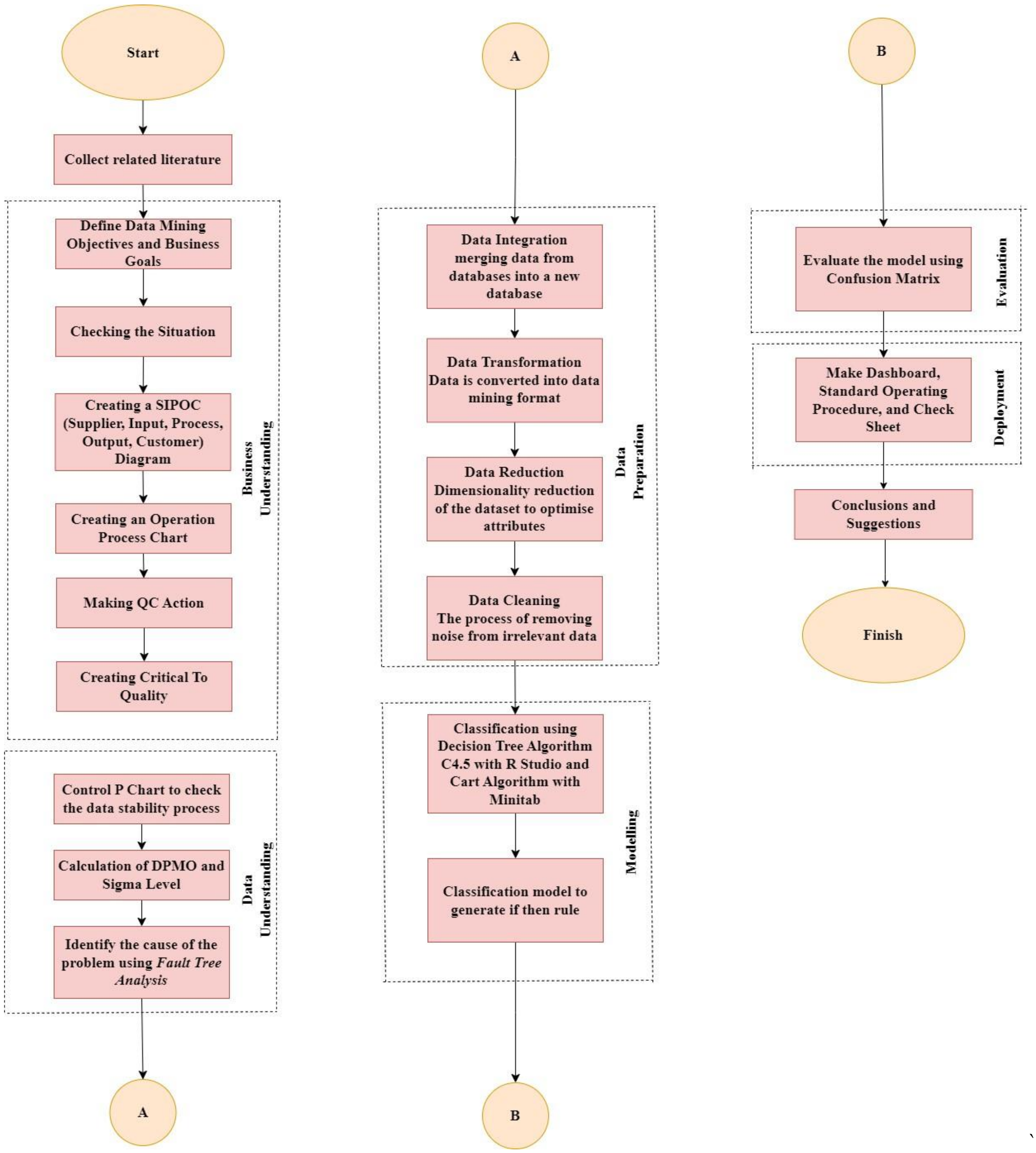


Figure 1. The Flowchart data processing

3.1.1. SIPOC Diagram

The SIPOC diagram is a diagram that describes the production process activities in detail [24]. The SIPOC diagram is made from the prepared raw materials to the finished product, ready to be sent to the customer. The SIPOC diagram of M lithos product. In Figure 2, the PT P Supplier consists of suppliers of bottles, cappers, cartons, and sticker labels. After that, there are Bottle, Capper, Carton, Sticker Label inputs to enter the production process with the stages of unscramble machine, orienter, labelling, filler, capper machine and induction seal, laser batch, robotic, carton erector, weigher, carton sealer, marking boxes, palletizer with the output in the form of product lithos m and cutomer in the form of distributors.

Figure 2 shows a spot diagram consisting of the supplier, input, process, output, and customer. Supplier The PT P consists of bottle suppliers with “PT Abadi Plastik”, “PT Bumi Mulia Indah Lestari (BIL)”, “PT Lyhock Batavia Plastik”, “PT Karlina MP”, “PT Usaha Bersama Sukses”. Capper supplier with "PT Dinito

Jaya Sakti". Carton supplier with "PT Multi Box Indah", "PT Intan Ustrix", "PT Cakra Walam Mega Indah (CMI)". Sticker label suppliers with "PT Surya Baru", "PT Primasindo MK", "PT Satia Mitra LP", "PT Subuer Berkah", "PT Anugerah Prima Printing", "PT Master Lebel", "PT Aneka Rupa Tera". The inputs are bottle, capper, carton, and sticker label. Next, there is a process consisting of an unscramble machine, orienter, labeling, filler, capper machine and induction sealer, laser batch, divider, robotic, carton erector, weigher, carton sealer, marking box (batch carton), palletizer. Furthermore, the output is product lithos M, and the customer is the distributor.

