Quality improvement of NH1X36B pre-printed box with QM-CRISP DM approach at PT X

(*Peningkatan kualitas NH1X36B pre-printed box dengan pendekatan QM-CRISP DM di PT X*)

Tania Theodora[1], Rina Fitriana[2] & Anik Nur Habyba[3]
[1,2,3]Jurusan Teknik Industri, Fakultas Teknologi Industri, Universitas Trisakti, Jakarta, Indonesia

[3#)]Corresponding author: anik@trisakti.ac.id

Received 14 June 2021, Revised 17 November 2021, Accepted 24 November 2021, Published 30 November 2021

Abstract. PT X is a manufacturer of cardboard box whose products are indispensable for various fields. The problem identified in NH1X36B pre-printed box, which is a shoes box, is the high defect rate that exceed company target of 2%. This study aims to reduce the defect rate of the product. The Quality-Management (QM) and Cross Industry Standard Process for Data Mining (CRISP-DM) approach was conducted by integrating Six Sigma and data mining. The Business Understanding phase was intended to define business and data mining objectives, SIPOC (Supplier-Input-Process-Output-Customer) diagram, and Critical-to-Quality (CTQ). In Data Understanding phase, it is known that the Defects Per Million Opportunities (DPMO) value is 1210.12. Data preparation phase was carried out with data cleaning, reduction, and discretization. Based on the Modeling result using Decision Tree C4.5 and FP-Growth algorithm, it is known that the dominant attributes causing high rejection are smeared ink, white spots, uneven varnish, and delamination. Decision Tree model accuracy of 90.24% indicates that the model is performing well. Analysis using FMEA yielded priority correction to the causes of smeared ink, uneven varnish, and delamination. Process improvement in Deployment phase was the application of plate cleaning and mounting form, printing process checklist, and SOP for sheet inspection. The improvement plans managed to improve the quality by rising sigma level from 4.533 to 4.648 sigma and decrease defect rate to 1.559%.

Keywords: data mining, decision tree, FP-Growth, QM-CRISP DM, six sigma.

Abstrak. *PT X merupakan produsen kardus box yang produknya sangat diperlukan untuk berbagai bidang. Masalah yang teridentifikasi pada kotak pracetak NH1X36B yaitu kotak sepatu adalah tingginya tingkat kecacatan yang melebihi target perusahaan sebesar 2%. Penelitian ini bertujuan untuk mengurangi tingkat kecacatan produk. Pendekatan Quality-Management (QM) dan Cross Industry Standard Process for Data Mining (CRISP-DM) dilakukan dengan mengintegrasikan Six Sigma dan data mining. Tahap Pemahaman Bisnis dimaksudkan untuk menentukan tujuan bisnis dan data mining, diagram SIPOC (Supplier-Input-Process-Output-Customer), dan Critical-to-Quality (CTQ). Pada tahap Data Understanding diketahui nilai Defects Per Million Opportunities (DPMO) sebesar 1210,12. Tahap persiapan data dilakukan dengan pembersihan, reduksi, dan diskritisasi data. Berdasarkan hasil pemodelan dengan menggunakan algoritma Decision Tree C4.5 dan FP-Growth diketahui bahwa atribut dominan yang menyebabkan rejeksi tinggi adalah corengan tinta, bercak putih, varnish tidak rata, dan delaminasi. Akurasi model Decision Tree sebesar 90.24% menunjukkan bahwa model tersebut berkinerja baik. Analisis menggunakan FMEA menghasilkan koreksi prioritas pada penyebab tinta tercoreng, pernis tidak rata, dan delaminasi. Perbaikan proses pada tahap Deployment adalah penerapan formulir pembersihan dan pemasangan pelat, checklist proses pencetakan, dan SOP pemeriksaan lembar. Rencana perbaikan tersebut berhasil meningkatkan kualitas dengan menaikkan level sigma dari 4,533 menjadi 4,648 sigma dan menurunkan tingkat kecacatan menjadi 1,559%.*

Kata kunci: data mining, decision tree, FP-Growth, QM-CRISP DM, six sigma.

1 Introduction

Process improvement is a task that can be supported with data mining and quality management approach. Data mining is the process of finding something meaningful from a new correlations, patterns and trends that exist by way of sifting through large amounts of data stored in the repository, using a variety of algorithms (Larose & Larose, 2014). Classification is a data mining study that finds a model that describes and differentiates data into classes.

The task of association rule mining is to discover if there is a frequent itemset or pattern in the database that contains interesting relationships between those frequent itemsets, which can be used for further decision-making activities (Man et al., 2016). One of the widely used algorithms in association task is FP-Growth or Frequent Pattern-Growth. FP-Growth emplys a divide and conquer strategy and an FP-tree data structure to achieve a condensed database representation (Han et al., 2012).

