American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-7, Issue-10, pp-12-20
www.ajer.org
Research Paper Open Access

Decision Tree and Chi-Square Analysis to Determine Student's Criteria Choosing Study Program

Deny Jollyta[*1], M. Wahyudi[1], Hengki T. Sihotang[1], Sri Wahyuni[1], Muhammad Zarlis[2], Syahril Efendi[2]

[1](Program Doctoral, Department of Computer Science and Information Technology, Universitas Sumatra Utara, Indonesia)

[2](Department of Computer Science and Information Technology, Universitas Sumatra Utara, Indonesia)

Corresponding Author: Deny Jollyta

ABSTRACT : *Appropriate information about the criteria of students who choose a study program at a college can facilitate the socialization of relevant study program in the community and help policy-making related to the development of study program in the future. In this research, database of Informatics students is needed as samples. Through the database of Informatics students at a college, several criteria were selected to support data training using C4.5 Algorithm of Decision Tree techniques in data mining, such as gender, school origin, major at school and total test scores. The data will be trained to produce a decision tree that describes the criteria of students who choose the Informatics study program in the form of rule. The training results showed that there were seven rules that have been simplified through independent tests using chi-square with an error rate of 9.19%.*

KEYWORDS*-database of informatics students, decision tree, c4.5 algorithm, chi-square, rules*

Date of Submission: 02-10-2018 Date of acceptance: 13-10-2018

I. INTRODUCTION

The Background of prospective students accepted in Informatics Engineering program has an effect on easiness in learning. Difficulties are usually realized when the learning process takes place and the value obtained by students. For some people, especially in Riau, Informatics Engineering is the favorite program. The decline in student achievement resulted in the decreasing performance of the program because the output was not in line with expectations.

The success of student achievement is influenced by several factors among these factors is the educational background that is owned by a student before [1]. Knowing that kind of students are accepted in Informatics Engineering program can facilitate the program to determine and provide appropriate techniques or learning methods. In this research, the criteria of prospective students are taken from gender, school origin, majors at school and the total test scores obtained when admission to the Informatics engineering program. Criteria processed using Data Mining with Decision Tree technique. The Decision Tree is a classification and prediction method that is powerful and famous. Decision Tree method change the fact that a very large into a decision tree that represents the rule. Rules can be easily understood by the natural language [2]. Data Mining is a computer science that can identify the criteria of students who enter the appropriate of study program. Using the Decision Tree technique and C4.5 Algorithm allows it to be happened.

Various research related to education has been done by using classification technique. Research on the classification of students who are loyal based on external factors is done by Kakavand [3]. The study compared three classification techniques to see the strongest external factors affecting student loyalty. These results may still be developed taking into account unequal external factors in each college.

The Classification of student performance submitted by Raut [4] and Niswatin [1] which aim to produce a recommendation to the college, enable to improve student ability. In addition, Hamsa [5] also conducted a similar study by comparing two techniques namely classification and fuzzy. Academic records and initial academic information serve as predictions. Prediction is not possible if student details are not entered or not valid and Mesaric [6] was talked about the academic criteria that make a student successful. Others research are by Sudarma [7], Hlaing [8] and Waguih [9] which each discusses about using Decision Tree for network traffic incident, network intrusion detection and detection of Denial of service attack. Different research is also shown by [10, 11]. Their researches performed that some algorithms in the classification technique are used for burn area detecting. Some of classification techniques are also compared to produce the most appropriate algorithm for detecting hotspots caused by fires. Some of researches are mentioned have demonstrated the use of decision tree performed in all areas.

Based on previous research, it is still possible to do new research by utilizing database of Informatics students as sample of study. Training result is a decision tree in rule form. To support the knowledge or information generated, the rule is simplified using chi-square. The chi-square table and the independent test results of each attribute show a simpler rule. The results are also supplemented with predictions of error in the training. Benefit and contributions of this research are helping new admissions committee in the selection of new students based on students criteria. Appropriate information about the criteria of students who choose a study program at a college can facilitate the socialization of relevant study program in the community and help policy-making related to the development of study program and performance of college in the future.

