Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, Number 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4208

Analysis of Red Brick Product Quality Improvement at UD. Batu Bata Bulan Using CRISP-DM and C4.5 Algorithm

Widya Septiana[1], Farhan Tsani[2], Silvia Firda Utami[3*], Rina Meri Andani[4]
[1,2,3,4]Industrial, Sumbawa University of Technology, Sumbawa, Indonesia.
[1]widya7553@gmail.com, [2]farhantsani58@gmail.com, [3]silvia.firda.utami@uts.ac.id, [4]meriandanir@gmail.com

open access

***Corresponding Author**

Article History:
Submitted: 01-07-2024
Accepted: 30-10-2024
Published: 29-11-2024

Keywords:
Red Brick Defects; data mining; CRISP-DM and Algorithm C.45; UD. Batu Bata Bulan.

ABSTRACT

UD Batu Bata Bulan is a home industry in Batu Bulan Village that produces red bricks. The industry has been operating for 20 years and has been producing bricks every day. The home industry is facing problems related to the quality of red bricks that require appropriate action to improve to meet the desired quality. In addition, the home industry is still experiencing difficulties in conducting product quality inspections due to the lack of inspection technology to help the process. The problems faced can be detrimental to UD. Batu Bata Bulan. Therefore, it is necessary to analyze the causes of defective products in the process of producing red bricks and improve the quality of red brick products. Therefore, the researcher conducted an analysis to address the problems that occurred in the company related to product quality. The solution that can be given is to classify the type of defective product using the role of data mining. In this study, the standard Cross Industry Standard Process For Data Mining (CRISP-DM) procedure and C.45 Algorithm were used in data processing.The result of this research indicate significant knowledge in classifying black color defect in data this could facilitate the quality inspection department in making accurate decisions.

INTRODUCTION

The construction using red bricks is one of the most common construction methods used in Indonesia. Red bricks are used as the main material in the construction world, becoming a key element in development. Red bricks are one of the building elements used in the construction of buildings made of soil or a mixture of other materials that are burned (Permatasari, 2019). In the manufacturing industry, red bricks have production houses that function as a place to produce and sell bricks. Sumbawa Besar is one of the areas that has a red brick industry.

UD Batu Bata Bulan is one of the home industries in Batu Bulan Village that produces bricks. This industry has been established for 20 years and carries out the production process every day. Every day, 600 pieces of red bricks are printed. In one week, UD. carries out production for 5 days. So for 1 week, the total production of red bricks is 3,000 pieces. Then for 1 month, there are 4 weeks so the number of red brick production obtained is 12,000 pieces. The selling price of 1 piece of red brick is 1,000.

Based on the results of interviews with the owner of UD. Batu Bata Bulan this home industry is facing problems related to the quality of red bricks which require appropriate action to improve product quality to meet the desired product quality. Product quality is the ability of a product to carry out its function, consisting of durability, reliability, accuracy, ease of operation and repair, and other valuable attributes (Tirtayasa, Lubis, & Khair, 2021). In addition, this home industry is still experiencing difficulties in the process of checking the quality of red bricks because no inspection technology can help the process.

Over time, the demand for the quality of red bricks must be better and more competitive, this greatly affects the sustainability of the production process of UD. Batu Bata Bulan. Because of the long age, the production process at UD. Batu Bata is increasingly experiencing problems that result in many defective red brick products. Defects that occur in red bricks such as recycling, non-standard sizes, cracks, asymmetry, and black colors have not been classified. Quality is a product image and the company's responsibility for products related to customer perceptions of quality (Putera & Wahyono, 2019). This is in line with complaints from customers regarding the quality of the red bricks chosen for the business. There is a need for quality control efforts on these products so that they can continue to progress.

