Implementation ID3 Algorithm to Predict Children Achievement in Response (Case Study Children Playgroup School)

Nursikuwagus Agus, Melian Lusi and Permatasari Deasy
Department of Information System, Faculty of Technic and Computer Science,
Indonesia Computer University, Bandung, Indonesia

Abstract: In decision support system, many methods which conducted based on decision tree. The algorithm that has implemented, one of them is ID3 (Iterative Dichotomized 3) algorithm. To predict the future situation is complicate and need the time. Student response is one activity to measure the growth of knowledge and adaptation children in the environment, especially in education. In association with that statement, the purpose in this research provides explanation about the decision tree that has been predicting student achievement response with ID3 algorithm. Response is the subject in the playgroup school that has measured to see the speed of cognitive response in every child. The system has been classifying student that has a high response. On 45 student data which used three parameters, the system has been selecting where the children who have strong response. Parameters are used to classify the student in this research such as attendance, personal attitude and creativity. Meanwhile, the target parameter in this research is response parameter. In algorithm ID3, the total student has to 45 students precisely. Including the data, the data has been created by the parameter above have measure such as bad, medium and good. The result of the system that has been constructed is the decision tree rule based on IF-THEN inference and stated good or bad response. The accuracy of the system has created is 100%. On the conclusion, the using of ID3 algorithm in this research, it can be noticed that personal attitude have significant influence to a children response in cognitive perspective.

Key words: ID3 algorithm, predict, achievement, response, classifying

INTRODUCTION

Data mining is the extraction of a knowledge which is of a very large data to determine the pattern of such data, (Ogunde and Ajibade, 2014). Each data held by any institution, aims to find patterns, predictions or classification of the existing data (Ogunde and Ajibade, 2014). Writing is activity whose demand a lot of people. In addition can be done anywhere and anytime, income is not bad. Now, writing became a profession that is in demand by various circles. As for the children themselves, the ability to write is usually started without them knowing it. But do not rule out the possibility, they started for some reason. For example: used to seeing parents who work as writers, see their peers have published a book or maybe because he was not satisfied with any of the stories they read. Therefore, to accommodate the ideas and creativity of the students in the world of literacy and writing then SDIT Lukmanul Judge Bandung opened extracurricular "Young Writers Club" (KPC). Although, at first, KPC is only for students in grade three to grade five. But considering the response of the students who enough enthusiastic, so the school opened the opportunity for the first and second grade students to participate. Class for 6th grade students are only allowed to follow the activities study addition during the first semester only. In its regular activities, supervisor KPC offers a method of "free, intelligent and creative". Free to perform other activities, the origin is not disturbed. Able to understand the material presented, intelligent. In addition, he was also invited to keep creativity widest. Schneider (1980) said that attitude is description of a person's thoughts or mental responses of circumstances. The behavior itself is knowledge impact and understanding and trust of an object encountered. Attitude also can result from a person's pleasure, the value received from an object and also the philosophy of learning (Schneider, 1980; Liu and Dong, 2016; Bandmir and Mehrpouyan, 2016).

The school atmosphere is quiet and comfortable. So, it becomes a means of refreshing among the tight schedule and school activities. Nevertheless, it is possible for the students to remain creative and accomplished in accordance with the interests and talents of each. Even

Corresponding Author: Nursikuwagus Agus, Department of Information System, Faculty of Technic and Computer Science,
Indonesia Computer University, Bandung, Indonesia

supervising KPC has implemented the method of "free, intelligent and creative" but in fact often a problem in practice. As for problems that often arise, among others: how teacher KPC to deliver the material, how is the response of the material is given, How to implement Iterative methods Dichotomized 3 (ID3) in the process of determining the KPC student's response to teacher.

To make easier in research and data collection, the definition of the issues to be discussed only involves determining the student's response, based on the criteria established in advance such as: attendance, skills and creativity, student data is taken only the data of students who take extracurricular activities KPC. The purpose of this research was to determine the extent of the student's response KPC against the material provided mentors over the years. To find out what criteria could improve the student's responses. To apply ID3 method in the process of determining is the best value. The benefit of this research is a measurement of success in conveying the material KPC. As one means of supporting the school in improving student's response to a variety of subjects who given.

Literature review

Data mining: Data mining is a series of processes to discover knowledge or patterns in a database (Tan, 2006). Data mining is also used to predict the existing data in the database. As for the stages that exist in data mining, among others: data cleaning, is the process of deleting data that is outlier and inconsistent. Data integration is a merge of multiple data sources. Data selection, a collection of data that will be used from the data source data transformation, is the process by which data is transformed into a form suitable for the process of data mining (Ogunde and Ajibade, 2014; Tan, 2006).

Data mining is an important process and involves a method for generating a pattern of data. Pattern evaluation is a process for testing the correctness of the data patterns representing knowledge in the data itself (Ramanathan, 2013; Tan, 2006; Nuchso *et al.*, 2016; Owusu, 2016). Presentation knowledge is a process visualization which has presenting knowledge to be employed within display the results of knowledge mining to the user. A mining process which finding a model or a data class for the purpose of estimating the class of an object that the label is not known (Tan, 2006). It can also be said as learning (classification) which maps an element (item) data into one of the classes that have been defined (Bhardwaj and Vatta, 2013).

