Data Mining Application for Classification of Online Transportation Customer Satisfaction Using C4.5 Algorithm

Arie Restu Wardhani [1], Ryan Avrilio Irawan [2] , Fhadillah Ain Marpaung [3] , Idris Ivan Saputra [4] , Anastasia L Maukar [5]

[1]Department of Industrial Engineering, University of Widyagama Malang , Jl. Borobudur No. 35 Malang , Indonesia
[2,3,4]Department of Informatics Engineering, University of Widyagama Malang , Jl. Borobudur No. 35 Malang , Indonesia
[5]Department of Industrial Engineering, President University , JABABEKA, Indonesia

Article Info

Article history:

Received January 11, 2024
Revised January 29, 2024
Accepted February 14, 2024

Keywords:

Data Mining
C.4.5 Algorithm
Transportation Optimization
Data Analysis
Online Motorcycle Taxi

ABSTRACT

In the era of increasing business competition, transportation companies are required to enhance the efficiency and effectiveness of their services. One method that can be employed to optimize fleet management is through Data Mining analysis. This study focuses on optimizing Ojek online transportation services using the C.4.5 Algorithm method. The aim of this research is to group customers and areas based on service demand patterns, thus improving fleet distribution and reducing waiting times. The data used in this study includes location, demand, and trip frequency information. The analysis results show that the C.4.5 algorithm method effectively groups the data, providing optimal fleet distribution and enhancing service performance. This research demonstrates that applying data mining through the C.4.5 algorithm method can be an effective strategy for improving management and operational efficiency in Ojek online transportation services, offering competitive advantages in service efficiency and customer satisfaction.

Corresponding Author:

Ryan Avrilio Irawan
Department of Informatics Engineering
Faculty of Engineering, University of Widyagama Malang
Jl. Borobudur No. 35 Malang, East Java
Email: ryanaprilio1204@gmail.com

1. INTRODUCTION

The online transportation industry in Indonesia continues to grow with the presence of various service providers such as Maxim, FastGo , BitCar , Bonceng , and Anterin.id, in addition to Gojek and Grab. Online transportation services aim to move people and goods from one location to another. Maxim offers the GoBike service, which allows users to get a delivery or pick-up service using a motorbike. To improve the quality of service, Maxim provides a rating feature in its application, which allows customers to provide feedback on the services they receive. However, Maxim faces several challenges in measuring the level of customer satisfaction, especially GoBike users, caused by factors such as price, facilities, driver service, and company loyalty to customers [1] .

Along with the development of technology, online transportation services such as Maxim have made it easier for people to meet their mobility needs. The presence of this service is very helpful, especially in urban areas that often experience congestion due to dense population and excessive use of private vehicles. However, although Maxim has become an important choice for many people, there are a number of problems that affect customer satisfaction, such as punctuality, fleet comfort, and ease of ordering [2] .

ISSN: xxxx-xxx

The tight competition between online transportation service providers, such as Gojek , Grab, and Maxim, is also a challenge in itself. To remain competitive, Maxim must continue to innovate and provide excellent service that can meet customer expectations. This study aims to measure the level of Maxim user satisfaction using quantitative methods and data analysis through the C4.5 algorithm, which shows that security factors, service availability, and waiting time greatly affect user satisfaction.

Maxim Transportation Online faces various significant challenges in its efforts to classify the level of customer satisfaction, especially for GoBike service users. Although Maxim has tried to provide the best service, there are a number of problems that hinder the achievement of optimal levels of satisfaction. Some of the main factors that cause customer dissatisfaction include prices that are considered less competitive, inadequate facilities compared to competitors, driver services that have not met expectations, and the level of company loyalty that is considered still lacking in maintaining long-term relationships with customers . Based on data obtained in 2021, around 17.53% of customers expressed their dissatisfaction with various aspects provided by Maxim [1] .