II. MATERIAL AND METHOD

2.1 Decision Tree

Identification criteria can be done by using one of the data mining techniques, such as Decision Tree. The decision tree algorithm is based on a divide-and-conquer approach for classifying a problem. The algorithm works from top to bottom, searching at each stage of the attribute to divide it into the best part of the class, and recursively processing sub problems generated from the division. This strategy generates a decision tree that can be converted into a set of classification rules [12]. The decision tree is a classification method that uses a representation of a tree structure in which each node represents an attribute, its branch represents the value of the attribute, and the leaf represents the class. The topmost node of the decision tree is root node [13].

In the decision tree, there are 3 types of nodes, namely Root Node, Internal Node, and Leaf Node. Root Node is the topmost node, in this node there is no input and can have no output or have more than one output. Internal Node is a branching node, in this node there is only one input and has at least two outputs. The leaf node or terminal node is the final node. At this node there is only one input and no output [14], such as given by Fig. 1.

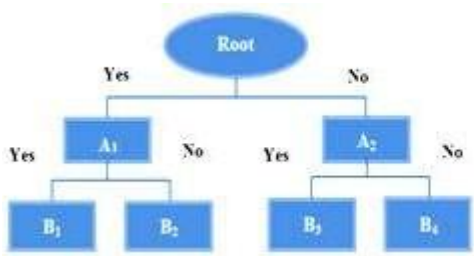

Fig 1: Structure of Decision Tree

Based on Fig. 1, the decision tree depends on the if-then rule, but does not require parameters and metrics. Simple and interpretable structures allow the decision tree to solve multi-type attribute problems. Decision tree can also manage missing values or noise data [14]. Many algorithms can be used in the formation of Decision Tree, such as ID3, CART, and C4.5 [15]. The C4.5 algorithm and decision tree are two indivisible models, because to construct a decision tree, the C4.5 algorithm is required [16]. Building a classification with Decision Tree using Algorithm C4.5, requires some work processes. First, the data is arranged in the form of tables with attributes and records that state a parameter created as a criterion in the formation of a tree. One of the attributes represents per-item solution data, is called the target attribute. Attributes have values called instances. Second, convert data into decision tree by extracting data using entropy. And third, convert the decision tree into a rule that can be simplified [17]. The Entropy formula is as follow.

Entropy (S) = -p+ log2 p+ - p- log2 p-

Description: Entropy (S) = sample space; p+ = number of positive solution; p- = number of negative solution. Entropy is the number of bits that are thought to be needed to extract a class from a number of random data in a sample room [17]. Selected Entropy is one that has the smallest value of sample numbers and from sample attributes numbers.

2.2 Classification Stages

This study included experimental research using new student data at a college that amounted to 87 data. Determination of data called Knowledge Discovery in Database (KDD). Here are the steps in KDD [18].

Fig 2: Classification Stages

Then Fig. 2 is explained in the following.

Data Preparation

Classification stages on Fig. 2 are starting with preparing data training. Data training is the data that taken from the database of new informatics students of a college. The database is selected through the process contained in the data ma or Knowledge Discovery in Database (KDD) [18]. The following describes the steps in preparation of the intended data:

- Selection: The activities undertaken are the selection or segmentation of data based on certain criteria. In this activity, data is selected from many data to be re-selected.
- Pre-processing: At this stage, data cleaning is done, where unsupportable field can be discarded. Additionally the data is reconfigured to ensure the format remains consistent.
- Transformation: is an activity that transforms data so that data can be used and traced. In other words, transformation performs complex data mapping.

Data Training

Data training is the second stage on Fig. 2. This stage is the process of pattern extraction from the selected data, which is the process of KDD to form the pattern.

Decision Tree into the Rule

The last stage on Fig. 2 is to change the decision tree into a rule in accordance with branching that occurs. There are 2 ways to make changes:

- Interpretation and Evaluation: is a process of pattern interpretation into knowledge that can be used for decision making.
- Simplification with Chi-Square: is a process to simplify the rule so that resulting knowledge as expected.

III. MATERIAL AND METHOD

The first Training begins with the preparation of a database of new informatics students. The selected databases amount to 87 data with four criteria or attributes, instance and target attributes, as shown in Table 1.

