Classification Model C.45 on Determining the Quality of Customer Service in Bank BTN Pematangsiantar Branch

Meilin Widyastuti¹, Agnes Gracella Fepdiani Simanjuntak¹, Dedy Hartama², Agus Perdana Windarto², Anjar Wanto³

¹Students of STIKOM Tunas Bangsa Pematangsiantar, North Sumatera, Indonesia
²STIKOM Tunas Bangsa Pematangsiantar, North Sumatera, Indonesia
*agus.perdana@amiktunasbangsa.ac.id

Abstrak—This study aims to classify the quality of customer service at Bank BTN Pematangsiantar Branch. Data is obtained from the results of the customer questionnaire of Bank BTN Pematangsiantar Branch. Attributes used are 7, namely Age (C1), Job (C2), Old to Customer (C3), Tangible (C4), Reliability (C5), Assurance (C6), Responsiveness (C7). Moted used in this research is C4.5 Algorithm and assisted by Rapidminer software to make decision trees. From the results of the study, there were 5 rules for classification in determining the quality of customer service with 3 rules with satisfaction status and 2 rules with dissatisfied status. The C.45 algorithm can be used in the case of determining customer service quality of Bank BTN Pematangsiantar Branch with an accuracy of 77.78%. With this analysis, it is expected that Bank BTN can improve the quality of service to customer satisfaction.

1. Introduction

Customer satisfaction is very important in assessing the level of management and services provided by the bank to its customers [1]. Republic of Indonesia Law Number 10 of 1998 concerning Banking (Article 1 and 2) Bank is a business entity that collects funds from the public in the form of deposits and distributes them in the form of credit and/or other forms with the aim of improving the lives of many people.[2] The existence of banking services in the community is indeed more profitable, especially in the economic sector, where economic actors are more flexible in carrying out their economic activities to support survival. Banking service businesses in the community that promote good service in order to gain trust from the community as their customers will face various conditions or views that arise from the community as an expression of satisfaction or dissatisfaction with the services they receive from the bank they trust[3]. Many branches of computer science can solve complex problems. This is evidenced by several studies in the field of datamining [4][5–8][9][10], field of artificial neural networks [11–15], in the field of decision support systems [16–20][21].

The C4.5 algorithm is a widely used decision tree classification algorithm because it has the main advantages of other algorithms [9][22]. The C4.5 algorithm is the development of ID3 [23]. Some of the developments carried out in C4.5 are as follows, including overcoming missing values, being able to overcome continuous data, and training. The advantages of the C4.5 algorithm can produce decision trees that are easy to interpret, have an acceptable level of accuracy, are efficient in handling discrete type attributes and can handle discrete and numeric type attributes. In constructing a tree, the C4.5 algorithm reads the entire training data sample from storage and loads it into memory[24]. This is one of the weaknesses of the C4.5 algorithm in the "scalability" category. This algorithm can only be used if the training data can be stored as a whole and at the same time in memory. By using C.45 Algorithm the application of Data Mining will result in a process of determining the quality of customer service at Bank BTN Pematangsiantar Branch. Based on these data, the variables that will be used in the research are determined. The independent variables used are Tangible, Reliability, Assurance, Responsiveness.
While the response variable is satisfaction, namely customers who are satisfied and dissatisfied with the service of Bank BTN Pematangsiantar Branch.

2. Methodology

2.1. Data Mining

Data mining is the process of using large data sets to gather important hidden knowledge. It is divided into seven steps like data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation[9]. Data mining is used for any kind of data, including database data, data warehouse data or transactional data[25].

2.2 Classification

Classification by decision tree, a classification that is very reliable as a means of classification and prediction. In general, a decision tree is a modeling picture of a subject that is composed of a series of decisions that provide a solution. Classification rules can be easily presented in easy language so that users can understand it, or in its database as Microsoft Excel, Microsoft Access and SQL. In some applications, and prediction accuracy of classification is important, for example, predict potential customers in the business market. There are various algorithms to build a decision tree, the most popular is the ID3 and C4.5[23].

