IJISIT

http://ejournal.enlightenlearner.com/index.php/ejournal

Data Mining Design in Teacher Teaching Assessment Based on Student Satisfaction Index

[1]Seriani loi, Universitas Prima Indonesia, Indonesia
[2]Janiali Sirait, Universitas Prima Indonesia, Indonesia
[3]Elsa Naibaho, Universitas Prima Indonesia, Indonesia
Correspondence: E-mail: serianiloi@gmail.com

Article Info

Article history:

Received 06 02, 2022
Revised 06 16, 2022
Accepted 06 28, 2022

Keywords:

Data Mining
C4.5 Algorithm
Teaching

ABSTRACT

Teacher performance appraisal is an analytical process in order to produce good teaching The process of assessing teacher performance at SMP Kartika Pematang siantar is not so detailed (detailed). With the non-detailed assessment, it is feared that there will be a subjective assessment (based on personal interests), which can cause social jealousy for teachers who do not accept achievement. In the assessment of teacher teaching, data on student satisfaction with the teacher is needed. In collecting data using a questionnaire and the variables used in this research there are 4, namely discipline, teaching quality, behavior, and social relations. This study aims to assist the school in improving the performance of teachers, especially in the way of teaching teachers based on student satisfaction by using the C4.5 algorithm.

1. INTRODUCTION

Data mining is defined as a process that uses various data analysis tools to find patterns and data relations so that they can be used to make accurate predictions [1][2]. Technique methods or algorithms in data mining vary widely. The selection of the right method or algorithm is very dependent on the objectives and the overall process.

The C4.5 algorithm is a decision tree technique that is often used which produces several rules and a decision tree with the aim of increasing the accuracy of the predictions being made, in addition, the C4.5 algorithm is an easy-to-understand algorithm [3]. The C4.5 algorithm is a decision tree classification algorithm that is widely used because it has major advantages over other algorithms. The advantages of the C4.5 algorithm are that it can produce a decision tree that is easy to interpret, has an acceptable level of accuracy, is efficient in handling discrete type attributes and can handle discrete and numeric type attributes..

In teaching and learning activities there are several causes of student and student

dissatisfaction with teacher performance, including teachers who often arrive late, are not good at delivering material, and teachers who pay less attention to the development of students and students.

To get quality teachers, of course, the school will conduct an assessment for every teacher who teaches at the school by measuring the level of student satisfaction in order to improve and improve the quality of the school. Every student and student certainly wants to get good educational services. That is why it is necessary to know whether or not students are satisfied with the teacher's teaching method provided by the school. The purpose of this research is to analyze whether or not students are satisfied with the teacher's teaching method using the RapidMiner Studio application.

The results of this study were conducted to provide a rule model in analyzing whether students are satisfied or not with the teacher's teaching method using the C4.5 algorithm.

2. METHODS

2.1. Data Mining

Data mining is the process of analyzing data from different perspectives and summing it up into Data mining can be said to be a process of extracting knowledge from a large amount of available data. The knowledge generated from the data mining process must be new, easy to understand, and useful.

2.1.1 Data Mining Function

According to Han and Kamber (2018), data mining functions and the kinds of patterns that can be found, namely:

1. Concept/Class Description
2. Mining Frequent
3. Classification and Prediction
4. Cluster
5. Outer Analysis
6. EvolutionAnalysis

2.2. Decision Tree Concept

Categorical type features for features whose va lues are of categorical type (nominal or ordinal) can have several different values.

Decision Tree or Decision Tree is a tree that is used as a reasoning procedure to get answers to the problems entered. The tree that is formed is not always a binary tree. If all the features in the data set use 2 kinds of categorical values, then the tree form obtained is in the form of a binary tree. If the feature contains more than 2 kinds of categorical values or uses a numeric type, the tree form obtained is usually not a binary tree [8]. Decision Tree there are 3 types of Nodes, namely [9]:

1. Root Node, is the top node, at this node there is no input and can have no output or have more than one output.
2. Internal Node, is a branching node, at this node there is only one input and has at least two out puts.
3. Lead Node or Terminal Node, is the end node, at this node there is only one input and has no output

2.3. C4.5 Algorithm

The C4.5 algorithm was introduced by Quinlan (1996) as an improved version of ID3. In ID3, Decision Tree induction can only be performed on features of categorical type (nominal or ordinal), while numerical types (interval or ratio) cannot be used [8]. The algorithm which is the development of ID3 can classify data using a decision tree method which has advantages can process numeric (continuous) and discrete data, can handle missing attribute values, produce rules that are easy to interpret, and are the fastest among algorithms that use main memory in computers [10]. The important thing in Decision Tree induction is how to state the conditions for testing nodes consisting of 3 groups [8]:

1. Binary Features

Features that only have two different values, the test conditions when this feature becomes a node (root or internal) only have two branch choices. An example of the solution can be seen in Figure 1

In general, there are 2 features of categorical type, namely binary splitting and multi splitting. The combination is presented as Figures 2 and 3 where the solution only

allows binary solving, such as the CART Algorithm, so it gives the number of possible combinations as S^(K-1), where k is the number of distinct values in the feature.

