Journal of Sisfotek Global

Versi Online: http://journal.stmikglobal.ac.id/index.php/sisfotek
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
Vol. 11, No. 2, September 2021, pp. 127-134

C45 Algorithm for Motorcycle Sales Prediction On CV Mokas Rawajitu

Debby Alita[1], Setiawansyah[2], Ade Dwi Putra[3]

[1,2,3]Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
E-mail: [1]debbyalita@teknokrat.ac.id, [2]setiawansyah@teknokrat.ac.id, [3]adedwiputra@teknokrat.ac.id

ARTICLE HISTORY

Received : *23 August, 2021*
Revised : *7 September, 2021*
Accepted : *20 September, 2021*

KEYWORDS

C45 Algorithm
Cross Validation
Rapid Miner
Prediction
Motorcycle Sales

ABSTRACT

CV Mokas Rawajitu is a company that sells various types of used motorbikes both in cash and on credit. In sales, the problem that occurs is the frequent occurrence of ups and downs in motorcycle sales due to the mismatch of the available motorcycle variants with consumer interests so that motorcycle sales often do not reach the target. The role of data mining is needed to analyze consumer purchasing patterns at CV Mokas Rawajitu which can produce information, namely knowing what types of motorbikes most in-demand by consumers are and which are most in-demand in the market by predicting using the C4.5 algorithm based on the sales transaction data they have. from previous periods. The study used a dataset of motorcycle sales at CV Mokas Rawajitu from 2017-2019 with a total data volume of 1,411 data. The attributes used are the motorbike category, the motorbike brand, the motorbike price, and the year of production. The tools used in this research are Rapid Miner. The results of the application of the C4.5 Algorithm can be used as a prediction of sales at CV Mokas Rawajitu because the results of the accuracy of testing data and models using 9-Fold Cross Validation reach a value of 87.95% where the 9th fold reaches the highest accuracy value with a Sensitivity level of 97, 15%, 69.05% Specificity, 86.57% Precision, 12.05% Error (Error Rate) and 30.95% False Positive Rate.

1. INTRODUCTION

The use of motorcycles in Indonesia is very popular because the prices are relatively cheap, affordable for some people and the use of fuel is economical and the operational costs are also very low. As time goes by, consumer demand for motorcycles continues to increase so companies must be able to provide sales products to meet market demand.

CV Mokas Rawajitu is a company that sells various types of used motorcycles, both in cash and on credit. In sales, the problem that occurs is that there are frequent fluctuations in motorcycle sales due to the incompatibility of motorcycle variants available with consumer interest so that motorcycle sales often do not meet market demand and are detrimental to the company because many motorcycle products are not sold. To meet market demand, the role of data mining is needed. According to previous research conducted by data mining is an exploration and analysis activity, from large amounts of data to find useful patterns and rules. The purpose of data mining is to make it easier for companies to improve marketing, sales, and customer service through a better understanding of customers [1].

One technique in data mining is prediction, prediction technique is a technique that aims to identify patterns that appear repeatedly in a data and classify them based on behavior or values that are expected in the future [2]. Predictions can be applied using the C4.5 algorithm. The C4.5 algorithm is one of the algorithms used to form a decision tree [3]. The decision tree method converts very large facts into decision trees that represent rules that can be easily understood by natural language [4]. One technique in data mining is prediction, a prediction technique is a technique that aims to identify patterns that appear repeatedly in a data and classify them based on behavior or values expected in the future. C4.5 is one of the algorithms used to form decision trees.

The more competitive the company in producing motorcycles makes motorcycle showrooms need analysis of prediction changes, namely the tendency of customer behavior in choosing motorcycles [5].

Based on this fact found a problem how to predict the pattern of buying on motorcycles. Data mining applications created can predict purchasing patterns with the C4.5 Algorithm method [6]. With the field of the brand, year, and price on the motorcycle table calculated by gain, the field with the largest value becomes the root of the tree. Tree results can illustrate the predictive tendency of buying patterns. It is expected that the specified rules can provide information to the Manager in conducting analysis to determine stock and target market share.

CV Mokas Rawajitu has transaction data stored in its database. The transaction data is increasing day by day. As the amount of data on CV Mokas Rawajitu increases, the role of the C4.5 Algorithm is needed which forms a decision tree to analyze consumer buying patterns on CV Mokas Rawajitu which can produce information, namely knowing what types of motorcycles most in demand by consumers are and which sell best. in the market based on sales transaction data held from previous periods. By using the C4.5 algorithm, it can be seen the attributes that affect consumer buying interest, so that in the future the company can supply motorcycle variants in accordance with consumer interests [7]. That way the company can minimize the possibility of not achieving sales targets.

The tool used to apply the association rule to the research data used in this study is Rapid Miner 7.1. Rapid Miner is an open-source software for analyzing data mining, text mining, and predictive analysis. Rapid Miner is a solution for analyzing data mining using descriptive and predictive techniques in providing insights to users so that they can make the best decisions.

2. METHODS

This research was conducted using predictive data mining methods. According to Witten, Ian H, and Frank in [8]s definition of data mining is the process of extracting data (previously unknown, implicit, and considered useless) into information or knowledge or patterns from large amounts of data. Data that is considered "garbage" because it is not patterned/unstructured and useless, is processed (filtered) so as to form useful new information or knowledge or patterns. In simple terms, data mining can be interpreted as the process of extracting or digging the existing knowledge in a set of data. The information and knowledge obtained can be used in many fields, such as business management, education, health, and so on [9]. the stages of the research can be seen in the following picture.