Quality Management (QM) or quality management is an action that oversees all activities to maintain the level of excellence (quality) desired (Haviana & Hernadewita, 2019). Six Sigma is a process improvement method and a statistical concept that seeks to define the variation inherent in processes. Since the 1980s, it has been widely recognized as the most effective methodology for obtaining a breakthrough improvement (Ghosh & Maiti, 2014). Technically, the process of Six Sigma (level 6 sigma), there are only 3.4 million defects per one million opportunities (Schroeder et al., 2008). The application of six sigma in this study was carried out by applying the Critical to Quality (CTQ) in identifying disability attributes and calculating the sigma level of the company.

More and more companies are now willing to start integrating data mining techniques in process analysis dan improvement (Rogalewicz & Sika, 2016). If a methodology is used, most companies prefer the Cross Industry Standard Process for Data Mining (CRISP-DM) because of its stepwise procedure and general applicability (Chapman et al., 2000).

In conducting this research, QM-CRISP DM or the combination of Quality Management and Cross Industry Standard Process for Data Mining were applied by integrating six sigma and data mining. The integration of the two was carried out by implementing data mining studies and processing data by applying various quality tools. Data mining, or to be precise C4.5 decision tree algorithm (classification task) and FP-Growth (association task), were used to obtain rules that identify the main attributes causing high rejection.

Decision Tree C4.5 generates a decision tree by splitting the nodes according to the data record attributes. The C4.5 is an algorithm developed by J. Ross Quinlan, which improves the previous ID3. In case of the ID3 algorithm, numerical attributes cannot be used, and when the attribute category value is large, there is a disadvantage in that the number of values of child nodes becomes very large. C4.5 is an algorithm that complements these disadvantages and adds new functionalities (Quinlan, 1993). The previous study discussed the nature and implications of data mining techniques and their implementations on product design and manufacturing. Using the data mining technique, the authors create an intelligent tool for automatically extracting useful information, enabling the engineers and the managers to understand the complex manufacturing data easily (Wang, 2007).

The most related works to our QM-CRISP DM approach are the data mining experimentation framework to improve six sigma projects (Fahmy et al., 2018) and data mining approach for production processes by synthesizing CRISP-DM and Quality Management (Schäfer et al., 2018). CRISP-DM (Cross-Industry Standard Process for Data Mining) is a framework in data mining that consists of 6 stages, namely Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Business Understanding is an initial phase that aims to understand business. Data Understanding is the stage of identifying data to be used, including data collection, selection, and description. Data Preparation is the stage of preparing data so that it is ready to be modeled. Modeling is the data modeling stage by the selected data mining and algorithm studies. Evaluation is the process of evaluating the model, and Deployment is the deployment phase of the model (Chapman et al., 2000).

SIPOC (Supplier-Input-Process-Output-Customer) diagram is a tool to map processes in a visual form that makes can be easier to understand the whole process (Montgomery, 2013). In addition, a control chart is used to represent data on the process to be evaluated to determine whether the process is under statistical control. Ishikawa diagrams are used to identify the main causes of disability based on man, machine, method, materials, measurement, and environmental factors. Failure Mode and Effect Analysis (FMEA) is a failure analysis method by calculating the value of the Risk Priority Number (RPN) based on the multiplication of the values of Severity (S), Occurrence (O), and Detection (D). The highest RPN value gets the priority scale for improvement. Proposed tool to help improve the quality of this research is a check sheet and Standard Operating Procedure (SOP). The check sheet is a tool to help check the form's quality of the form that holds the data for the specific purpose, which later can be converted into useful information. Standard Operating Procedure (SOP) is a written document used to implement specific work procedures.

PT X is engaged in providing integrated cardboard packaging as a solution to meet customer needs. One of the product categories with a high percentage of defects is the pre-printed box. This product is a type of corrugated cardboard whose surface is printed beforehand. The object of this research is the NH1X36B pre-printed box, which is a shoe box. The average percentage of reject products from September to November 2020 was 2.346%, with 35,123 products rejected. With a defect allowance set by the company at 2%, it can be acknowledged that the defect rate of the product exceeds the tolerance limit each month. The high number indicates that the company is urgent in taking more quality control actions to reduce defects.CRISP-DM, a data mining framework, was chosen as this research methodology.

The purpose of this study is to reduce the defect rate of NH1X36B pre-printed box products with the following stages:

1) Identifying business objectives, as well as Critical to Quality (CTQ).
2) Determining the sigma level of the NH1X36B pre-printed box production process.
3) Identifying most influential attributes causing high rejection of NH1X36B pre-printed Box by forming a decision tree and association rules from product defects data set.
4) Identifying the root causes of defects in NH1X36B pre-printed Box production based on Failure Mode and Effect Analysis.
5) Formulating a proposal for improvements to the production process of the NH1X36B pre-printed box in reducing potential causes of defects.
6) Determining the level of sigma after the implementation of the proposed improvements.

