

Analysis Iterative algorithms Dichotomizer (ID3): The Satisfaction Study in Computer Laboratory

Ibnu Rasyid Munthe¹, Sumitro Sarkum², Volvo Sihombing³
ibnurasyidmunthe@gmail.com¹, ,sumitro.ulb@gmail.com²,volvo_lecture@yahoo.co.id³.

Akademi Manajemen Informatika Labuhan Batu^{1,2,3}

Abstract. The purpose of this article to explore the information level of student satisfaction in a computer laboratory. This study was used algorithms *Iterative Dichotomizer* (ID3) to represent the concepts in the form of a decision tree, the existence a decision tree to produce a good algorithm training data to help decide the completion of good problems in laboratory facilities. The research method was used the *Cross Industry Process For Data Mining* (CRISP-DM). RapidMiner as an application in the making of a decision tree (*Decision Tree*) and test level data of the student's satisfaction. The Result of of student's satisfaction in a computer laboratory facilities were processed on a Rapid miner has been constructed a decision tree. The Decision trees were tested used the Confusion Matrix to produce value accuracy of the decision tree. The sample was used a students of AMIK Labuhanbatu and the results showed that student satisfaction in a computer laboratory facilities about 92.86%. The value Precision of the decision tree for student satisfaction in the computer laboratory facilities AMIK Labuhan Batu about 91.18% by a positive class: Satisfied. The value recall of a decision tree for student satisfaction in the computer laboratory facilities AMIK Labuhan Batu about 100.00% by a positive class: Satisfied and produce AUC value about: 0.933 0.133 +/- (micro: 0.933) (positive class: Satisfied) Average Value AUCbetween 0.90 - 0.10 with a classification of Very Good.

Keyword: Decision Tree, Iterative Dichotomize (ID3), (CRISP-DM), Rapidminer

1 Introduction

To showed the satisfaction of both the products and services were in the assessment of the characteristic of product or service, or the product itself, which provides consumers the level of excitement related to the fulfillment of the needs of consumers. The Customer satisfaction is feeling happy or disappointed who emerged after comparing the performance (yield) of the product thought to the expected performance [1], In this case the customer / consumer is a student at an institution of higher education, so if A student feel a satisfied with the value given by the laboratory facilities (computer products) and supporting facilities of the laboratory, it is very likely students are recommended to the environment to select campus good in the provision of facilities will become a natural promotion to increase new student to join at the campus.

Iterative algorithms Dichotomizer (ID3) for make a decision tree so that can be extracted knowledge of the problems that occurred. The ID3 algorithm is the learning useful because it

can be build a good decision trees and efficient[2], the method of ID3 Decision Tree is the best method if compared with other classification algorithms[3].

2 Methodology

2.1 Algortima Iterative Dichotomizer (ID3)

The most popular technique classification was used a decision tree. The Decision Tree was also a method of classification and prediction a very powerful and famous, The Decision tree was a classification method that used a representation of the structure of the tree (*tree*) where each *node* represents an attribute, the branch represents from value attributes, and leaves represent the class [4].

2.2 Cross Industry Proress For Data Mining (CRISP-DM)

The Steps will be implemented as a research framework. The each step was based on the methodology used. The research method was used the Cross Industry Process For Data Mining (CRISP-DM)[5].

2.2.1 Business Understanding

The data in this study was used a the primary data. The primary data was student of AMIK Labuhan Batu, The data was collected the data of student satisfaction on the facilities and laboratory facilities AMIK Labuhan Batu.

2.2.2 Data Understanding

The first steps, in determining the data to be processed. From it, the data has been gotten of student satisfaction on the facilities and laboratory facilities AMIK Labuhan Batu. The data will not be processed all because this research only do research How to create a decision tree in AMIKLabuhan Batu student satisfaction with the method of iterative algorithms Dichotomizer (ID3) in the computer laboratory facilities .The second steps, Choosing a Student Satisfaction Data on Facilities and Support Laboratory AMIK Labuhan Batu separates data incomplete due to the attributes were not documented or attributes that were not owned (Missing Value).The handling Missing Value achieved by splitting a blank record. The Third steps, to determining the attributes used on the first steps, the attributes were used, such as: one attributes supporting facilities (good, enough and less), both attributes computer laboratory (good, enough and less); tertiary attributes interest to visit again (visit and Not Visit) and the last target student satisfaction (satisfied and dissatisfied). The four Steps, doing the data conversion, data with attributes have been choosen andthen it will be converting to easily the process of data mining an attributes in accordance with the third steps, the data will be processed with the help of data mining tools.