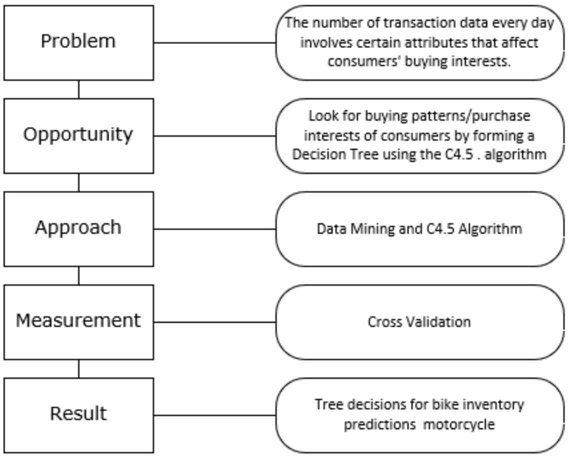

Problem
The number of transaction data every day involves certain attributes that affect consumers' buying interests.
Opportunity
Look for buying patterns/purchase interests of consumers by forming a Decision Tree using the C4.5 . algorithm
Approach
Data Mining and C4.5 Algorithm
Measurement
Cross Validation
Result
Tree decisions for bike inventory predictions motorcycle

Figure 1. Research Methods

The research stage by identifying the problem in the research that is being done. Problem identification refers to ever-increasing sales transaction data and search involvement of these attributes in attracting buyer interest. The next stage is to propose in the form of analysis of sales transaction data using C4.5 Algorithm to make decision trees. based on the attributes used. The next stage is to analyze data using data mining and C4.5 algorithm. Data testing conducted using Cross Validation, where Training data will be divided into several parts and 1 part as data testing. The results in this study will be in the form of a decision tree that can be used as a recommendation for motorcycle supplies on cv Mokas Rawajitu.

The process of determining consumer purchasing patterns is carried out using the C4.5 Algorithm. The C4.5 algorithm is a data classification algorithm with a decision tree technique that has advantages. These advantages, for example, can process numeric data (continuous) and discrete, can handle missing attribute values, produce rules that are easy to interpret, and the fastest among other algorithms [10].

The process of determining consumer purchasing patterns is carried out using sales transaction data that has been optimized to be able to classify data with the label of selling well and not selling well. The stages/flow of the C4.5 algorithm calculation starts from preparing a dataset that has been labeled then calculating the total entropy and then calculating the entropy and gain of each attribute [11]. The attribute with the largest gain value will be used as a node/root. The calculation will continue until all attributes are partitioned. For more details, see Figure 1.

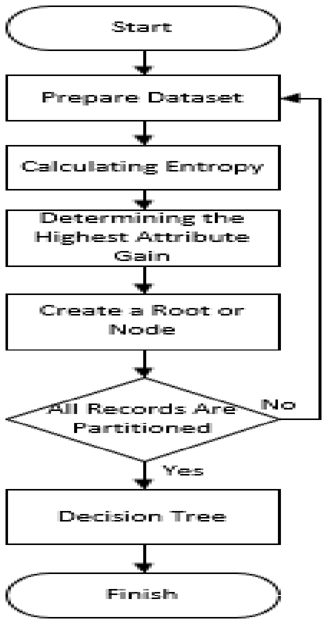


Figure 2. Algorithm C45

Based on the steps of the C4.5 algorithm above, the formula for calculating the entropy value is obtained, namely by Equation (1):

$$Entropy(S) = \sum_{i=1}^{n} - pi \text{ x } \log_2 pi \qquad (1)$$

S = case set

n = Number of partitions

Pi = Proportion of Si to S

Calculate the gain value using Equation

$$Gain(S) = Entropy\ (S) - \sum_{i=1} \frac{Si}{S} \times Entropy(Si) \quad (2)$$

S = Set of cases

A = Feature or attribute

n = Number of attribute partitions A

|Si| = Proportion of Si to S

|S| = Number of cases in S

Entropy (Si) = Entropy for samples that have the i value

3. RESULTS AND DISCUSSION

3.1. Implementation Rapid Miner

The results are in the form of research data that has been processed and poured in the form of tables, graphs, photos, or images. The discussion contains the results of the analysis and research results associated with an established knowledge structure (review of the literature referred to by the author) and raises new theories or modifications to existing theories [12].

Rapid Miner is open-source software for analyzing data mining, text mining, and predictive analysis[13]. Rapid Miner is a solution for analyzing data mining using descriptive and predictive techniques in providing insight to users so that they can make the best decisions[14]. The steps for implementing the C4.5 Algorithm on Rapid Miner are as follows:

1. Click Add Data then select the dataset storage location which will be used. Due to the author's dataset in the form of files that have the format *.xls and not comes from a database like SQL then select the location from My Computer.