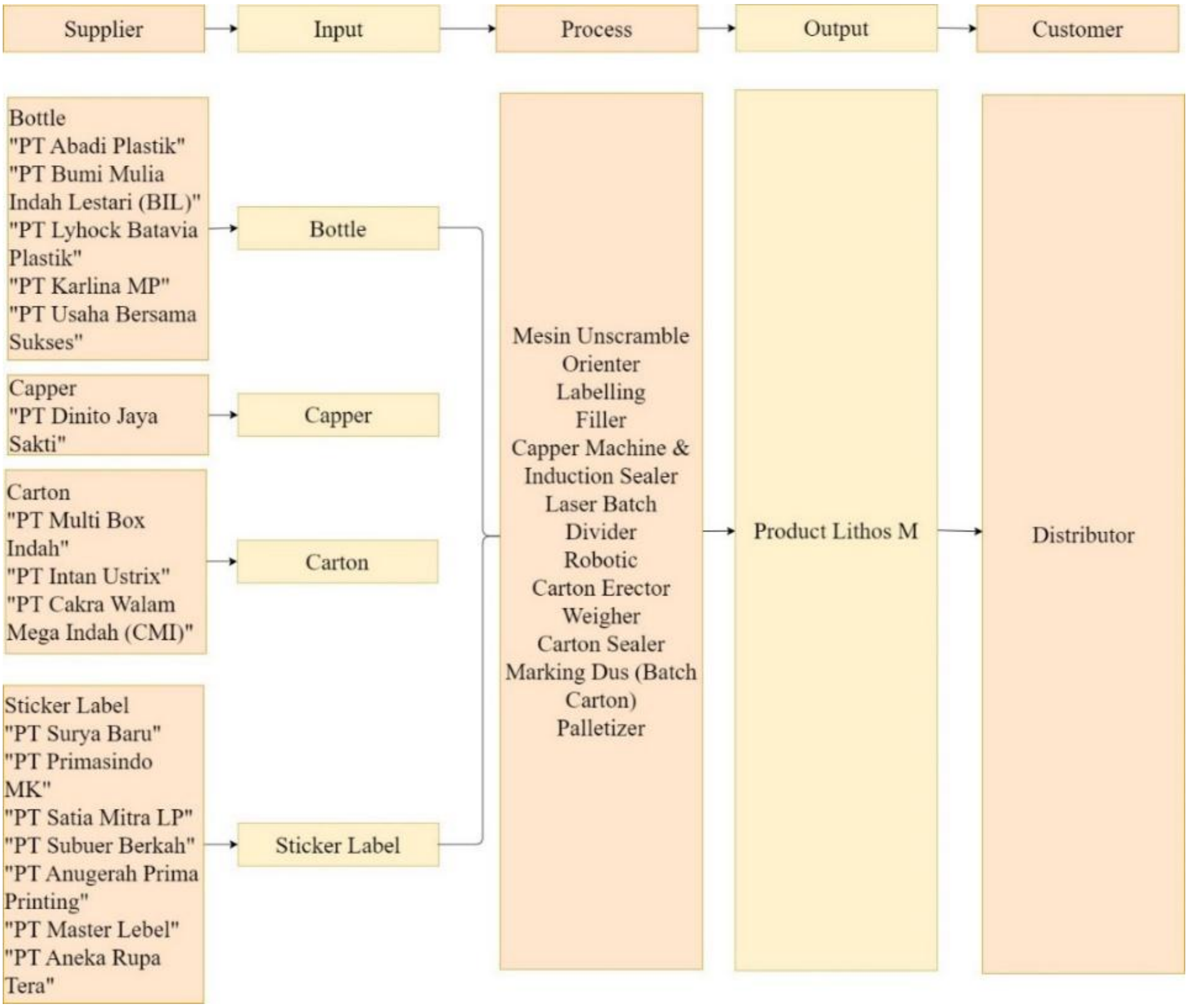


Figure 2. The SIPOC diagram

In Figure 3 is a chart of the operation process consisting of the first operation, namely the unscramble machine which performs the process of supplying and arranging lithos bottles, the second operation is the orienter machine to direct the same bottle position, the third operation is the labelling machine to attach the sticker to the bottle, the fourth operation is the filler machine to put the lubricant into the bottle, the fifth operation is the capper & induction seal machine to attach the capper to the bottle, the sixth operation is the laser batch machine to attach the batch code to the bottle cap, The seventh operation is a divider machine to divide the bottle position, the eighth operation is a robotic machine to put the bottle into the carton, the ninth operation is a carton erector machine to form the carton, the tenth operation is a weigher machine to weigh the weight of the bottle and carton, the eleventh operation is a carton sealer machine to seal the carton, the twelfth operation is a carton batch or box marking machine to attach the batch code to the carton, the thirteenth operation is a palletizer machine to arrange the finished products [25].

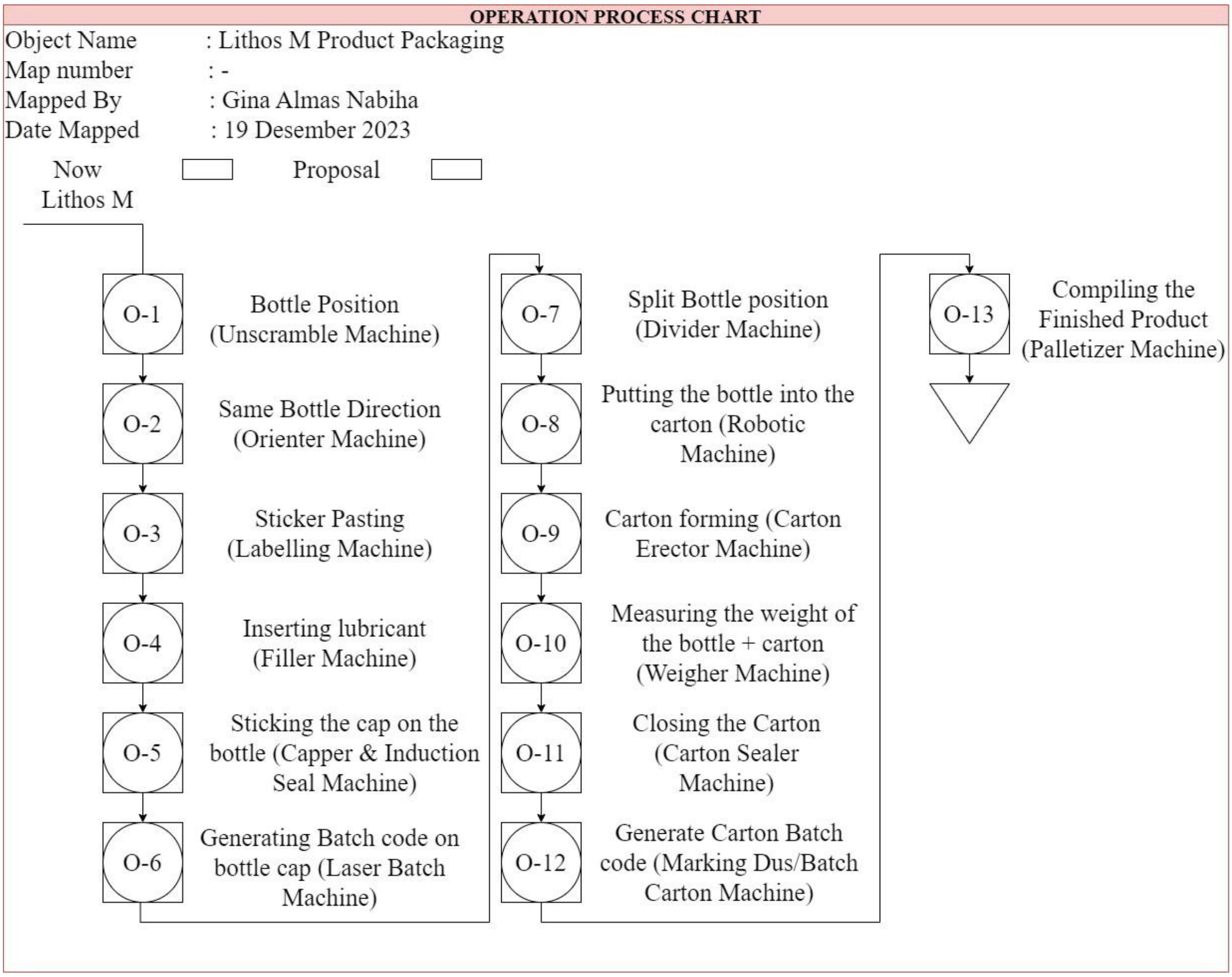


Figure 3. The operation process chart

Figure 4 shows the QC actions of unscrambling, labelling, filling, and capping. In unscramble and labelling there are actions, namely unscramble does not make the bottle defective, the direction of the bottle forward, the condition of the inside and outside of the bottle is clean and dry, the suitability of the product with the sticker used on both sides of the bottle, not tilted and not torn. In filling & cappering, the existing actions are writing the weight of the lubricant in the bottle, the surface of the lubricant tried to be above the tera line, writing the temperature of the lubricant at the time of filling a maximum of 40 degrees Celsius, the appearance/color of the lubricant product by the specifications carried out per 30 minutes, the suitability of the sticker, the suitability of the capper with the bottle.

Figure 5 shows The QC actions for the induction sealer and laser maker carton erector. In the induction sealer and laser maker, some actions show the outer condition of the bottle is clean and dry; the airfoil is tight, the Batch number is precise, aligned, and not tilted, and the oriented and divider machines are working correctly. In the carton erector, the actions include the condition of the carton in a clean and dry state, the suitability of the type and size of the carton with the product being produced, the temperature of the glue machine according to 180 degrees centigrade, and the number and arrangement of bottles in doos according to the service tag.

In Figure 6, there are The QC actions for carton sealer & printing doos, weighing & palletizing. In carton sealer & printing doos, there are QC actions, namely the temperature of the glue machine is 180 degrees Celsius, the configuration and spelling of the printing of the product name and batch number are appropriate, the printing results are legible and not skewed, the suitability of printing doos on the carton. In weighing and palletizing, there are The QC actions, namely writing the weight per box according to the range and no product exceeds the pallet surface, the arrangement of doos on the pallet is by the provisions, and the topmost arrangement is tied.