The problem of defects that occur quite a lot can cause losses for UD. Batu Bata Bulan. Therefore, further studies need to be carried out on the classification of types of defects in red brick products using CRISP-DM, a C45 algorithm decision tree. The CRISP-DM work stages are used in data mining by implementing six work stages, namely business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Dhewayani et al., 2022). A decision tree is one of the strong classification methods and can transform large facts into a decision tree that represents rules (Nurmuslimah, 2022). One study has been successfully carried out by (Andarista & Jananto, 2022). in classifying the results of motor vehicle testing by modeling using CRISP-DM to explain the results in accordance with the test with pass and fail.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4208

LITERATURE REVIEW

1. Product Quality Assurance
   Product quality assurance is an activity to ensure that the products produced meet customer needs and expectations. Product quality assurance is carried out by applying various methods and techniques to prevent and reduce product defects. Product quality is an important factor that needs to be considered by a company to increase product competitiveness in the market and provide customer satisfaction (Santoso, 2019).
2. Data Mining
   Data mining is a technique for extracting knowledge from large datasets with an interdisciplinary approach that includes statistical analysis, data visualization, pattern recognition, and database management (Nikmatun I.A., 2019).
3. Cross Industry Standard Process For Data Mining (CRISP-DM)
   The CRISP-DM method is a flexible methodology that can be adapted to the needs of data mining projects, from simple projects to complex ones (Suhanda , Kurniati, & Norma, 2020)(Christian, 2020).
4. Decision Tree Method C4.5
   The C45 algorithm is a learning algorithm used to build a hierarchical decision tree by representing the relationship between predictor variables and target variables (Andarista & Jananto (2022).

METHOD

The flowchart for this research can be seen in Figure 1 below.

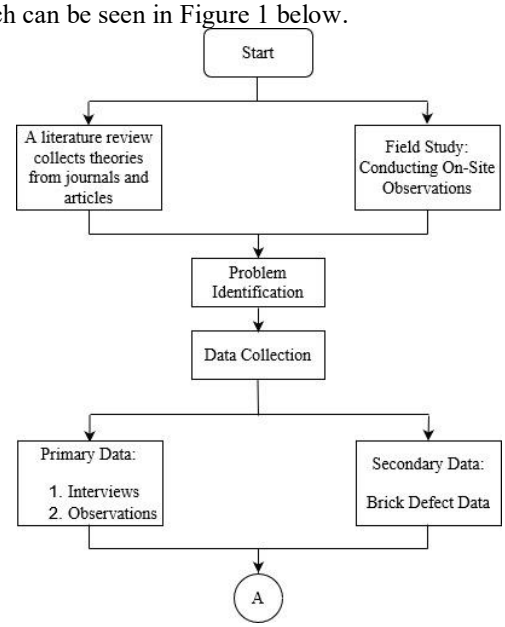


Figure 1. Research Flowchart

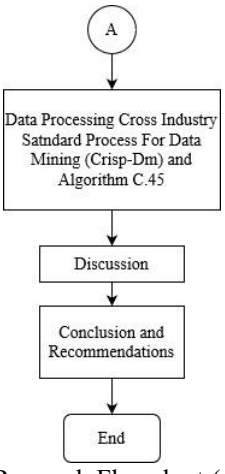


Figure 2. Research Flowchart (continued)

This research utilizes the Cross Industry Standard Process for Data Mining (CRISP-DM) method to classify defective red brick products and employs the C4.5 decision tree algorithm to follow up and obtain a decision tree.

The CRISP-DM framework is a data mining framework that implements six phases in its method to identify the

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, Number 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4208

basis and results of a process. These phases consist of business understanding, data understanding, data preparation, modeling, deployment, and evaluation (Christian, 2020)(Pambudi & Setiawan, 2019).

1. Business Understanding This stage aims to assess the condition of a business to gain an overview of the resources available and needed.
2. Data Understanding Through this stage, conclusions are drawn regarding the data to be used and the strategies that must be taken to address the data.
3. Data Preparation The data found in the previous step must be preprocessed, cleaned, and evaluated dimensionally.
4. Data Modeling At this stage, the model to be used is determined and applied.
5. Evaluation This stage involves evaluating the performance of the data model and re-examining the entire process that has been carried out to ensure that no data or steps are missed.

There are several steps to create a decision tree using the C4.5 algorithm (Suiroh, Astuti, & Basysyar, 2024)(Fatimah & Saepudin, 2022):

1. Prepare training data. Training data is usually taken from previously classified historical data.
2. Calculate the root of the tree. The root will be taken from the selected attribute by calculating the gain value of each attribute. Before calculating the gain value, calculate the entropy value first. Then, the most dominant gain value will become the first root.
3. Calculate the gain value.
4. Repeat steps 2 and 3 until all records are partitioned.
5. The decision tree partitioning process will stop if:
   a. All records in node N have the same class.
   b. There are no more attributes in the partitioned record.