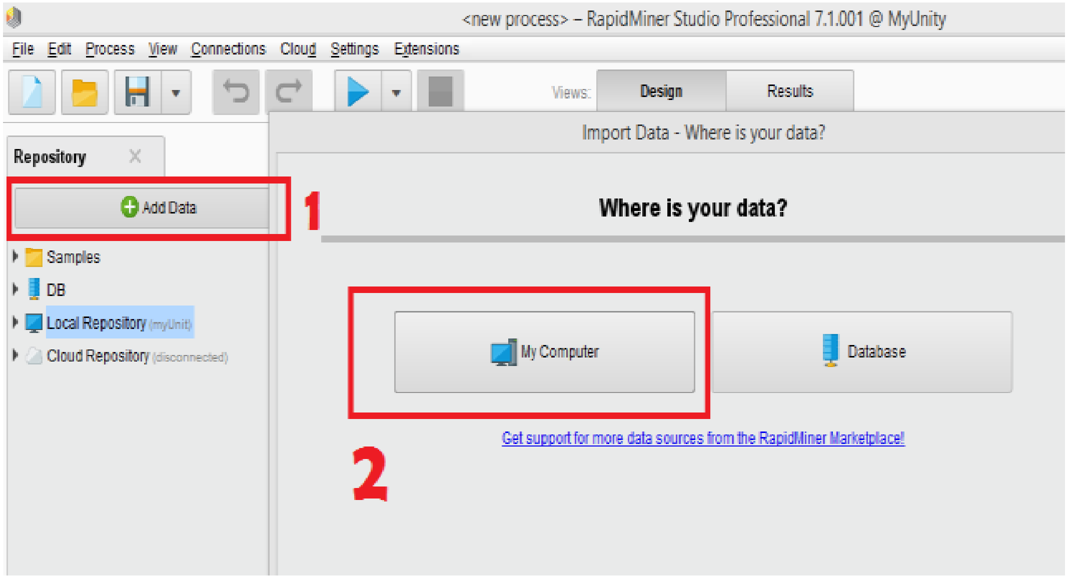
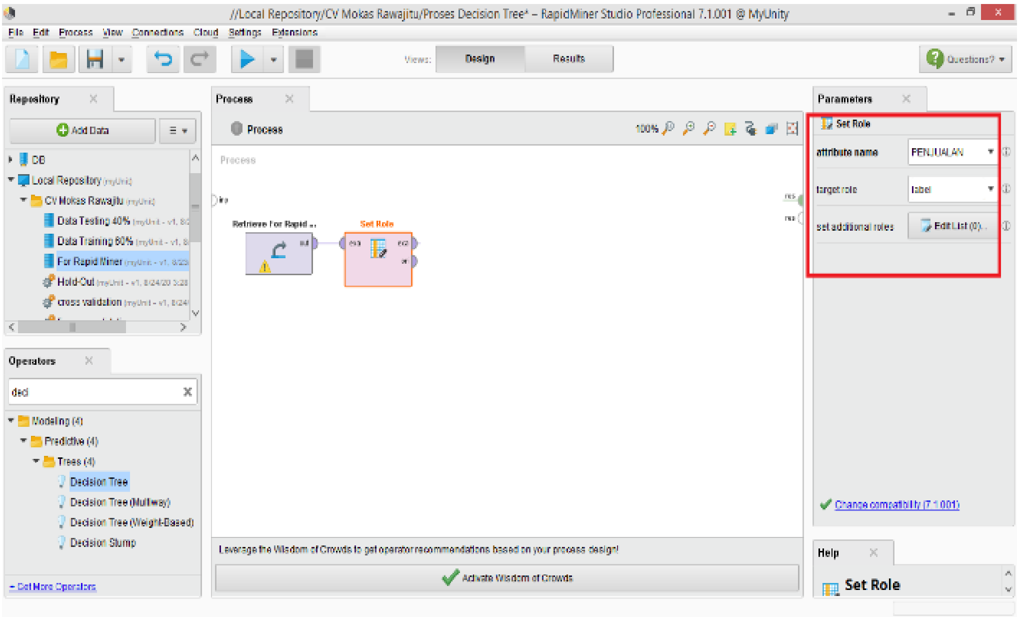

<new process> – RapidMiner Studio Professional 7.1.001 @ MyUnity
File
Edit
Process
View
Connections
Cloud
Settings
Extensions
Views:
Design
Results
Import Data - Where is your data?
Repository
Add Data
1
Samples
DB
Local Repository (myUnit)
Cloud Repository (disconnected)
Where is your data?
My Computer
Database
Get support for more data sources from the RapidMiner Marketplace!
2

Figure 3. Import Data Rapid Miner

2. Next, do a search on the operator Set Role, Set Role is used to define goals that will look for the factors that affect them. Where in this research is the goal? which will be searched for the factors that influence Sales then Sales is the goal, then the target role is changed to label.


//Local Repository/CV Mokas Rawajitu/Proses Decision Tree* – RapidMiner Studio Professional 7.1.001 @ MyUnity