2 Method

Research requires precise and systematic steps in order to get good results. In this study, several stages of implementation were carried out, namely preliminary research, problem identification, literature review, research objectives, data collection, data processing, and results analysis. To clarify the data processing method, a flow chart is made that contains the stages of data processing as shown in Figure 1. This figure describes in detail the data processing based on the CRISP-DM phases; Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

Data processing was carried out systematically by using CRISP-DM (Cross-Industry Standard Process for Data Mining) framework as follows:

1) At the Business Understanding stage, business objectives and data mining objectives were defined. SIPOC (Supplier, Input, Process, Output, Customer) diagrams are made to identify the production process's activities more clearly. Critical-to-Quality (CTQ) is used to identify the type of product defects.
2) Data Understanding stage, data selection, process stability measurement with control charts, and DPMO value and sigma level were carried out.
3) At the Data Preparation stage, data pre-processing was performed in data cleaning, data reduction, and data transformation. The result of this stage was a new data format that was ready to be modeled.
4) Data modeling was carried out with data mining algorithms at the Modeling stage, namely, the Decision Tree C4.5 algorithm and FP- Growth using the Rapid Miner software.
   The C4.5 algorithm is a derivative algorithm from CLS (Canonical Labeling System) and ID3 (Iterative Dichotomiser 3), which can classify in the form of decision trees and rules that can be easily understood. This algorithm uses two heuristic criteria in determining the ranking, namely information gain and gain ratio. C4.5 is also widely used because it can be used on numeric and nominal attributes (Wu et al., 2008). FP-Growth (Frequent Pattern-Growth) is an a priori algorithm development that has succeeded in eliminating candidate generation (Han et al., 2000). This algorithm works by first compressing the database representing frequently

occurring items into the FP tree, storing all critical information, and dividing the compressed database into a set of conditional databases (Wu et al., 2008). FP-Growth has been used in several cases to identify strong relationships between different defects using an extensive manufacturing database (Wongwan & Laosiritaworn, 2018).

5) At the evaluation stage, the evaluation of the model was carried out based on the level of accuracy obtained and further analysis of the main causes using Ishikawa and FMEA table.
6) At the Deployment stage, improvement proposals were formulated and implemented based on data analysis. The deployment stage included an evaluation of implemented process improvement plans.

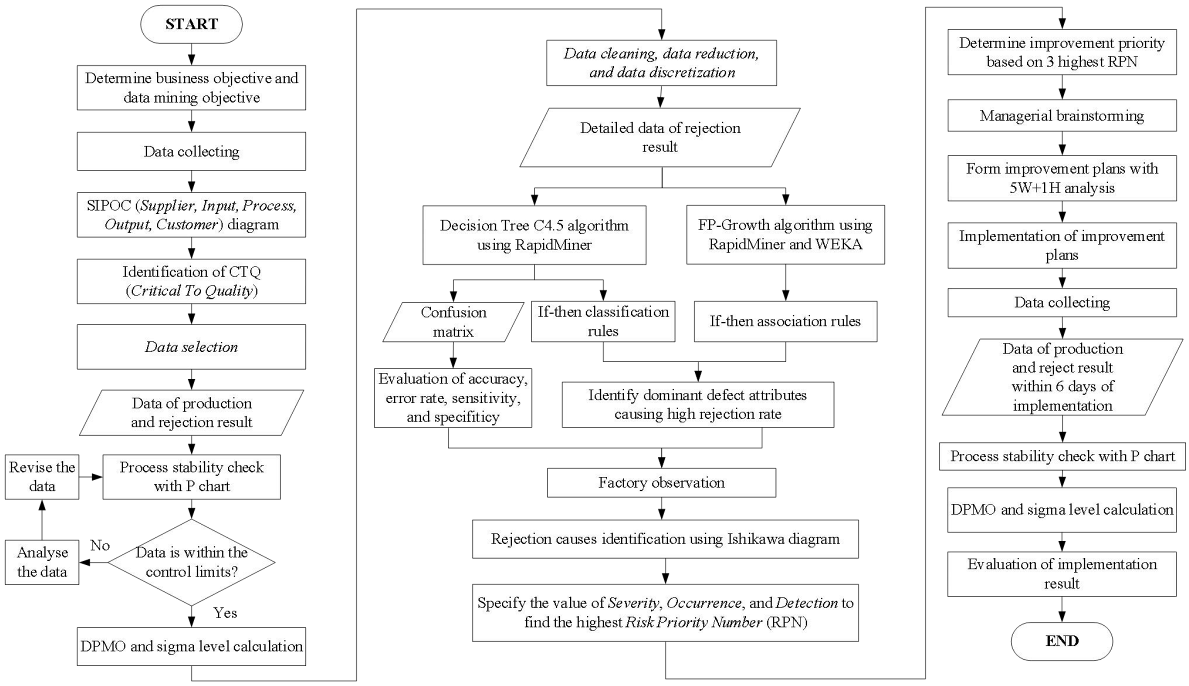

Figure 1 Data processing flowchart.

3 Result and Discussion

Business Understanding

In this research, the determined business objective was to reduce the number of rejected products, while the objective of data mining was to explore the knowledge of defect patterns that cause high rejection. In addition, the SIPOC diagram showed nine production processes for the NH1X36B pre-printed box, namely material inspection, printing, corrugating, cutting, inspection, final inspection, counting, packing, and storage. As shown in Table 1, the results showed 19 CTQ identification of defect attribute types.