2.3. C4.5 Algorithm

C4.5 algorithm is an algorithm to form a decision tree by counting the value of the gain, where the biggest gains are to be used as an initial node or the root node. C4.5 algorithms step in building a decision tree[26]. There are several steps in making a decision tree in the C4.5 algorithm, namely:

a) Prepare the training data. Training data is usually taken from historical data that had happened before or called past data and has been grouped into certain classes [22] [26].

b) Calculate the roots of a tree. The root will be taken from the attribute that will be selected, by calculating the gain value of each attribute, the highest gain value which will be the first root. Before calculating the gain value of an attribute, first, calculate the entropy value. To calculate the entropy value, the formula is used:

$$Entropy(S) = \sum_{i=1}^{n} - p_i \log_2 p_i$$

Information:

- $S$ = case set
- $n$ = number of partitions $S$
- $p_i$ = the proportion of $S_i$ to $S$

c) Calculate the Gain value using the equation:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \cdot Entropy(S_i)$$

Information:

- $S$ = case set
- $A$ = feature
- $n$ = number of partition attributes $A$
- $|S_i|$ = The proportion of $S_i$ to $S$
- $|S|$ = number of cases in $S$

d) Repeat step 2 and step 3 until all records are partitioned

e) The decision tree partition process will stop when:

1. All records in N node get the same class.
2. There is no attribute in the record that is partitioned again. There are no records in an empty branch.

3. Result and Discussion

This study includes customers of Bank BTN Pematangsiantar Branch. The results of the recapitulated questionnaire can be seen in the following table:
| Alternative | A         | B             | C | D | E | F | G   | Response |
|-------------|-----------|---------------|---|---|---|---|-----|----------|
| A1          | Adult     | Private employees | CL | SB | SB | SB | C   | Satisfied |
| A2          | Adult     | Entrepreneur    | BR | B  | B  | B  | B   | Satisfied |
| A3          | Adult     | Private employees | BR | SB | B  | C  |     | Satisfied |
| A4          | Adult     | Entrepreneur    | L  | SB | B  | SB | SB  | Satisfied |
| A5          | Adult     | Private employees | BR | SB | B  | SB | B   | Satisfied |
| A6          | Adult     | Private employees | BR | SB | SB | SB | B   | Satisfied |
| A7          | Teenagers | Student        | BR | SB | B  | SB | SB  | Satisfied |
| A8          | Adult     | BUMN employee   | CL | SB | SB | SB | B   | Satisfied |
| A9          | Adult     | BUMN employee   | L  | SB | SB | SB | SB  | Satisfied |
| A10         | Adult     | Private employees | BR | SB | B  | B  | C   | Satisfied |
| A11         | Teenagers | Student        | BR | SB | B  | B  | B   | Satisfied |
| A12         | Adult     | Private employees | CL | B  | B  | B  | B   | Satisfied |
| A13         | Adult     | Entrepreneur    | BR | B  | B  | B  | B   | Satisfied |
| A14         | Adult     | Entrepreneur    | CL | B  | B  | B  | B   | Satisfied |
| A15         | Old       | BUMN employee   | L  | SB | SB | SB | B   | Satisfied |
| A16         | Adult     | BUMN employee   | CL | SB | SB | SB | SB  | Satisfied |
| A17         | Adult     | PNS            | CL | C  | B  | C  | B   | Satisfied |
| A18         | Adult     | Private employees | BR | C  | B  | C  | C   | Not satisfied |
| A19         | Old       | Private employees | CL | C  | C  | B  | C   | Not satisfied |
| A20         | Adult     | Student        | BR | B  | C  | C  | B   | Not satisfied |
| A21         | Teenagers | Student        | BR | B  | C  | C  | C   | Not satisfied |
| A22         | Adult     | Private employees | BR | C  | C  | C  | C   | Not satisfied |
| A23         | Adult     | Student        | BR | C  | B  | C  | B   | Not satisfied |
| A24         | Teenagers | Student        | CL | B  | SB | B  | SB  | Satisfied |
| A25         | Old       | BUMN employee   | L  | SB | B  | SB | B   | Satisfied |
| A26         | Teenagers | Student        | L  | SB | SB | SB | SB  | Satisfied |
| A27         | Adult     | PNS            | L  | B  | B  | B  | B   | Satisfied |
| A28         | Teenagers | Student        | BR | B  | B  | B  | B   | Satisfied |
| A29         | Teenagers | Student        | CL | B  | SB | B  | SB  | Satisfied |
| A30         | Old       | PNS            | L  | B  | B  | B  | B   | Satisfied |