Set Role

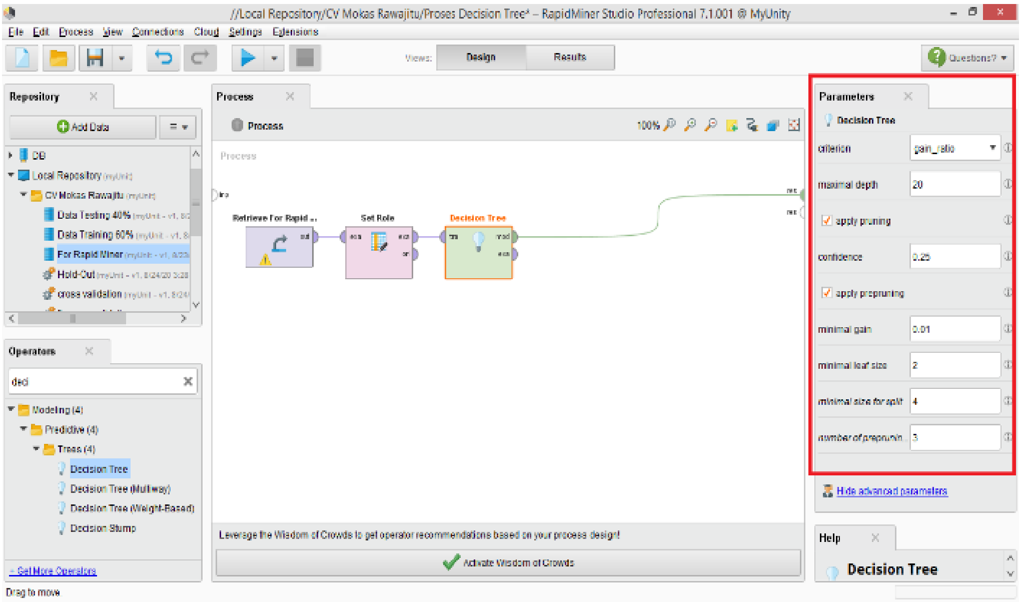
attribute name: PENJUALAN

target role: label

set additional roles: Edit List (0)...

Figure 4. Setting Rule Rapid Miner

3. Applying the C4.5 Algorithm (Decision Tree) Search for Decision Tree on the Operator and drag it to the side of Set Rol then connect it to Set Role and res.


//Local Repository/CV Mokas Rawajitu/Proses Decision Tree* – RapidMiner Studio Professional 7.1.001 @ MyUnity

Decision Tree

criterion: gain_ratio

maximal depth: 20

☑ apply pruning

confidence: 0.25

☑ apply prepruning

minimal gain: 0.01

minimal leaf size: 2

minimal size for split: 4

number of prepruning: 3

Figure 5. Implementation Algorithm C45 Rapid Miner

4. Apply a minimum gain of 0.01 to trim the branches so that the decision tree that is formed is not too dense. After that, run the process for getting the decision tree results from the C4.5 Algorithm. From the implementation of the C4.5 Algorithm on Rapid Miner using the Decision Tree and Decision Tree (Multiway) operators by applying pruning or pre-running from a minimum gain = 0.01, the results are in the form of a decision tree. The results of the Decision Tree from the application of the C4.5 Algorithm on Rapid Miner can be seen in the image below.

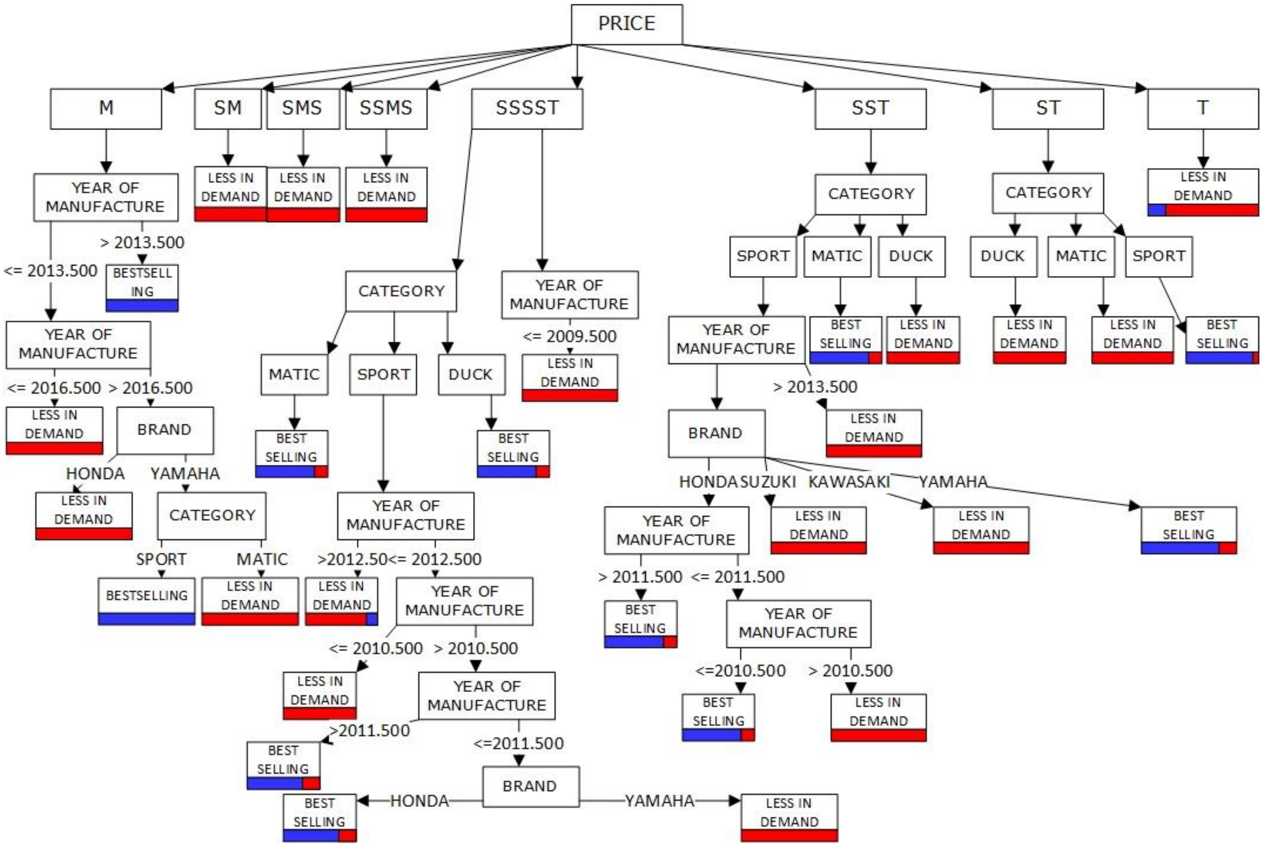
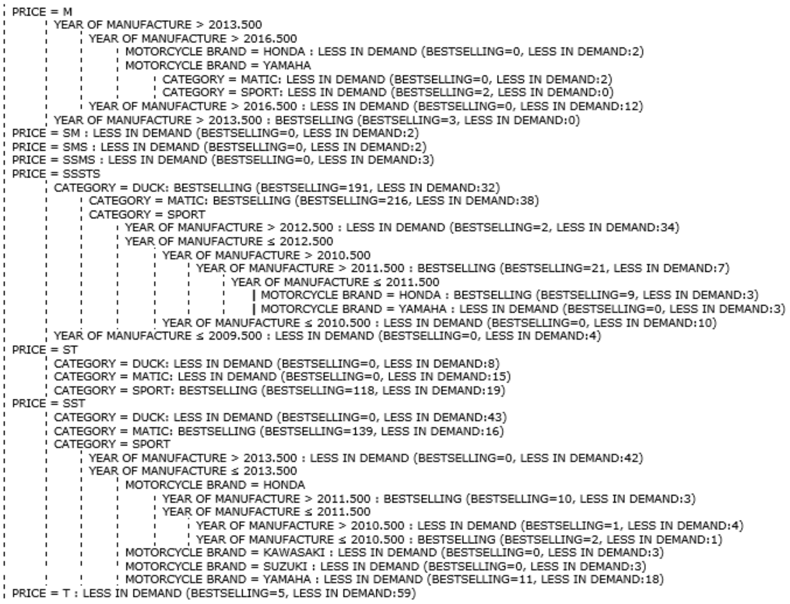