RESULT

The data processing in this research was carried out using the CRISP-DM (Cross Industry Standard Process for Data Mining) approach and the C4.5 algorithm with several stages, including Business Understanding, Data Understanding, Data Preparation, Data Modeling, Data Evaluation, and Data Deployment (Kamagi & Hansun, 2019) (Jollyta, Siddik, Mawengkang, & Efendi, 2021) (Sulistiyawati & Supriyanto, 2021).

Business Understanding

Business understanding is the initial stage in the CRISP-DM process. Based on interviews with the owner of the home industry, the goal is to improve the quality of red brick products to meet standards. However, this home industry is experiencing difficulties in making quality improvements, especially related to actions that have not been appropriate to assist in the inspection process of red brick production. As a result, there is a lack of further action regarding the problem. To achieve business goals, the company needs appropriate actions to reach the right solutions.

Data Understanding

Data understanding is an important stage in the CRISP-DM process, where the process of understanding the data to be used in the data mining project takes place. In this research, the data used comes from UD. Batu Bata Bulan and focuses on defect data in the production process. The results of data collection reveal that this data is in the form of a dataset consisting of 5 observed attributes, namely recycle 60, recycle 100, non-standard size 60, non-standard size 100, crack 60, crack 100, not symmetrical 60, not symmetrical 100, black color 60, and black color 100. Based on the defect count dataset, black is the color that has the highest number of defects compared to other types of defects. The standard number of black colors is 46 and the non-standard is 64. Furthermore, the data will be preprocessed. The preprocessing stage is the initial step to prepare the data so that it is ready to be analyzed or processed further.

Data Preparation

Data preprocessing is the process of transforming raw data into data that is ready for further analysis. This process is done by cleaning the data from errors, incompleteness, and duplication. The data selected based on the required attributes and classes is then processed using Rapidminer software.

Data Transformation

The data transformation process is carried out to change raw data into data that is ready to be used in the Decision Tree C4.5 algorithm. The cleaned and selected data is then imported into Rapidminer software. The data transformation can be seen in the following figure.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4208

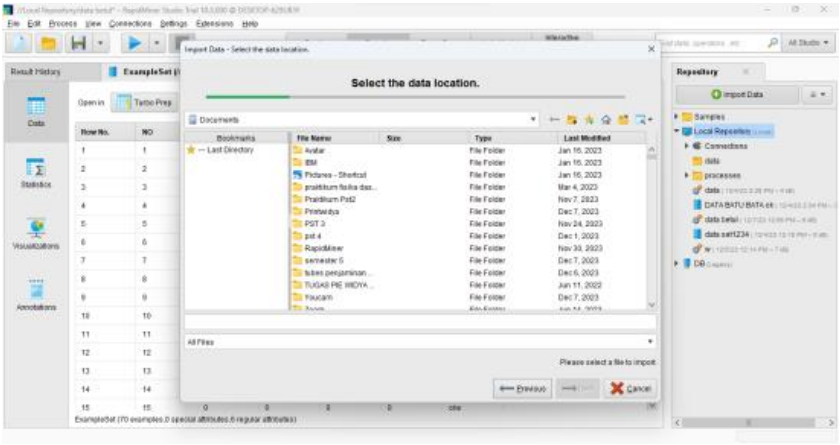


Figure 3. Tranformation Data

Based on the image above, the first stage in carrying out data transformation is by clicking the import data menu and selecting the brick data file that will be processed using the Rapidminer application.

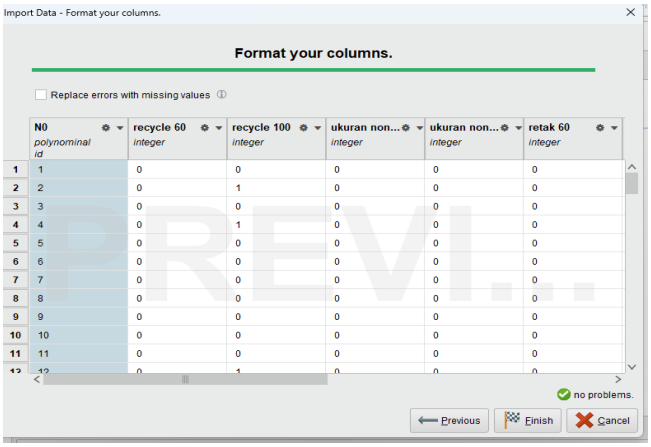


Figure 4. Tranformation Data (continued)

After selecting the data to be processed, the data display will appear with the polynomial ID column format in the number column, then in the brick defect data type column select the integer format and in the results column select the binomial label format.