Data Understanding

At this stage, the production data set was used to measure process stability using a control chart of one of the QM tools. P chart was created using production data set from September to November 2020 with total of 31 production data per number of work orders. Figure 2a. shows the p control chart graph, where two points are outside the control limit, so the data needs to be revised. Those two data were outside the control limits because of their particular cause of variation: poor printing machine performance (first spot), and there was a decrease in numbers of production and reject rate so that the proportion of defects decreased significantly (second spot). As the printing machine was not performing well, it was immediately treated by conducting maintenance. After those two data are deleted, all data had been within control limits as shown in Figure 2b.

Table 1 Critical to quality of NH1B36X pre-printed box

No.	Occur in process	Defect attributes	Details
1	Pre-printing	Hickes	Mini white spots spread on top of the printed sheet
2		Uneven varnish	Non-glossy finishing look
3		Not standard color	The printed result is not on the color range (light, standard, dark)
4		Topliner wrinkles	Wrinkles because of the raw material itself
5		Uneven color	Different blue colors on certain areas of printed sheet
6		Smeared ink	Ink splash on certain area of printed sheet
7		Dirty top liner	Dirt on printed sheet
8	Corrugating	Delamination	Unglued cardboard waves, waves are not attached to top and bottom liner
9		Oily	Oil stain on corrugated sheet
10		Blister	The orange-skin pattern on cardboard
11		Striped pattern	Striped pattern on bard board
12		Overlap	Nonuniformity of cardboard waves
13		Corr wrinkles	Wrinkles on cardboard waves
14		Dirty corr	Dirt on cardboard waves
15		Dented corr	Deep cardboard grooves
16	Die-cutting	Torn creasing	A torn part on creasing of the folded area
17		Dented die cut	Cardboard grooves
18		Misregistered	Misalignment cutting
19		Torn	A torn area on finished goods

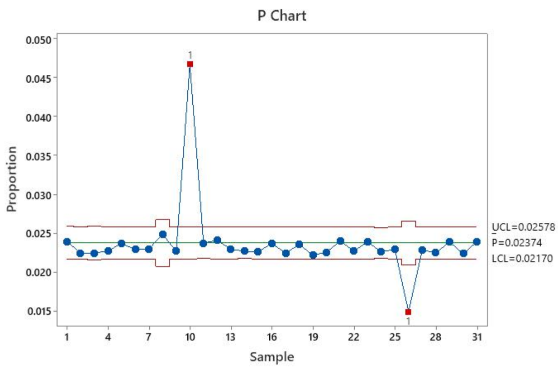

Figure 2a. P Control Chart

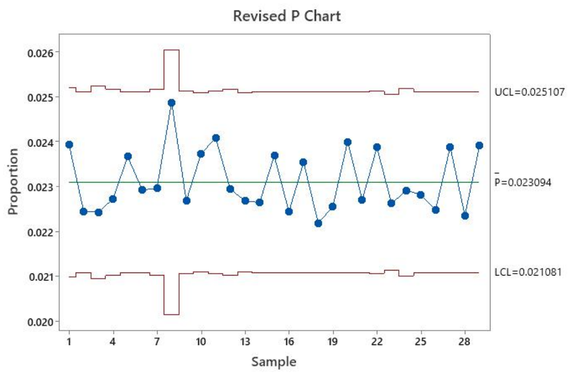

Figure 2b. Revised P Chart

The value of Defects per Million Opportunities (DPMO) and the sigma level were calculated to determine the process capability in the calculation stage. Based on the amount of production, CTQ, and the number of defects from September to November, the DPMO value is 1210.12 with a sigma level of 4.533. Of course, the company wants to gradually improving for better performance and look for ways that can help reduce the number of reject products.

Data Preparation

There were three data preprocessing steps used in this study: data cleaning, data reduction, and data discretization. Data cleaning was very important because data sets are "dirty". Data cleaning was done by correcting typing errors or unsafe writing formats and cleaning data from unnecessary columns. Data rows containing 0% reject percentage were deleted in data reduction because the data discussion was only focusing on reject classification. Furthermore, the data were transformed from numeric data into categorical and binominal data with a total of 274 rejection data.

Modeling

One source data set was used to create two different models in this study, namely the classification and association models. The data used for modeling was the reject record dataset, which contains 274 historical data of quality inspection per batch from January to November 2020, with detailed defect attributes in each inspection batch. Decision Tree C4.5 aims to classify the reject level based on the type of defect attribute characteristics. The association model (FP-Growth) is made to ensure the conformity of the results of the rules obtained by looking deeper into the association pattern.

1. Decision Tree C4.5

Decision Tree modeling was conducted using Rapid Miner. By performing parameter optimization, the Decision Tree results are obtained as shown in Figure 3. This figure shows the decision tree of the product defect data set for January to November 2020 with the reject classification output. Based on the decision tree, there were four types of defect attributes that were most influential on the classification of reject products, which are smeared ink (printing process), white spots (printing process), dented (diecutting process), and delamination (corrugating process).