Figure 6. Decision Tree Results

Based on the decision tree formed, then get a description like the following



```
PRICE = M
      YEAR OF MANUFACTURE > 2013.500
           YEAR OF MANUFACTURE > 2016.500
                MOTORCYCLE BRAND = HONDA : LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:2)
                MOTORCYCLE BRAND = YAMAHA
                     CATEGORY = MATIC: LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:2)
                     CATEGORY = SPORT: LESS IN DEMAND (BESTSELLING=2, LESS IN DEMAND:0)
           YEAR OF MANUFACTURE > 2016.500 : LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:12)
      YEAR OF MANUFACTURE > 2013.500 : BESTSELLING (BESTSELLING=3, LESS IN DEMAND:0)
PRICE = SM : LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:2)
PRICE = SMS : LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:2)
PRICE = SSMS : LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:3)
PRICE = SSSTS
      CATEGORY = DUCK: BESTSELLING (BESTSELLING=191, LESS IN DEMAND:32)
           CATEGORY = MATIC: BESTSELLING (BESTSELLING=216, LESS IN DEMAND:38)
           CATEGORY = SPORT
                YEAR OF MANUFACTURE > 2012.500 : LESS IN DEMAND (BESTSELLING=2, LESS IN DEMAND:34)
                YEAR OF MANUFACTURE ≤ 2012.500
                     YEAR OF MANUFACTURE > 2010.500
                          YEAR OF MANUFACTURE > 2011.500 : BESTSELLING (BESTSELLING=21, LESS IN DEMAND:7)
                               YEAR OF MANUFACTURE ≤ 2011.500
                                   | MOTORCYCLE BRAND = HONDA : BESTSELLING (BESTSELLING=9, LESS IN DEMAND:3)
                                   | MOTORCYCLE BRAND = YAMAHA : LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:3)
                     YEAR OF MANUFACTURE ≤ 2010.500 : LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:10)
      YEAR OF MANUFACTURE ≤ 2009.500 : LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:4)
PRICE = ST
      CATEGORY = DUCK: LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:8)
      CATEGORY = MATIC: LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:15)
      CATEGORY = SPORT: BESTSELLING (BESTSELLING=118, LESS IN DEMAND:19)
PRICE = SST
      CATEGORY = DUCK: LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:43)
      CATEGORY = MATIC: BESTSELLING (BESTSELLING=139, LESS IN DEMAND:16)
      CATEGORY = SPORT
           YEAR OF MANUFACTURE > 2013.500 : LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:42)
           YEAR OF MANUFACTURE ≤ 2013.500
                MOTORCYCLE BRAND = HONDA
                     YEAR OF MANUFACTURE > 2011.500 : BESTSELLING (BESTSELLING=10, LESS IN DEMAND:3)
                     YEAR OF MANUFACTURE ≤ 2011.500
                          YEAR OF MANUFACTURE > 2010.500 : LESS IN DEMAND (BESTSELLING=1, LESS IN DEMAND:4)
                          YEAR OF MANUFACTURE ≤ 2010.500 : BESTSELLING (BESTSELLING=2, LESS IN DEMAND:1)
                MOTORCYCLE BRAND = KAWASAKI : LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:3)
                MOTORCYCLE BRAND = SUZUKI : LESS IN DEMAND (BESTSELLING=0, LESS IN DEMAND:3)
                MOTORCYCLE BRAND = YAMAHA : LESS IN DEMAND (BESTSELLING=11, LESS IN DEMAND:18)
PRICE = T : LESS IN DEMAND (BESTSELLING=5, LESS IN DEMAND:59)
```