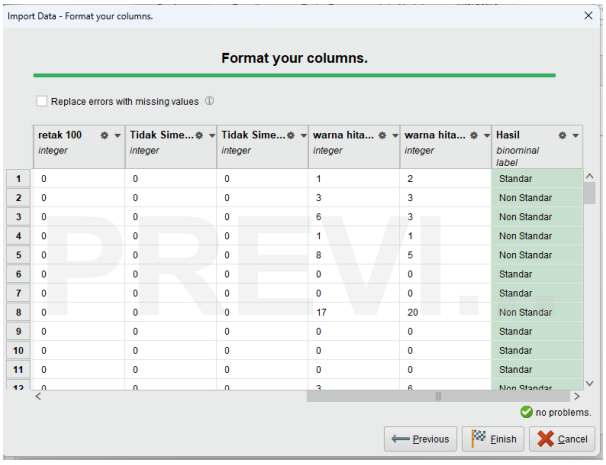


Figure 5. Tranformation Data (continued)

The data transformation process involves modifying data types to match the requirements of the Decision Tree algorithm and RapidMiner software. This transformation divides the data into four data types: polynomial, binomial, integer, and real. In the transformation process, two main components are considered: role ID and role label. The name

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, Number 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4208

is used as the role ID, while the result is used as the role label. The transformed data is then used as training data to calculate entropy, and information gain, and determine the attribute with the highest information gain value. This highest information gain value is used to build the model using the C4.5 algorithm.

Modeling

The modeling process utilizes data mining techniques with the C4.5 algorithm. The C4.5 algorithm calculates entropy and gain values to form a decision tree. The data used in the formation of the decision tree is the cleaned production process data. The decision variables used are the attributes recycle, non-standard size, crack, not symmetrical, and black color. The results of the C4.5 algorithm calculation are shown in Table 2. This table presents the entropy and gain values of each attribute. These values are used to build the optimal decision tree structure.

Table 1. Algorithm C.45

ATTRIBUTE		Product Count	Standard	Non-Standard	ENTROPY GAIN	GAIN
Total		110	64	46	0,5903	
Black Color						0,0387
	60	53	33	20	0,5756	
	100	57	40	17	0,0387	

The table above shows the modeling calculation using the C4.5 algorithm. It can be seen that the attribute with the highest gain value is the black color attribute with a value of 0.0387. Therefore, the black color attribute is the attribute that becomes the root node. The calculation results will be presented as follows:

Calculation of Entropy Value

Total entropy value of the attribute: = (-(64/110) * $\log_2$(64/110)) + (-(46/110) * $\log_2$(46/110)) = (-0.5818 * $\log_2$0.5818) + (-0.4181 * $\log_2$0.4181) = 0.2737 + 0.3167 = 0.5903

Entropy value of black color 60: = (-(33/53) * $\log_2$(33/53)) + (-(20/53) * $\log_2$(20/53)) = -0.6226 * $\log_2$0.6226 - 0.3773 * $\log_2$0.3773 = 0.2562 + 0.3194 = 0.5756

Entropy value of black color 100: = (-(40/57) * $\log_2$(40/57)) + (-(17/57) * $\log_2$(17/57)) = -0.7017 * $\log_2$0.7017 - 0.2982 * $\log_2$0.2982 = 0.2159 + 0.3134 = 0.5293

Calculation of Gain Value = (0.5903) - (53/110) * (0.5756) - (57/110) * (0.5293) = 0.0387

Based on these calculations, the overall entropy value is 0.5903, the entropy value for black color size 60 is 0.5756, and the entropy value for black color size 100 is 0.5293.

