UNIVERSITAS UTPADAKA SWASTIKA
JISI Jurnal Ilmiah Sistem Informasi
ISSN: XXXX-XXXX

Analysis of Customer Satisfaction in Computer Training Institutions Using C4.5 Algorithm

Asep Surahmat

Universitas Utpadaka Swastika, Tangerang, Indonesia
E-mail: asep.surahmat@utpas.ac.id

Diterima: 22 Oktober 2023 | Dipublikasikan 26 Oktober 2023

ABSTRACT

Customer satisfaction is an important reference in the sustainability of a business, as is the case with the computer training business at Radar Education Center which has competitors in its business. Until now there is no measuring tool for customer satisfaction at the Radar Education Center, this problem will be solved through the Decision Tree method using the C4.5 algorithm to analyze customer satisfaction. The data used comes from Radar Education Center customer data which consists of 80% training data and 20% testing data. After calculating accuracy using the Confusion Matrix, the Class Precision value was "yes" 83.33% and Class Recall "yes" was 100%.

KEYWORDS: Decision Tree, C4.5, Customer Satisfaction

1. Introduction

In this research journal, customer satisfaction at computer training institutions using the C4.5 algorithm will be explained. The C4.5 algorithm is a popular data mining technique used for classification and prediction.[1]

By applying the C4.5 algorithm, we can gain insight into the factors that influence customer satisfaction and develop strategies to improve it.

In the context of this research, the researcher refers to several relevant research papers and has applied the C4.5 algorithm in similar research. This research focuses on analyzing customer satisfaction in various industries, such as water supply services [1], automotive company [2], and financial institutions [3]. The results of this research show the effectiveness of the C4.5 algorithm in analyzing customer satisfaction and the factors underlying it.

By applying the C4.5 algorithm to this computer training institute research, it aims to contribute to existing knowledge about customer satisfaction analysis. This research can help training institutions to better understand the needs and preferences of their customers, thereby leading to improved service quality and customer retention.

2. Methodology

The research methodology for this journal focuses on analyzing customer satisfaction in computer training institutions using the C4.5 algorithm. The following steps outline the research design, data collection methods, and data analysis techniques employed in the study.

Research Design

This study adopts a quantitative research design, aiming to measure and analyze customer satisfaction levels in computer training institutions. The C4.5 algorithm is utilized to identify the key factors influencing customer satisfaction.

Data Collection

a. Survey Design

A structured questionnaire will be developed to collect data on various aspects of customer satisfaction, such as course quality, instructor competence, and learning outcomes. The questionnaire will be designed based on existing literature and expert opinions [4].

b. Data Collection

The surveys will be distributed to a sample of computer training institution customers, and the responses will be collected either through online platforms or in-person interviews, depending on the participants' preferences.

Data Analysis

a. Preprocessing

The collected data will be cleaned and prepared for analysis, including handling missing values and outliers.

b. Model Building

The C4.5 algorithm, a decision tree-based classification algorithm, will be applied to the prepared data to build a decision tree model that predicts customer satisfaction based on various input variables [5][6].

c. Model Evaluation

The performance of the decision tree model will be evaluated using appropriate metrics, such as accuracy, precision, recall, and F1 score.

d. Interpretation

The decision tree model will be interpreted to understand the key factors influencing customer satisfaction in computer training institutions.

e. Study Limitations

The limitations of this study include the potential bias in self-reported data, the sample size, and the focus on a specific context of computer training institutions. These limitations should be considered when interpreting the results.

By following this research methodology, the study aims to provide valuable insights into customer satisfaction in computer training institutions and contribute to the existing body of knowledge.

2.1 Data Mining

Based on developments in data mining the fast one is of course inseparable development of the role of information technology so big [7]. Data Mining is defined as an extraction process information from various data sets big. As one unit of the series process, data mining can be divided into in to several stages of the process. Stages It is interactive, involving the user directly or through intermediaries knowledge base [8].