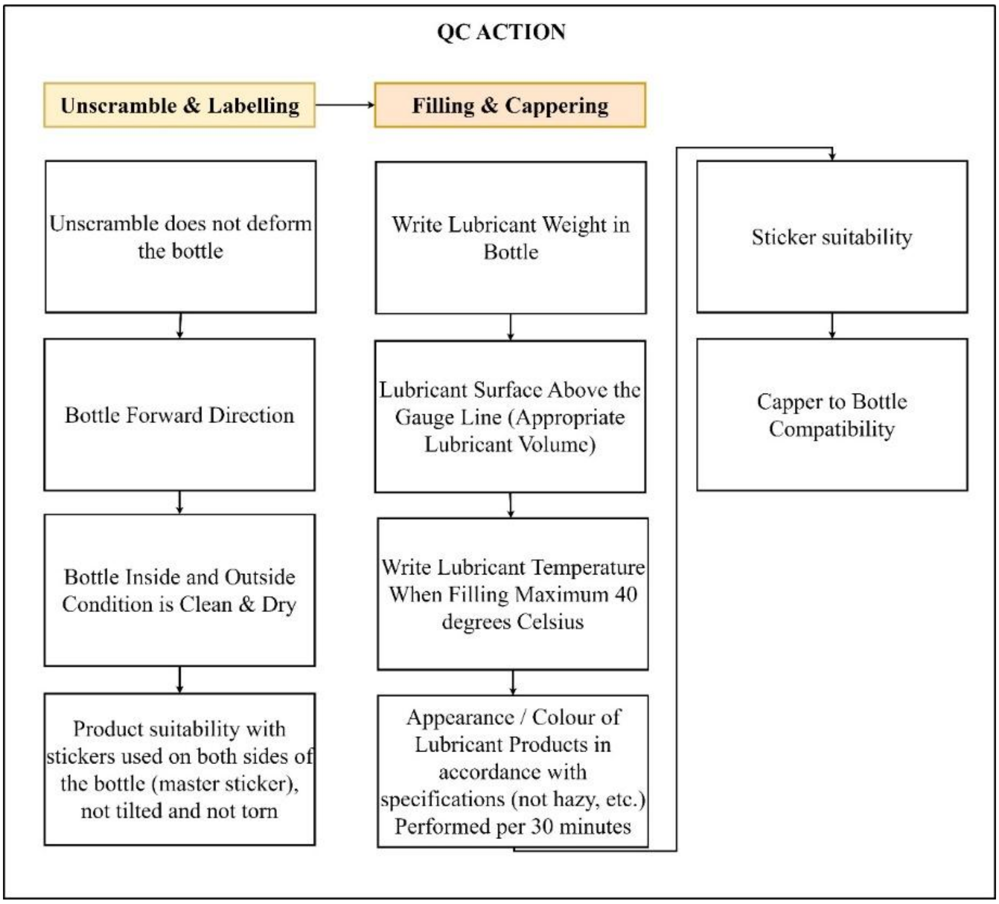


Figure 4. The QC action unscramble, labelling, filling & capping

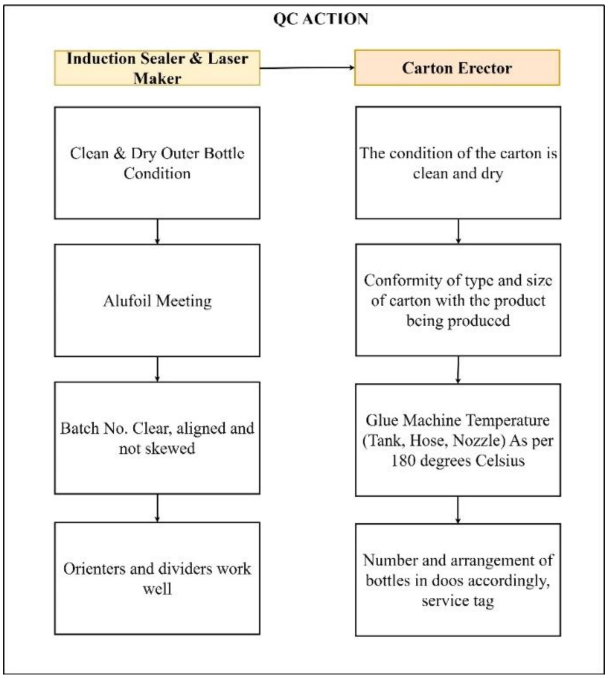


Figure 5. The QC action induction seal & laser maker, carton erector

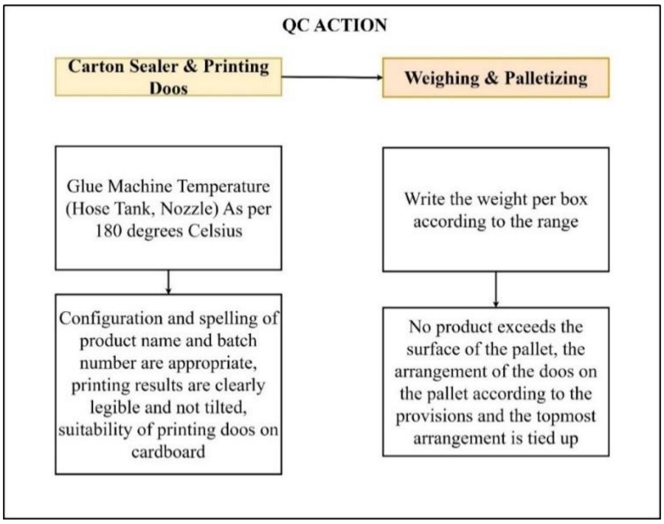


Figure 6. The QC action carton sealer & printing doos, weighing & palletizing

3.1.2. Critical To Quality

Critical to Quality is a quality characteristic that results in product conformance to specifications to improve customer satisfaction. The CTQ identification process determines the desired product quality characteristics categorized as defects. Quality characteristics of lithos M lubricant packaging products. A Pareto diagram determines the category level of an event that is most important and impactful [26]. From Figure 7, the three defects with the highest percentage of reject material are the bottle dent unscramble machine, Sticker 2nd Defect Internal, and Cap 2nd Defect Internal.

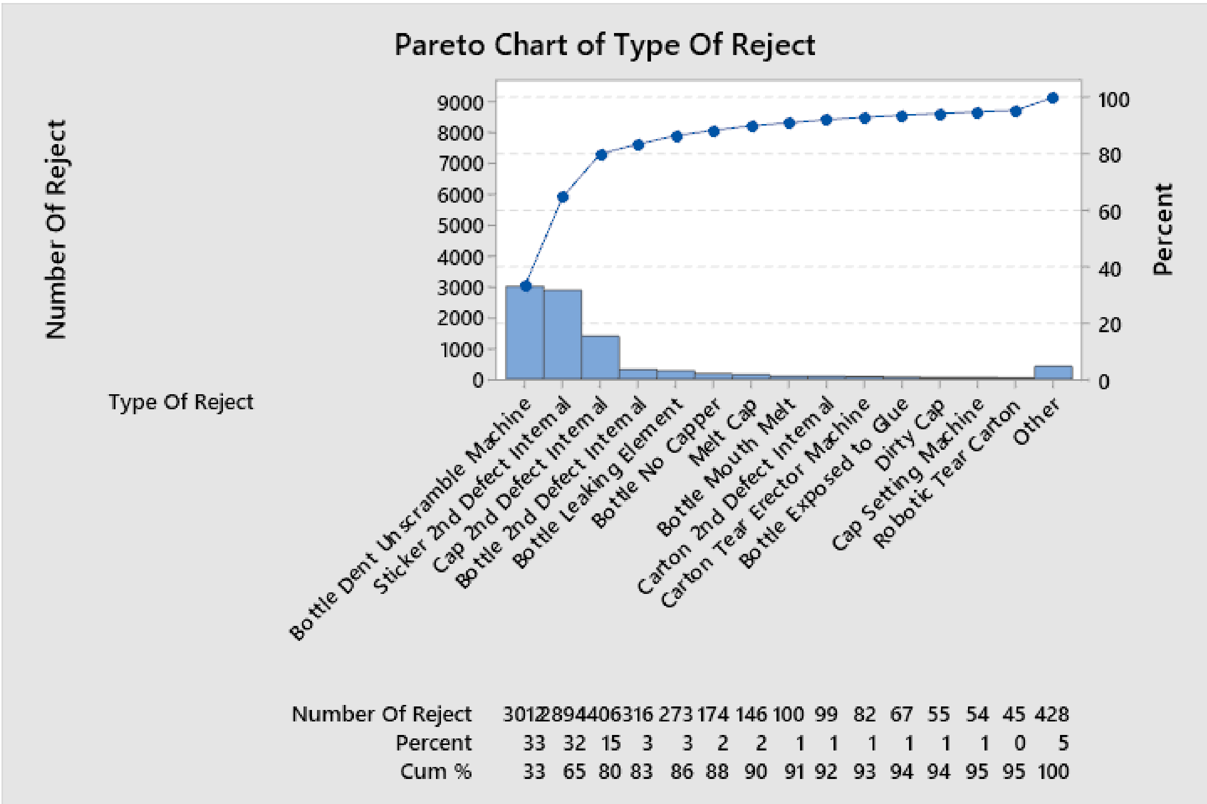


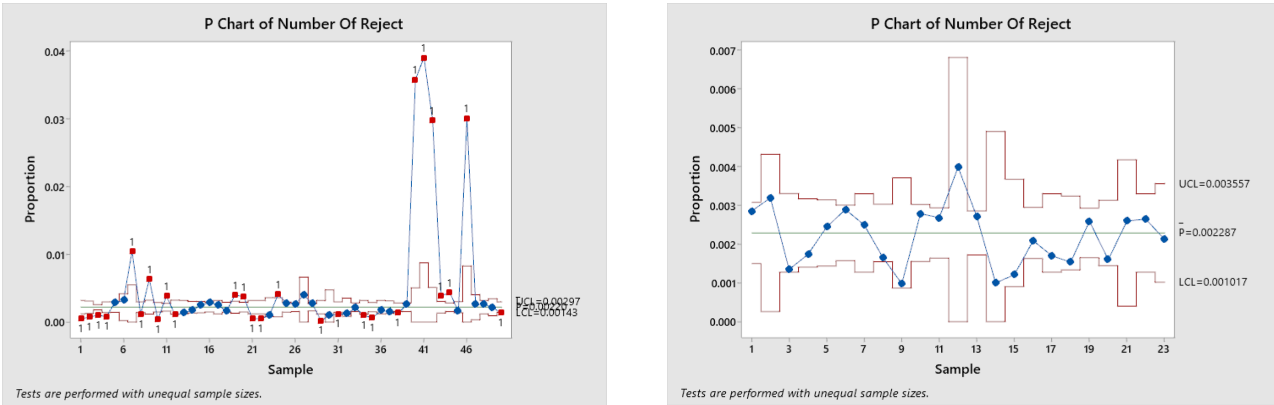
Figure 7. The Percent reject material lithos M

3.2. Data Understanding

Data Understanding is the stage of understanding data. Data recognition is essential for the research process. At this stage, the production data set is used to see if the production process is within the control limits using the control P chart. The Control P Chart diagram was created using in-process reject material data for June, July, August, September, and October 2023.