Modeling Using RapidMiner

The calculation process of the C4.5 algorithm using Rapidminer software consists of several stages, namely entering the dataset (transformation), training and testing the Decision Tree C4.5, and generating accuracy values. The stages in data processing using Rapidminer software are as follows:

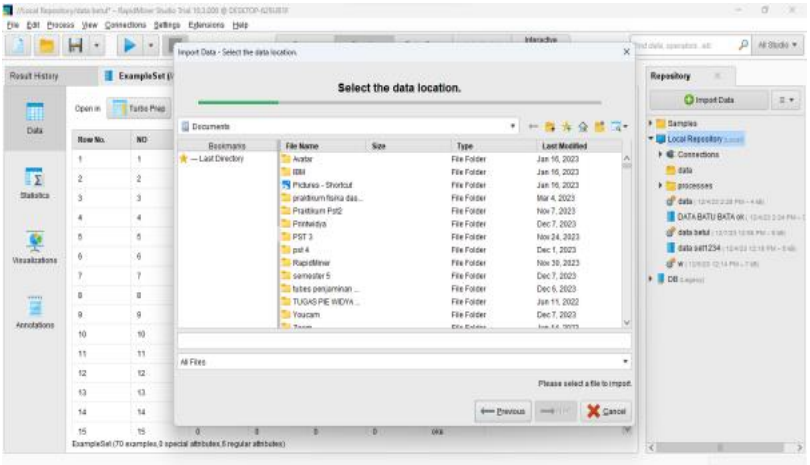


Figure 6. Read data excel view

Based on the image above, after entering the Rapidminer software, add the "read excel" operator to search for the data to be processed. Excel read data is used from the results of data that has been preprocessed which can be seen in the image.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, Number 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4208

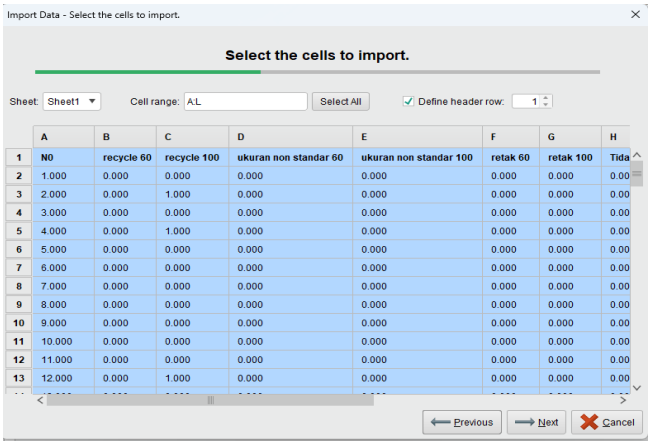


Figure 7. Dataset view

Based on Figure 7, a display will appear of the data set selected to be processed using Rapidminer software

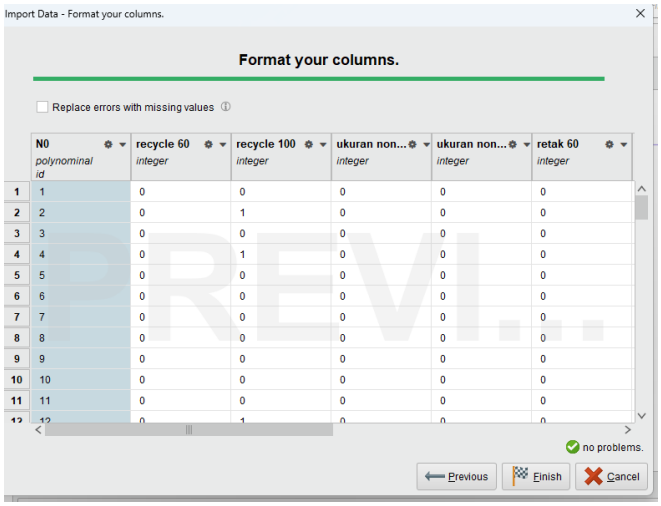


Figure 8. Import defect data

Next, the import data display will appear, so at this stage the data is adjusted based on the data type, such as changing the defect type to Standard and Non-Standard to binominal type. In the NO attribute, role is selected as the ID and in the result attribute, role is selected as the label, which can be seen in Figure 8 above