3. Numeric type features

For features of numeric type, the test conditions in nodes are expressed by comparison tests (A<_V) or(A≥V) with binary results, or for multi with results in the range of values in the form v_i≤A<v_i<v_(i+1,) for I = 1, 2, …..,k. For the multi method, the algorithm must examine all possible ranges of continuous values. The solution to numerical features is presented in Figure 4.

3. RESULTS AND DISCUSSION

3.1 Data analysis

In this study, the data used will be processed from the results of a questionnaire given to students of Kartika Middle School class 2018 and 2019. The data used consists of:

1. Discipline Variables,
2. Teaching Quality Variable,
3. Behavioral Variables, and
4. Social Relationship Variables.

Each factor consists of a question posed to a given student. The questionnaire that has been carried out then looks for the average of the factors used. The data used are descriptive statistics by giving questionnaires to YPI high school students Amir Hamzah. The questionnaire was given using linker 4 consisting of ST (Strongly Agree), S (Agree), KS (Disagree), and TS (Disagree). Then the data is processed using RapidMiner using Performance which functions as data validation and reliability to find data accuracy. Accurate data will then be processed data to find results from research problems using Rapid Miner and decisions are made from the processed results of Rapid Miner.

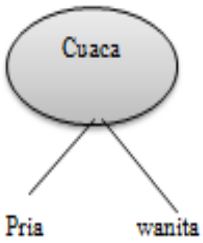

Figure 1. Binary feature tester requirements

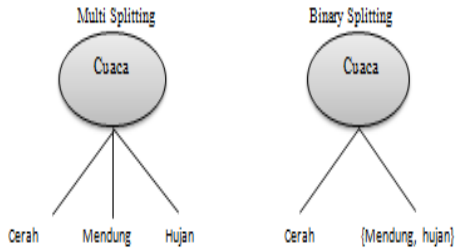

Figure 2. Requirements for Testing Nominal Type Features

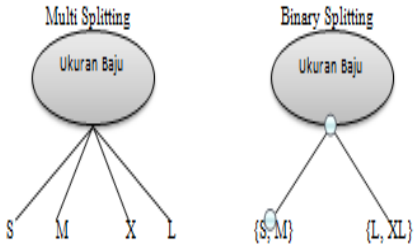

Figure 3. Conditions for Testing Ordinal Type Features

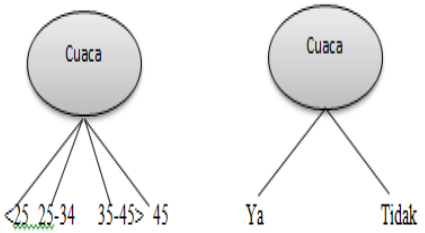

Figure 4. Numerical Edged Feature Testing Terms

Table 1. Recapitulation Result

No	Respondent	V1	V2	V3	V4	RESULTS
1	R1	S	S	S	S	Satisfied
2	R2	ST	S	S	ST	Satisfied
3	R3	ST	S	S	S	Satisfied
4	R4	KS	KS	KS	KS	Not Satisfied
5	R5	KS	KS	KS	KS	Not Satisfied

Information:
V1 = Discipline Variables
V2 = Teaching Quality Variables,
V3 = Behavior Variables, and
V4 = Relationship Variables Social

Calculates the number of cases, the number of cases for the satisfied result, the number of cases for the dissatisfied result, and the entropy of all cases and the cases are divided based on the Discipline Variable attribute.

.

Teaching Quality Variables, Behavioral Variables, and Social Relations Variables. After that, do the calculation of the gain for each attribute.

Table 2 Node 1. Calculations

Node	Number of cases		Satisfied (S1)	Not Satisfied (S2)	Entropy	Gain
1	Total	50	31	19	0.95804	
	VI					0.356416678
	ST	19	16	3	0.62925	

S	21	15	6	0.86312	
KS	10	0	10	0	
TS	0	0	0	0	
		V2			0.285711011
ST	7	4	3	0.98523	
S	30	25	5	0.65002	
KS	10	2	8	0.72193	
TS	3	0	3	0	
		V3			0.329033058
ST	13	8	5	0.96124	
S	28	23	5	0.67694	
KS	7	0	7	0	
TS	2	0	2	0	
		V4			0.335370829
ST	10	5	5	1	
S	29	25	4	0.57879	
KS	8	1	7	0.54356	
TS	3	0	3	0	

From the picture, it can be seen that the KM (Kilometer) attribute has 3 values, namely KS, S, and ST. Of these three values still require further calculations. The calculation can be seen in the following table :