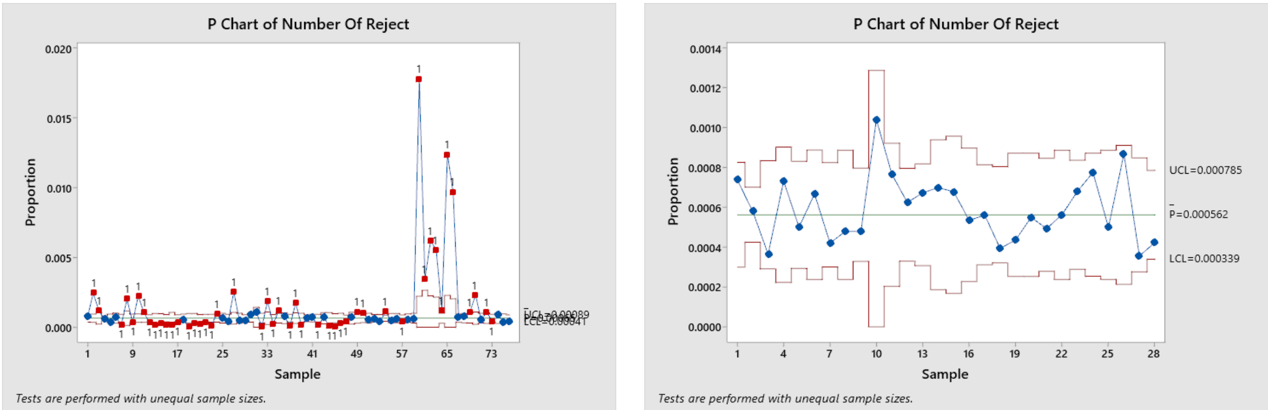
3.2.1. Control P Chart

A Control chart is a Statistical Process Control tool used in the quality control process [27]–[29]. In this study, due to attribute data, Figure 8-10 used the P chart. PT P has a data record of material reject products with attribute defect categories. The control chart is based on the proportion of product defects from June to October 2023.



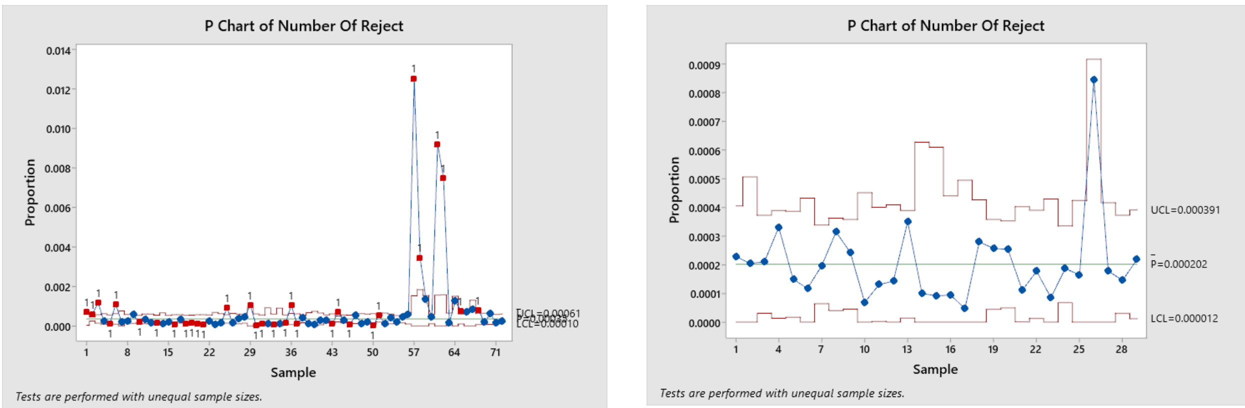
(a) Before (b) After

Figure 8. The P chart for the bottle dent unscramble machine



(a) Before (b) After

Figure 9. The P chart for the 2nd sticker defect internal



(a) Before (b) After

Figure 10. The P chart for the 2nd cap defect internal

- Bottle Dent Unscramble Machine

$$\text{DPO} = \frac{\text{number of defects}}{\text{production quantity x number of defect opportunities}} \tag{1}$$

$$= \frac{1360}{594680 \text{ x } 3} = \frac{1360}{1784040} = 0.00076231$$

DPMO = DPO x 1.000.000
$= 0.00076231 \text{ x } 1000000 = 762.31$

Sigma Level =NORMSINV $\frac{(1.000.000-\text{DPMO})}{1.000.000} + 1.5 = \text{NORMSINV } \frac{(1.000.000-762.31)}{1.000.000} + 1.5 = 4.66$

- Sticker 2nd Defect Internal

$$DPO = \frac{1052}{1870473 \text{ x } 3} = \frac{1052}{5611419} = 0.00018747$$

DPMO = DPO x 1.000.000
$= 0.00018747 \text{ x } 1000000 = 187.47$

Sigma Level =NORMSINV $\frac{(1.000.000-\text{DPMO})}{1.000.000} + 1.5 = \text{NORMSINV } \frac{(1.000.000-187.47)}{1.000.000} + 1.5 = 5.06$

- Cap 2nd Defect Internal

$$DPO = \frac{274}{1359494 \text{ x } 3} = \frac{274}{4078482} = 0.00006718$$

DPMO = DPO x 1.000.000
$= 0.00006718 \text{ x } 1000000 = 67.18$

Sigma Level=NORMSINV $\frac{(1.000.000-\text{DPMO})}{1.000.000} + 1.5 = \text{NORMSINV } \frac{(1.000.000-67.18)}{1.000.000} + 1.5 = 5.32$

3.2.2. Fault Tree Analysis

Fault tree analysis (FTA) aims to identify and know the causes of material defects. Figure 11 and Figure 12 identify potential failures with a top-down approach, top failures (top events), and essential events (essential events) [30], [31].

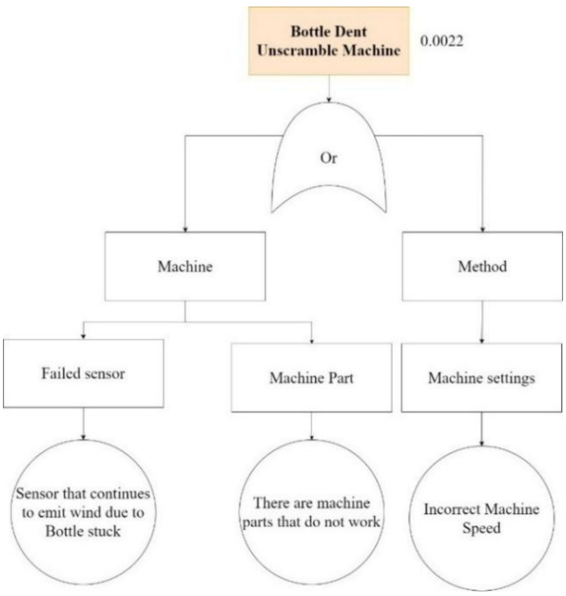


Figure 11. The fault tree analysis of bottle bottle-dent unscramble machine

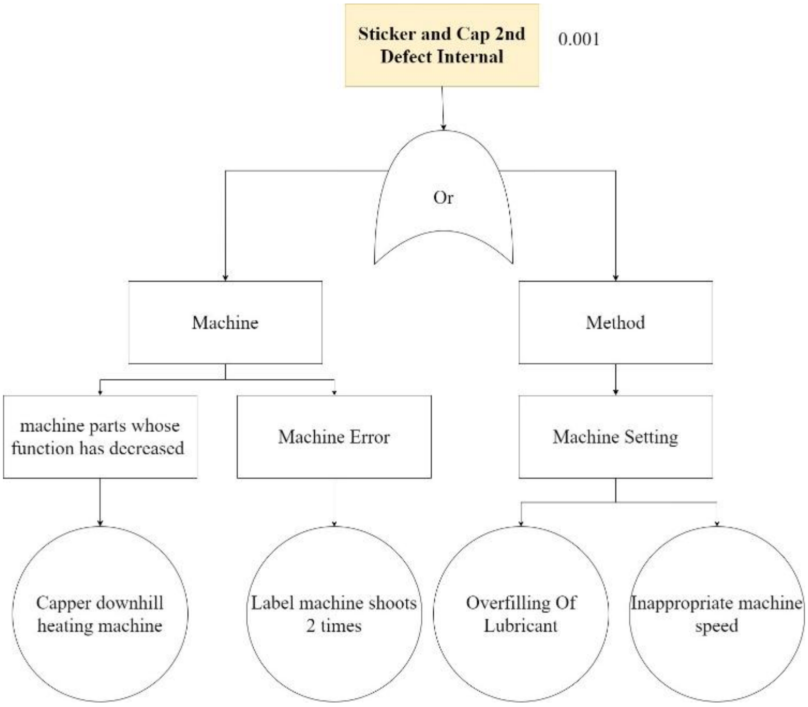


Figure 12. The fault tree analysis of sticker and cap 2nd defect internal

3.3. Data Preparation

Table 1 is the in-process reject material data carried out through data integration by combining in-process reject material data from June to October 2023. The data consists of date, month, year, number of rejects, and net usage. The number of rejects is the amount of packaging material supporting rejected lithos M products. In contrast, net usage is the amount of packaging material supporting accepted lithos M products.