Iterative Dichotomized 3 (ID3) algorithm: ID3 is a simple decision tree learning algorithm developed by Ross

Quinlan basic idea of ID3 algorithm is to construct the decision tree by employing a top down, greedy search through the given sets to test each attribute at every tree node. In to select the attribute that is most useful for classifying a given sets, we introduce a metric- information gain (Ogunde and Ajibade, 2014; Tan, 2006; Wu, 2008). To find an optimal way to classify a learning what we need to do is to minimize the questions asked (i.e., minimizing the depth tree). Thus, we need some function which can measure which questions provide the most balanced splitting. The information gain metric is such a function (Ogunde and Ajibade, 2014). A decision tree is a tree in which each branch node represents a choice between a number of alternatives and each leaf node represents a decision (Ogunde and Ajibade, 2014). They are commonly used for gaining information for the purpose of decision making. Decision tree starts with a root node on which it is for users to take appropriate actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each represents a possible scenario of decision and its outcome (Ogunde and Ajibade, 2014). ID3 is a simple decision tree learning algorithm developed by Ross Quinlan. The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. In order to select the attribute that is most useful for classifying a given sets, we introduce a metric-information gain (Ogunde and Ajibade, 2014). We have been using Eq. 1-4 to calculate information gain and entropy to know about heterogeneous of data classification (Bhardwaj and Vatta, 2013; Bhatt *et al.*, 2015; Ramanathan, 2013):

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

$$\text{Gain} (S, A) = \text{Entropy} (S) - \sum_{v \in (A)} \frac{S_A}{S} \times \text{Entropy} (S_A)$$
$$(2)$$

$$\text{Accuracy} = \frac{f_a + f_b}{f_a + f_b + f_c + f_d} \quad (3)$$

$$\text{Error rate} = \frac{f_c + f_d}{f_a + f_b + f_c + f_d} \quad (4)$$

The F_a and f_b are the correct prediction class result from ID3. Meanwhile, f_c and f_d are the wrong prediction class (Tan, 2006).

Student achievement: Student achievement is a measurement personal achievement that including about knowledge, attitude and beliefs (Schneider, 1980; Mustafidah, 2012). Student achievement can be measured speed response of student. The increase of response is depending on understanding about learning and teacher teaching (Schneider, 1980; Rahmani, 2014).

MATERIALS AND METHODS

Research model: At beginning research about student achievement, in Schneider (1980) told about significance the cognitive term influence to the response of student. An educational institution, it is importance to approximately prior knowledge of students to predict their performance in future academics (Adhatrao *et al.*, 2013). This helps them to identify promising students and also provides them an opportunity to pay attention to and improve those who would probably get lower grades. (Adhatrao *et al.*, 2013). In association with related past work (Adhatrao *et al.*, 2013; Schneider, 1980), in this research, we have proposed about research model. The research model has been built based on tree model. We can be seen at Fig. 1:

- IF paramter 1 = High and parameter 2 = Medium then decision parameter = Excellent
- IF parameter 1= Low and parameter 2 = Enough and parameter 3 = Low then decision parameter is worse
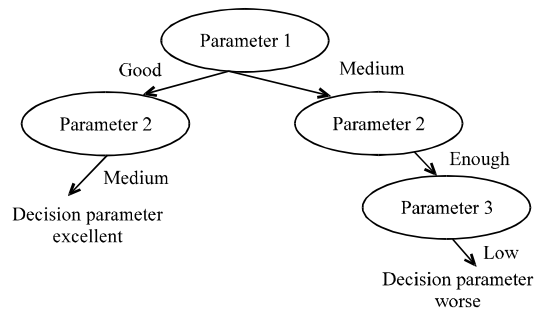

Fig. 1: Research model for Iterative Dichotomized 3 (ID3) (Tan, 2006)

RESULTS AND DISCUSSION

Data analysis: The data set in this research, we have been using data achievement from children school. At the 45 children performance, the attribute is spread of presence, student attitude and creativity. The target attribute to ID33 is response. At Table 1 can be shown about student performance.

In the ID3 algorithm process, we have been retrieving that the significant of attribute in student performance is attitude. Attitude attribute is one of parameter that research by Ramanathan (2013). The calculation, we can be seen that the response for each student is classified into 2 classes. The response student that have an excellent approximately 33 and worse is 12 student. The decision tree that we have, it can be written at Table 2.