Information:
A = Age          E = Reliability
B = Work         F = Assurance
C = Long Being a Customer   G = Responsiveness
D = Tangible

Preprocessing results obtained are then processed into the Decision tree using RapidMiner software to determine the classification of customer service quality determination [27]. Data that has been successfully imported is then processed into the Decision tree with the following data processing model [28]:

![Figure 1. Connection between Data and Decision Tree Models](image)

The results of data processing with the decision tree model, can be seen in Figure 2 as follows:
After the formation of rules/rules as many as 5 rules where there are 3 rules that are successfully classified as Satisfied, and the remaining rules after being classified are as many as 2 rules with Unsatisfactory values. The rules above prove that the C.45 algorithm can be applied in predicting the quality of service to customer satisfaction of Bank BTN Pematangsiantar Branch Office [28]. In research supported by the results of data processing using RapidMiner software, the accuracy of the C4.5 algorithm is 77.76%. To see the value of accuracy can be seen in Figure 4 below[29]:

| accuracy: 77.76% | true PUAS | true TIDAK PUAS | class precision |
|-----------------|-----------|----------------|-----------------|
| pred PUAS       | 7         | 2              | 77.76%          |
| pred TIDAK PUAS | 0         | 0              | 0.00%           |
| class recall    | 100.00%   | 0.00%          |                 |

**Figure 4.** Value of accuracy of the C4.5 algorithm

4. Conclusion

Based on the results and discussion it can be concluded that with the analysis in predicting the quality of service to customer satisfaction of Pematangsiantar Branch Stasse Savings Bank, the level of customer satisfaction at the Pematangsiantar Branch Savings Bank can be measured clearly. After analyzing in predicting the level of customer satisfaction with bank services to customers, in some aspects that become a benchmark, it is known that the reliability quality aspect is the most dominant of several aspects.

References

[1] M. Haddad and H. Fecht, “Behavior of C45 steel Subjected to Different High Pressure Torsion (HPT),” vol. 5, no. 1, pp. 144–153, 2015.

[2] T. Leppert, “SURFACE LAYER PROPERTIES IN DRY TURNING OF C45 STEEL.”

[3] P. D. Cahyani, “Tingkat Kepuasan Nasabah Terhadap Kualitas Layanan Perbankan Syariah di Yogyakarta,” Esensi, vol. 6, no. 2, pp. 151–162, 2016.

[4] A. P. Windarto, “Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method,” Int. J. Artif. Intel. Res., vol. 1, no. 2, pp. 26–33, 2017.

[5] U. R. Raval and C. Jani, “Implementing and Improvisation of K-means Clustering,” Int. J. Comput. Sci. Mob. Comput., vol. 5, no. 5, pp. 72–76, 2016.

[6] M. K. Arzoo, A. Prof, and K. Rathod, “K-Means algorithm with different distance metrics in spatial data
mining with uses of NetBeans IDE 8.2,” Int. Res. J. Eng. Technol., vol. 4, no. 4, pp. 2363–2368, 2017.

[7] S. Kumar and S. K. Rath, “Performance Evaluation of K-Means Algorithm and Enhanced Mid-point based K-Means Algorithm on Mining Frequent Patterns,” Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 4, no. 10, pp. 545–548, 2014.

[8] A. Yadav and S. Dhingra, “An Enhanced K-Means Clustering Algorithm to Remove Empty Clusters,” IJEDR, vol. 4, no. 4, pp. 901–907, 2016.

[9] B. Supriyadi, A. P. Windarto, T. Soemartono, and Mungad, “Classification of natural disaster prone areas in Indonesia using K-means,” Int. J. Grid Distrib. Comput., vol. 11, no. 8, pp. 87–98, 2018.

[10] R. Rahim et al., “Enhanced pixel value differencing with cryptography algorithm,” in MATEC Web of Conferences 197, 2018, vol. 3011, pp. 1–5.

[11] B. Febriadi, Z. Zamzami, Y. Yunefri, and A. Wanto, “Bipolar Function in Backpropagation Algorithm in Predicting Indonesia’s Coal Exports by Major Destination Countries,” in IOP Conference Series, 2018, pp. 1–8.