Table 1. The first data

Number	In Process Date	Type Of Reject	Number Of Reject	Total Net Usage
1	8-Jun	Bottle Dent Unscramble Machine	120	32440
2	8-Jun	Bottle Dent Unscramble Machine	15	100000
3	8-Jun	Cap 2nd Defect Internal	10	232440
4	8-Jun	Sticker 2nd Defect Internal	10	232440
5	5-Jun	Bottle Dent Unscramble Machine	10	20000
6	5-Jun	Sticker 2nd Defect Internal	22	45480
7	5-Jun	Sticker 2nd Defect Internal	32	27580
8	6-Jun	Sticker 2nd Defect Internal	114	31420
9	6-Jun	Sticker 2nd Defect Internal	45	33280
10	7-Jun	Bottle Dent Unscramble Machine	10	20000
331	5-Oct	Cap 2nd Defect Internal	20	2155
332	5-Oct	Sticker 2nd Defect Internal	27	2155
333	11-Oct	Cap 2nd Defect Internal	16	2118
334	11-Oct	Sticker 2nd Defect Internal	20	2118
335	11-Oct	Sticker 2nd Defect Internal	9	842
336	16-Oct	Cap 2nd Defect Internal	3	2401
337	16-Oct	Sticker 2nd Defect Internal	21	2401
338	17-Oct	Cap 2nd Defect Internal	32	1512
339	17-Oct	Sticker 2nd Defect Internal	42	1512
340	18-Oct	Bottle Dent Unscramble Machine	16	515

Table 2 is the in-process reject material data, which is calculated to find the value of the production amount by adding the number of rejects to the amount of net usage. Table 3 is the in-process reject material data, which is calculated to find the Percent reject value by dividing the number of rejects by the amount of production, then multiplying by one hundred Percent.

Table 2. The second data

Number	In Process Date	Type Of Reject	Number Of Reject	Total Net Usage	Production Quantity
1	8-Jun	Bottle Dent Unscramble Machine	120	32440	32560
2	8-Jun	Bottle Dent Unscramble Machine	15	100000	100015
3	8-Jun	Cap 2nd Defect Internal	10	232440	232450
4	8-Jun	Sticker 2nd Defect Internal	10	232440	232450
5	5-Jun	Bottle Dent Unscramble Machine	10	20000	20010
6	5-Jun	Sticker 2nd Defect Internal	22	45480	45502
7	5-Jun	Sticker 2nd Defect Internal	32	27580	27612
8	6-Jun	Sticker 2nd Defect Internal	114	31420	31534
9	6-Jun	Sticker 2nd Defect Internal	45	33280	33325
10	7-Jun	Bottle Dent Unscramble Machine	10	20000	20010
331	5-Oct	Cap 2nd Defect Internal	20	2155	2175
332	5-Oct	Sticker 2nd Defect Internal	27	2155	2182
333	11-Oct	Cap 2nd Defect Internal	16	2118	2134
334	11-Oct	Sticker 2nd Defect Internal	20	2118	2138
335	11-Oct	Sticker 2nd Defect Internal	9	842	851
336	16-Oct	Cap 2nd Defect Internal	3	2401	2404
337	16-Oct	Sticker 2nd Defect Internal	21	2401	2422
338	17-Oct	Cap 2nd Defect Internal	32	1512	1544
339	17-Oct	Sticker 2nd Defect Internal	42	1512	1554
340	18-Oct	Bottle Dent Unscramble Machine	16	515	531

Table 3. The third data

Number	In Process Date	Type Of Reject	Number Of Reject	Total Net Usage	Production Quantity	Percent Reject
1	8-Jun	Bottle Dent Unscramble Machine	120	32440	32560	0.37%
2	8-Jun	Bottle Dent Unscramble Machine	15	100000	100015	0.01%
3	8-Jun	Cap 2nd Defect Internal	10	232440	232450	0.00%
4	8-Jun	Sticker 2nd Defect Internal	10	232440	232450	0.00%
5	5-Jun	Bottle Dent Unscramble Machine	10	20000	20010	0.05%
6	5-Jun	Sticker 2nd Defect Internal	22	45480	45502	0.05%
7	5-Jun	Sticker 2nd Defect Internal	32	27580	27612	0.12%
8	6-Jun	Sticker 2nd Defect Internal	114	31420	31534	0.36%
9	6-Jun	Sticker 2nd Defect Internal	45	33280	33325	0.14%
10	7-Jun	Bottle Dent Unscramble Machine	10	20000	20010	0.05%
331	5-Oct	Cap 2nd Defect Internal	20	2155	2175	0.92%
332	5-Oct	Sticker 2nd Defect Internal	27	2155	2182	1.24%
333	11-Oct	Cap 2nd Defect Internal	16	2118	2134	0.75%
334	11-Oct	Sticker 2nd Defect Internal	20	2118	2138	0.94%
335	11-Oct	Sticker 2nd Defect Internal	9	842	851	1.06%
336	16-Oct	Cap 2nd Defect Internal	3	2401	2404	0.12%
337	16-Oct	Sticker 2nd Defect Internal	21	2401	2422	0.87%
338	17-Oct	Cap 2nd Defect Internal	32	1512	1544	2.07%
339	17-Oct	Sticker 2nd Defect Internal	42	1512	1554	2.70%
340	18-Oct	Bottle Dent Unscramble Machine	16	515	531	3.01%

Table 4 and Table 5 are data transformations, which convert data into a form according to the data mining format. Table 4 categorizes minor defects of 50 as low, 50 to 100 as medium, and significant defects of 100 as high. Table 5 classifies Percent reject 0 to 0.5 percent as low, Percent reject 0.5 to 1 percent as medium, and Percent reject greater than 1 percent as high.

Table 4. The number of reject

Range Number Of Reject	Number Of Reject
<50	Low
50-100	Medium
>100	High

Table 5. The percentage of reject

Range Percent Reject	Reject
0-0.5%	Low
0.5%-1%	Medium
>1%	High

Table 6 shows the input, process, and output. Input based on the variables being grouped, namely the number of defects, the percent reject and type of defect. Process is a variable grouping is carried out, and the output is the result of grouping.

Table 6. The input, activity, output

Number	Input	Activity	Output
1	Number Of Reject	Classifying data on the number of defects into low, medium, and high	Data on the number of defects in low, medium, and high
2	Percent Reject	Classifying Percent Reject into low, medium, and high	Reject data into Low, Medium, and High
3	Types of defects	Classifying the defect types into Bottle Dent Unscramble Machine, Sticker 2nd Defect Internal, and Cap 2nd Defect Internal	Defect Type Data into Bottle Dent Unscramble machine, Sticker 2nd Defect Internal and Cap 2nd Defect Internal

3.4. Modelling

Modeling is the stage of developing a prediction model that predicts variables in the data set. Modeling is used to have similarities with the objectives of the C4.5 decision tree classification model [32]. In Figure 17, the target variable is a low/medium/high number of defects, and the influencing variables are the reject and the cause of rejects [33], [34].

Table 7 is a data reduction by discarding unnecessary variables such as net usage and production amounts. Table 8 shows data cleaning by removing several variables not needed, such as the amount of net usage and the amount of production. It showed that all the data is not empty, so modeling can be done with the data.

Table 7. The fourth data

Number	Type Of Reject	Number Of Reject	Reject
1	Bottle Dent Unscramble Machine	>100	High
2	Bottle Dent Unscramble Machine	<50	Low
3	Cap 2nd Defect Internal	<50	Low
4	Sticker 2nd Defect Internal	<50	Low
5	Bottle Dent Unscramble Machine	<50	Low
6	Sticker 2nd Defect Internal	<50	Low
7	Sticker 2nd Defect Internal	<50	Low
8	Sticker 2nd Defect Internal	>100	High
9	Sticker 2nd Defect Internal	<50	Low
10	Bottle Dent Unscramble Machine	<50	Low
331	Cap 2nd Defect Internal	<50	Low
332	Sticker 2nd Defect Internal	<50	Low
333	Cap 2nd Defect Internal	<50	Low
334	Sticker 2nd Defect Internal	<50	Low
335	Sticker 2nd Defect Internal	<50	Low
336	Cap 2nd Defect Internal	<50	Low
337	Sticker 2nd Defect Internal	<50	Low
338	Cap 2nd Defect Internal	<50	Low
339	Sticker 2nd Defect Internal	<50	Low
340	Bottle Dent Unscramble Machine	<50	Low

Table 8. The data cleaning

	Type	Number	Reject
[1]	FALSE	FALSE	FALSE
[2]	FALSE	FALSE	FALSE
[3]	FALSE	FALSE	FALSE
[4]	FALSE	FALSE	FALSE
[5]	FALSE	FALSE	FALSE
[6]	FALSE	FALSE	FALSE
[7]	FALSE	FALSE	FALSE
[8]	FALSE	FALSE	FALSE
[9]	FALSE	FALSE	FALSE
[327]	FALSE	FALSE	FALSE
[328]	FALSE	FALSE	FALSE
[329]	FALSE	FALSE	FALSE
[330]	FALSE	FALSE	FALSE
[331]	FALSE	FALSE	FALSE
[332]	FALSE	FALSE	FALSE
[333]	FALSE	FALSE	FALSE

Figure 13 is the result of a decision tree image; the explanation related to the decision tree image is as follows:

- If Cap Type Internal Defect, Sticker Internal Defect (No), then Number of Defects Low (25%).
- If Cap Type Internal Defect, Sticker Internal Defect (No) → Botol Machine Dents Unscramble, Reject Low Then Defect Quantity Low (19%).