3.2 Data collection technique

Data collection methods used in this study are:

1. Documentation Data collection methods that record company reports and study library books related to the problem.
2. Question list method (questionnaire) According to Sugiyono (2004: 135) questionnaire is a data collection technique that is done by giving a set of questions or written statements to respondents to answer. G. Data analysis and analysis techniques

The data analysis used in this research is qualitative and quantitative analysis. Qualitatively, the likert scale is used with the respondents' answer choices as follows:

1. Student assessment for real performance is divided by the following criteria:
   a. Answer Strongly Agree (SS) with a score: 5
   b. Answer Agree (S): 4
   c. Answer Moderately Agree (CS): 3
   d. Answer Disagree (TS): 2
   e. Answer Strongly Disagree (STS): 1
2. Student assessment for expected performance is divided by the following criteria:
   a. Answer Strongly Agree (SS) with a score: 5
   b. Answer Agree (S): 4
   c. Answer Moderately Agree (CS): 3
   d. Answer Disagree (TS) :2
   e. Answer Strongly Disagree (STS): 1

3.3 Analysis technique

The analysis technique uses statistical equipment using the customer satisfaction equation (Fandy Tjiptono, 2001: 37) the formula used is: IKP = (PP - EX)

Description:
IKP : Customer Satisfaction Index
PP : Real Performance (Perceived Performance)
EX: Expected Performance (Expectation)
The measurement results range from negative, zero or positive values with the following criteria:
1. Negative, if the actual performance felt by students is smaller than the expected performance.
2. Zero, if the actual performance is the same as the expected performance.
3. Positive, if the real performance is greater than the expected performance.
In other words:
1. Students are stated to be very satisfied if the real performance they feel is greater than the expected performance.
2. Students are declared satisfied if the perceived real performance is the same as the expected performance.
3. Students are declared dissatisfied if the perceived real performance is less! Than expected performance. The calculation process in finding the entropy and gain values at the next nodes is the same as the calculation for finding the entropy and gain values at node 1. The number of cases used is adjusted to the number of cases for each value in attribute V1. The calculation will continue until all the existing attributes partitioned out.

4. CONCLUSION

Based on the previous discussion, it can be concluded that the application of Data mining using the C4.5 algorithm in determining evaluating teacher performance in the teaching and learning process can be applied. The source of the data used in this study was data obtained directly from YPI high school students Amir Hamzah through a questionnaire. The number of test data is 50 students using two classes. From the calculation results of the C4.5 Algorithm, a classification with the dominant variable is obtained, namely V1 (Discipline Variable).

Testing data on Rapidminer 5.3 using naive bayes can display two classes of classification results with an accuracy rate of 82.00%. and can be categorized as excellent.

5. ACKNOWLEDGEMENT

Thank you for all lecturers, students and families who support this research

6. REFERENCES

[1] G. Wahyuningtyas, I. Mukhlash, And Soetrisno, Aplikasi Data Mining Untuk Penilaian Kredit Menggunakan Metode Fuzzy Decision Tree, J. Sains Dan Seni Pomits, Vol. 2, No. 1, Pp. 1–6, 2014.
[2] C. Fadlan, S. Ningsih, And A. P. Windarto, Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra, Jutim, Vol. 3, No. 1, Pp. 1–8, 2018.
[3] L. Navia Rani, Larissa Navia Rani , Fakultas Il Mu Komp Uter , Vol. 2, No. 2, Pp. 33–38, 2015.
[4] A. Saleh, Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga, Citec J., Vol. 2, No. 3, Pp. 207–217, 2015
[5] A. P. Windarto, “Implementation Of Data Mining On Rice Imports By Major Country Of Origin Implementation Of Data Mining On Rice Imports By Major Country Of Origin

Using Algorithm Using K-Means Clustering Method," Int. J. Artif. Intelegence Res., Vol. 1, No. 2, Pp. 26–33, 2017.

[6] P. Soepomo, Penerapan Data Mining Untuk Klasifikasi Prediksi Penyakit Ispa (Infeksi Saluran Pernapasan Akut) Dengan Algoritma Decision Tree (Id3), Vol. 2, 2014.

[7] B. M. Metisen, H. L. Sari, Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila, J. Media Infotama, Vol. 11, No. 2, Pp. 110–118, 2015.

[8] R. I. E. Saragih and O. Simatupang, "Regression Genetic Algorithm (RGA) Based Approach For Optimizing Bank Deposit", IEEE publisher., 2019.

[9] R. I. E. Saragih and Rhamadani. "Context Mobile Application (CMA) for Increasing the Service of Koperasi in Indonesia" JTKSI. 2022.

[10]R. I. E. Saragih. "Fuzzy Logic Controller (Flc) dalam Penentuan Mutation Rate Algoritma Genetika". USU. 2015