This problem becomes even more complex considering that Maxim needs to classify customer satisfaction levels more effectively. So far, customer satisfaction level classification has been done manually using applications such as Microsoft Office Excel. This process groups customers based on satisfied and dissatisfied categories regarding various aspects such as price, facilities, driver service, and company loyalty. However, this manual process has limitations, especially in managing large amounts of data and analyzing the relationship between interrelated variables. Therefore, to increase the effectiveness of managing and analyzing customer satisfaction data, Maxim requires a more sophisticated and automatic system.

The solution that can be applied to overcome this problem is to build a data mining-based application system that utilizes the C4.5 algorithm. The C4.5 algorithm is one of the effective methods for big data classification and can be used to analyze customer data more quickly and accurately. By using this algorithm, Maxim can produce a decision tree that describes the relationship between factors that influence customer satisfaction levels, such as price, facilities, driver service, and company loyalty. Through this system, Maxim is expected to be able to classify customer satisfaction levels more efficiently, so that it can make more appropriate decisions in an effort to improve service quality and retain customers [1] .

Maxim Transportation Online, as one of the online transportation service providers, faces various obstacles in classifying customer satisfaction levels, especially for GoBike service users who use motorbikes as a means of transportation. Some of the main problems that arise in measuring this level of satisfaction include customer dissatisfaction with various aspects, such as prices that are considered less competitive, inadequate facilities, inconsistent service quality from drivers, and lack of company loyalty in maintaining good relationships with customers. Based on the analysis conducted

2. METHOD

Concepts in mutually exclusive research related , depiction variable One with other Can connected in detail and systematically , Figure 1.

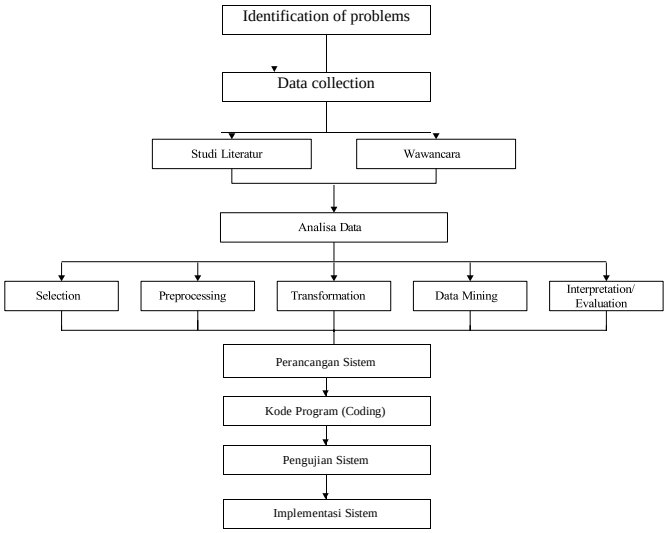


Figure 1. Research Stages

Description framework Work study :

2.1 Identification Problem

Step beginning in do study is identify problem . At the stage this , formulation problem must done with appropriate for the application of data mining using C4.5 algorithm for classify satisfaction customer online transportation can give optimal benefits for Maxim Transportation Online. This is covers observation to factors or constraint in classify level satisfaction user GoBike , so that Maxim Transportation Online can avoid mistakes and difficulties in analyze as well as classify satisfaction customers , in particular customer online transportation .

2.2 Data Collection

a. Study Literature
Data obtained from various source like books electronic (ebook), journal scientific , articles research , and internet site searches . The results of studies literature This is the collection relevant references with formulation problem For strengthen problem , building base theory , and become reference in design control and simulation flow in data mining applications .

b. Interview
Data collection was carried out through interview direct with Maxim Order section on Maxim Transportation Online. Respondents main in interview This is Mr. Muhammad Farizi as head of Maxim Order.

2.3 Data Analysis

After the data is complete collected , steps next is do processing or data analysis . Research This use C4.5 algorithm for data analysis . In In general , the C4.5 algorithm has a number of step in build tree decision :

1. Choose attribute as root .
2. For branch For every mark attributes .
3. For case to in each branch .

ISSN: xxxx-xxx

4. Repeat the process for every branch until all case in branch own same class .
5. Election attribute as root based on the highest Gain value from available attributes . Gain calculation is performed with use formula following :
   Gain(S,A) = Entropy (S) - ∑ n |S i | *Entropy(S) [3]
   I =1 |S| I
6. Before count Gain value , need counted Entropy value first before . Entropy is used For determine how much informative A attribute in produce decision . Formula The basis of Entropy is as following :
   Entropy(S,A)= - ∑n Pi* Log2Pi [4]

   Information :

   S : Set case .
   A : Attributes .
   n : Amount partition on S.
   P_ i : Proportion from a subset of .