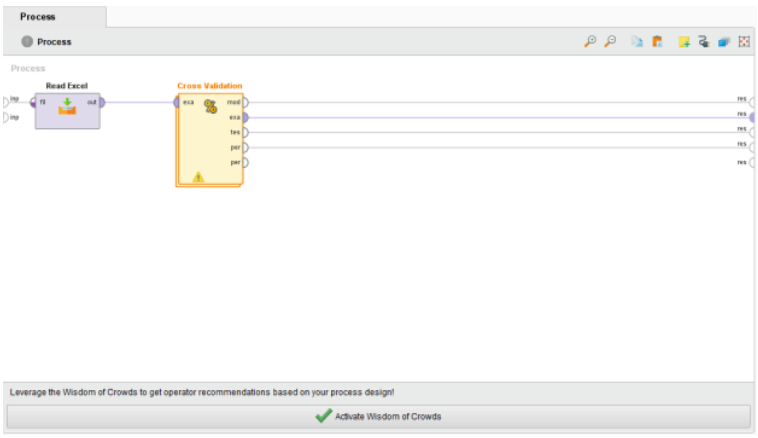


Figure 9. Read excel and cross validation view

Based on Figure 9 above, the Read Excel process in the Rapidminer application is modeled. In building a data collection model with the Decision Tree algorithm, there are several operators that will be used, namely Read Exel and Cross Validation. The validation process using the cross validation feature can help increase the accuracy of the

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 4, Number 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4208

decision tree results.

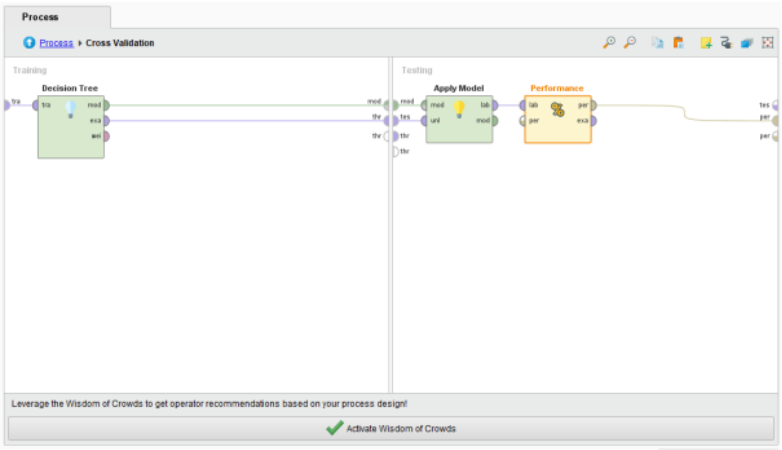


Figure 10. Cross Validation Decision Tree C4.5 model

In cross validation modeling which can be seen in the image above, there are two process parts, namely the training part is used to find out the results of the decision tree classification algorithm and the data testing process is the part using the apply model feature to apply the model to testing data and performance features. to display the confusion table, which is used to display accuracy, recall and precision results.

After analyzing the data in RapidMiner, the next step is to generate an output in the form of a decision tree model. This model is a visual representation of the rules or formulas that have been discovered based on the previously processed data. This model presents valuable information in the form of rules that can assist companies in making decisions regarding defective products. Through the decision tree model, companies can identify the factors that most influence the occurrence of defects in products. This knowledge can be used to improve product quality, reduce the number of defects, and increase overall efficiency and performance. Thus, this decision tree model becomes an effective tool in improving the production process and optimizing product quality at UD. Batu Bata Bulan. The results of the decision tree for production process defects can be seen in the following figure.

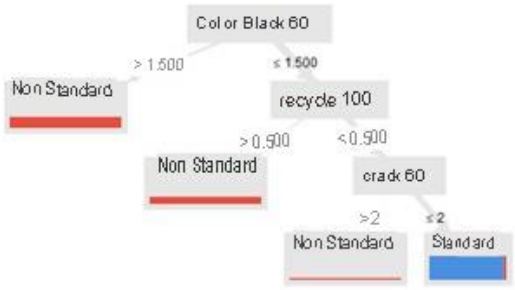


Figure 11. Defect Model on Red Bricks Using Decision Tree C4.5

Tree

```
warna hitam 60 > 1.500: Non Standar {Standar=0, Non Standar=28}
warna hitam 60 ≤ 1.500
|   recycle 100 > 0.500: Non Standar {Standar=0, Non Standar=17}
|   recycle 100 ≤ 0.500
|   |   retak 60 > 2: Non Standar {Standar=0, Non Standar=2}
|   |   retak 60 ≤ 2: Standar {Standar=62, Non Standar=2}
```