In the ID3, we have calculated the accuracy and error rate. The accuracy can be counted by Eq. 3 and error rate is used by Eq. 4. The result of performance, we can be seen as a confusion matrix which has been created from ID algorithm. We need confusion matrix to calculate the accuracy of model and error rate. At Table 3, we are shown about result of confusion matrix. The accuracy and error rate of model is:

$$Accuracy = \frac{33+12}{33+12+0+0} \times 100\% = 100\%$$

$$Error\ rate = \frac{0+0}{33+12+0+0} \times 100\% = 0\%$$

The effectiveness of ID3 in this research can be proved. As a final result, accuracy have a 100 and 0% error rate. We can be said, the data which have involved in the processing is little significant value because the other attribute like presence and creativity are not significant value to the response attribute. The variety of data still more leak. For the comparison, we used

Table 1: An example of data set about student performance and target attribute

Names	Class	Attendance	Attitude	Creativity	Response
Asiya Afifah Zuhri	1B	Good	Medium	Low	Excellent
Syauqina Qonita	1B	Good	Medium	Low	Excellent
Alifia Fidela Firtinita	2C	Good	Medium	Low	Excellent
Ashfa Nadhira Rabbaniyyah	2C	Good	Medium	Low	Excellent
Alya Khairunnisa	3A	Low	Enough	Low	Worse
Hani Qurrata A'Yuni	3A	Good	Medium	Medium	Excellent
Nida Fakhirah	3A	Good	Medium	Medium	Excellent
Nurul Maghfirah	3A	Good	Medium	Medium	Excellent

Table 2: Confusion matrix (Tan, 2006)

Attributes	Excellent	Worse
Excellent	33	0
Worse	0	12

Table 3: The comparison between ID3 and other methods

Evaluations on training set summary	ID3	J48	MLP	Random tree
Correctly classified instances	45	45	45	45
Incorrectly classified instances	0	0	0	0
Kappa statistic	1	1	1	1
Mean absolute error	0	0	0.0109	0
Root mean squared error	0	0	0.0157	0
Relative absolute error (%)	0	0	2.7445	0
Root relative squared error (%)	0	0	3.5385	0
Total number of Instances (%)	45	45	45	45
Accuracy (%)	100	100	100	100
Error rate (%)	0	0	0	0

other method to see about accuracy and error date. At Table 3, we can show the comparison each to other in method.

CONCLUSSION

The research has been accomplished, we can be concluded that the attribute is have significant to the response attribute is attitude. There are accuracy reaches 100% and error rate is 0% on 45 data. Indeed, we can be said, in the preschool just attitude which has been influencing to the response. Because the response is variable measurement for each research that include about cognitive aspect.

REFERENCES

Adhatrao, K., A. Gaykar, A. Dhawan, R. Jha and V. Honrao, 2013. Predicting students performance using ID3 and C4. 5 classification algorithms. Intl. J. Data Mining Knowl. Manage. Process, 3: 39-52.

Bandmir, S.M.A.A. and A. Mehrpouyan, 2016. Designing and implementing knowledge industry: An integrative framework for universities. J. Adv. Hum. Soc. Sci., 1: 28-40.

Bhardwaj, R. and S. Vatta, 2013. Implementation of ID3 algorithm. Intl. J. Adv. Res. Comput. Sci. Software Eng., 3: 845-851.

Bhatt, H., S. Mehta and L.R. D'mello, 2015. Use of ID3 decision tree algorithm for placement prediction. Int. J. Comput. Sci. Inf. Technol., 6: 4785-4789.

Liu, H. and X. Dong, 2016. Influence of social relations on knowledge conflicts-an empirical study from Chinese universities. J. Administrative Bus. Stud., 2: 8-18.

Mustafidah, H.A., 2012. Student learning achievement prediction based on motivation, interest an dicipline using fuzzy Inference system. Juita, 2: 1-7.

Nuchso, N., S. Tuntivivat and P. Klayklueng, 2016. The effect of learning organization and servant leadership on child-centered teaching behavior with the mediating role of knowledge sharing in education of chanthaburi diocese schools. Int. J. Humanities Arts Soc. Sci., 2: 181-188.

Ogunde, A.O. and D.A. Ajibade, 2014. A data mining system for predicting university students graduation grades using ID3 decision tree algorithm. J. Comput. Sci. Inf. Technol., 2: 21-46.

Owusu, N.O., 2016. Collaborative knowledge management: Examining the challenges in the rural banking industry in the eastern region of Ghana. Intl. J. Humanities Arts Soc. Sci., 2: 111-120.

Rahmani, B.A., 2014. Early model of student's graduation prediction based on neural network. Telkomnika, 12: 465-474.

Ramanathan, L.E.A., 2013. Predicting students performance using modified id3 algorithm. Intl. J. Eng. Technol., 5: 2491-2497.

Schneider, L.A.F., 1980. Child care teachers attitudes, beliefs and knowledge regarding science and the impact on early childhood learning opportunities. Master Thesis, Oklahoma State University-Stillwater, Stillwater, Oklahoma, USA.

Tan, P.S., 2006. Introduction to Data Mining. Pearson Education, New York, USA., ISBN:978-81-317-1472-0, Pages: 771.

Wu, X., V. Kumar, J.R. Quinlan, J. Ghosh and Q. Yang *et al.*, 2008. Top 10 algorithms in data mining. Knowledge Inform. Syst., 14: 1-37.