2.2.3 Data prepatation

On Steps here,it covers all activities to prepare data to be incorporated into the modeling tool, the data is processed from raw data early. The processed data into Ms.Excel by setting attributes (parameters first set) and grade or label and study data so that data expectations have better knowledge of the data processed. It will assist in the selection of the data used for the process of data mining.

2.2.4 Modeling

The method in this research was used iterative algorithms Dichotomizer (ID3) to did the measurement accuracy in this study used the software RapidMiner. Here was a picture of the process Research

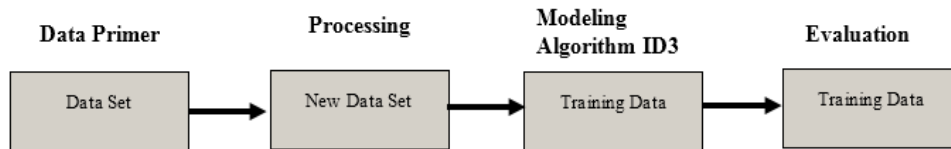


Fig. 1. The research process used

2.2.5 Evaluation

This steps doing validation and measurement accuracy of the results achieved by using software model confusion matrix for the measurement of the accuracy of the model

3 Results and Discussion

3.1 Data Set (Data Primer)

The raw data received from the data of student satisfaction, and choosing of the parameters to be analyzed. The parameters is taken an attribute before made by class or label "satisfied" and "dissatisfied" attribute will be the parameter or variable *input*. The Data Student Satisfaction In computer lab AMIK Labuhan Batu explained that Primary data student has done filling questionnaire on student satisfaction AMIK Labuhan Batu on means computer labs, attributes supporting facilities is the attribute to the category of facilities (Good) when expectations have been met, and the label will answered by students if all the attributes answered by whether students feel "satisfied" or "dissatisfied" in supporting facilities, computer labs, are interested to come back

3.2 Processing

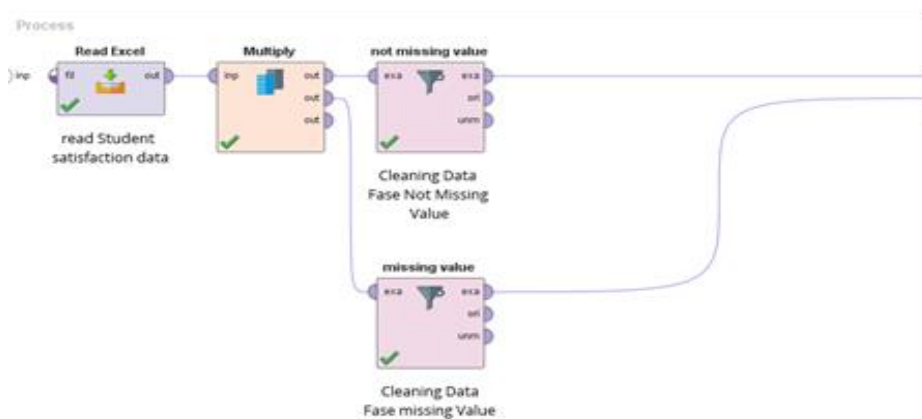


Fig. 2. Processing Cleaning Data

In figure 2, the software rapid miner perform data processing from primary data read student satisfaction in a computer lab with a filter to separate the blanks missing value intended for data processing at a later stage is going well. The Cleaning the data serves as the correct data and will be used on stage while dry Data2 next as data is missing and can not do at the next step.

3.3 Modeling Algorithm ID3

In Figure3. Training Algorithm ID3 data is the step of making the training data with modeling ID3 algorithm generates decision trees on Student Satisfaction Data on a Computer Laboratory Facility. From the results of the training data ID3 algorithm can be knowledge.