Table 1: Database New Informatics Students

No	Name of Student	Gender	School Origin	Major	T-TS	Accepted
1	R. Ertantyo	M	SHS	NS	MD	Yes
2	Vinsentwijaya	M	VHS	CNE	MD	Yes
3	Wendy sutriono	M	VHS	CNE	MD	Yes
	-					
	-					
87	Wan Berry	M	SHS	SS	MD	Yes

Table 1 consists of student names as sample test, attributes or criteria consisting of gender, where L is for male and P is for female instances. Furthermore, there are attributes of school origin that consist of SHS (Senior High School) and VHS (Vocational High School) as an instance, majors at schools consisting of three instances, namely NS (Natural Science), SS (Social Sciences) and CNE (Computer and Network Engineering). As the last attribute, there is a total test score obtained when a prospective student enters college. The total test score consists of three instances, namely high (HG), medium (MD) and low (LW). Accepted is called by target attributes and consist of two answers, Yes and No. Target Attributes show whether or not students receive based on database of new informatics students.

3.1 Root Note Determination

As a first step in training using Decision Tree, a root node search is done by comparing entropy values from the four attributes or criteria. The attribute that has the smallest Entropy value will be selected to be the root node.

Table 2: Data Attribute of Gender

Gender	Accepted	Number
Male	+	64
Male	-	11
Female	+	9
Female	-	3
Total		87

It starts with a gender attribute as shown by Table 2. Male instances received as Informatics students (+) amounted to 64 people, while those not accepted (-) amounted to 11 people. Table 2 is also shows that the instance of female students received (+) is 9 people, while those who are not accepted as students of Informatics amount to 3 people. Then this data is processed using the Entropy formula, where -p + log2 p + - p-log2 p- with the following calculation:

Attribute = Gender

Gender = Male = q1

Then q1 = - (64/75)*Log ((64/75), 2)–(11/75)*Log((11/75), 2)

= 0.60144

Gender = Female = q2

Then q2 = - (9/12) * Log ((9/12),2) – (3/12) * Log ((3/12), 2)

= 0.81128

Entropy of Gender = (75/87)*q1 + (12/87)*q2

= 0.6303792

Furthermore, the preparation of the table as well as Table 2 for the attributes of school origin, the major at school, and the total test score. Data trained to obtain entropy value. Here is shown the arrangement of tables and their entropy values shown in tables 3, 4, and 5.

Table 3, 4, 5 show that the amount of school origin data that accepted or not accepted then data is processed using the entropy formula.

Table 3: Data Attribute of School Origin

School Origin	Accepted	Number
SHS	+	37
SHS	-	3
VHS	+	36
VHS	-	11
Total		87

Table 4: Data Attribute of Major at School

Major at School	Accepted	Number
NS	+	28
NS	-	2
SS	+	24
SS	-	7
CNE	+	21
CNE	-	5
Total		87

Table 5: Data Attribute of Total Test Scores

Total Test Scores	Accepted	Number
High	+	3
High	-	0
Middle	+	52
Middle	-	1
Low	+	18
Low	-	13
Total		87

Table 3, 4 and 5, obtained entropy value of each of 0.6007711, 0.6075104 and 0.4318694.

3.2 Tree Compilation

From the four attributes seen that the attribute of Total Test Scores have the lowest value, is 0.4318694.

Mean that, the Total Test Scores attribute is selected as Root Node in the initial tree compilation as Fig. 3.

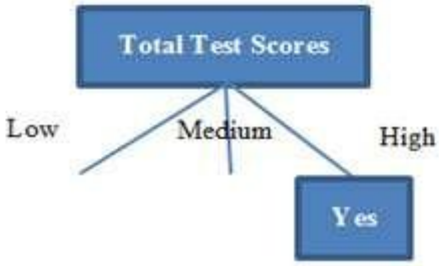

Fig 3: The Initial Tree Compilation

The next node is called the leaf node. The determination of attributes to be leaf nodes is based on instances or branches that have positive (+) and negative (-) values. In Fig. 3, of the three instances or branches, only the medium and low instances that meet the conditions. That is, this instance will bring up the next leaf node.

To compile the next leaf node, the calculation of the entropy value is performed by tracing the total test scores data being medium and low instances. Total test scores becomes a marker attribute to start the calculation of the next leaf node. Based on data grouping from total test scores, the results of medium instance produced kind of leaf nodes. The training data for the total test scores of medium instance as shown in Table 6, 7, and 8.