- If Cap Type Internal Defect, Sticker Internal Defect (No) → Bottle Dents Machine Unscramble, Reject Low, Then Number of Defects Medium (6%).
- If Cap Type 2nd Defect Internal, Sticker 2nd Defect Internal (Yes) → Cap 2nd Defect Internal, Sticker 2nd Defect Internal, Then the Number of Defects Low (75%).

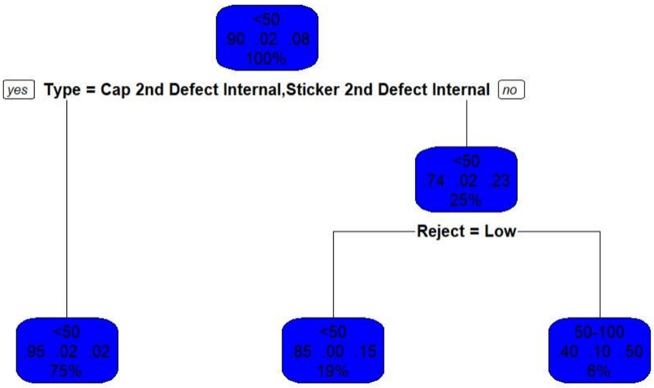


Figure 13. The decision tree from R Studio

Figure 14 is the result of a decision tree image; the explanation related to the decision tree image is as follows:

- If Cap Type 2nd Defect Internal, 2nd Defect Sticker Internal, Then Defect Count High.
- If Internal 2nd Defect Cap Type, Internal 2nd Defect Sticker, Reject Medium, Then Defect Count is High.
- If Internal 2nd Defect Cap Type, Internal 2nd Defect Sticker, Reject High, Low, Then Defect Count is Low.
- If Bottle Type Dented Unscramble Machine, Then Defect Count Medium.

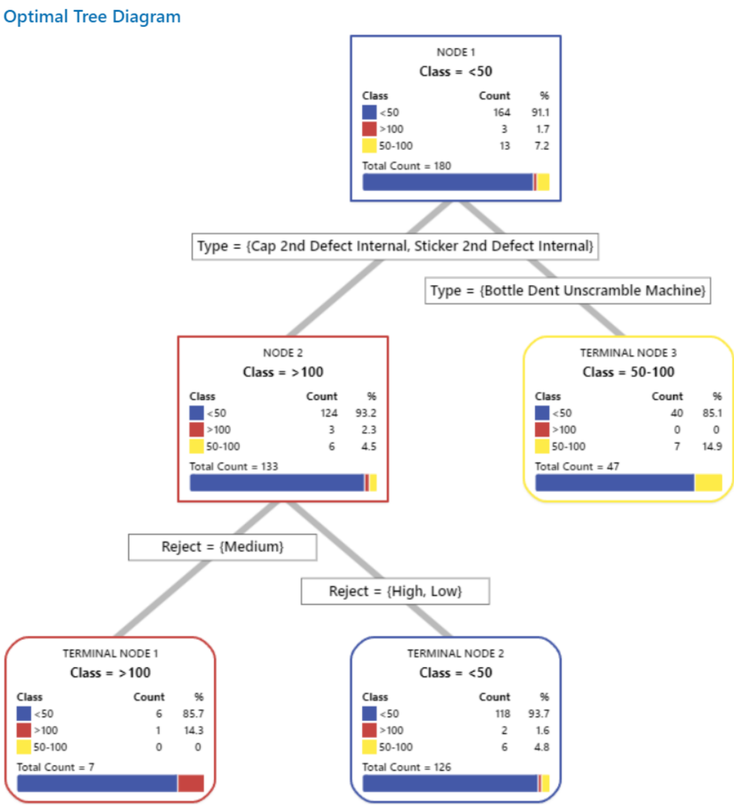


Figure 14. The decision tree from minitab

3.5. Evaluation

The evaluation stage is the stage used to check the level of Accuracy and precision of a model. The evaluation stage in this study uses a Confusion Matrix to measure the performance of the classification method [35], [36]. Based on Figure 15, the Accuracy is 0.89, the Sensitivity of defects is low at 0.9869, the number of defects is high at 0.0, and the number of defects is medium at 0.0769.

```
Confusion Matrix and Statistics

          Reference
Prediction <50 >100 50-100
    <50    151    2     12
    >100     0    0      0
    50-100   2    2      1

Overall Statistics

               Accuracy : 0.8941
                 95% CI : (0.8378, 0.936)
    No Information Rate : 0.9
    P-Value [Acc > NIR] : 0.65996

                  Kappa : 0.1476

 Mcnemar's Test P-Value : 0.01098

Statistics by Class:

                     Class: <50 Class: >100 Class: 50-100
Sensitivity              0.9869     0.00000      0.076923
Specificity              0.1765     1.00000      0.974522
Pos Pred Value           0.9152         NaN      0.200000
Neg Pred Value           0.6000     0.97647      0.927273
Prevalence               0.9000     0.02353      0.076471
Detection Rate           0.8882     0.00000      0.005882
Detection Prevalence     0.9706     0.00000      0.029412
Balanced Accuracy        0.5817     0.50000      0.525723
```

Figure 15. The confusion matrix (R studio)

Based on Table 9, there are two processes, namely the training and testing processes. The training process is a process that uses training set data or sample data that already knows the attributes of the sample data to build a model. The testing process is a process used to determine the accuracy of the model created in the training process to classify its attributes. Accuracy describes how accurate the model is in classifying correctly, as seen through the results of Minitab in the training class, which obtained high class with 33.3 percent, low with 72 percent, and medium with 53.8 percent. The high value is obtained in class testing at 0 percent, low at 81.7 percent, and medium at 76.9 percent.

Table 9. The confusion matrix (minitab)

		Predicted Class (Training)			
Actual Class	**Count**	**<50**	**>100**	**50-100**	**% Correct**
<50	164	118	6	40	72.0
>100	3	2	1	0	33.3
50-100	13	6	0	7	53.8
All	180	126	7	47	70.0

		Predicted Class (Test)			
Actual Class	**Count**	**<50**	**>100**	**50-100**	**% Correct**
<50	142	116	2	24	81.7
>100	5	1	0	4	0.0
50-100	13	3	0	10	76.9
All	160	120	2	38	78.8

From Table 10, judging from the results of the two confusion matrices, it can be seen that the Accuracy of using R studio to create a decision tree is higher than that of Minitab, which is 0.89.

Table 10. The accurate R studio and minitab

	R Studio	Minitab
Accurate	0.8941	0.788

3.6. Deployment

The results of this research are in the form of a dashboard display as a visualization of information, SOPs, and a check sheet as proposed improvements to product reject materials [37]. Figure 16, Figure 17, and Figure 18 shows that the dashboard display consists of three views, namely define problem, capability process, and decision tree. Define problem is a display that is used to see the problems experienced by the company. Capability process is a display to see the stability of the production process and Decision Tree is a display using data to form a decision tree C4.5 algorithm with R Studio and Cart algorithm with Minitab.