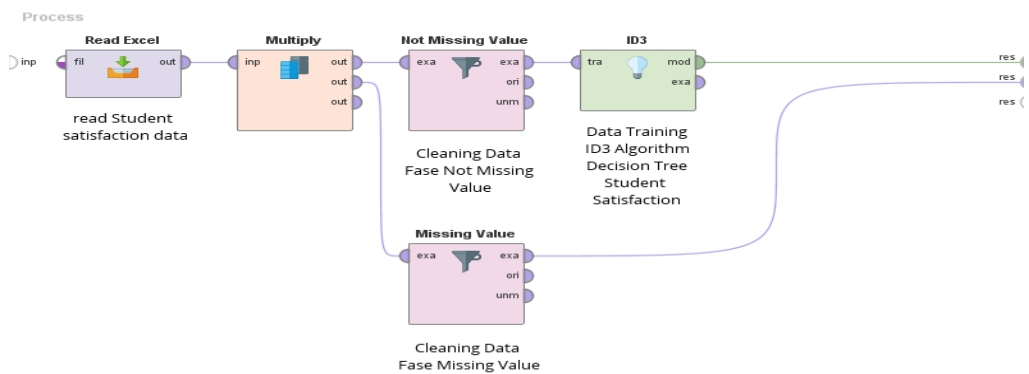


Fig. 3. Data Training Algorithm ID3

Data that has been carried out training on ID3 algorithm can be seen in Figure 4. Decision Trees Student Satisfaction In Computer Lab Facility AMIK Labuhan Batu generate data in the form of a decision tree. At the first data Students choose the supporting facilities = good student satisfaction against the computer lab facilities with selecting the data of 25 people with a total ratio was 59.52% are satisfied. The two students are choosing Data Supporting Facilities Laboratory = Enough and computer = Neither the satisfaction of students to the computer lab facilities with a data select as two people with a total ratio about 4.76% are satisfied. Three students chose Data Supporting Facilities and Computer Laboratory = enough = Adequate and interests of students to come back = been the satisfaction of students to the computer lab facilities to choose as many as four people were satisfied and 2 are not satisfied with the total ratio was 14.29%. Data to four students who opt = sufficient supporting facilities and computer labs = less then the satisfaction of students to the computer lab facility to choose the data by 4 people with a total ratio was 9.52% dissatisfied. And five students are choosing Data Supporting Facilities = less then the satisfaction of students to the computer lab facilities with a data selecting as many as 5 people with a total ratio about 11.90% were dissatisfied.

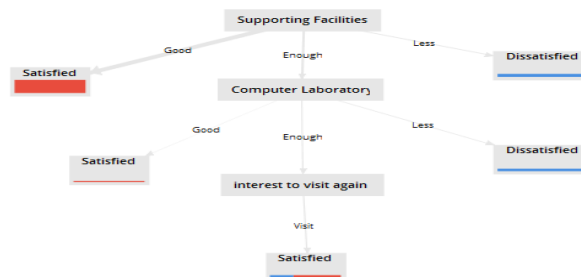


Fig. 4. Decision Trees Student Satisfaction In Computer Lab.

3.4 Evaluation

To evaluate the decision tree student satisfaction on laboratory means AMIK Labuhan Batu computer by using confusion matrix. Confusion matrix is information about the actual classification and prediction in the classification. The system Performanceclassification in the general classification using the data in the matrix[6], Accuracy in classification is the percentage of accuracy of records of data that are classified correctly after testing on the classification results. The Formula Accuracy can be defined on the following formula:

$$\text{accuracy} = \frac{TP+TN}{P} \quad [7]$$

Note: TP is the number of records / tuple positive is assigned correctly by the model classification algorithm, TN is the number of records / tuple negative is labeled as one of the classification algorithm model, and P is the total of all records that were evaluated.

accuracy: 92.50% +/- 10.00% (mikro: 92.86%)

	true Dissatisfied	true Satisfied	class precision
pred. Dissatisfied	8	0	100.00%
pred. Satisfied	3	31	91.18%
class recall	72.73%	100.00%	

precision: 91.67% +/- 10.54% (mikro: 91.18%) (positive class: Satisfied)

recall: 100.00% +/- 0.00% (mikro: 100.00%) (positive class: Satisfied)

Fig. 5. Confusion Matrix algoritma ID3

From Figure 5. Confusion Matrix ID3 algorithm was tested using rapid miner software generate value accuracy of the decision tree for student satisfaction in the computer laboratory facilities AMIK Labuhan Batu 92.50%. While Precision ID3 algorithm was tested using the software rapid miner produces precision value of the decision tree for student satisfaction in the computer laboratory facilities AMIK Labuhan Batu 91.67% with a positive class: Satisfied. Whereas Recall ID3 algorithm was tested using the software generates rapid miner recall value of a decision tree for student satisfaction in the computer laboratory facilities AMIK Labuhan Batu 100.00% with positive class: Satisfied

AUC: 0.833 +/- 0.211 (mikro: 0.833) (positive class: Satisfied)