Table 6: Data Attribute of Gender of Medium Instance

Gender	Accepted	Number
Male	+	44
Male	-	1
Female	+	8
Female	-	0
Total		53

Table 7: Data Attribute of School Origin of Medium Instance

School Origin	Accepted	Number
SHS	+	27
SHS	-	0
VHS	+	25
VHS	-	1
Total		53

Table 8: Data Attribute of Major at School of Medium Instance

Major at School	Accepted	Number
NS	+	19
NS	-	0
SS	+	18
SS	-	0
CNE	+	15
CNE	-	1
Total		53

In the same way, the data are trained and show that some instances have a value of 0. It causes some instances not being computable or should be ignored. For an example that Table 6 has amount of instance female is 0. Automatically, female instance should be ignored. The same thing applied to Table 7 for instance SHS and Table 8 for instance NS and SS. Next step is to calculate the entropy value for each attribute. Table 6, 7 and 8, obtained entropy value of each of 0.1305358, 0.1153779 and 0.1018234.

The training result of all medium instances, generate the smallest entropy on the major at school attributes. This attribute becomes the next leaf node. Further training is done on data that has positive and negative values. The same thing is done on the total test scores of low instance. The calculation process keeps repeating until it does not generate branching anymore. The resulting entropy is 0.8999158. Overall, the final tree compilation formed from the training results is shown in Fig. 4.The Tree Compilation as shown in Fig. 4 shows information that allows someone to be accepted as informatics students. Then the tree structure formed from each leaf node is changed to rules.

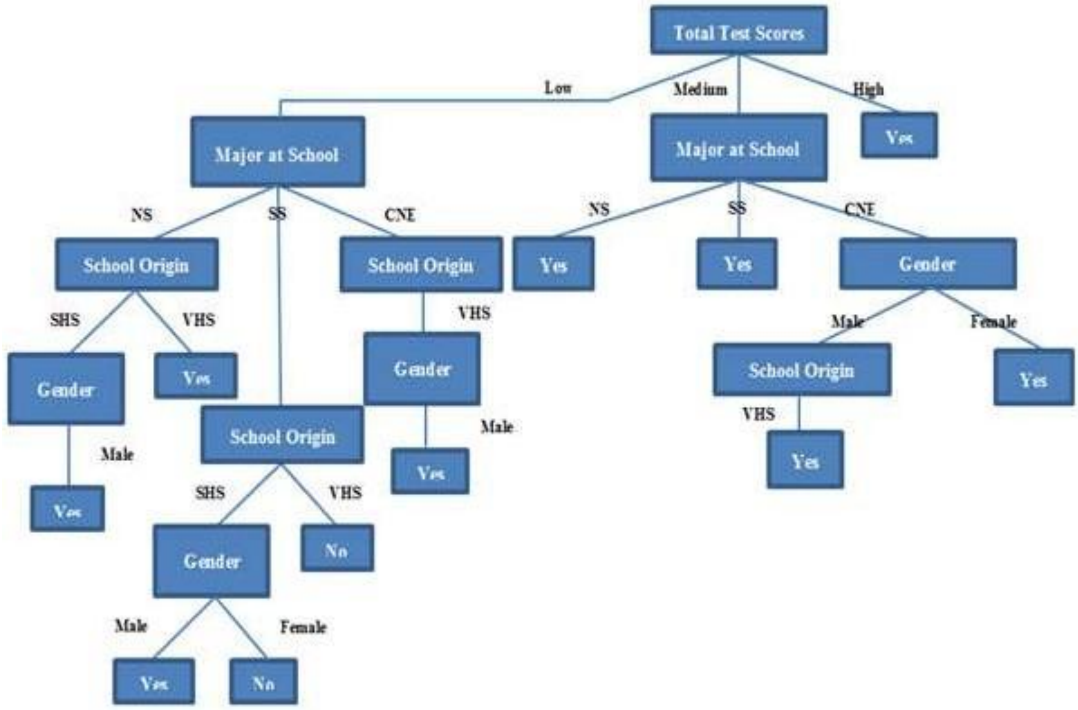

Fig 4: Final of Tree Compilation

3.3 Decision Tree into the Rule

Changing a tree to a rule means creating a rule or rule that matches to the structure of the tree. The rule that is formed based on the final tree compilation in Fig. 4. The determination of the rule starts from root node to each leaf node and the rules are:

R1 : If Total Test Scores = High then Accepted = Yes
R2 : If Total Test Scores = Medium ^ Major at School = NS then Accepted = Yes
R3 : If Total Test Scores = Medium ^ Major at School = SS then Accepted = Yes
R4 : If Total Test Scores = Medium ^ Major at School = CNE ^ Gender = Male ^ School Origin = VHS then Accepted = Yes
R5 : If Total Test Scores = Medium ^ Major at School = CNE ^ Gender = Female then Accepted = Yes
R6 : If Total Test Scores = Low ^ Major at School = NS ^ School Origin = SHS ^ Gender = Male then Accepted = Yes
R7 : If Total Test Scores = Low ^ Major at School = NS ^ School Origin = VHS then Accepted = Yes
R8 : If Total Test Scores = Low ^ Major at School = SS ^ School Origin = SHS ^ Gender = Male then Accepted = Yes
R9 : If Total Test Scores = Low ^ Major at School = SS ^ School Origin = SHS ^ Gender = Female then Accepted = No
R10 : If Total Test Scores = Low ^ Major at School = SS ^ School Origin = VHS then Accepted = No
R11 : If Total Test Scores = Low ^ Major at School = CNE ^ School Origin = VHS ^ Gender = Male then Accepted = Yes

3.4 Simplification and Rule Testing

The training results show that the number of rules formed is 11. The simplification of the rule or tree pruning (Tree Pruning) is needed to avoid the number of prediction model caused by the development of hypothesis by the algorithm used so as to reduce the error of sample training. In this research, simplification and rule testing were performed using chi-square. The preparation of testing with chi-square is as follows [6]:

- Create an integrated distribution table (contingency) by declaring all event values in each rule.
- Calculates the level of independence between criteria on a rule, ie between attributes and target attributes.
- Eliminating unnecessary criteria, which are high independence.

The required data in the integrated distribution table is taken from tables 2, 3, 4, and 5 which are attribute tables of training data. Furthermore the data in tables 2, 3, 4, and 5 are grouped into contingency tables, as in tables 9, 10, 11 and 12.

Table 9: Contingency of Gender

	Male	Female	Marginal
Accepted	64	9	73
No	11	3	14
Marginal	75	12	87

Table 10: Contingency of School Origin

	SHS	VHS	Marginal
Accepted	37	36	73
No	3	11	14
Marginal	40	47	87

Table 11: Contingency of Major at School

	NS	SS	CNE	Marginal
Accepted	28	27	21	76
No	2	4	5	11
Marginal	30	31	26	87

Table 12: Contingency of Total Test Scores

	High	Medium	Low	Marginal
Accepted	3	52	18	73
No	0	1	13	14
Marginal	3	53	31	87

Contingency tables that shown in Table 9, 10, 11, and 12 are obtained by calculating the target attribute that answers Yes and No to each instance of each attribute. The next step is testing the data in contingency table using chi-square distribution. Each contingency table will be calculated to obtain degrees of freedom value (degree of freedom or DF), confidence level (α / alpha)), and value of chi-square table ($X^2\alpha$). Here is the independence test for the gender contingency table.

From Table 9, DF value is:

DF = (number of rows–1)*(Number of coloum-1)
= (2-1) * (2-1)
= 1

Confidence level (α) is = 0.05

$X^2\alpha$ value is: 3.89

Test table of Contingency of Gender Attribute is given by Table 13.

Table 13: Test of Contingency of Gender Attribute

	Male	Female	Marginal
Accepted	62.93103	10.06897	73
No	12.06897	1.931034	14
Marginal	75	12	87

The calculations of Table 13 are described as follows:

To Male Accepted = (73 / 87) * 75 = 62.93103

To Female Accepted = (73 / 87) * 12 = 10.06897

To Male No = (14 / 87) * 75 = 12.06897

To Female No = (14 / 87) * 12 = 1.931034

The value of independence test (chi-square) is =

(((64-62.93103)^2)/62.93103)+

(((9-10.06897)^2)/010.06897)+

(((11-12.06897)^2)/12.06897)+

(((3-1.931034)^2)/1.931034) = **0.81807.**

Based on calculations above, the value of independence test (X2) is smaller than value of chi-square table (X2α). This means that if X2 is smaller than X2α, then the attributes of gender is independent or can be removed. Furthermore, the same calculations are performed on the other three contingency table attributes. In Table 14 summarized independence test results for all attributes.