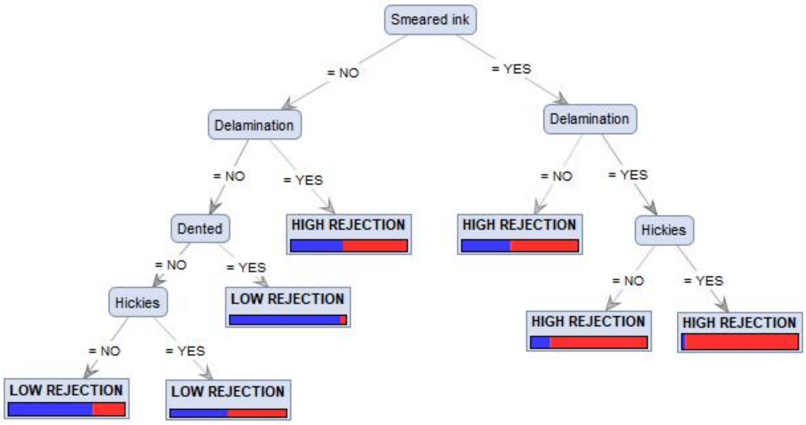

Figure 3 Reject classification decision tree

The decision tree produces 7 functions of "If-Then" rules as shown in Table 2. This study focuses the results on rules with HIGH REJECT decisions.

Table 2 If-then rules of the decision tree

No.	Jika (*If*)	Maka (*Then*)
1	SMEARED INK = NO, DELAMINATION = NO, DENTED = NO, HICKIES = NO	LOW REJECTION
2	SMEARED INK = NO, DELAMINATION = NO, DENTED = NO, HICKIES = YES	LOW REJECTION
3	SMEARED INK = NO, DELAMINATION = NO, DENTED = YES	LOW REJECTION
4	SMEARED INK = NO, DELAMINATION = YES	**HIGH REJECTION**
5	SMEARED INK = YES, DELAMINATION = NO	**HIGH REJECTION**
6	SMEARED INK = YES, DELAMINATION = YES, HICKIES = YES	**HIGH REJECTION**
7	SMEARED INK = YES, DELAMINATION = YES, HICKIES = NO	**HIGH REJECTION**

2. Association Rules using FP-Growth

The second model is the identification for association relationships based on association rules using the FP- Growth algorithm. In the FP- Growth algorithm, the software will extract data to get frequent itemsets that produce the best association rules with the highest confidence. The result of association rules produced 10 best rules based on the highest confidence value. Confidence is in the range of 0-1, where the higher the value, the more the rules will appear. Modeling using WEKA and RapidMiner showed the same result. As shown in Table 3, these if-then rules show the relationship pattern of each attribute. Based on the formed rules, functions 1 - 5 and 6 - 10 produce a pattern containing the attribute "% ACTUAL REJECT = HIGH REJECT". From the nine combinations, it is known that the pattern contains 4 dominant defect attributes that are associated with high rejection, namely smeared ink, delamination, uneven varnish, and hickies. All confidence values are above 0.91, which means that the confidence in the resulting pattern is high. All lift ratios are more than 1, so it can be concluded that the pattern of association has a high strength.

Table 3 If-then rules of FP-Growth algorithm

No.	If	Then	Confidence	Lift
1	% ACTUAL REJECT = HIGH REJECTION, SMEARED INK = YES, UNEVEN VARNISH = YES, HICKIES = YES	DELAMINATION = YES	0,964	1,83
2	DELAMINATION = YES, SMEARED INK = YES, UNEVEN VARNISH = YES, HICKIES =YES	%ACTUAL REJECT = HIGH REJECTION	0,964	1,64
3	DELAMINATION = YES, SMEARED INK = YES, HICKIES =YES	%ACTUAL REJECT = HIGH REJECTION	0,955	1,62
4	%ACTUAL REJECT = HIGH REJECTION, DELAMINATION = YES, UNEVEN VARNISH = YES, HICKIES = YES	SMEARED INK = YES	0,947	1,8
5	SMEARED INK = YES, UNEVEN VARNISH = YES, HICKIES = YES	%ACTUAL REJECT = HIGH REJECTION	0,933	1,58
6	SMEARED INK = YES, UNEVEN VARNISH = YES, HICKIES = YES	DELAMINATION = YES	0,933	1,77
7	SMEARED INK = YES, HICKIES = YES	%ACTUAL REJECT = HIGH REJECTION	0,932	1,58
8	DELAMINATION = YES, SMEARED INK = YES, UNEVEN VARNISH =YES	%ACTUAL REJECT = HIGH REJECTION	0,926	1,57
9	DELAMINATION = YES, UNEVEN VARNISH = YES, HICKIES = YES	%ACTUAL REJECT = HIGH REJECTION	0,919	1,56
10	DELAMINATION = YES, SMEARED INK = YES	%ACTUAL REJECT = HIGH REJECTION	0,913	1,55

Based on the results of modeling using Decision Tree C4.5 and FP- Growth, it could be concluded that the four defect attributes that can cause high rejection are smeared ink, uneven varnish, hickies, and delamination. Therefore, problem analysis is focused on these attributes.