Figure 12. Description of Decision Tree C4.5 Modeling

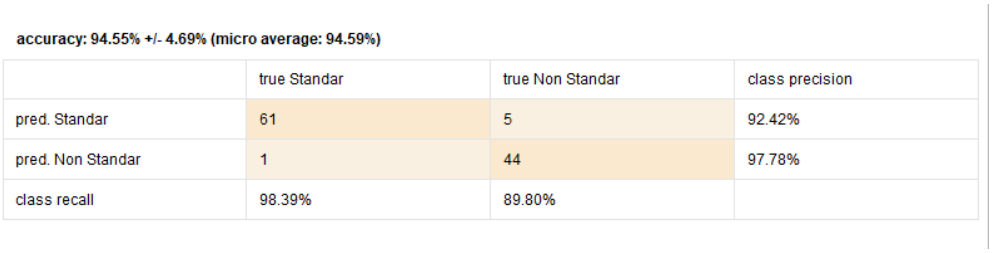
The figure above shows that many defective products resulting from the production process are discarded. This is due to the lack of decision-making in classifying defective products. Classification of defective products can help simplify the inspection process so that the quality inspection section does not need to have difficulty in finding the limit of defective products or making daily inspection reports.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4208

DISCUSSION

The result of the decision tree is a representation of the decision-making process based on the given data. This decision tree explains how the classification of production process defect data is determined based on its attributes. After the data is processed, the accuracy level of the data is tested by applying the C4.5 algorithm model which uses the Receiver Operating Characteristic (ROC). The results of the Decision Tree C4.5 Accuracy can be seen in the following figure.



accuracy: 94.55% +/- 4.69% (micro average: 94.59%)

	true Standar	true Non Standar	class precision
pred. Standar	61	5	92.42%
pred. Non Standar	1	44	97.78%
class recall	98.39%	89.80%	

Figure 14. Decision Tree C4.5 Accuracy Results

The calculation of training data accuracy from 110 data resulted in an accuracy of 92.42%, where there were 61 data predicted as Standard and were True OK, but 5 data were True Non-Standard. Then, for the Non-Standard predicted data, 1 of them was True Standard, and 44 were True Non-Standard. The Recall value is 89.80%, and the precision is 97.78%. The resulting Confusion Matrix can be seen in the following table.

Table 2. Confusion Matrix

		Actual Value	
		TRUE	*FALSE*
Predicted Value	*TRUE*	TP 61	FP 5
	FALSE	FN 1	TN 44

Based on the Confusion Matrix above, there are four cells: 1. True Positive (TP): The number of observations that actually belong to the Non-Standard class and are also correctly predicted as the Non-Standard class. The value is 61. False Positive (FP): The number of observations that belong to the Standard class, but are incorrectly predicted as the Non-Standard class. The value is 5. True Negative (TN): The number of observations that actually belong to the Standard class and are also correctly predicted as the Standard class. The value is 44. False Negative (FN): The number of observations that actually belong to the Non-Standard class, but are incorrectly predicted as the Standard class. The value is 1.

Accuracy Calculation: Accuracy = (TP + TN) / (TP + TN + FP + FN) * 100% = (61 + 44) / (61 + 44 + 5 + 1) * 100% = 105 / 111 * 100% = 0.9459 * 100% = 94.59%

Recall Calculation: Recall = TN / (TN + FP) * 100% = 44 / (44 + 5) * 100% = 44 / 49 * 100% = 0.8979 * 100% = 89.79%

Precision Calculation: Precision = TN / (TN + FN) * 100% = 44 / (44 + 1) * 100% = 44 / 45 * 100% = 0.9777 * 100% = 97.77%

CONCLUSION

Based on the analysis results in the data mining process using the C4.5 algorithm carried out at CV. Batu Bata Bulan using black defect data, it can be concluded that the results of the defect data classification show that the attribute with the highest gain value is the black color attribute with a value of 0.0387. Therefore, the black color attribute is the attribute that becomes the root node. Based on the evaluation, testing, and data processing that has been carried out from the results of data mining classification using the C4.5 decision tree algorithm, it can be concluded that the confusion matrix produces an accuracy of 94.59%, Recall of 89.80%, and precision of 97.78%, which means these values are quite high. This indicates that the algorithm can perform accurately and is considered Excellent Classification. Based on the use of Rapidminer software, it is proven that the results of the decision tree as a C4.5 algorithm can provide significant knowledge in the classification of black defect data. This makes it easier for the quality inspection section to make decisions and carry out appropriate improvements. The suggestion in this research for researchers is to find more attribute value data to maximize the data processing process in determining the prediction of red brick defects. The suggestion given to the company is to carry out regular improvements in the production process to avoid excessive product defects so as not to cause significant losses.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4208