Table 14: All Attribute Independence Test Results

	Gender	School Origin	Major at School	T-Test-Scores
DF	1	1	2	2
A	0.05	0.05	0.05	0.05
$X^2\alpha$	3.89	3.89	6.27	6.27
X^2	0.81807	4.04812	3.13013	23.8298
Comparison	$X^2 < X^2\alpha$	$X^2 > X^2\alpha$	$X^2 < X^2\alpha$	$X^2 > X^2\alpha$
Decision	Independent Removed	Dependent Permanent	Independent Removed	Dependent Permanent

From Table 14, the independence test result of all attributes shows that there are two attributes that have X^2 value smaller than $X^2\alpha$ value which resulted in both attributes can be removed from the rule. The simplified rule is as follows:

R1 : If Total Test Scores = High then Accepted = Yes
R2 : If Total Test Scores = Medium then Accepted = Yes
R3 : If Total Test Scores = Medium ^ School Origin = VHS then Accepted = Yes
R4 : If Total Test Scores = Low ^ School Origin = SHS then Accepted = Yes
R5 : If Total Test Scores = Low ^ School Origin = VHS then Accepted = Yes
R6 : If Total Test Scores = Low ^ School Origin = SHS then Accepted = No
R7 : If Total Test Scores = Low ^ School Origin = VHS then Accepted = No

3.5 Prediction Result

There are some instances that have a value of 0 which results in ignoring or omitting calculations. Uncounted instances reduce the number of branches and branch variations that are generated. In addition, the training process shows different predictions with the initial database. The difference is in the attribute purpose. Differences show the level of error occurring in the training process. Table 15 contains a comparison of initial databases with predictions (training results).

Table 15: Prediction Result

No	Name of Students	Gender	School Origin	Major at School	T-Test-Scores	Accepted	Prediction
1	R. ErtantyoEp	M	SHS	NS	MD	Yes	Yes
2	VinsentWijaya	M	VHS	CNE	MD	Yes	Yes
3	Wendy Sutriono	M	VHS	CNE	MD	Yes	Yes
	-						
10	Tommy Christian Wijaya	M	SHS	NS	LW	No	Yes
24	-						
26	IlhamAbadiTarihoran	M	VHS	CNE	LW	No	Yes
29	-						
31	Bobby Erlangga	M	VHS	CNE	LW	No	Yes
47	Fendi Santo Wijaya	M	VHS	NS	MD	Yes	Yes
48	DheaAnanda	F	VHS	SS	MD	Yes	Yes
49	RekiKurniawan	M	VHS	CNE	LW	No	Yes
51	-						
60	Andini Elisabeth	F	SHS	SS	LW	Yes	No
68	-						
70	StephanusSamosir	M	VHS	CNE	LW	No	Yes
71	HeriantoMendrofa	M	VHS	NS	MD	Yes	Yes
72	FajarAdiPrasetyo	M	VHS	CNE	MD	No	Yes
73	KrismanDaeli	M	VHS	SS	MD	Yes	Yes
74	IlhamFikri	M	SHS	NS	LW	No	Yes
75	-						
87	Wan Berry Pranata	M	SHS	SS	MD	Yes	Yes

From the Table 15, can be seen clearly comparison between the sample data from the database of new Informatics students with the results of training or prediction. There are 8 of 87 sample data that the results are not equal to the training results (blue) with the percentage of error is:
Percentage of Error = (8 / 87) * 100% = 9.19%

The difference is caused by the intervention of experts or colleges concerned when selecting students. Therefore, we need to know the comparison by using the percentage of error rate.

When compared with previous studies [1], no simplification of the rule was made. Rule needs to be simplified as it will be a reference in making conclusions or information. This study gives different results because it shows comparison of decision tree results before and after simplified. The more criteria and sub criteria that are compared will result in many rules. Decision makers need more simple and accurate information.

IV. CONCLUSION

Training begins Implementation of Decision Tree in determining the criteria of students received in a Informatics Engineering program is very helpful in maintaining the sustainability of the program. This research has resulted in knowledge or information about the criteria of students who choose Informatics Engineering program through training conducted. The resulting rules have been tested by chi-square.