Figure 7. Decision Tree Results Description

Based on the decision tree formed and the description above, the attributes that affect sales in CV Mokas Rawajitu's sales transactions are the SSSTS Price with the Automatic and Duck Category, the ST Price with the Sports Category, the STS Price with the Automatic and Sports Category with the Honda Brand and the Year of Manufacture 2011.

3.2. Performance Measurement Classifier

The X-Validation test method or commonly called K-Fold Cross Validation is a test method used to evaluate the performance of the model [15], which will determine the level of accuracy of the application of the C45 algorithm to motorcycle sales data at CV Mokas Rawajitu. the use of the K-fold Cross-Validation test method can reduce computational time while maintaining the accuracy of the estimate [16]. The accuracy results from testing k=2, k=3,…k=10 fold cross-validation can be seen in Table 4.5 below. Based on the Table of Cross-Validation test results, it is known that the test using k=2, k=3…k=10 Fold Cross Validation gets the highest accuracy results at fold k=9.

Table 1. Comparison of Accuracy

K-Fold	Accuration	Recall	Precision	Error Rate	False True Rate
Fold-2	95,54%	94,42%	85,58%	14,46%	32,68%
Fold 3	86,25%	95,26%	85,85%	14,46%	32,68%
Fold 4	78,24%	96,42%	86,24%	12,76%	31,60%
Fold 5	87,10%	96,48%	85,81%	12,90%	32,90%
Fold 6	87,60%	96,73%	86,44%	12,40%	31,17%
Fold 7	86,39%	95,79%	85,76%	13,61%	32,90%
Fold 8	87,60%	96,73%	86,44%	12,40%	31,17%
Fold 9	87,95%	97,15%	86,57%	12,05%	30,95%
Fold 10	87,31%	96,52%	86,25%	12,69%	31,60%

Based on this comparison, the calculation from the accuracy to the fold-9's False Positive Rate is as follows:

1. Accuracy is the result of the calculation of all the correct classification values divided by the total number of data. The accuracy value will be said to be good if it is closer to or equal to 1 or 100% in percent numbers.

$$accuration = \frac{\mathrm{TP+TN}}{TP+TN+FN+FP} \times 100\% \qquad (3)$$

$$accuration = \frac{922+319}{922+319+27+143} \times 100\% = 87{,}95\%$$

2. Recall is the number of true positive classifications divided by the total number of true positive and false classes. The sensitivity value can be said to be good if the value is getting closer to or equal to 1.

$$recall = \frac{\mathrm{TP}}{TP+FN} \times 100\% \quad (4)$$

$$recall = \frac{922}{922+27} \times 100\% = 97{,}15\%$$

3. Precision is the number of true positive classifications divided by the total number of true positives and false positives. The precision value can be said to be good if the value is getting closer to or equal to 1.

$$precision = \frac{\mathrm{TP}}{TP+FP} \times 100\% \qquad (5)$$

$$precision = \frac{922}{922+143} \times 100\% = 86{,}57\%$$

4. Error Rate is the result of the calculation of all misclassified values divided by the total number of data. The error rate is said to be good if the value is close to or equal to 0.

$$error\ rate = \frac{\mathrm{FP+FN}}{TP+TN+FN+FP} \times 100\% \qquad (6)$$

$$error\ rate = \frac{143+27}{922+319+27+143} \times 100\% = 12{,}09\%$$

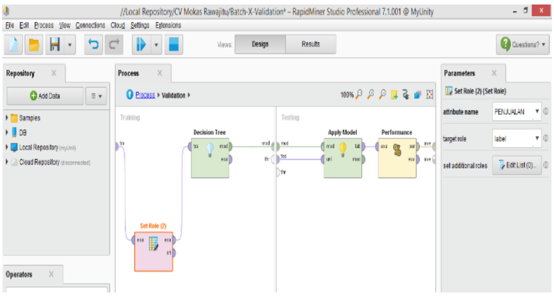
5. The false-positive rate is obtained by counting the number of false-positive classifications divided by the number of negative classes. The value of the false positive rate will be said to be good if it is close to or equal to 0.

$$false\ positive\ rate = \frac{\mathrm{FP}}{TN+FP} \times 100\% \qquad (7)$$

$$precision = \frac{143}{27+143} \times 100\% = 30{,}95\%$$

3.3.Testing Using Batch X Validation

This validation method is used to find the effect of each attribute on the sales attribute. The following is a statistical picture of each attribute. Here are the steps for implementing Batch X Validation on Rapid Miner.



attribute name PENJUALAN

target role label

set additional roles Edit List (0)...

