

onf._Ser.__Mater._Sci._Eng._879
_012072_INCITEST_Afrianto_dkk
.pdf
by

Submission date: 12-Apr-2023 12:28PM (UTC+0700)

Submission ID: 2062296831

File name: onf._Ser.__Mater._Sci._Eng._879_012072_INCITEST_Afrianto_dkk.pdf (741.55K)

Word count: 5193

Character count: 26808

PAPER • OPEN ACCESS

13

Decision Tree Method with C4.5 Algorithm for Students Classification Who is Entitled to Receive Indonesian Smart Card (KIP)

To cite this article: E Afrianto *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **879** 012072

1

View the [article online](#) for updates and enhancements.

You may also like

- [Analysis of data mining classification by comparison of C4.5 and ID algorithms](#)

R. Sudrajat, I. Irianingsih and D. Krisnawan

16

- [Implementing binary particle swarm optimization and C4.5 decision tree for cancer detection based on microarray data classification](#)

A C Pradana, Adiwijaya and A Aditsania

- [Classification of Farmer's Eligibility as Recipients of Subsidized Fertilizer](#)

9

[Assistance with C4.5 Algorithm](#)
Diah Nur Yunita, Anna Hendri Soleliza Jones and Dewi Soyusiawaty



*Benefit from connecting
with your community*

ECS Membership = Connection

ECS membership connects you to the electrochemical community:

- Facilitate your research and discovery through ECS meetings which convene scientists from around the world;
- Access professional support through your lifetime career;
- Open up mentorship opportunities across the stages of your career;
- Build relationships that nurture partnership, teamwork—and success!

Join ECS!

Visit electrochem.org/join



¹³ **Decision Tree Method with C4.5 Algorithm for Students Classification Who is Entitled to Receive Indonesian Smart Card (KIP)**

E Afrianto¹, J E Suseno², B Warsito³

¹Master of Information System, School of Postgraduate Studies, Diponegoro University, Jl.Imam Bardjo SH, No.5, Pleburan, Semarang, Indonesia

²Department of Information Technology, Faculty of Science and Mathematics, Diponegoro University, Jl.Prof.H.Soedarto SH, Tembalang, Semarang 50275, Indonesia

³Department of Statistics, Faculty of Science and Mathematics, Diponegoro University, Jl.Prof.H.Soedarto SH, Tembalang, Semarang 50275, Indonesia

Email: *ekofrogie@gmail.com, **jatmikoendro@gmail.com, ***budiwst2@gmail.com

Abstract. The research intends to create an application program which is able to make analysis data in a school to classify students who is entitled to accept Indonesian Smart Card. It was a government program that aims to finance school education free of charge, for children from poor families. Problems arose when the distribution to students, some children from poor families did not get it but rich families actually got it. Analysis and the right way are needed so that the distribution can be right on target to eligible students. This study used a decision tree method with C4.5 algorithm to classify students who are entitled to receive the Indonesian Smart Card. This application is developed with C4.5 algorithm to determine the decision tree method. Research data was carried out at Junior High School One of Jatibarang, Brebes. Data record of 300 students, with 240 as training data while 60 as testing data. From the test results were accuracy was 97%, it is proven this method has a high accuracy so that the application can help the decision makers solved the distribution problems to eligible students. The impact is that poor students will get the right to receive Indonesian Smart Card.

1. Introduction

Education and poverty are problems from the past which have occurred in developing countries. The relationship between education and poverty is very large due to the high burden of education costs, so that people who are unable can be difficult to reach, even though obtaining education will provide the ability for students to develop through the mastery of knowledge and skills [1]. It is worried that children who do not attend school and drop out of school will later be carried out in crime. Children do not have the provision of education, so it will be easily affected by negative relationships in the environment that does not get education at school [2]. Education is one of the three main keys in the Human Development Index, namely the health index, the education index and the standard of living index. The Human Poverty Index is in contrast to the Human Development Index. With a high gap where only rich people can send their children to school, while people who are economically unable find it difficult to send their children to school. Therefore, the difference between human development and poverty is a picture of the problems that must be evaluated and resolved by the government [3]. To overcome this government has now created an Indonesian Smart Card (KIP) program that aims to provide free education costs for