7. If There is attributes that have Lots value , need calculate Gain Ratio. Before that , it is necessary calculated Split Information, which is formulated as following :
   Split Info (S,A)= ∑c Si Log2 Si [5]
   i =1 S S

   Information :

   S : Data set used For training .
   A : Attributes .
   S_ i : Amount sample for subset .
   : Number of subsets generated from breakdown use attributes .

After the Split Information value is calculated , the Gain Ratio can be counted use formula :
Gain Ratio (S,A)= $\frac{\text{Gain (S,A)}}{\text{SplitInfo (S,A)}}$ [6]

Steps This ensure that algorithm choose the most informative attribute For build tree decision with high efficiency and accuracy .

Steps data analysis with The C4.5 algorithm is :
1. Data Selection
Selecting level data satisfaction customer online transportation (especially user GoBike on Maxim Transportation Online). Data selected based on relevant attributes For reduce room scope research [7] .
2. Pre-Processing
Cleaning level data satisfaction customers , namely delete information that is not relevant , check data inconsistencies , and fix them error [8] .
3. Transformation
Convert data to in a format that can be processed in a way mathematical For ensure data validity and avoid data corruption during the classification process [9] .
4. Data Mining
Do data classification with C4.5 algorithm for find pattern or rules that can used For predict level satisfaction customers [10] .
5.Interpretation/Evaluation
Implement and evaluate the resulting pattern from the data mining process [11] . The results of the analysis served in an easy format understood by the parties user system , namely Maxim Order on Maxim Transportation Online.

2.4 Design System

Design system is a process that includes various activity For explain in detail how system will function . Modeling system done using UML (Unified Modeling Language), which consists of from use case to describe functionality system , activity diagram for visualize channel Work system , and class diagram as representation objects used as base in development system .

2.5 Program Code (Coding)

Stage making program code aims For simplify Name column during the data entry process (entering) or tabulate data). The coding process This use Language PHP programming and utilizing MySQL database.

2.6 Testing System

Testing system aiming For verify functionality overall system and ensure Handling error Already in accordance with what is expected . Testing beginning done on the localhost server, such as : http://localhost/customer-satisfaction-classification/ .

2.7 Implementation System

Implementation is stage end in the process of implementation system new , where the system prepared For operated . Stage This aiming For to realize the system that has been designed previously .

3. RESULTS AND DISCUSSION

Data is obtained from the Maxim Order section as input with a total of 155 data. This data is divided into two parts:

- 100 data for training.
- 55 data for testing.

Table 1. Scale *Likert* Satisfaction Customer

Choice	*Star*	Abbreviation	Score
Very Satisfied	★★★★★	SP	5
Satisfied	★★★★	P	4
Enough Satisfied	★★★	CP	3
No Satisfied	★★	TP	2
Very No	★	STP	1

The data selection stage is carried out to obtain data that meets the needs in classifying satisfaction. customer user transportation on line, needed data classification results satisfaction customer Which in can from previous data collection. In this data selection stage, the types of classification attributes are first determined with the aim of matching the needs of the system to be built, Price (A1), Facilities (A2), Service (A3), Loyalty (A4), Category (A5). After determining the attributes in classifying customer satisfaction of online transportation users, the next step is to group the attributes .

In the data mining process, there is often incomplete data, which can be caused by missing values or errors in attributes. In the analysis of the classification of customer satisfaction levels of Maxim online transportation service users, some attributes may be irrelevant, so a *data preprocessing stage is needed* to adjust the database to the specified standards. The C4.5 algorithm, which is used to build a decision tree ,

ISSN: xxxx-xxx

involves two types of attributes, namely input attributes and target attributes. These attributes are used to make comparisons in the calculation of *Gain* and *Ratio* .