//Local Repository/CV Mokas Rawajitu/Batch-X-Validation* – RapidMiner Studio Professional 7.1.001 @ MyUnity
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Questions?
Repository
Add Data
Samples
DB
Local Repository (myUnit)
Cloud Repository (disconnected)
Operators
Process
Process ▸ Validation ▸
100%
Training
Testing
Decision Tree
Apply Model
Performance
Set Role (2)
Parameters
Set Role (2) (Set Role)

Figure 8. Process in Batch X Validation

Enter the Set Role operator (change attribute name become Sales and target roles become labels) and Decision Tree into the Data Training process. After that input Apply Model and Extract operator Performance in the Data Testing process.

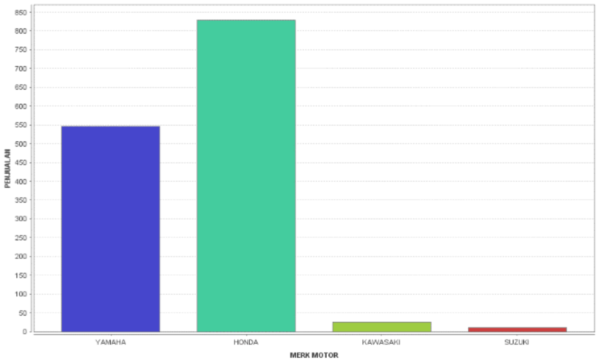


Figure 9. Motorcycle Brand Statistics Against Sale

Based on the graph, it can be concluded that the motorcycle brands that dominate sales are the Yamaha brand with sales > 500 units and the Honda brand with sales > 800 units.

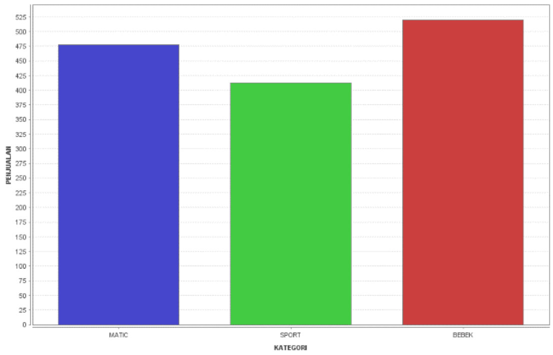


Figure 10. Category Statistics on Sales

From the graph, it can be seen that almost all motorcycle categories (Automatic, Duck, Sport) have high sales, but the duck category is the category that dominates sales, with sales of >500 motorcycle units.

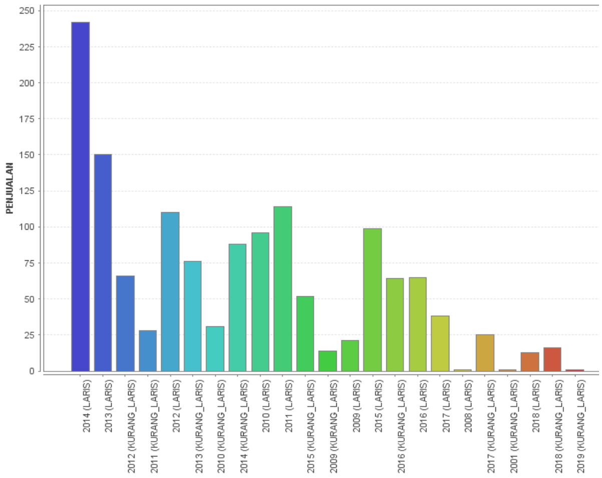


Figure 11. Manufacture Year Statistics Against Sales

Based on the graph, it can be seen that the motorcycles made in 2014 and 2013 were the best selling motorcycles in terms of sales, with sales of > 225 units for motorcycles produced in 2014 and sales of almost 150 units for motorcycles produced in 2014.

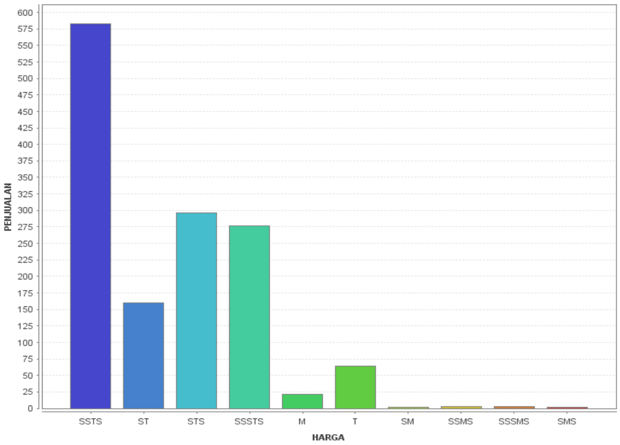


Figure 12. Price Statistics Against Sales

Based on the graph, it can be said that the motorcycle is priced at SSTS with a price range between IDR 7,000,001 - IDR. 10,000,000 is the most dominating motorcycle sales where the total sales reached > 575 units of motorcycles.

The results of this Batch-X-Validation test show that:

1. Motorcycle brand is Honda and Yamaha is brand motorcycles that dominate sales.
2. Almost the entire content of the Category attribute dominates sales, but the highest sales of Category attribute is Duck motor.
3. Motorcycles with the most 2013-2014 Years of Manufacture dominate sales.
4. Motorcycle with SSSTS sales price is the predominant price class.