⁵ Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Published under licence by IOP Publishing Ltd

poor and vulnerable poor families who want to send their children aged 6 to 21 years. Students who obtain KIP from the government will be given regular cash funds stored in the KIP card for free tuition fees at school. With the KIP program from the government it is expected that the number of out of school and dropping out of school can drop dramatically, and can attract back students who have dropped out of school to return to school [4]. From the results of observations and interviews of researchers with several parents of students and the school, researchers get the facts in the field that for the provision of the Indonesian Smart Card which is the Indonesian Smart Program (PIP) given by the government for underprivileged students, it turns out there are still wrong targets. Poor students who are entitled to get educational assistance funds from the government through KIP apparently do not get the card, while students who are able to get KIP to get education assistance funds. Based on these problems, it requires an application that is able to provide decisions automatically and accurately in predicting classification students who is entitled to receive an Indonesian Smart Card (KIP), using data and information that already exist at school.

Research using the decision tree with C4.5 algorithm has been used in previous studies. Decision tree C4.5 algorithm was one method of data mining to predict the selection of scholarship grantee to senior high school student. It can be seen from internal and external factors of the students, the problems in getting scholarship were some criteria such as parents' income, parental responsibility, and majors, academic and non-academic achievement. Decision tree C4.5 algorithm test results and analysis was accurately applied for prediction with 94.7368% accuracy of final grades of senior high school students [5]. Decision tree algorithm C4.5 was used to construct decision tree according to the actual situation and the non-intelligence factors on English learning achievement of college students. The test results indicated that the prediction rule can accurately predict of average students and underachievers in English learning achievement. It can be used as a good predictor to meet the requirement of teaching assistance and for the students with average or poor achievement in English learning [6]. Decision tree method with algorithm C4.5 was constructed an assertive reasoning selection methodology by improving acquisition on the balance coefficient using emergency knowledge elements. It was the key to emergency management divided by two levels model selection through hierarchical representation.

They were top model construction process based on assertive reasoning and underlying model selection process based on decision tree algorithm C4.5. From the results, this method improved the accuracy and timeliness. And also it reduced the requirements for emergency domain knowledge [7]. The Decision tree with C4.5 algorithm can be used to detect and diagnose errors in network connected photovoltaic systems. The data set has been divided into two parts, where 66% is used for training data and the rest is for testing data. Next, new data are recorded to evaluate durability, effectiveness and efficiency of the two models. The test results show that the model has a high predictive performance in detection with high accuracy while the diagnosis model has an accuracy of 99.80% [8]. The Decision tree C4.5 algorithm can be used in a process to detect and classify VT (Ventricular Tachycardia) and VF (Ventricular Fibrillation) arrhythmias using temporal, spectral, and statistical features. The assessment of this process was carried out on 57 electrocardiograms (ECG) signal test data. Testing and the results showed that the proposed method achieved a sensitivity of 90.97%, specificity 97.86% and an accuracy of 97.02% in the C4.5 classification, which was better than the results obtained by the Support Vector Machine algorithm which has an accuracy of 92.23%. Informative features and C4.5 algorithm can be a help for doctors to detect correctly [9]. One of the data mining methods will be applied to build this application, namely the Decision Tree method with C4.5 Algorithm for data classification and predictive nature. Decision tree is an important technique in data mining for the form of classification and prediction of data classes. Methods that use decision trees will present the rules.

The data in the decision tree will be expressed in tabular form with attributes and records. The parameters formed state the attributes that are used as criteria in compiling a decision tree. The research on the classification of students who are entitled to receive Indonesian Smart Card with the decision tree method with the C4.5 algorithm has not existed before, so researcher need to examine this because of the distribution of the wrong target. So that poor students who are entitled to do not even get it, while

students who are rich there who got it. Therefore, with this research, poor students will be entitled to get Indonesian Smart Card.

Based on the results of previous studies above, it can be clearly said that the decision tree method has a high enough accuracy value in terms of predicting classification what will happen in the future. This is due to the fact that the way it works is collecting existing data, studying the data and the results can provide assistance to a precise and accurate decision making. The intention of this study is that decision makers in classification students who were entitled to receive an Indonesian Smart Card at school can be assisted in making a final decision regarding the classification of students to be entitled. Decisions taken must be precise and accurate so that the distribution Indonesian Smart Card (KIP) is not wrong target.