REFERENCES

Andarista R.R., & Jananto A. (2022). Penerapan Data Mining Algoritma C4.5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor," Jurnal Tekno Kompak, vol. 16, no. 2.

Christian Y. (2020). Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran pada Startup Early Stage dengan Menggunakan CRISP-DM," JURIKOM (Jurnal Riset Komputer), vol. 9, no. 4, p. 966. doi: 10.30865/jurikom.v9i4.4486.

Dhewayani F.N., et al. (2022). Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. Jurnal Teknologi dan Informasi, vol. 12. doi: 10.34010/jati.v12i1.

Fatimah A.I., & Saepudin S. (2022). "Penerapan Data Mining Dengan Metode Apriori Pada Penjualan Sembako (Studi Kasus: Grosir Sembako Lina. (Vol. 8, Issue 2). https://rekayasa.nusaputra.ac.id/index

Jollyta D., Siddik M., Mawengkang H., & Efendi S. (2021). Teknik Evaluasi Cluster Solusi Menggunakan Python Dan Rapidminer .1st ed. Yogyakarta: Deepublish.

Kamagi D.H., & Hansun S. (2019). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," J. Ultim., vol. 6, no. 1, pp. 15–20. doi: 10.31937/ti.v6i1.327.

Nikmatun I.A. (2019). Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. Jurnal SIMETRIS, vol. 10, no. 2.

Nurmuslimah S. (2022). Sistem Pendukung Keputusan Pada Teknologi Informasi. PT. Global Eksekutif Teknologi.

Pambudi R. H., & Setiawan B.D. (2019). Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal," Jurnal Pengembangan Tekonologi Informasi dan Ilmu Komputer, vol. 2, no. 7, pp. 2637–2643.[Online]. Available: http://j-ptiik.ub.ac.id

Permatasari S. (2019). Pengaruh Bahan Tambah Batu Bata Merah Terhadap Kuat Tekan Beton fc'21 Menggunakan Agregat Kasar PT. Amr dan Agregat Halus Desa Sunggup Kota Baru.

Putera A.K., & Wahyono. (2019). Pengaruh Kualitas Pelayanan,Citra Merek,Dan Kualitas Produk Terhadap Loyalitas Konsumen Melalui Kepuasan Konsumen. Management Analysis Journal, vol. 7, no. 1. [Online]. Available: http://maj.unnes.ac.id

Santoso J.B., (2019). Pengaruh Kualitas Produk, Kualitas Pelayanan, Dan Harga Terhadap Kepuasan Dan Loyalitas Konsumen (Studi Pada Konsumen Geprek Bensu Rawamangun). Jurnal Akuntansi dan Manajemen, vol. 16, no. 01.

Suhanda Y., Kurniati I., & Norma S. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. Jurnal Teknologi Informatika dan Komputer, vol. 6, no. 2, pp. 12–20. doi: 10.37012/jtik.v6i2.299.

Suiroh S., Astuti R., & Basysyar F.M. (2024). Implementasi Algoritma K-Means Pada Pengelompokan Data Penerimaan Peserta Didik Baru Di Smkn 1 Balongan. JATI (Jurnal Mahasiswa Teknik Informatika), 8(1), 1–8. https://doi.org/10.36040/jati.v8i1.8335

Sulistiyawati A., & Supriyanto E. (2021). Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan. Jurnal Teknokompak. https://doi.org/10.33365/jtk.v15i2.1162

Tirtayasa S., Lubis A. P., & Khair H. (2021). Keputusan Pembelian: Sebagai Variabel Mediasi Hubungan Kualitas Produk dan Kepercayaan terhadap Kepuasan Konsumen. Jurnal Inspirasi Bisnis dan Manajemen vol. 5, no. 1, pp. 2579–9312. [Online]. Available: http://jurnal.unswagati.ac.id/index.php/jibm