Based on the test results, the attributes that affect sales on CV Mokas Rawajitu sales transactions are the Price of "SSSTS" with the Category "Automatic" and "Duck", the Price of "ST" with the Category "Sport", the Price of "STS" with the Category "Automatic" and "Sport" with the Brand "Honda" and the Year of Manufacture "2011". With the rules that have been formed, and after testing data using Cross Validation with an accuracy result that reaches 87.32% it can be concluded that the resulting rules using the C4.5 Algorithm can be used as a prediction of motor sales on the CV Mokas Rawajitu. Based on the results of the application of the C4.5 algorithm, the recommendations that can be applied are as a supply of motors that must be ensured the availability of its stock is a motor with a price is SSSTS or with a price range of 4,000,000–7,000,000 for the Automatic and Duck Category, STS or with a price range of Rp10,000,001 - Rp13,000,000 for the Automatic and

Sport Category with the Honda Brand and with the 2011 Manufacturing Year.

4. CONCLUSIONS

The conclusions of research using the C4.5 Algorithm, that can be drawn are as follows is Based on the application of Batch-X-Validation obtained statistical results of each attribute against Sales attributes are as the motorcycle brand is Honda and Yamaha is the motorcycle brand that dominates the sales, almost the entire content of Category attributes dominates sales, but the highest sales of Category attributes are Duck motorbikes, motorcycles with the Year of Manufacture 2013-2014 dominate the sales, motorcycles with SSSTS selling prices are the most dominating price class.

The attributes that are the result of the application of the C4.5 Algorithm can be used as predictions for sales on CV Mokas Rawajitu because the results of the accuracy of data and model testing using K-Fold Cross Validation achieve the value of 87.95% with level, Recall 97.15%, Precision 86.57%, error (Error Rate) 12.05% and False Positive Rates 30.95%.

Cross-Validation testing using Rapid Miner proves that the 9th fold value produces a more accurate accuracy value than the 2nd fold, 10th fold value. So, it is proven that the k-fold value is good for used is 9.

REFERENCES

[1] C. Rygielski, J.-C. Wang, and D. C. Yen, "Data mining techniques for customer relationship management," *Technol. Soc.*, vol. 24, no. 4, pp. 483–502, 2002.

[2] B. Bakhshinategh, O. R. Zaiane, S. ElAtia, and D. Ipperciel, "Educational data mining applications and tasks: A survey of the last 10 years," *Educ. Inf. Technol.*, vol. 23, no. 1, pp. 537–553, 2018.

[3] S. D. Hapid, M. I. Dzulhaq, and T. Mulyono, "Sistem Pendukung Keputusan Penyeleksian Supplier Bahan Produksi Dengan Metode Simple Additive Weighting (SAW)," *vol*, vol. 10, pp. 33–37, 2020.

[4] B. Novianti, T. Rismawan, and S. Bahri, "Implementasi Data Mining Dengan Algoritma C4. 5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak)," *Coding J. Komput. dan Apl.*, vol. 4, no. 3, 2016.

[5] J. Eska, "Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4. 5," 2018.

[6] P. Purwadi, "Implementasi Data Mining Untuk Memprediksi Pola Pembelian Sepeda Motor Pada Showroom CV. Viva Mas Motors Dengan Metode Algoritma C4. 5," *J. Sist. Inf. Kaputama*, vol. 2, no. 2, 2018.

[7] N. Azwanti, "Analisa Algoritma C4. 5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning," *Inform. Mulawarman J. Ilm. Ilmu Komput*, vol. 13, no. 1, p. 33, 2018.

[8] J. Suntoro, *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Elex Media Komputindo, 2019.

[9] A. N. Khormarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komput*, pp. 1–12, 2016.

[10] C. B. Andrianto, K. Kusrini, and H. Al Fatta, "Analisis Sistem Pendukung Keputusan Penerima Beasiswa Di Smp Muhammadiyah 2 Kalasan," *Respati*, vol. 12, no. 34, 2017.

[11] J. S. D. Raharjo, D. Damiyana, and L. Steven, "Perancangan sistem pakar diagnosa penyakit jantung dengan metode forward chaining berbasis android," *J. Sisfotek Glob.*, vol. 7, no. 2, 2017.

[12] R. Tullah, S. M. Mustafa, and A. Rochim, "Sistem Pakar Pendeteksi Penyakit Diabetes Mellitus Menggunakan Algoritma Fuzzy Logic Takagi Sugeno Kang," *J. SISFOTEK Glob.*, vol. 9, no. 2, 2019.

[13] L. Elvitaria, "Memprediksi Tingkat Peminat Ekstrakurikuler pada Siswa SMK Analisis Kesehatan Abdurrab Menggunakan Algoritma C4. 5 (Studi Kasus: SMK Analis Kesehatan Abdurrab)," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 2, no. 2, pp. 110–124, 2017.

[14] D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di Twitter," *J. Tekno Kompak*, vol. 14, no. 2, pp. 86–91, 2020.

[15] R. Agusli, L. F. Gustomi, and G. Prasetyo, "Sistem Penunjang Keputusan Dalam Pemilihan Siswa Berprestasi Menggunakan Metode Promethee," *J. SISFOTEK Glob.*, vol. 9, no. 1, 2019.

[16] E. Prasetyo, "Data mining mengolah data menjadi informasi menggunakan matlab," 2014.