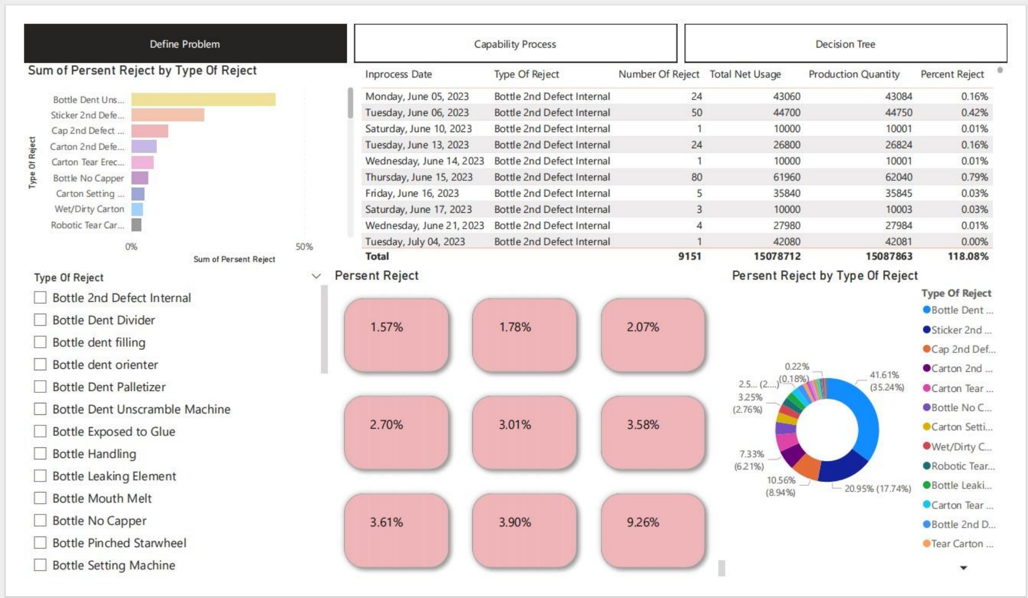


Figure 16. The dashboard define problem

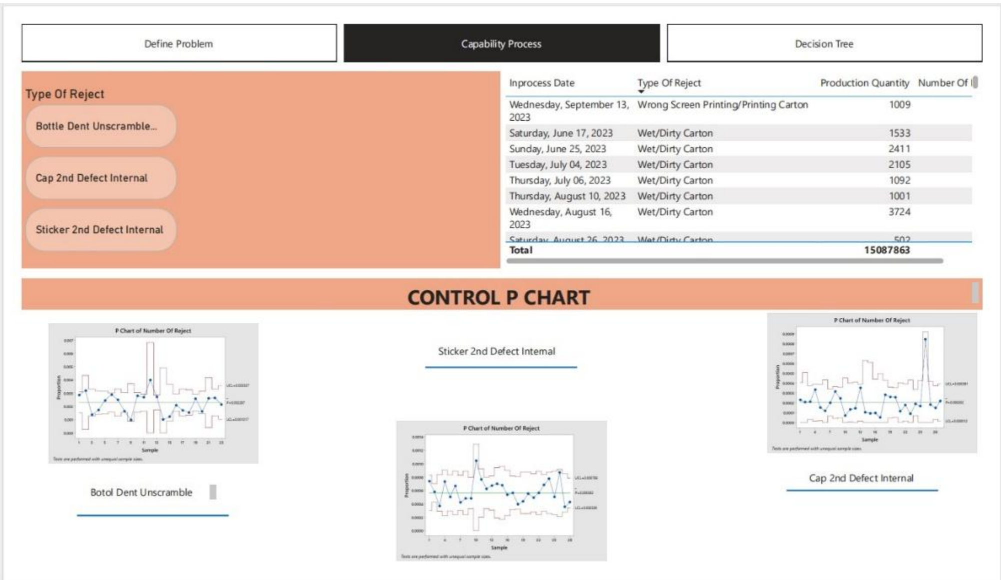


Figure 17. The dashboard capability process

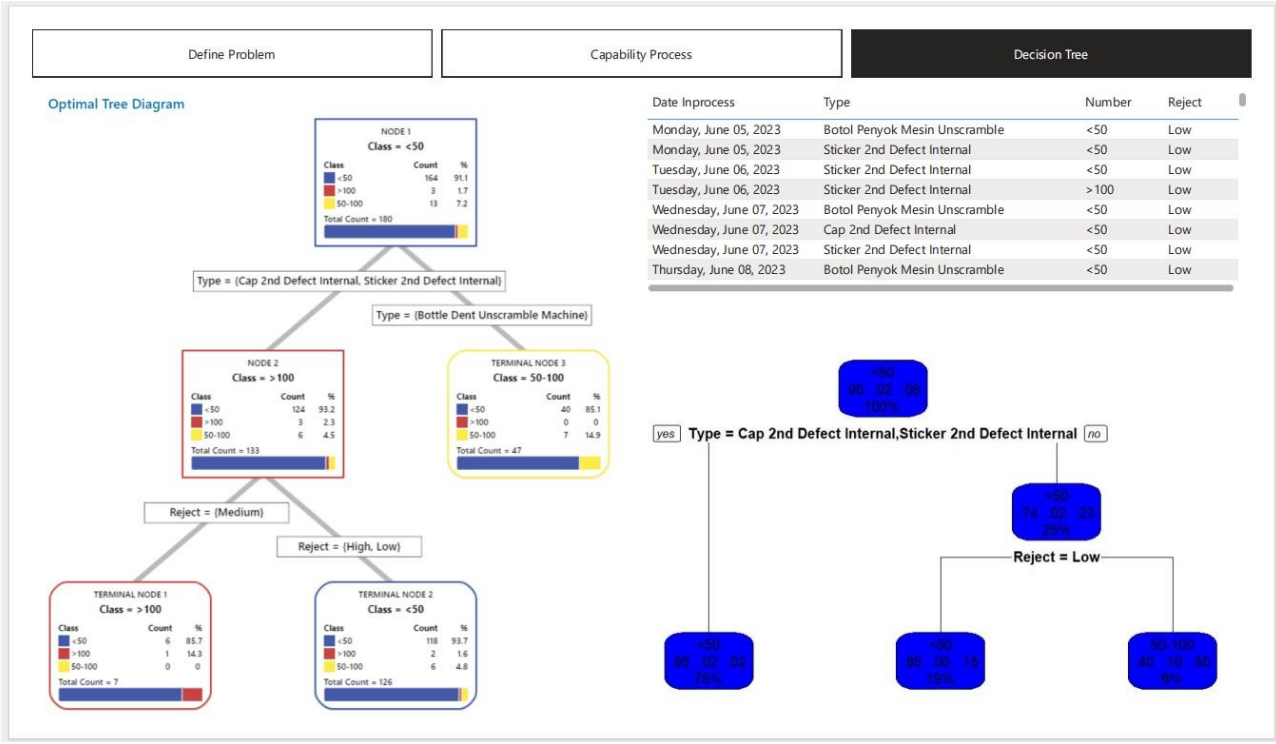


Figure 18. The dashboard decision tree

Table 11 and Table 12 show the standard operating procedure and check sheet as proposed improvements for PT P to improve the Quality of supporting materials for packaging lubricant products.

Table 11. The standard operating procedure

Standard Operating Procedure
Procedure
Prepare personal protective equipment (PPE) consisting of a helmet and gloves. Cleaning the Machine
Operation
Starting the Machine Prepare Materials Operate the Machine
Machine Maintenance
Cleaning machine parts Cleaning the nozzle of each machine from dirt Cleaning the machine from dust
Turn off the machine.
Press the Stop Button Clean up the work site.

Table 12. The check sheet

Preparation Check	**Day/Date:** **Operator :**		
Machine	Compatibility Yes	 No	Action
Cleaning the Machine			
Checking Machine Cable Condition			
Checking the Sensor			
Checking Machine Parts			
Flushing Every Machine			
Precise and appropriate machine settings			

4. CONCLUSION

Problems are caused by the three highest product defects, namely dented bottles, unscrambled machine, sticker second internal defect, and cap second internal Defect. The control chart to determine the production process on the bottle-dented unscramble machine with a sigma level value of 4.66, sticker second internal Defect with a value of 5.06, and cap second internal defect with a value of 5.32, which means that the production process is quite good and at this stage there is a fault tree analysis to find out the causes of defects in packaging materials supporting lithos M products, on the bottle dented unscramble machine with a probability of 0.0022 from the machine aspect, there is a sensor error due to a bottle that is stuck and there are machine parts that do not work. A decision tree is formed with the C4.5 and CART algorithms with an if-then rule. The confusion matrix is given the result of the C4.5 decision tree algorithm with R Studio is 89 percent, while the CART algorithm with Minitab is 78 percent. The dashboard displays the overall information using the powerful business intelligence application from the results of data modelling.

REFERENCES

[1] Z. Liao, J. Chen, X. Chen, M. Song, “Digital platform capability, environmental innovation quality, and firms’ competitive advantage: The moderating role of environmental uncertainty”, *Int. J. of Production Economics*, vol. 268, pp. 2024, doi: 10.1016/j.ijpe.2023.109124

[2] R. D. Safitri and A. N. Khairi, “Quality control of dry noodle product processing in Semarang Central Java,” *J. Agri-Food Sci. Technol.*, vol. 2, no. 1, pp. 88–95, 2022, doi: 10.12928/jafost.v2i1.4775.

[3] S. S. Salvi and S. S. Kerkar, “Quality assurance and quality control for project effectiveness in construction and management,” *Int. J. Eng. Res.*, vol. V9, no. 02, pp. 26–29, 2020, doi: 10.17577/ijertv9is020028.

[4] R. Elyas and W. Handayani, “Statistical Process Control (SPC) untuk pengendalian kualitas produk mebel di ud. ihtiar jaya,” *Bisma J. Manaj.*, vol. 6, no. 1, p. 50, 2020, doi: 10.23887/bjm.v6i1.24415.