The error rate that occurred in the training result was 9.19% not only because of the weakness of the attribute value but also influenced by the involvement of the college in determining the accepted student. The high percentage of errors can reduce the quality of students due to lax process of acceptance of prospective students who are not in accordance with the consistency of attributes that have been determined. This means that students should be accepted, not accepted or otherwise. Error percentage calculation is expected to decrease every year. Consistency of college is required in the process of admission of new prospective students.

The results of this study are expected to be a reference for colleges. Seven rules produced indicate that Informatics Engineering program accepted students based on that rules. Knowing and understanding the criteria of students who choose study program in place, will facilitate socialization and admissions of new students process. In addition, this research can still be developed through the addition of training attributes and the number of sample according to the type of program.

ACKNOWLEDGEMENTS

The author would like to thank to Universitas Sumatra Utara who supported this research.

REFERENCES

[1]. Niswatin, R.K., and Wulanningrum R, Prediction of college student achievement based on educational background using decision tree methods, Indonesian Journal of Electrical Engineering and Computer Science,4(2), 2016,429-438.doi: 10.11591/ijeecs.v4.i2.pp429-438.
[2]. Kusrini and Luthfi E. T, Algoritmadata mining(Yogyakarta: Andi Offset, 2009).
[3]. Kakavand S, Mokfi T, and Jafar M.T., Prediction the loyal student using decision tree algorithms, Information and Communication Technology Research,4(1), 2014, 32-37.
[4]. Raut A. B., and Nichat A. A., Students performance prediction using decision tree technique. International Journal of computational Intelligence Research,13(7), 2017, 1735-1741.
[5]. Hamsa H, Indiradevi S, and Jubilant J.K., Student academic performance prediction model using decision tree and fuzzy genetic algorithm, Procedia Technology, 25(2016), 2016, 326-332.
[6]. Mesaric J and Sebalj D, Decision trees for predicting the academic success of students. Croation Operational Research Review, 7(2016), 2016, 367-388.doi: 10.17535.
[7]. Sudarma, M and Hostiadi D. P., The establishment of decision tree model in network traffic incident using c4.5 method, International Journal of Informatics and Communication Technology, 2014, 23-29.
[8]. Hlaing T, Feature selection and fuzzy decision tree for network intrusion detection, International Journal of Informatics and Communication Technology,2012, 109-118.
[9]. Waguih H, A data mining approach for the detection of denial of service attack, IAES International Journal of Artificial Intelligence, 2013, 99-106.
[10]. Sitanggang, I. S., Yaakob R, Mustapha N, and Ainudin, Burn area processing to generate false alarm data for hotspot prediction models,TELKOMNIKA,2015, 1037-1046.doi: 10.12928/TELKOMNIKA.v13i3.1543.
[11]. Thariqa P, Sitanggang, I. S., and Syaufina L, Comparative analysis of spatial decision tree algorithms for burned area of peatland in rokan hilir riau,TELKOMNIKA,2016, 684-691.doi: 10.12928/TELKOMNIKA.v14i1.3540.
[12]. Witten I. H, Frank E, Hall, M. A, Data mining practical machine learning tools and techniques,(3rd ed), (USA: Elsevier, 2011).
[13]. Gorunescu, F, Data mining concept model and techniques(Berlin: Springer,ISBN 978-3-642-19720-8., 2011).
[14]. Dua and Xian, Data mining and machine learning in cybersecurity (CRC Press: Taylor & Francis Group, 2011).
[15]. Larose D.T, and Chandal D, Discovering knowledge in data, introduction to data mining, 2nd Edition (New Jersey: John Wiley & Sons, Inc, 2014).
[16]. Han and Kamber, Concept and technique, 3nd Edition(2012).
[17]. Basuki A and Iwan S, Decision tree (Politeknik Elektronika Negeri Surabaya, 2003).
[18]. Fayyad, Shapiro, and Smyth, From data mining to knowledge discovery in database(AAAI, 1996).

Deny Jollyta "Decision Tree and Chi-Square Analysis to Determine Student's Criteria Choosing Study Program "American Journal of Engineering Research (AJER), vol. 7, no. 10, 2018, pp. 12-20