2. Method

The method used in this research is the data of recipient students and those who did not receive the Indonesian Smart Card (KIP) from students of SMP Negeri 1 Jatibarang, Brebes. Data collection was carried out surveys and interviews directly with the Principal and School Operators by asking for and retrieving data from Microsoft Excel students directly from the school computer used as a school data storage. The data in this research material relates to the data of student ⁴ who received and did not receive the Indonesian Smart Card (KIP). The dataset that has been obtained is then divided into 2 types of data, namely training data and testing data. The data that has been shared later will be entered into machine learning, which means that the machine is designed for the learning phase (train). Therefore, in order to know whether machine learning is learning in accordance with the training data and the performance provided is in accordance with the desired level of accuracy, after the machine is taught (training), it is then necessary to be tested (testing). Usually as a general rule (rule of thumb) for the proportion of training data by 80%, while testing data by 20%. There is also a set proportion, training data by 75%, while testing data by 25%. The data used in this study were 300 student records divided into training data of 240 student records, while the testing data were 60 student records [10]. This research was conducted in several stages that were worked on, in getting accurate and optimal results. These stages are divided into several sections, namely Planning, Determination of Samples, Data Collection, Determination of Training Data and Testing Data, Use of Decision Tree Method with C4.5 Algorithm and Data Processing.

The stage of making this classification first determines and plans the criteria ³ in the KIP class. Within these criteria there are criteria needed to determine the KIP class, namely father's education, father's occupation, mother's education, mother's occupation, number of family dependents, property tax bills and electricity bills. Next to facilitate the process of system calculation, the criteria and sub-criteria will be converted and transformed into categories / classes and in the form of notations or symbols as in table 1 and table 2.

Table 1. Criteria in Determining Variables for KIP Recipient Students

No	Code	³ Criterias	Information
1	A	Father's Education	Variabel Input
2	B	Father's Occupation	Variabel Input
3	C	Mother's Education	Variabel Input
4	D	Mother's Occupation	Variabel Input
5	E	Number of Family Dependents	Variabel Input
6	F	Property Tax Bills	Variabel Input
7	G	Electricity Bills	Variabel Input
8	H	Entitled to Receive KIP	Variabel Input
		Not Entitled to Receive KIP	

Table 2. KIP Criteria and Sub Criteria for Student Recipients

No	Criteria	Code	Sub Criteria	Category / Class	Symbol
³ 1	Father's Education	A1	Dropped out of SD, SD/Equivalent, SMP, MTs	Low	0
		A2	SMA, SMK, MA	Intermediate	1
		A3	D1, D2, D3, D4, S1, S2	High	2
2	Father's Occupation	B1	Farm Workers, Small Traders, Low Employee, Small Entrepreneur	Low	0
		B2	Sharecroppers, Middle Traders, Middle Employee, Middle Entrepreneur PNS/TNI/POLRI, Landowner	Medium	1
		B3	Farmers, Top Employee, Wholesalers	High	2
³ 3	Mother's Education	C1	Dropped Out of SD, SD/Equivalent, SMP, MTs SMA, SMK, MA	Low	0
		C2	D1, D2, D3, D4, S1, S2	Intermediate	1
		C3		High	2
4	Mother's Occupation	D1	Housewife, Farm Workers, Small Traders, Low Employee, Small Entrepreneur Sharecroppers, Middle	Low	0
		D2	Traders, Middle Employee, Middle Entrepreneur PNS/TNI/POLRI, Landowner Farmers, Top Employee,	Medium	1
		D3	Wholesalers	High	2
⁵ 5	Number of Family Dependents	E1	>3 Members of Family	Many	0
		E2	3 Members of Family	Moderate	1
		E3	¹² 2 Members of Family	Few	2
6	Property Tax Bills	H1	Rp.10.000 – Rp.50.000	Cheap	0
		H2	Rp.50.000– Rp.150.000	Moderate	1
		H3	Rp.150.000– Rp.300.000	Expensive	2

No	Criteria	Code	Sub Criteria	Category / Class	Symbol
7	Electricity	I1	Rp.10.000 – Rp.60.000	Cheap	0
	Bills	I2	Rp.60.000 – Rp.160.000	Moderate	1
		I3	Rp.160.000 – Rp.300.000	Expensive	2

2.1. Data Mining

Data mining is the process of determining a pattern and information from large amounts of data. The source data is in the form of a database, data warehouse, web and other information storage places or dynamic data in the system. Data mining is an analytical database observation to summarize data in a new way or method that is useful and can be understood by data owners. Basically, the process carried out by data mining is as follows: prediction, description, classification, estimation, clustering and association [11].