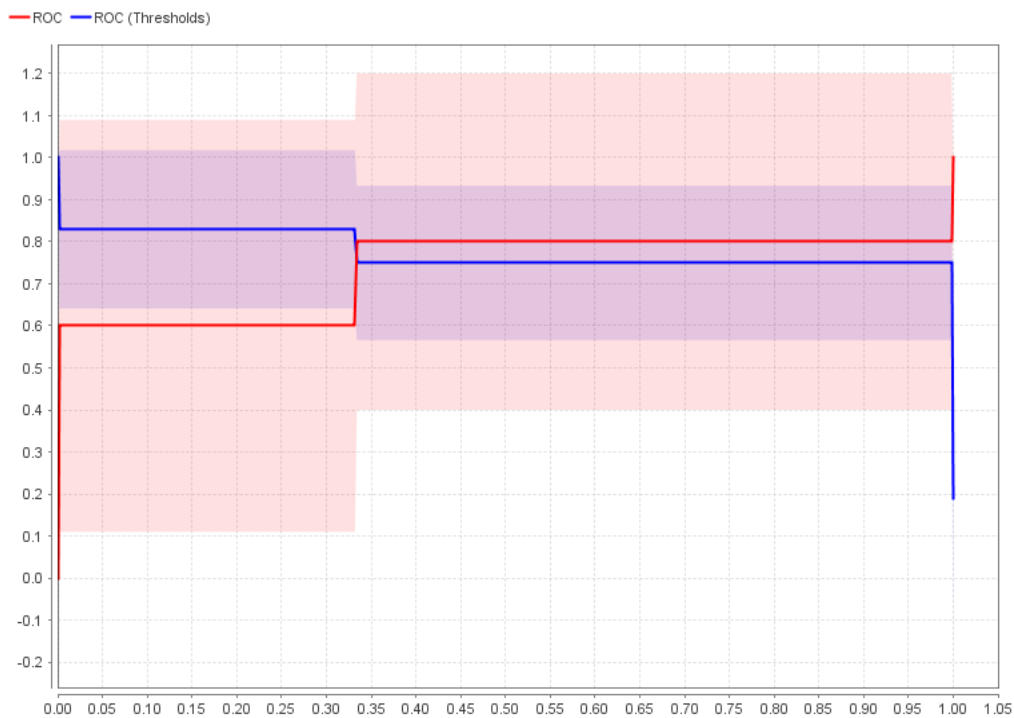


Fig. 6. Curve ROC

From Figure 6. ROC curve with accuracy at 5 fold cross validation software was tested using rapid miner produces AUC value: 0.833 +/- 0.211 (micro: 0.833) (positive class: Satisfied) Average Value AUC between 0.80 – 0.90 with a classification of Good.[8]

4 Conclusion

The Results obtained from the data of student satisfaction in a computer laboratory facilities are processed on a Rapid miner has constructed a decision tree. The Decision trees were tested using the Confusion Matrix generate value accuracy of the decision tree for student satisfaction in the computer laboratory facilities AMIK Labuhan Batu 92.50%. Precision value of the decision tree for student satisfaction in the computer laboratory facilities AMIK Labuhan Batu about 91.67% with a positive class: Satisfied. recall value of a decision tree for student satisfaction in the computer laboratory facilities AMIK Labuhan Batu 100.00% with positive class: Satisfied and produces AUC value: 0.833 +/- 0.211 (micro: 0.833) (positive class: Satisfied) Average Value AUC between 0.80 – 0.90 with a classification of Good.

References

- [1] P. Kotler and K. L. Keller, *Marketing Management*, vol. 22, no. 4. 2009.
- [2] S. Appavu alias Balamurugan and R. Rajaram, "Effective solution for unhandled exception in

- decision tree induction algorithms,” *Expert Syst. Appl.*, vol. 36, no. 10, pp. 12113–12119, Dec. 2009.
- [3] M. J. A. Berry and G. S. Linoff, *Data Mining Techniques For Marketing, Sales, and Customer Relationship Managemen*, vol. 38, no. 4. 2005.
- [4] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*. 2011.
- [5] P. Chapman *et al.*, *Crisp-Dm 1.0 Step-by-step data mining guide*. 2000.
- [6] A. K. Santra and C. J. Christy, “Genetic Algorithm and Confusion Matrix for Document Clustering,” *IJCSI Int. J. Comput. Sci. Issues*, vol. 9, no. 1, pp. 322–328, 2012.
- [7] Mochammad Yusa, E. Utami, and E. T. Luthfi, “Evaluasi Performa Algoritma Klasifikasi Decision Tree ID3 , C4,5 dan Cart pada Dataset Readmisi Pasien Diabetes,” *InfoSys J.*, vol. 4, no. 1, pp. 23–34, 2016.
- [8] Y. Elmande and P. Widodo, “Pemilihan Criteria Splitting dalam Algoritma Iterative Dichotomiser 3 (ID3) untuk Penentuan Kualitas Beras: Studi Kasus Pada Perum Bulog Divre Lampung,” *J. Telemat. MKOM*, vol. 4, no. 1, 2012.