Evaluation

After modelling, it is necessary to evaluate data mining results based on the resulting performance, namely the level of accuracy. The results of the performance evaluation of the decision tree C4.5 model in the form of a table view confusion matrix can be seen in Table 4.

Table 4 Confusion Matrix

	true LOW REJECTION	true HIGH REJECTION
pred. LOW REJECTION	13	0
pred. HIGH REJECTION	4	24

Each confusion matrix box interprets different meanings. 13 is the True Positive (TP) value, 24 is the True Negative (TN) value, 0 is the False Positive (FP) value, and 4 is the False Negative (FN) value. These four values can be used as the basis for calculating accuracy, error level, sensitivity, and specificity. Based on the iterative modeling experiment, this model produces the highest accuracy, 90.24%, splitting 85% for data training and 15% for data testing. This shows fairly high accuracy and the model is performing well. The error rate shows that the percentage of wrong predictions is 9.76%. Sensitivity is a true positive rate which able to predict low rejection correctly. The sensitivity value of 76.47% indicates that out of 13 low reject data, 4 data are wrongly predicted. Specificity is a true negative rate which able to predict high rejection correctly. With a value of 100%, all test data with high rejection are correctly predicted.

Determination of the Causes of Defects Using Ishikawa Diagram

The model evaluation results were analyzed using Six Sigma tools in the form of Ishikawa diagrams and Failure Mode and Analysis (FMEA). This aims to identify more deeply of the defect root causes. Based on Figure 4, smeared ink defects can occur due to machine and method factors. The cause of the method factor is the absence of an appropriate impression control method. As for the machine factor, smeared ink occurs due to the condition of the plates that are not clean and peeled off, dirty rubber rolls, and worn-out chamber blades.

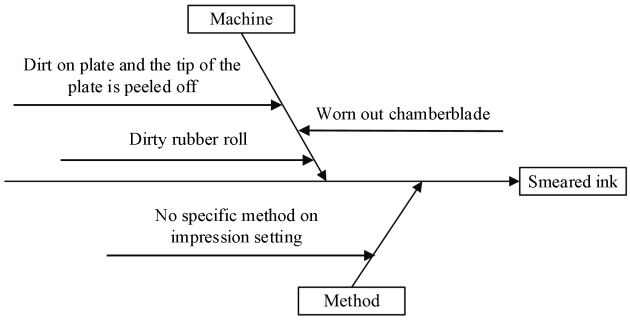

Figure 4 Ishikawa diagram of smeared ink defect

As seen in Figure 5, the uneven varnish cause is similar to smeared ink because both are executed with the same process.

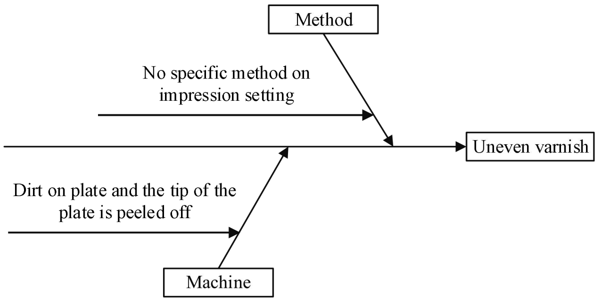

Figure 5 Ishikawa diagram of uneven varnish defect

Based on Figure 6, hickies defect was occurred due to the materials and method factors. Differences in temperatures can affect the level of paint dryness and transfer between units, resulting in hickies. In the materials factor, hickies might happen because of the dusty condition of the paper roll so that the ink cannot be transferred completely.

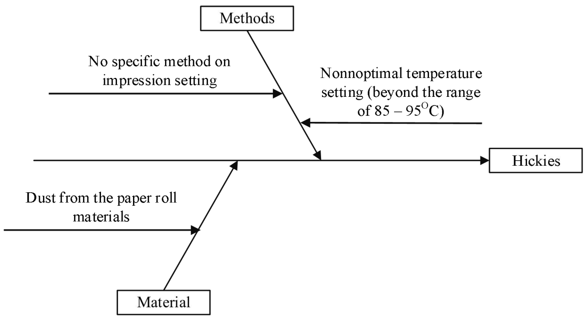

Figure 6 Ishikawa diagram of hickies defect

Delamination defects occurred in the corrugating process, where the cardboard wave pattern of the medium paper did not adhere well to the top and bottom layers of paper. In the man factor, if the product is fluting type A cardboard, the operator presses the glue for type E fluting, then glue content will be different. The temperature too high will also prevent the glue from affecting the moisture of the cardboard. In addition, paper scraps from the machining process can fall into the glue area and prevent the paper from sticking.

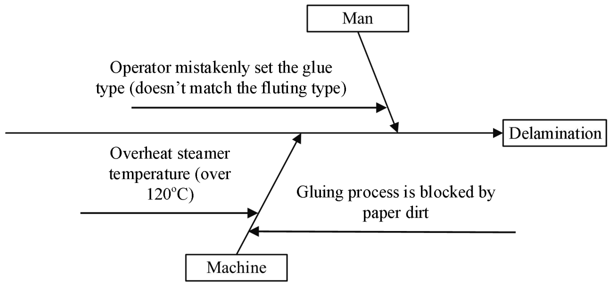

Figure 7 Ishikawa diagram of delamination defect.