2.2. Classification

Classification is a process of finding a model or function that distinguishes concepts or classes of data that have the aim of predicting the class of objects whose class labels are unknown based on the analysis of training data or training data (data objects whose class is known). Algorithms used for the classification process are C4.5 Algorithm, key nearest neighbour, naïve bayes, genetic algorithms, rule based methods and Support Vector Machines (SVM) [12].

2.3. Decision Tree

Decision tree is a decision support method that uses a decision model that is shaped like a tree or tree. Decision tree can overcome a problem by mapping various possible alternatives, and there are also possible factors that can affect the alternative along with the existing estimates. Decision tree is one method that can be used to display an algorithm which only contains conditional control statements. Decision tree is a structure like a flowchart where each internal node represents the possibilities that exist in the attribute, each branch represents the result of that possibility, and each leaf node represents the class label (the decision is taken after counting all the attributes)[13]. Examples of images from the decision tree process can be seen and explained their explanation in Figure 1.

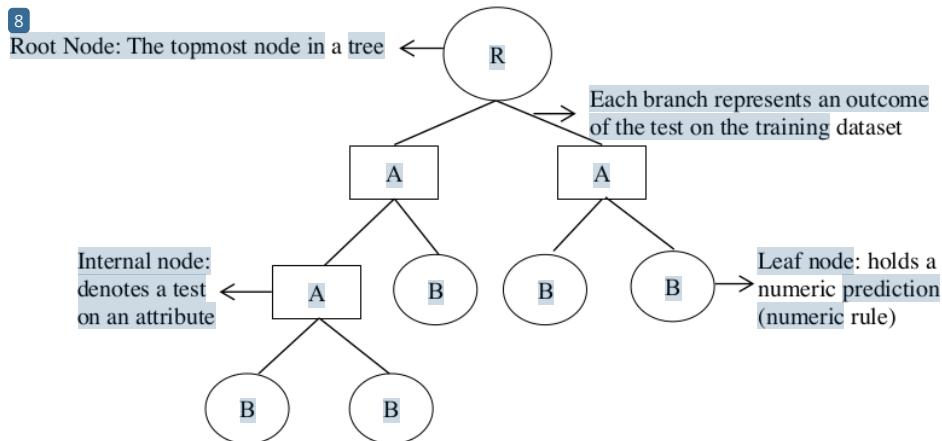


Figure 1. Decision Tree

14 2.4. Algorithm C4.5

C4.5 Algorithm is a type of algorithm for classification and prediction of data using decision tree techniques that have well-known advantages and are very popular compared to other classification and prediction techniques, such as naive bayes, random forest and Key Nearest Neighbour (KNN). The advantages of the C4.5 algorithm with this decision tree technique, for example; can handle missing attribute values, can process discrete and numeric data (continuous) which results in rules that are easy to interpret and have the fastest ability among other algorithms that use main memory on a computer [14]. To select an attribute as the root node based on the highest gain value of the existing attributes. To calculate the entropy and gain using the following formula:

2.4.1. Entropy

Entropy is a measure that can determine the characteristics of impurity and homogeneity of a data set from information theory [15]. The entropy formula is defined as follows:

$$\text{Entropy (S)} = \sum_{i=1}^n (-p_i) * \log_2 (p_i) \quad (1)$$

Informations:

E : Entropy

S : Case Set

n : Number of S Partition

p_i : Number of Sample for Class i / Proportion of S_i to S

2.4.2. Gain

Gain is a measure of the effectiveness of an attribute in classifying data, after the entropy value obtained for a data set [16]. The gain formula is defined as follows:

$$G(S,A) = \text{Entropy (S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy (S}_i) \quad (2)$$

Informations:

G : Gain

S : Case Set

A : Attribute

n : Number of Partition Attribute A

$|S_i|$: Number of Sample Partition for - i

$|S|$: Number of Sample in S

3. Results and Discussion

3.1. Implementation Calculation Using Algorithm C4.5

These are the examples of calculation entropy and gain in C4.5 algorithm. The calculations on training data totalling 240 students with the C4.5 algorithm based on training data are as follows:

The formula determines the entropy of all attributes (total):

$$\text{Entropy (S)} = \sum_{i=1}^n (-p_i) * (\log p_i)$$

Calculations of node 1 (root)

Is known:

- Number of cases (S) = 240
- Number of cases - Eligible (S1) = 160
- Number of cases - Not Eligible (S2) = 80
- Entropy Total (S) = ...