[12] N. Nasution, A. Zamsuri, L. Lisnawita, and A. Wanto, “Polak-Ribiere updates analysis with binary and linear function in determining coffee export in Indonesia,” IOP Conf. Ser. Mater. Sci. Eng., vol. 420, no. 12089, pp. 1–9, 2018.

[13] Sumijan, A. P. Windarto, A. Muhammad, and Budiharjo, “Implementation of Neural Networks in Predicting the Understanding Level of Students Subject,” Int. J. Softw. Eng. Its Appl., vol. 10, no. 10, pp. 189–204, 2016.

[14] A. Wanto, M. Zarlis, Sawaluddin, and D. Hartama, “Analysis of Artificial Neural Network Backpropagation Using Conjugate Gradient Fletcher Reeves in the Predicting Process,” J. Phys. Conf. Ser., vol. 930, no. 1, pp. 1–7, 2016.

[15] A. P. Windarto, L. S. Dewi, and D. Hartama, “Implementation of Artificial Intelligence in Predicting the Value of Indonesian Oil and Gas Exports With BP Algorithm,” Int. J. Recent Trends Eng. Res., vol. 3, no. 10, pp. 1–12, 2017.

[16] A. N. D. J. D. Fadhilah, “Perancangan Aplikasi Sistem Pakar PENyakit Kulit Pada Anak Dengan Metode Expert System Development Life Cycle,” J. Algoritm. Sekol. Tinggi Teknol. Garut, vol. 9, no. 13, pp. 1–7, 2012.

[17] S. Fekri-Ershad, H. Tajalizadeh, and S. Jafari, “Design and Development of an Expert System to Help Head of University Departments,” Int. J. Sci. Mod. Eng., vol. 1, no. 2, pp. 45–48, 2013.

[18] M. Min, “A rule based expert system for analysis of mobile sales data on fashion market,” 2013 Int. Conf. Inf. Sci. Appl. ICISA 2013, 2013.

[19] M. Mohammadi and S. Jafari, “An expert system for recommending suitable ornamental fish addition to an aquarium based on aquarium condition,” arXiv Prepr. arXiv1405.1524, vol. 3, no. 2, pp. 1–7, 2014.

[20] I. Chen and B. L. Poole, “Performance Evaluation of Rule Grouping on a Real-Time Expert System Architecture,” vol. 6, no. 6, pp. 883–891, 2014.

[21] Sudirman, A. P. Windarto, and A. Wanto, “Data mining tools | rapidminer: K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia,” IOP Conf. Ser. Mater. Sci. Eng., vol. 420, p. 12089, 2018.

[22] R. H. Pambudi, B. D. Setiawan, and Indriat, “Penerapan Algoritma C4.5 Dalam Program Untuk Memprediksi Kinerja Siswa Sekolah Menengah,” J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 2, no. 7, pp. 2637–2643, 2017.

[23] I. A. Ganiyu, “Data Mining : A Prediction for Academic Performance Improvement of Science Students using Classification,” Int. J. Inf. Commun. Technol. Res., vol. 6, no. 4, 2016.

[24] R. Shaun, J. De Baker, and J. E. B. Eds, “Educational Data Mining 2008 The 1st International Conference on Educational Data Mining,” Network, no. January, p. 187, 2008.

[25] R. Revathy and R. Lawrance, “Comparative Analysis of C4.5 and C5.0 Algorithms on Crop Pest Data,” Int. J. Innov. Res. Commun. Comput. Eng., vol. 5, no. 1, pp. 50–58, 2017.

[26] S. K. Chaturvedi, “Attack Detection over Network based on C45 & RF Algorithms,” vol. 15, no. 9, pp. 68–73, 2015.

[27] J. Kratochvil, J. Petru, M. Pagac, J. Holubjak, and J. Mrazik, “EFFECT OF CHIP BREAKERS ON THE CUTTING FORCE DURING THE MACHINING OF STEEL C45,” vol. 11, no. March, pp. 173–178, 2017.

[28] M. Turnip, “Penerapan algoritma e4.5 untuk penentuan tingkat konsumsi konsumen pada medan solusindo,” no. February, 2017.

[29] S. Sudirman, A. P. Windarto, and A. Wanto, “Data Mining Tools | RapidMiner : K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops In Indonesia,” IOP Conf. Ser. Mater. Sci. Eng., vol. 420, no. 12089, pp. 1–8, 2018.