Failure Mode and Effect Analysis (FMEA)

After the causes of the problem were identified using the Ishikawa diagram, the FMEA table was made to describe each of the causes of the existing defects. The score was filled based on the reference RPN table and from interview result with one of the production supervisors. The results of the RPN calculation are in Table 5. The highest RPN values of 280 and 245 arere found in the smeared ink and uneven varnish defects with the exact failure cause, which is the plate condition was not suitable for use. The RPN value of 245 is also found in delamination defects because of paper flakes that prevent the glue from sticking. The two causes that led to the high value of the RPN will be a priority for problem-solving by formulating improvement proposals. As one example of the calculations for the value of the highest RPN is as follows:

$$RPN=Severity(S)xOccurence(O)xDetection(D) \quad (1)$$

$$RPN=8x7x5=280$$

Table 5 Summary of FMEA

Failure type	Process	S	Cause of failure	O	Control	D	RPN
Smeared ink	Printing	8	Dirt on the plate and the tip of the plate is peeled off	7	Regularly clean and change the plate after a certain period of time	5	**280**
			No specific method on impression setting	6	Re-calibrate the machine after first run approval is done	4	192
			Dirty rubber roll	5	Wipe the dirty area with the duster	3	120
			Worn out chamberblade	5	Maintenance every two weeks	6	240
Hickies	Printing	7	Dust from the paper roll materials	5	Turn on the dust cleaner machine	6	210
			No specific method on impression setting	6	Re-calibrate the machine after first run approval is done	4	168
			Nonoptimal temperature setting (beyond the range of 85-95 C)	4	Reset the temperature based on first run approval result	3	84
Uneven varnish	Printing	7	No specific method on impression setting	6	Re-calibrate the machine after first run approval is done	4	168
			Dirt on the plate and the tip of the plate is peeled off	7	Regularly clean and change the plate after certain period	5	**245**
Delamination	Corrugating	7	The operator mistakenly set the glue type (does not match the fluting type)	4	Operator training	3	84
			Overheat steamer temperature (over 120 C)	4	Adjust the speed	7	196
			Gluing process is blocked by some paper dirt from the machine	5	Sheet inspection while the process is still running, warn the operator to stop and clean the machine	7	**245**

Deployment

Based on the QM-CRISP DM approach carried out in the previous stage, the improvement proposals are formulated to be implemented at this stage. Proposed improvements to NH1X36B pre-printed Box production process can be seen in Table 6.

Table 6 Proposed improvements based on FMEA

Failure type	Cause of failure		Improvement proposal
1. Smeared ink 2. Uneven varnish	The plate is not in proper condition to be used	Dirt on the plate surface	Plate cleaning and mounting form
		The tip of the plate his already peeled off	Printing process checklist
3. Delamination	Gluing process is blocked by some paper dirt from the machine	The QC inspector was late in informing the operator that there had been a lot of delamination QC inspector missed the quality inspection of delamination defect	Training & SOP of corrugated sheet

Due to the limited research time and COVID-19 pandemic, the implementation of these proposed improvements was carried out within 6 days of production, resulting in 8 production records. The inappropriate plate condition is the main cause of smeared ink and uneven varnish defects in the printing process. Based on the analysis of the causes, the first proposed improvement is the application of plate cleaning and mounting form so that the officer can conduct activities regarding conformance checking points. Along the production days before this improvement was implemented, the officers had carried out the cleaning and mounting activities quite well, but they were often not careful and thorough enough to check the plate condition. Therefore, the form can be used as a record and a tool to check the conformity results. Form filling was equipped with instructions as a guide for officers to fill in the correct steps.

In addition to the plate cleaning and mounting form, the second recommendation is a checklist for the printing process. Checklist is used to check the suitability of each point of standard printing process. With this check sheet application, it is expected that the operator will be able to check the feasibility of each component more carefully so that the printing machine is in optimal condition to minimize the potential for defects.

The process improvement recommendation for delamination defect is by providing training and SOP for corrugated sheet inspection. Training is given to QC staff to carefully inspect all sides of the sheet based on the resulting defect attributes. The QC Supervisor carries out training by emphasizing the inspection procedures according to the SOP, work discipline, responsibilities, and work vigilance. Operators and QC staff are obliged to read the SOP before carrying out activities. With this SOP, it is expected that QC staff will be able to more carefully inspect each sheet based on the type of defect to provide information to the operator if there is a defect that requires immediate corrective action, such as machine shutdown.

