



Decision Support System for Classification of Child Intelligence Using C4.5 Algorithm

Marlia Purnamasari

Information Technology Faculty – Informatics Eng. Dept
Universitas Serang Raya (UNSERA)
Banten Province - INDONESIA

Sulistiyo

Information Technology Faculty – Informatics Eng. Dept
Universitas Serang Raya (UNSERA)
Banten Province - INDONESIA

Abstract: Constraints for education to produce quality graduates are still many schools that have a traditional mindset in the running process of learning, where school only emphasizes the logic and language ability regardless of the types of intelligence possessed by the child. This study aims to implement the algorithm C4.5 to classify the type of intelligence of children of primary school age, and mapping types of intelligence they have. The research data is data type of child intelligence gathered from several elementary schools in two different districts, namely Serang and Cilegon to see and compare the types of intelligence of children in each school. The data obtained were divided into training data and test data, which will use the evaluation algorithm C4.5 (J48) using Weka application as an analytical tool. The results of the analysis using the Weka decision tree form that will serve as a rule for the child's intelligence determines the type of new data that will be tested. Tests on the new data showed that, the classification of the type of intelligence of children have a level of accuracy of 88%. The results are expected to be used as a decision support for teachers or parents to provide appropriate learning method for primary school children in accordance with the kind of intelligence they have.

Keywords: decision support system, C4.5 algorithm, child intelligence, weka, classification

I. INTRODUCTION

Many parents are complaining because the results of intelligence tests (IQ) of their children showed unsatisfactory results and beyond their expectations, because many parents still regard the child's intelligence lies only in the ability. Such a misconception that often leads to conflict between parents and children. The tendency of parents to impose its will on the other hand may result in the child will feel depressed, lose enthusiasm for learning so that children will tend to be lazy school.

Parents or teachers often have difficulty in recognizing the type of intelligence. This is because not all tests related to intelligence of children will show the same results. When a child following types of intelligence test to test using a test psychotest orally, in writing or through an online though even then it would get different results from each test are followed so that there is no one thing that can be used as a reference for determining the type of intelligence.

Based on a survey of research that has been conducted in several schools in Serang and Cilegon, found approximately 40% of gifted children, but are not able to perform on par with the actual capacity owned. As a result, even highly capable, children are lazy to excel. Surely this could affect the development and handling of the child, in the future. Improper handling of the different types of intelligence that dimiliki primary school age children, can result in 40% of gifted children, it becomes 60% the opposite of the other children who are not talented.

A decision support, of course, is needed to encapsulate all kinds of children's intelligence test that can be used as a reference and provide convenience for parents or teachers to determine their type of intelligence. Supporting this decision, of course, must be able to read the type of intelligence of a child (in this case children of primary school age) to be able to determine what action should be done against the child in accordance with each type of intelligence. One of the

methods that can be used to determine the type of intelligence is the C4.5 algorithm.

C4.5 algorithm is basically intended for classification and prediction [1], [2] and widely used in data mining study. Use of C4.5 staple due to several reasons, namely, C4.5 have high levels of robustness and accuracy that is better than the other algorithms [3], [4], but it also, C4.5 is considered more efficient than the other algorithms [5]. Many applications are developed using the C4.5 algorithm approach, for example, the concept of mapping the distribution of development region, which is able to classify the area of development in accordance with certain variables [6], the classification of IP traffic of computer networks [5] and detection of disorders of the data traffic on the network [7]. In the medical field, C4.5 algorithm is used for classification of patients Sufferers of breast cancer [8], testing of clinical guidance for clinical decision support systems associated with cases of chronic disease [9]. In addition, this algorithm can also be used to identify an algorithm based on the scheme of its source code [10], the classification for prediction needs the talent and abilities to choose the right person occupying a particular position [11], [12] analyzes the possibility of withdrawal of candidates a college student [13] non-active students at a prediction [14] as well as for determining the level of tsunami hazard [15].

This study aims to implement the algorithm on the classification of types of intelligence C4.5 primary school age children by forming a decision tree that can be used as a rule base to determine the type of intelligence of children of primary school age. The variables used in this study include gender, the value of learning, personality and behavior of students.