Entropy (Total) = ((-Number of Eligible / Number of Cases (S) * log₂ (Number of Eligible / Number of Cases (S)) + (-Number of Not Eligible / Number of Cases (S) * log₂ (Number of Not Eligible / Number of Cases (S)))

$$\begin{aligned}\text{Entropy (Total)} &= ((-160/240 * \log_2(160/240) + (-80/240 * \log_2(80/240))) \\ &= ((-0,666666667 \times (-0,5849625) + (-0,333333333) \times (-1,5849625)) \\ &= 0.389975 + 0.528320833\end{aligned}$$

$$\text{Entropy (Total)} = 0,918295834$$

After calculating the total entropy, then the entropy calculation for each attribute is performed. Calculation of each attribute as follows:

Attributes of Father Education (Low Value)

$$\text{Entropy (S)} = \sum_{i=1}^n (-p_i) * \log_2(p_i)$$

$$\text{Entropy (Father's Education, Low)} = \sum_{i=1}^n (-p_i) * \log(p_i)$$

$$\begin{aligned}\text{Entropy (Father's Education, Low)} &= ((-\text{Number of entitled} / \text{Number of cases (S)} * \log_2(\text{Number of} \\ &\quad \text{entitled} / \text{Number of cases (S)}) + (-\text{Number of eligible} / \text{Number} \\ &\quad \text{of cases (S)} * \log_2(\text{Number of non-entitled} / \text{Number of cases} \\ &\quad \text{(S)})) \\ &= ((-144/152 * \log_2(144/152) + (-8/152 * \log_2(8/152))) \\ &= ((-0,947368421) \times (-0,078002512) + (-0,0526315789) \times (- \\ &\quad 4.24792751)) \\ &= 0.0738971166 + 0.223575132\end{aligned}$$

$$\text{Entropy (Father's Education, Low)} = 0,297472249$$

From the entropy calculation of each attribute then a gain calculation is performed, while the example of the gain calculation is explained as below. The calculation for each gain attribute is then the same, to form a temporary decision tree from the node 1 table.

Calculation of Root Gain

$$G(S,A) = \text{Entropy (S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy (S}_i)$$

$$\begin{aligned}\text{Gain (Total, Father's Education)} &= \text{Entropy Total} - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy (S}_i) \\ &= \text{Entropy Total} - ((\text{Number of Low Scores} / \text{Number of Cases (S)}) * \\ &\quad \text{Low Value Entropy}) - (\text{Number of Intermediate Values} / \text{Number} \\ &\quad \text{of Cases (S)}) * \text{Intermediate Value Entropy}) - (\text{Number of High} \\ &\quad \text{Values} / \text{Number of Cases (S)}) * \text{High Value Entropy))\end{aligned}$$

$$\begin{aligned}
&= 0,918295834 - ((76 / 120) * 0,297472249) - (32 / 120) \\
&\quad * 0,811278124) - (12 / 120) * 0)) \\
&= 0,918295834 - (0,633333333 * 0,297472249) - (0,266666667 * \\
&\quad 0,811278124) - (0,1 * 0) \\
&= 0,918295834 - (0,188399091 - 0,216340833 - 0) \\
&= 0,918295834 - (-0,027941742)
\end{aligned}$$

Gain (Total, Father's Education) = 0.946237576

Table 3. Calculation of node 1

Node	Attribute	Value	Number of Case (S)	Entitled (S1)	Not Entitled (S2)	Entropy	Gain
1	Total		240	160	80	0.918295834	0.946237576
		Father's Education					
		Low	152	144	8	0.297472249	
		Intermediate	64	16	48	0.811278124	
	Father's Occupation	High	24	0	24	0	0.918295834
		Low	142	142	0	0	
		Medium	28	0	28	0	
		High	50	0	50	0	
	Mother's Education						0.913478625
		Low	156	150	6	0.235193382	
		Intermediate	68	8	60	0.522559375	
		High	16	0	16	0	
	Mother's Occupation						0.316689088
		Low	200	160	40	0.721928095	
		Medium	18	0	18	0	
		High	22	0	22	0	
	Number of Family Dependents						1.252629898
		Many	104	104	0	0	
		Moderate	86	56	30	0.933025295	
		Few	50	0	50	0	
	Property Tax Bills						0.918295834
		Cheap	164	164	0	0	
		Moderate	26	0	26	0	
		Expensive	50	0	50	0	
	Electricity Bills						0.918295834
		Cheap	160	160	0	0	
		Moderate	28	0	28	0	
		Expensive	52	0	52	0	