4 Conclusion

Based on data processing and analysis of the results obtained from this research with QM - CRISP-DM approach, the following conclusions can be drawn:

1) At the Business Understanding phase, the business objective is to reduce the product rejection rate, and the data mining objective is to explore the pattern of defects attributes to gain more knowledge about attributes causing high rejection. Based on the printing, corrugating, and die-cutting processes, there are 19 defect attributes as Critical to Quality, for instance, hickies, uneven varnish, unstandardized color, top liner wrinkles, striped colour, smeared ink, dirty top liner, delamination, oily, blister, striped pattern, overlap, wrinkled sheet, dirty corrugated sheet, dented corrugated sheet, broken creasing, dented in diecutted result, misregistered, and torn sheet.
2) At the Data Understanding phase, it is known that the DPMO of NH1X36B pre-printed Box NH1X36B from September to November 2020 is 1210.12 with the level of sigma of 4.533.

3) Based on the Modeling stage with the Decision Tree C4.5 algorithm, it is known that 3 defects causing high rejection are smeared ink, delamination, and hickies. Those defects also appeared in the FP-Growth algorithm, with the addition of 1 attribute that is uneven varnish. From these two algorithms, it can be concluded that smeared ink, delamination, hickies, and uneven varnish will be analyzed further as the 4 dominant factors causing high rejection.
4) Based on the FMEA table, the three highest RPN were obtained in the printing and corrugating process, with two main defect causes, dirt on the plate and tip of the plate is peeled off, and gluing process is blocked by some paper dirt from the machine.
5) At this stage of deployment, three proposed improvements to increase the product quality are the application of plate cleaning and mounting form, printing process check sheet, and SOP for corrugated sheet inspection to minimize the potential defect causes.
6) The implementation results showed a positive impact on quality improvement, marked by a decrease in DPMO to 820.75 and an increase in the sigma level to 4.648 sigma from 4.533.

This research was conducted during the COVID-19 pandemic with many limitations. The use of data mining can be used to explore Ishikawa diagrams more in determining the critical manufacturing process. This will help the company to know the defect classification based on the causative factors.

Acknowledgments

The authors pay homage to the Faculty of Industrial Technology, Universitas Trisakti, for continuous encouragement, guidance, and support for their work. This paper is the research funded by Universitas Trisakti in the academic year 2020/2021.

References

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM -Cross-Industry Standard Process for Data Mining- 1.0 Step-by-step data mining guide. *CRISP-DM Consortium*.

Fahmy, A. F., Mohamed, H. K., & Yousef, A. H. (2018). A data mining experimentation framework to improve six sigma projects. *ICENCO 2017 - 13th International Computer Engineering Conference: Boundless Smart Societies*, *2018-January*. https://doi.org/10.1109/ICENCO.2017.8289795

Ghosh, S., & Maiti, J. (2014). Data mining driven DMAIC framework for improving foundry quality-a case study. *Production Planning and Control*, *25*(6). https://doi.org/10.1080/09537287.2012.709642

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. https://doi.org/10.1016/C2009-0-61819-5

Han, J., Pei, J., & Yei, Y. (2000). Mining Frequent Patterns without Candidate Generation. *ACM SIGMOID Record*, *29*(2), 1–12. https://doi.org/10.1145/335191.335372

Haviana, E., & Hernadewita, H. (2019). Productivity improvement in the rubber production process using value stream mapping method to eliminate waste. Operations Excellence: Journal of Applied Industrial Engineering, 11(2), 119-130. doi:http://dx.doi.org/10.22441/ oe.v11.2.2019.023

Larose, D. T., & Larose, C. D. (2014). Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition. In *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition* (Vol. 9780470908747). https://doi.org/10.1002/9781118874059

Man, M., Abu Bakar, W. A. W., Abdullah, Z., Jalil, M. A., & Herawan, T. (2016). Mining association rules: A case study on benchmark dense data. *Indonesian Journal of Electrical Engineering and Computer Science*, *3*(3). https://doi.org/10.11591/ijeecs.v3.i2.pp546-553

Montgomery, D. (2013). Introduction to Statistical Quality Control. In *New York: John Wilwy & Sons, Inc*.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

Rogalewicz, M., & Sika, R. (2016). Methodologies of knowledge discovery from data and data mining methods in mechanical engineering. *Management and Production Engineering Review*, *7*(4). https://doi.org/10.1515/mper-2016-0040

Schäfer, F., Zeiselmair, C., & Becker, J. (2018). Synthesizing CRISP-DM and Quality Management : A Data Mining Approach for Production Processes. *2018 IEEE International Conference on Technology Management, Operations and Decisions, ICTMOD 2018*, 190–195.

Schroeder, R. G., Linderman, K., Liedtke, C., & Choo, A. S. (2008). Six Sigma: Definition and underlying theory. *Journal of Operations Management*. https://doi.org/10.1016/j.jom.2007.06.007

Wang, K. (2007). Applying data mining to manufacturing: The nature and implications. *Journal of Intelligent Manufacturing*, *18*(4). https://doi.org/10.1007/s10845-007-0053-5

Wongwan, K., & Laosiritaworn, W. (2018). Application of association rules in woven wire mesh defects analysis. *2018 7th International Conference on Industrial Technology and Management, ICITM 2018*, *2018-January*, 325–329. https://doi.org/10.1109/ICITM.2018.8333970

Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z. H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1–37. https://doi.org/10.1007/s10115-007-0114-2