II. CHILD INTELLIGENCE

Intelligence is a property of the mind that includes many related mental abilities, such as the capacity to think, plan, solve problems, think abstractly, to understand the ideas and

language, and learn. Types of intelligence, based on Wechler standards are as follows:

- 1) *Very superior* : ≥ 128
- 2) *Superior* : 120-127
- 3) *Bright normal* : 111-119
- 4) *Average* : 91-110
- 5) *Dull normal* : 80-90
- 6) *Borderline* : 66-79
- 7) *Mentally defective (terbelakang mental)* : ≤ 65 .

III. C4.5 ALGORITHM

C4.5 algorithm uses the fact that each attribute of data that can be used to make the decision to split the data into smaller subsets. To make a decision, use the attribute with the highest information gain. Predictable attribute known as the dependent variable, because its value depends on the values of all other attributes that help in predicting the dependent variable, which is known as the independent variable in the data set.

Follow C4.5 decision tree algorithm, the following is a simple algorithm. To classify a new item, it first needs to make a decision tree based on the attribute values of the data sets available. The formula used to calculate the grades shown in the formula below :

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

where,

- S : the set of cases
 A : attribute
 n : number of partitions of attributes A
 $|S_i|$: the number of cases in the i-th partition
 $|S|$: number of cases in S

Before getting the value of Gain must first be sought Entropy values. Entropy is used to determine how informative an attribute input to produce output attributes. Basic equation for the determination of entropy is shown by the following equation:

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i \quad (2)$$

where,

- S : the set of Case
 n : number of partitions S
 p_i : proportion of S_i to S

IV. RESEARCH METHODOLOGY

The initial phase of the study by collecting data from multiple types of intelligence of children of primary school in Serang and Cilegon by means of a questionnaire on students' test pass / elementary school student. A total of 200 children and sample data will be used as a kind of intelligence research data. In this study, the classification of the type of children's intelligence be carried out using the method of C4.5 algorithm. Of the 200 children sample data, and then divided into two types of data, the first data, the training data for the algorithm C4.5 (J48) as many as 175 students of data, while the second data, test data that is taking

the data sample of 25 student data. The variables used were gender (JK), the value of learning (NB), personality (KP) and behavior (PR). Furthermore, the method of calculation algorithm C4.5 (J48) that will distinguish each type of intelligence in the child student / student elementary school. The determination of the rule base using Weka application, after the determination of the rule base data sets, then tested again against the test data to determine the type of intelligence possessed primary school age children.

V. RESULT AND DISCUSSION

This chapter discusses the results achieved in the study. Description and discussion of the results of systematic composed of, among others, the source of the data used, the analysis of the data using weka application and testing the results of data analysis so as to determine the type of children's intelligence Based on the existing variables.

A. Source of Data

The data in this study were drawn from the results of the questionnaire were collected from approximately 38 primary schools located in Serang and Cilegon. Here are some sample data collected (see table 1).

B. Data Analysis Using Weka

Before analyzed using Weka application, the data is processed in accordance with the first attribute / variable specified then stored in a file format *.csv. The next step is, dividing the data into two types of data, the first data set consists of 175 as training data, while the remaining 25 sets as test data. Training data and then inserted into the Weka application to establish rules klasifikasi the type of intelligence. From the results of the initial analysis, the distribution of types of intelligence obtained by illustrations that are distributed to each of the variables forming intelligence (see Figure 1).