14

From the results of table 3. It can be seen that the attribute with the highest gain is the Number of Family Dependents, amounting to 1.252629898. Thus, the Number of Family Dependents as the root node. There are three values of the Total Family Dependents attribute, namely many, medium and few. Little attribute value has classified the case into one decision, namely the decision "Not Entitled", the attribute value of many has classified the case into one decision, namely the decision "Entitled" so that no further calculation is needed. But for the attribute value is still need to be calculated again, because there are still "Entitled" and "Not Entitled". After calculating the entropy and gain of each attribute, it will produce nodes in tabular form which will be implemented into the stages of the decision tree form until the final decision tree is formed. The final decision tree is in Figure 2.

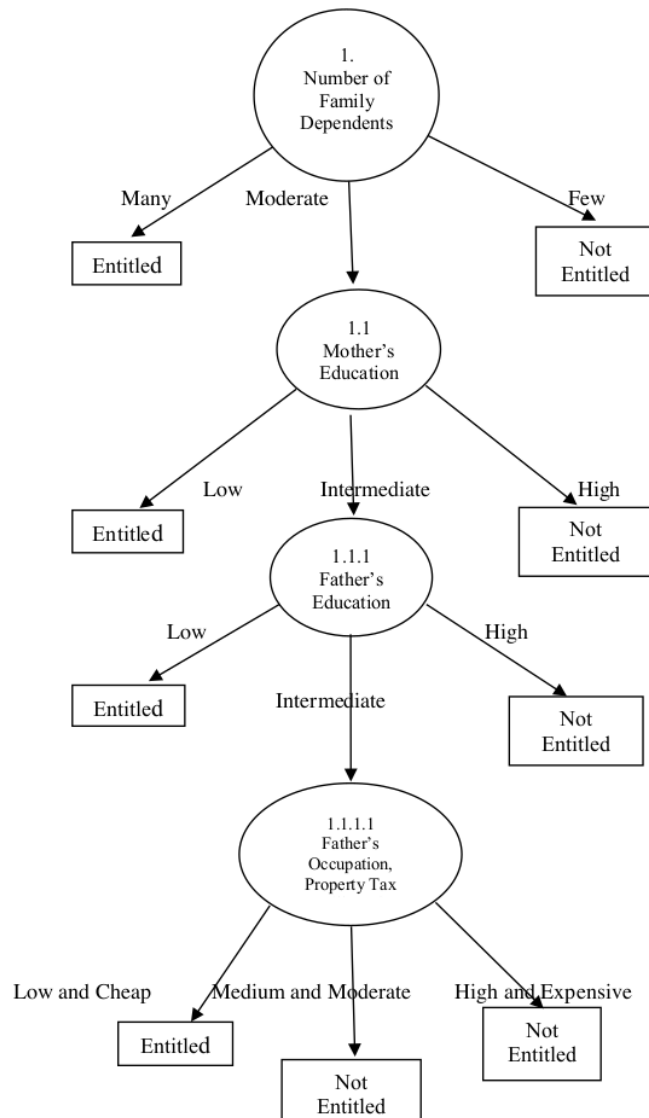


Figure 2. Decision Tree Result

From the results of the decision tree a classification rules will be formed that are used for the classification rules on the system. This is remarks on the results of the decision tree:

1. If there are many dependents, you are entitled to a KIP card
2. If the number of dependents is moderate, the mother's education is low, and then she is entitled to a KIP card
3. If the number of dependents is moderate, mother's education is low, father's education is low, and then you are entitled to a KIP card
4. If the number of dependents is moderate, mother's education is intermediate, father's education is low, father's occupation is low, property tax bills and electricity bills are cheap, and then are entitled to KIP
5. If the number of dependents is moderate, mother's education is intermediate, father's education is intermediate, father's job is medium, property tax bills and electricity bills are medium, and then it is not eligible for KIP
6. If the number of dependents is moderate, mother's education is intermediate, father's education is high, father's occupation is high, property tax bills and electricity bills are expensive, and then it is not eligible for KIP
7. If the number of family dependents is few, they are not entitled to a KIP card
8. If the number of dependents is moderate, the mother's education is high, and then she is not entitled to a KIP card
9. If the number of dependents is moderate, mother's education is high, father's education is high, and then she is not entitled to obtain a KIP card
10. If the number of dependents is moderate, mother's education is intermediate, father's education is high, father's occupation is high, property tax bills and electricity bills are expensive, and then it is not eligible for KIP

3.2. Validate Testing

The process of calculating the validity testing using the Confusion Matrix method. Confusion matrix is a method for calculating accuracy in C4.5 decision tree algorithm. Validity testing is carried out on the results of testing data of students totaling 60 students consisting of attributes and attribute values that will be determined the decision is entitled to or not entitled to receive KIP [17].

Table 4. Confusion Matrix

C4.5 Classification	Positive Classification	Negative Classification
Entitled	TP (True Positive) = 40	FN (False Negative) = 2
Not Entitled	FP (False Positive) = 0	TN (True Negative) = 18

Based on the results of data validation testing with the calculation of the formula above with the confusion matrix method, the results obtained are 97% system classification accuracy, the result of classification recall is 95%, the precision results are 100% and the error rate is 3%. The conclusion obtained from the validation of the decision tree method with the C4.5 algorithm for the classification of students who are entitled to receive the Indonesia Smart Card (KIP) is declared valid and the decision tree C4.5 algorithm method can be used as a classification method for KIP acceptance because of its large accuracy.

4. Conclusion

Based on the results of research and discussion of the classification system of students who are entitled to receive the Indonesian Smart Card (KIP), conclusions can be obtained from the case study conducted at SMP Negeri 2 Jatibarang, in the classification of students who are entitled to receive KIP at school using the decision tree method with the C4.5 algorithm illustrates that the classification system can be used by users because of its excellent performance. From this research, it can validate with confusion matrix method with classification results are accuracy is 97%, recall is 95%, precision is 100% and error rate is 3%. It can be concluded that decision tree method using C4.5 algorithm for students' classification that is entitled to receive an Indonesian Smart Card (KIP) is valid and a system which is able to determine students who are entitled to get a smart Indonesian card is in accordance with their classification for students who are poor or economically incapable.

Acknowledgement

Researcher gives many thanks to Mrs. Sri Yulianti, the Principal of SMP Negeri 1 Jatibarang Brebes, who has allowed to conduct research and process of collecting data at the school. The researcher also gives thank all the school teachers and staffs who has facilitated this research. This research is expected to be useful for teacher or official at school in determining students who is entitled to receive Indonesian Smart Card (KIP).

References

- [1] Tarabini, A and Jacovkis J. 2012. The poverty reduction strategy papers: An analysis of a hegemonic link between education and poverty. *International Journal of Educational Development*, **32**(4), pp. 507-516.
- [2] Nguyen, H T. 2019. Do more educated neighbourhoods experience less property crime? Evidence from Indonesia. *International Journal of Educational Development*, **64**, pp. 27-37.
- [3] Cremin, P and Nakabugo, M G. 2012. Education, development and poverty reduction: A literature critique. *International Journal of Educational Development*, **32**(4), pp. 499-506.
- [4] Direktorat Pembinaan Sekolah Menengah Atas. 2015. *Petunjuk Teknis Program Indonesia Pintar* (PIP), Subdit Kelembagaan dan Peserta Didik, Jakarta.
- [5] Sugiyarti E, Jasmi K A, Basiron B, Huda M, Shankar K and Maselena A. 2018. Decision support system of scholarship grantee selection using data mining. *International Journal of Pure and Applied Mathematics*, **119**(15), pp. 2239-2249.
- [6] Li, M .2018. A study on the influence of non-intelligence factors on college students' English learning achievement based on C4.5 algorithm of decision tree. *Wireless Personal Communications*, **102**(2), pp. 1213-1222.
- [7] Han L, Li W and Su Z. 2019. An assertive reasoning method for emergency response management based on knowledge elements C4. 5 decision tree. *Expert Systems with Applications*, **122**, pp. 65-74.
- [8] Benkercha, R and Moulahoum, S. 2018. Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system. *Solar Energy*, **173**, pp. pp. 610-634.
- [9] Mohanty, M., Sahoo, S., Biswal, P., and Sabut, S. 2018. Efficient classification of ventricular arrhythmias using feature selection and C4.5 classifier. *Biomedical Signal Processing and Control*, **44**, pp. 200-208.
- [10] Gola J, Britz D, Staudt T, Winter M, Schneider A S, Ludovici M and Mücklich F. 2018. Advanced microstructure classification by data mining methods. *Computational Materials Science*, **148**, pp. 324-335.
- [11] Issad H A, Aoudjit R and Rodrigues J J. 2019. A comprehensive review of Data Mining techniques in smart agriculture. *Engineering in Agriculture, Environment and Food*.
- [12] Meng Z and Shi Z. 2020. On rule acquisition methods for data classification in heterogeneous incomplete decision systems. *Knowledge-Based Systems*, pp. 105472.
- [13] Meng X, Zhang P, Xu Y and Xie H. 2020. Construction of decision tree based on C4. 5 algorithm