[5] M. A. Abdillah, A. Setyanto, and S. Sudarmawan, “Implementasi decision tree algoritma c4.5 untuk memprediksi kesuksesan pendidikan karakter,” *Respati*, vol. 15, no. 2, p. 59, 2020, doi: 10.35842/jtir.v15i2.349.

[6] A. R. M. T. Islam, M. A. A. Mamun, M. Hasan, M. N. Aktar, M. N. Uddin, M. A. B. Siddique, M. H.Chowdhury, M. S. Islam, A. B. M. M. Bari, A.M. Idris, V. Senapathi, "Optimizing coastal groundwater quality predictions: A novel data mining framework with cross-validation, bootstrapping, and entropy analysis". *Journal of Contaminant Hydrology*, Vol. 269, 2025.

[7] L. Shangguan, Z. Sun. "Enhancing data quality and real-time sharing performance in water informatics through decision tree mining algorithm", *Measurement: Sensors*, Volume 31, 2024.

[8] R. A. Casonatto, T. D. P. G. Souza, A. M. Mariano, "Quality and risk management in data mining: A CRISP-DM perspective", *Procedia Computer Science*. Vol. 242, 2024, pp. 161-168

[9] A. Fitzner, A. Kassoumeh, M Krause, J. Wonneberger, L. Grabinski, "Data mining approach using cluster analysis and decision trees for optimizing electrode paste quality in lithium-ion battery production", *Procedia CIRP*, Vol. 130, pp. 449-455, 2024.

[10] D. Wang, K. Yin, H. Wang, "Intelligent classification of construction quality problems based on unbalanced short text data mining", *Ain Shams Engineering Journal*, vol. 15, no.10, 2024. doi: 10.1016/j.asej.2024.102983

[11] T. L. Palgunadi, R. Fitriana, A. N. Habyba, Y. C. Liang, "Optimizing demand forecasting method with support vector regression for improved inventory planning", *J. Optimasi Sistem Industri*, vol.3, no.2, pp. 149-166, 2024. doi: 10.25077/josi.v23.n2.p149-166.2024

[12] N.K. Auliyaanisa, R. Fitriana, E.F. Harahap, "The support vector regression method with the grid search algorithm in forecasting sales of milk product". In: Mohd Saffe, S.N., Ismail, S.Z., Rosyidi, C.N., Tokhi, M.O. (eds) *Proceedings of the 7th Asia Pacific Conference on Manufacturing Systems and 6th International Manufacturing Engineering Conference—Volume 1. IMEC-APCOMS 2024. Lecture Notes in Mechanical Engineering.* Springer, Singapore., 2025. doi: 10.1007/978-981-96-4353-0_2

[13] F. Graciela, T. S. Dewayana, A. N, Habyba, "Enhancing customer satisfaction towards the service quality of Soekarno-Hatta International Airport using servqual and binary logistic regression", *Opsi*, Vol. 17,No. 2, 2024. doi: 10.31315/opsi.v17i2.11930

[14] T. C. Tyas, I. Giyanti, "Quality control of garment product using dmaic six sigma", *Opsi*, Vol. 17, No. 1, 2024. doi: 0.31315/opsi.v17i1.7107

[15] A. Sofiana1, E. P. Safitri, "Quality control related to inventory loss of animal feed raw materials using i-mr control map (case study: PT Cargill Indonesia, Plant Semarang)", *Opsi*, vol 16, no. 1, 2023, doi: 10.31315/opsi.v16i1.8897

[16] N. Rahmasuci, J. Saragih, I. Mayusda, "Quality improvement of tin ingot product using six sigma method at PT Timah Tbk", J. OPSI, vol. 16, no. 1, 2023, doi: 10.31315/opsi.v16i1.7434

[17] L. Fitria, D.Tauhida, A. Sokhibi, "Quality control with six sigma method to minimize polyester fabric product defects at PT Sukuntex", *Opsi*, vol 16, no. 1, 2023, doi: 10.31315/opsi.v16i1.6786

[18] M. Harahap, N. Hidayati, S. Panjaitan, E. Tambunan, and J. Sihombing, "Teacher quality affects on graduation of study programming data approach there with CRISP-DM method," *Sinkron*, vol. 8, no. 4, pp. 2275–2282, 2023, doi: 10.33395/sinkron.v8i4.12762.

[19] M. D. Anies and H. A. Primandari, "Implementasi CRISP-DM model menggunakan metode decision tree dengan algoritma cart untuk prediksi lila ibu hamil berpotensi gizi kurang," *J. Ilm. Pendidik. Mat.*, vol. 10, no. 2, pp. 367–380, 2022, doi: 10.31941/delta.v10i2.2172

[20] A. Khumaidi, "Data mining for predicting the amount of coffee production using CRISP-DM method," *J. Techno Nusa Mandiri*, vol. 17, no. 1, pp. 1–8, 2020, doi: 10.33480/techno.v17i1.1240.

[21] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[22] D. Andriansyah and L. Nulhakim, "The application of power business intelligence in analyzing the availability of rental units," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012019

[23] M. Sitanggang, E. Simamora, and F. D. Mobo "Increasing accuracy of classification in C4. 5 algorithm by applying principle component analysis for diabetes diagnosis,"*Jurnal Matematika Dan Pendidikan Matematika,* vol. 6, no.2, pp. 175–186, 2022, doi: 10.25217/numerical.v6i2.2610

[24] I. Sudirman and I. D. Sudirman, "Machine learning approach to uncover customer plastic bag usage patterns in a grocery store," *Int. J. Data Netw. Sci.*, vol. 7, no. 3, pp. 1125–1130, 2023, doi: 10.5267/j.ijdns.2023.5.011.

[25] P. Mishra and R. K. Sharma, "A hybrid framework based on SIPOC and Six Sigma DMAIC for improving process dimensions in supply chain network," *Int. J. Qual. Reliab. Manag.*, vol. 31, no. 5, pp. 522–546, 2014, doi: 10.1108/IJQRM-06-2012-0089

[26] A. I. Pratiwi, A. Hakim, and R. Y. Santosa, "Development of operation process map and analysis of inventory control based on material requirement planning in assembly line," *J. Ind. Eng. Halal Ind.*, vol. 1, no. 1, pp. 30–38, 2020, doi: 10.14421/jiehis.1800.

[27] J. Odunayo Braimah, O. J. Braimah, and M. B. Abdulsalam, "On the use of P-charts in the quality control of students performance in tertiary educational system using MINITAB," *Am. J. Math. Stat.*, vol. 5, no. 5, pp. 259–264, 2015, doi: 10.5923/j.ajms.20150505.06.

[28] N. C. Weeraratne, "The effect of average sample size method on p control chart with a variable sample size," *Imp. J. Interdiscip. Res.*, Vol.2, no.8 , 2016.

[29] R. Fitriana, D. Kemala Sari, and A. Nur Habyba, *Pengendalian dan penjaminan mutu*. Wawasan Ilmu, 2021.

[30] F. O. Ehiagwina, O. O. Kehinde, A. Sidiq Nafiu, L. O. Afolabi, and I. Olatinwo, "Fault tree analysis and its modifications as tools for reliability and risk analysis of engineering systems-An overview" *Int. J. Res. Publ. Rev.*, vol. 3, no. 1, pp. 383–396, 2022.

[31] P. R. Sonawane, S. Bhandari, R. B. Patil, "Reliability and criticality analysis of a large-scale solar photovoltaic system using fuzzy-fault tree analysis approach", *J. of King Saud University - Engineering Sciences*, 2024. doi:10.1016/j.jksues.2024.09.001

[32] R. Rahim et al., "C4.5 classification data mining for inventory control," *Int. J. Eng. Technol.*, vol. 7, no. March, pp. 68–72, 2018, doi: 10.14419/ijet.v7i2.3.12618.

[33] S. Celik and O. Yilmaz, "Prediction of body weight of Turkish tazi dogs using data mining Techniques: Classification and Regression Tree (CART) and multivariate adaptive regression splines (MARS)," *Pak. J. Zool.*, vol. 50, no. 2, pp. 575–583, 2018, doi: 10.17582/journal.pjz/2018.50.2.575.583.

[34] R. Fitriana, A. Nur Habyba, and E. Febriani, *Data mining dan aplikasinya*. Wawasan Ilmu, 2022.

[35] A. Jha, M. Dave, and S. Madan, "Performance evaluation of binary and multi-class dataset using ensemble classifiers," *Int. J. Eng. Res. Technol.*, vol. 11, no. 03, pp. 425–430, 2022.

[36] V. R. Sari, N. Hayatin, and Y. Azhar, "Classifying cyberbullying data on Indonesian social media feeds utilizing sentiment analysis technique with decision tree model," *AIP Conf. Proc.*, vol. 2453(1):030011, 2022, doi: 10.1063/5.0094675.

[37] R. Hadhoud and W. A. Salameh, "How business intelligence can help you to better understand your customers," *Int. J. Bus. Intell. Res.*, vol. 11, no. 1, pp. 50–58, 2020, doi: 10.4018/IJBIR.2020010104.