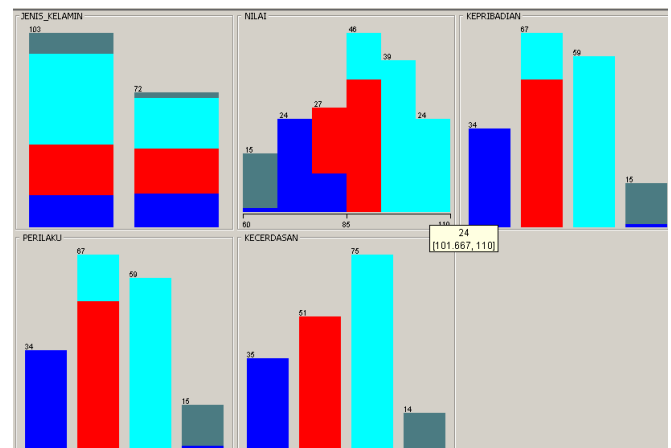


Figure 1. Distribution of types of intelligence according to the constituent variables

Figure 1 shows the distribution of the four types of children's intelligence to variable Gender, Values, Personality and Behavior. Furthermore, an analysis using the C4.5 approach in order to establish classification rules to be used as a determination of the child's intelligence. Keep in mind, the algorithm C4.5 in Weka applications, known as J48. Based on the analysis using the J48 in Weka application, obtained seven classification rules that can be used. The seventh rule is:

- 1). If personality = enough, then intelligence = borderline
- 2). If personality = good, and the value of ≤ 90 , then intelligence, = dull normal
- 3). If personality = good, and values > 90 , then intelligence = average
- 4). If personality = satisfactory, then intelligence = average
- 5). If personality = less, and value ≤ 90.5 , and behavior = excellent then intelligence = average
- 6). If personality = less, and value ≤ 90.5 , and behavior = less, then intelligence = borderline
- 7). If behavior = less and values > 90.5 , then intelligence = average

Table I. Preliminary data from the study (in part)

NO	Complete Name	Class	School	POINT	Remark
1	Ibnu Ramdhan Tio	5	SDN 2 CILEGON	79	Low
2	Muhammad Fadly Anizar	3	SDN 2 CILEGON	72	Low
3	Fahrul Zikri	3	SDN 2 CILEGON	72	Low
4	Phoundra Ginantaka Apriliano	3	SDN CIBEER	90	Normal
5	Salsabillah Naqiyah	3	SDN 2 CILEGON	80	Low
6	Anisa Salma Safira	3	SDIT INSANTAMA	88	Low
7	Nazifah	1	SDIT INSANTAMA	101	Normal
8	Indah K Wardianti	4	SDIT INSANTAMA	88	Low
9	Shafa J	6	SDN 2 CILEGON	90	Normal
10	Fadya	6	SDN 2 CILEGON	82	Low

Table II. The results of the processing of the type of children's intelligence (partly)

No	Complete Name	Sex	Point	Personality	Behavior	Remak
1	Ibnu Ramdhan Tio	L	79	Cukup	Less	Borderline
2	Muhammad Fadly Anizar	L	72	Cukup	Less	Borderline
3	Fahrul Zikri	L	72	Cukup	Less	Borderline
4	Phoundra Ginantaka Apriliano	L	90	Baik	Good	Dull Normal
5	Salsabillah Naqiyah	P	80	Baik	Good	Dull Normal
6	Anisa Salma Safira	P	88	Baik	Good	Dull Normal
7	Nazifah	P	101	Memuaskan	Excellent	Average
8	Indah K Wardianti	P	88	Baik	Good	Dull Normal