The stages in data mining are as follows:

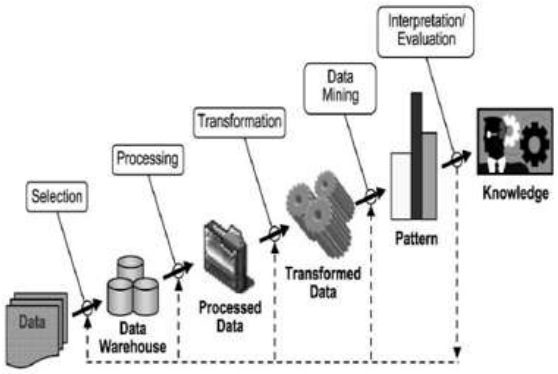


Figure 1. Data Mining Process [9]

a. Data Cleaning

Data cleaning stages are useful for clean up inconsistent data. Apart from that, this process is also used for eliminate inappropriate data attributes to the hypothesis used.

b. Data Integration

Stages this used in combining data from several sources that exist and are put together as data main.

c. Data Transformation

This data transformation is carried out with how to change data into a form or format which is in accordance with research needs there is.

d. Data Mining Engineering Applications

This stage is part of wrong one data mining process and must be considered that in a collection of techniques that already there can guarantee a solution in the process of implementing data mining certain.

e. Pattern Evaluation

In the evaluation phase patterns are found can be used to find patterns others

with certain unique characteristics or other predictions that have some value.

f. Pattern Presentation

At this presentation stage the pattern is can be found used for produce action on research what is being done.

2.2 Decision Tree

Decision Tree is a step to find a model to explain a data class concept which aims to determine the class of an object whose label is not yet known. The model itself is in the form of "if then". Decision tree is a classification method that is widely used because it is easy to interpret [10].

This Decision Tree is known as a tree decision with the so-called algorithm with the C4.5 algorithm. Where is this C4.5 is one such algorithm used based on training data. Algorithm this method uses quite a lot of classifications and predictions used.

2.3 Algorithm C4.5

The concept of a Decision Tree is changing data into decision trees and rules decisions or what are called decision rules and is a development of the algorithm C4.5. There are several stages used in implementing Decision Tree with

Algorithm C4.5, follows the stages [11]:

a. Prepare the training data that will be carried out used, this data is taken from the data grouped past into predetermined categories.
b. Calculate the root value of the tree. Mark The root will be taken from that attribute chosen, the way it is done calculation of gain from each attribute, the result of the highest gain value will be the first root in decision tree is created. Before calculate the gain value of the attribute, First calculate the entropy value with using the following formula:

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

Information:
S = Case Set
n = Number of partitions S
pi = Proportion Si to S

Then proceed with calculating the gene value with the following formula:

$$\text{Gain (S,A)} = \text{Entropy (S)} - \sum_{j=1}^{n} \frac{Si}{S} * \text{Entropy } (S_i$$

Information:
S = Case Set
A = Feature
n = Number of attribute partitions A
Si = Proportion of Si to S
|S| = Number of cases in S

c. Repeat step 2 and step 3 so that all records are partitioned.
d. The decision tree partitioning process will stop when:
All records in node N get the same class.
- There are no attributes in the record which is partitioned again.
- There are no records in the branch which is empty.

2.4 Confusion Matrix

Tabel 1. Error! No text of specified style in document.1 Confusion Matrix

		Observed	
		TRUE	FALSE
Predicted Class	TRUE	True Positive (TP)	False Positive (FP)
	FALSE	False Negative (FN)	True Negative (TN)

Confusion matrix is a model testing tables that have 4 (four) combination of predicted values (predicted values) and actual values In table 1 is a confusion table matrix where:

a. TP is the number of positive data correctly classified by the system.
b. TN is negative data correctly classified by the system.
c. FN is negative data however incorrectly classified by the system.

d. FP is positive data however incorrectly classified by the system.

In the confusion matrix the accuracy value is a comparison between data correctly classified with all the data. The accuracy value is calculated by equality:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \; x \; 100\%$$

Followed by searching for precision values which is the number of positive category data correctly classified divided by amount data classified as positive with using the equation:

$$Precision = \frac{TP}{TP + FP} \; x \; 100\%$$

The recall value is the percent of positive category data that is classified correctly by the system by calculating the equation:

$$Recall = \frac{TP}{TP + FN} \; x \; 100\%$$

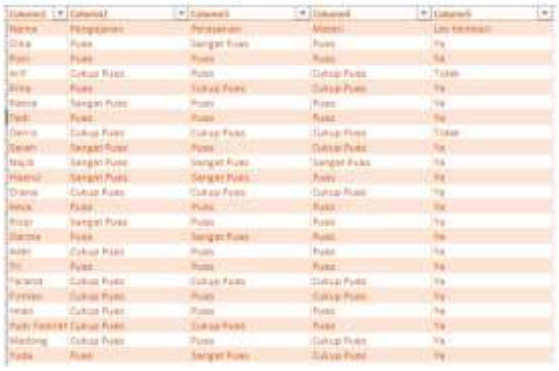
Followed by the error value, namely incorrectly identified cases in a number of data, so you can see how much the magnitude of the error rate in the system in use. To count the error percentage can be done via the following formula equation:

$$Error = \frac{FP}{TP} \; x \; 100\%$$

3. Results and discussion

In this research, survey data was used customer satisfaction from the Radar Education Center as many as 30 customers with uses three variables, namely about teaching, about that service provided, as well as customer materials and data tutoring again at a later date. For The data table can be seen in table 2 below:

Tabel 1. Error! No text of specified style in document. 2 Customer Survey Data Table



For the application of the method This C4.5 uses the Rapid Miner application as a tool to help with design schematics as follows:

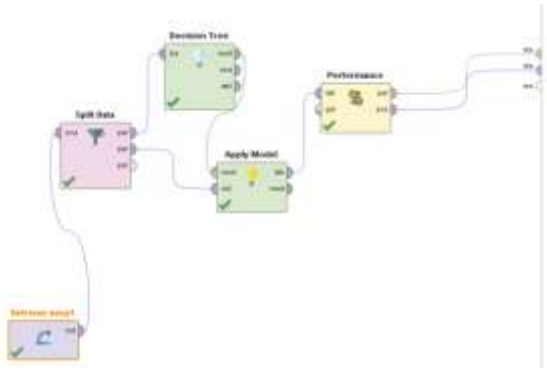


Figure 2. Design Process Rapid Miner

In the process design of this rapid miner Split Data is used which splits the data The main data becomes training as much as 80% and testing data as much as 20%. Then Decision Tree classification is used as well apply model to connect with performance test with confusion matrix.

Tabel 2. Confusion Matrix Result

The results of the confusion matrix test based on the testing data used obtained class recall of 100% and class precision of 83.33%.

References

[1] T. H. Sinaga, A. Wanto, I. Gunawan, S. Sumarno, and Z. M. Nasution, "Implementation of Data Mining Using C4.5 Algorithm on Customer Satisfaction in Tirta Lihou PDAM," *J. Comput. Networks, Archit. High-Performance Comput.*, vol. 3, no. 1, pp. 9–20, 2021, doi: 10.47709/cnahpc.v3i1.923.

[2] Supangat, A. Pratama, and T. Rahmawati, "Implementation of C4.5 Algorithm for Analysis of Service Quality in Companies of PT. XYZ," *Proc. 2nd Int. Conf. Bus. Manag. Technol. (ICONBMT 2020)*, vol. 175, pp. 38–44, 2021, doi: 10.2991/aebmr.k.210510.008.

[3] R. Muttaqien, M. G. Pradana, and A. Pramuntadi, "Implementation of Data Mining Using C4.5 Algorithm for Predicting Customer Loyalty of PT. Pegadaian (Persero) Pati Area Office," *Int. J. Comput. Inf. Syst.*, vol. 2, no. 3, pp. 64–68, 2021, doi: 10.29040/ijcis.v2i3.36.

[4] O. Villacampa, "Feature Selection and Classification Methods for Decision Making : A Comparative Analysis," *Nov. Southeast. Univ.*, no. 63, p. 488, 2015.

[5] A. Brunello, E. Marzano, A. Montanari, and G. Sciavicco, "J48SS: A novel decision tree approach for the handling of sequential and time series data," *Computers*, vol. 8, no. 1, pp. 1–28, 2019, doi: 10.3390/computers8010021.

[6] W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, "C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012005.

[7] H. K. Bhuyan and V. Ravi, "Analysis of Subfeature for Classification in Data Mining," *IEEE Trans. Eng. Manag.*, no. August, 2021, doi: 10.1109/TEM.2021.3098463.

[8] V. Anestiviya, A. Ferico, and O. Pasaribu, "Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : Sman 1 Natar)," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 80–85, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI.

[9] A. S. Candra Naya, "Analisa Kepuasan Pelanggan Menggunakan Algoritma C4.5 Pada PT. Cahaya Indotama Engineering," *J. Teknol. Pelita Bangsa*, vol. 12, no. 4, pp. 33–40, 2021.

[10] S. Febriani and H. Sulistiani, "Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5," *89Jurnal Teknol. dan Sist. Inf.*, vol. 2, no. 4, pp. 89–95, 2021.

[11] A. Surahmat and D. Kurnaedi, "Junior Class Preparedness Classification Faces A National Exam Using A C.45 Algorithm With A Particle Swarm Optimization Approach," *bit-Tech*, vol. 3, no. 1, pp. 11–20, 2020, doi: 10.32877/bt.v3i1.169.