Steps in determine node root the is as following:

a. Calculation *Entropy*

Step beginning algorithm C4.5 is look for *entropy* value , use formula equality :

$$Entropy(S) = (-\frac{S1}{S} * Log_2(\frac{S1}{S})) + (-\frac{S2}{S} * Log_2(\frac{S2}{S})) \quad [12]$$

b. Calculation *Information Gains*

Results *information* calculation *gain* Information Gains Attribute Price = 0.123, Information Gains Attribute Facility = 0.003, Information Gain Service Attribute = 0.007, Information Gain Loyalty Attribute = 0.016.

c. Calculation *Split Information*

Split information used as divider from *Gains (A)* Which will produce *gain ratio* . calculation *split info* from every attribute with use formula equality following This

$$Split\ Info = -\sum_{I=1}^{n} \frac{|S_i|}{|s|} Log_2 \frac{|S_i|}{|s|} \quad [13]$$

Results calculation Split Information Attribute Price = 1,867, Split Information Attribute Facility =1.801, Split Information Service Attributes

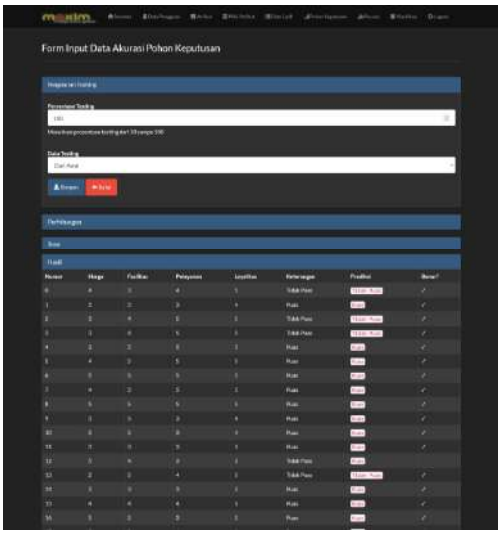
= 1.966, Split Information Attribute Loyalty = 2.009.

d. Calculation *Gains* Calculation *Ratio gain ratio* in classify satisfaction customer transportation on line:

$$GainRatio(A) = \frac{Gain(A)}{SplitInfoA(D} \quad [10]$$

Calculation results GainRatio Attribute Price = 0.066, GainRatio Attribute Facilities = 0.002, GainRatio Attribute Service = 0.003, GainRatio Attribute Loyalty = 0.008.

Implementation system , the program is hosted on the server so that it can be accessed . For example on the localhost server: http://localhost/klasifikasi-kepuasan-pelanggan-maxim/. Implementation of Accuracy Test Result Output Tree The decision is appearance Implementation of the output form for accuracy test results classification satisfaction customer user online transportation



Picture1. mplementation of Decision Tree Accuracy Test Result Output

The data mining application to classify customer satisfaction of online transportation users has been successfully implemented in the system by applying the *C4.5 Algorithm* , the results of the classification of customer satisfaction of online transportation users, the number of online transportation users with the classification of "Dissatisfied" is 91 customers, while those who are "Dissatisfied" are 9 customers.

e. CONCLUSION

System classification satisfaction customer user service this online transportation use data mining method with implementation C4.5 Algorithm . The process carried out covering calculation mark entropy , information split info, gain, and gain ratio, which then produce tree decision as output for determine level satisfaction Customer . Application This can give information in the form of results analysis as well as calculation based on method C4.5, accompanied by with appearance tree prepared decision for admin. Testing done using 100 training data consisting from four attribute classification , namely price , facilities , service , and loyalty . From the results said , obtained tree decisions that describe level satisfaction customer user online transportation .

ACKNOWLEDGEMENTS

Assalamualaikum Wr. Wb First of all, for the preparation of this journal, we received many challenges, obstacles and things that made it difficult for us, therefore we are grateful to Allah SWT because of the abundance of His grace we were able to complete this research well.

We would also like to thank and appreciate our lecturer, Mrs. Fitri Marisa, S.Kom., M.Pd., Ph.D., because she supported, helped, and gave us direction so that we could complete this research journal well.