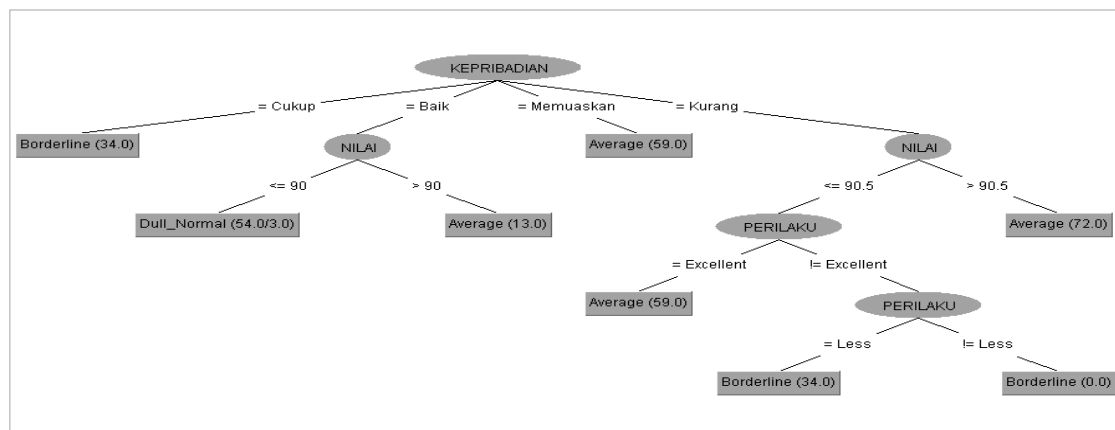


Figure 2. Visualization of the results of the classification rule C4.5 (J48)

Table III. Test data (in part)

No.	Complete Name	Sex	Point	Personality	Behavior
1	M. Dedi Wahyudin	L	90	Baik	Good
2	Inok Heryati	L	96	Memuaskan	Excellent
3	Romli Maulana	L	95	Memuaskan	Excellent
4	Sukron	L	86	Baik	Good
5	Naji Sobri	L	102	Memuaskan	Excellent
6	Supriyati	L	95	Memuaskan	Excellent
7	Bayu Saputra	L	98	Memuaskan	Excellent
8	Jumaenah	L	98	Memuaskan	Excellent
9	Sarwani	L	96	Memuaskan	Excellent
10	Zopian Zahuri	L	102	Memuaskan	Excellent

Table IV. Results of Classification Rules Based Testing (part)

No.	Complete Name	Sex	Point	Personality	Behavior	Results
1	M. Dedi Wahyudin	L	90	Baik	Good	Dull Normal
2	Inok Heryati	L	96	Memuaskan	Excellent	Average
3	Romli Maulana	L	95	Memuaskan	Excellent	Average
4	Sukron	L	86	Baik	Good	Dull Normal
5	Naji Sobri	L	102	Memuaskan	Excellent	Average
6	Supriyati	L	95	Memuaskan	Excellent	Average
7	Bayu Saputra	L	98	Memuaskan	Excellent	Average
8	Jumaenah	L	98	Memuaskan	Excellent	Average
9	Sarwani	L	96	Memuaskan	Excellent	Average
10	Zopian Zahuri	L	102	Memuaskan	Excellent	Average

Illustration / visualization of rules that are formed based on the training data are analyzed using Weka application (see Figure 2). The analysis also showed that the level of accuracy of the rules established by 97.71% with the data that formed the invalidity rate of classification of 2:28%. Kappa statistics values, the mean absolute error, and root mean square error are respectively 0.96, 0.021 and 0.106. While the percentage absolute error of 6:37% with a percentage of the mean square error of 25.77% (see figure 3)

Figure 3. Calculation of the accuracy of the classification results of the analysis of Weka

C. Testing Data To Recognize Type Child Intelligence

The results of the classification rules data 175 is formed, then tested against the data 25 elementary school students to see what kind of intelligence they have. 25 to the data of the students had any previous intelligence types predefined, so at this stage, is not only intended to determine the type of student intelligence alone, but also see the end result comparison to see the level of accuracy that is owned by the classification rules are formed. Here are the data of 25 students that will be tested (see table 3).

Based on the analysis of the data in Table 3, the results are as follows (see table 4). The results show, there are three students who showed different types of intelligence between the initial data with test data using the classification rules terbentuk. However, the accuracy of the test results showed 88% accurate and 12% less accurate, because there are differences between the results of the initial data with test data on students in the name of M. Dedi Wahyudin, Ade Setiadi and Rohmah.

Correctly Classified Instances	171	97.7143 %
Incorrectly Classified Instances	4	2.2857 %
Kappa statistic		0.9668
Mean absolute error		0.0219
Root mean squared error		0.1067
Relative absolute error		6.374 %
Root relative squared error		25.7777 %
Total Number of Instances		175

VI. CONCLUSION

Research conducted, has managed to establish a classification rule to determine the type of children's intelligence based on four variables, namely gender, test scores, behavior and personality. The results of the analysis can be used as a decision making tool for determining the type of intelligence of children, both for parents and educators, so as to provide proper treatment to the students according to the type of intelligence they have. Research can proceed with the analysis results to develop a stand alone application that is ready to be used for schools or other agencies in order to help identify the type of intelligence of children of primary school age. In addition, the variable forming classification rules, may be added, for example, by adding the age variable, the variable verbal and non-verbal, or other variables that might be used to help analyze the type of intelligence of children so that the results are more accurate.

VII. REFERENCES

- [1] Jos'e Martinez and Olac Fuentes, Using C4.5 As Variable Selection Criterion In Classification Tasks, Proceeding of the Ninth IASTED International Conference Artificial Intelligence And Soft Computing, September 12-14, 2005, Benidorm, Spain
- [2] Gaurav L. Agrawal and Hitesh Gupta, Optimization of C4.5 Decision Tree Algorithm for Data Mining Application, International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 3, March 2013
- [3] Payam Emami Khoonsari and AhmadReza Motie, A Comparison of Efficiency and Robustness of ID3 and C4.5 Algorithms Using Dynamic Test and Training Data Sets, International Journal of Machine Learning and Computing, Vol. 2, No. 5, October 2012
- [4] Harvinder Chauhan and Anu Chauhan, Implementation of decision tree algorithm c4.5, International Journal of Scientific and Research Publications, Volume 3, Issue 10, October 2013
- [5] Kuldeep Singh and S Agrawal. Performance Evaluation of Five Machine Learning Algorithms and Three Feature Selection Algorithms for IP Traffic Classification. IJCA Special Issue on Evolution in Networks and Computer Communications (1):25-32, 2011
- [6] Tb. Ai Munandar, Harsiti, Roy,. 2012. "Mapping concept of equitable regional development using C 4.5 Clasification Method". International Journal of IJARCS, Volume : 3, No. 2.
- [7] Radhakrishna Naik, Vivek Kshirsagar and B S Sonawane. Article: New Strategy for detecting Intrusion by using c4.5 algorithm.IJCA Proceedings on International Conference in Computational Intelligence (ICCIA 2012) ICCIA(8):-, March 2012
- [8] K.Rajesh and Sheila Anand, Analysis of SEER Dataset for Breast Cancer Diagnosis using C4.5 Classification Algorithm, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 2, April 2012
- [9] Jean-Baptiste Lamy, Anis Ellini, Vahid Ebrahimiinia, Jean-Daniel Zucker, Hector Falcoff and Alain Venot, Use of the C4.5 machine learning algorithm to test a clinical guideline-based decision support system, Stud Health Technol Inform. 2008; 136: 223–228
- [10] Ahmad Taherkhani, Recognizing Sorting Algorithms with the C4.5 Decision Tree Classifier, 18th IEEE International Conference on Program Comprehension, 2010
- [11] Hamidah Jantan, Abdul Razak Hamdan and Zulaiha Ali Othman, Human Talent Prediction in HRM using C4.5 Classification Algorithm, International Journal on Computer Science and Engineering, Vol. 02, No. 08, 2010
- [12] Ni. G. A.P Harry sapta rini, 2012. "Penggunaan algoritma C4.5 dan Fuzzy logic untuk klasifikasi talenta Karyawan". Journal Matrix, Volume. 2, No. 2, Hal. 95-100.
- [13] Kusriani, Sri Hartati, Retantyo, Agus, 2009. "Perbandingan metode nearest neighbor dan algoritma C4.5 untuk menganalisis kemungkinan pengunduran diri calon mahasiswa di STMIK Amikom Yogyakarta." Jurnal DASI. Vol 10. No. 1 2009.
- [14] Hastuti,, Khafiizh, 2012. "Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa Non aktif". Jurnal sematik 2012
- [15] Zezen Zainal Abidin, 2011. "Implementasi alagoritma C 4.5 untuk menentukan tingkat bahaya Tsunami". Journal semnasFI 2011, Hal. 29-36.