- for online voltage stability assessment. *International Journal of Electrical Power & Energy Systems*, **118**, pp. 105793.
- [14] Quinlan J R. 2014. *C4. 5: programs for machine learning*. Elsevier.
- [15] Hssina B, Merbouha A, Ezzikouri H and Erritali M. 2014. A comparative study of decision tree ID3 and C4. 5. *International Journal of Advanced Computer Science and Applications*, **4**(2), pp. 13-19.
- [16] Han J, Pei J and Kamber M. 2011. *Data mining: concepts and techniques*. Elsevier.
- [17] Xu J, Zhang Y and Miao D .2020. Three-way confusion matrix for classification: a measure driven view. *Information Sciences*, **507**, pp. 772-794.

ORIGINALITY REPORT

18%

SIMILARITY INDEX

14%

INTERNET SOURCES

18%

PUBLICATIONS

9%

STUDENT PAPERS

PRIMARY SOURCES

1	faculty.uobasrah.edu.iq Internet Source	4%
2	S L B Ginting, Y R Ginting, Sutono, A Rakhman. "Data Mining: The Classification Method to Predict the Types of Motorcycle Spare Parts to be Restocked", IOP Conference Series: Materials Science and Engineering, 2019 Publication	3%
3	shodhganga.inflibnet.ac.in Internet Source	1%
4	ejournal.nusamandiri.ac.id Internet Source	1%
5	gfzpublic.gfz-potsdam.de Internet Source	1%
6	Monalisa Mohanty, Santanu Sahoo, Pradyut Biswal, Sukanta Sabut. "Efficient classification of ventricular arrhythmias using feature selection and C4.5 classifier", Biomedical Signal Processing and Control, 2018 Publication	1%

7	eprints.utm.my Internet Source	1 %
8	Submitted to Rajiv Gandhi Proudhyogiki Vishwavidyalaya Student Paper	1 %
9	Ruidan Liu, Yu Dong. "Fault Diagnosis of Jointless Track Circuit Based on ReliefF-C4.5 Decision Tree", Journal of Physics: Conference Series, 2022 Publication	1 %
10	Rabah Benkercha, Samir Moulahoum. "Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system", Solar Energy, 2018 Publication	1 %
11	2020.mmbdconf.org Internet Source	1 %
12	gudangilmu79.blogspot.com Internet Source	1 %
13	F Syukmana, E Wahyudi, W Gata, H Wahono, N I Febianto, A Y Kuntoro, R O Nitra, L Effendi, D D Saputra, O R Sulaeman. "Predicting Relegation Clubs in Italian Serie A with Method based C4.5 Decision Tree Algorithm", Journal of Physics: Conference Series, 2020 Publication	1 %

14 infor.seaninstitute.org 1 %
Internet Source

15 Gary Neil Marks. "Are Father's or Mother's Socioeconomic Characteristics More Important Influences on Student Performance? Recent International Evidence", Social Indicators Research, 2007 1 %
Publication

16 A C Pradana, Adiwijaya, A Aditsania. "Implementing binary particle swarm optimization and C4.5 decision tree for cancer detection based on microarray data classification", Journal of Physics: Conference Series, 2019 1 %
Publication

Exclude quotes On Exclude matches < 1%
Exclude bibliography On

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13