We also do not forget to thank the entire team involved in this research, because of their dedication this research journal can be completed, once again we say thank you WassalamualaikumWr.Wb

REFERENCES

[1] E. Jannah, V. Sihombing, and M. Masrizal, “Application of Data Mining Classification of Online Transportation Customer Satisfaction Using C4.5 Algorithm,” *MEANS (Media Inf. Anal. and Sist.* , vol. 8, no. 1, pp. 1–7, 2023, doi: 10.54367/means.v8i1.2569.

[2] Suherman, D. Maulana, and V. Mustikaningtyas, “Application of Data Mining for Online Transportation Customer Satisfaction Classification (Online Motorcycle Taxi) using C4.5 Algorithm,” *Pros. SAINTEK* , vol. 1, no. 1, pp. 165–175, 2022.

[3] A. Febriyani, GK Prayoga, and O. Nurdiawan, “Informa Customer Satisfaction Index Using the C.45 Algorithm,” *JURIKOM (Jurnal Ris. Komputer)* , vol. 8, no. 6, p. 330, 2021, doi: 10.30865/jurikom.v8i6.3686.

[4] AS Sinaga, MH Munandar, and AS Sitio, “Machine learning algorithm to identify fraud emails with feature selection,” *IOP Conf. Ser. Mater. Sci. Eng.* , vol. 1088, no. 1, p. 012011, 2021, doi: 10.1088/1757-899x/1088/1/012011.

[5] AH Nasrullah, “Application of C4.5 Method for Classification of Students with Potential to Drop Out,” *Ilk. J. Ilm.* , vol. 10, no. 2, pp. 244–250, 2018, doi: 10.33096/ilkom.v10i2 .300.244-250.

[6] E. Elisa, “Analysis and Application of C4.5 Algorithm in Data Mining to Identify Factors Causing Construction Accidents at PT. Arupadhatu Adisesanti,” *J. Online Inform.* , vol. 2, no. 1, p. 36 , 2017, doi: 10.15575/join.v2i1.71.

[7] SM Putri and SA Arnomo, “Application of C4.5 Algorithm for Predicting Service Quality on Consumer Satisfaction (Case Study: Hinet Batam),” *J. Inf. Syst. Res.* , vol. 1, no. 2, pp. 70–76, 2020, [Online]. Available: http://ejurnal.seminar-id.com/index.php/josh/article/view/69

[8] S. Al Syahdan and A. Sindar, “Data Mining of Product Sales Using the Apriori Method at Indomaret Galang City,” *J. Nas. Computation and Technol. Inf.* , vol. 1, no. 2, 2018, doi: 10.32672/jnkti.v1i2.771.

[9] S. Dutalia, A. K. Lalo, P. Batarius, Y. Carmeneja, and H. Siki, “Implementation of C4 Algorithm . 5 for Sales Classification,” vol. 06, pp. 1–12, 2021.

[10] S. Takalapeta, “Application of Data Mining to Analyze Consumer Satisfaction Using C4.5 Algorithm Method,” *JIMP - J. Inform. Merdeka Pasuruan* , vol. 3, no. 3, pp. 34–38, 2018, doi: 10.37438/jimp.v3i3.186.

[11] Y. Bastian *et al.* , “Analysis of the Implementation of the C4.5 Algorithm in Measuring the Level of Satisfaction of Indihome Customers in Pematangsiantar City,” *KESATRIA (Journal of Implementation of Inf. System (Computer and Management))* , vol. 2, no. 1, pp. 62–69, 2021.

[12] HS Putra, H. Okprana, and RK Sormin, “Implementation of Decision Tree Algorithm in Determining Sales Location of Merchandise at Serbelawan Simalungun Market Service,” *SmartEDU* , vol. 1, no. 2, pp. 61–69, 2022, [Online]. Available: https://ejournal.abivasi.id/index.php/SmartEDU

ISSN: xxxx-xxx

[13] MS Ummah, “No主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析 Title,” *Sustain.* , vol. 11, no. 1, pp. 1–14, 2019, [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI