Application of Classification Method C4.5 on Selection of Exemplary Teachers

Hasudungan Siahaan¹, Herman Mawengkang¹, Syahril Efendi¹, Anjar Wanto², Agus Perdana Windarto²
¹Universitas Sumatera Utara, Medan - Indonesia
²STIKOM Tunas Bangsa, Sudirman Street Pematangsiantar, Medan - Indonesia

Email: hasudungansiahan@gmail.com, hmawengkang@yahoo.com, syahril1@usu.ac.id, anjarwanto@amiktunasbangsa.ac.id, agus.perdana@amiktunasbangsa.ac.id

Abstract. The study aims to analyze the selection of exemplary teachers using C4.5 algorithm, which is one of the existing decision tree methods in data mining theory. Teacher data was obtained from the school in SMA Negeri 2 Pematangsiantar (2010-2016). The data used contains information about teacher history and teacher assessment data. This research uses interview technique and questionnaire in obtaining data. There are 11 attributes used in the assessment process: NUPTK, name, age, education, status, Appointment Letter, competence, award, work, teaching load, personality, Position, and label. The system has been tested using the Rapidminer application with 16 data samples. Where the rules are obtained as many as 24 rules. The level of accuracy of the system states that four of the ten attributes have a very significant correlation to the model of relationship rules in determining the proposed model teacher, such as (Position, competence, education, and personality). The four attributes (Position, competence, education, and personality) contribute 82.8% to the model of the rule of connectedness rules in determining the best teacher.

1. Introduction
The development of information technology and computers have experienced very rapid growth, including the development of databases. As one of the public schools in Pematangsiantar city, SMA Negeri 2 Pematangsiantar has its own qualification in choosing teachers who will become teachers in their school. With qualified teachers SMA Negeri 2 Pematangsiantar expected to be able to create children who excel in both academic and non-academic. One of the tasks of a teacher is to guide students in learning activities. Teachers who teach in SMA Negeri 2 Pematangsiantar also selected to be able to teach at the school. Many qualifications must be owned by prospective teachers who want to teach in SMA Negeri 2 Pematangsiantar, so that teachers are able to guide well in accordance with their respective abilities. Due to the high dedication of a teacher in developing a quality human resource level, the government decided to award an outstanding teacher in accordance with Law No.14 of 2005 on Teachers and Lecturers. Article 36 Paragraph (1) which states that "outstanding teachers, exceptional dedication, and / or duty in a special region are entitled to an award". Achieving teachers are assessed by looking at some of their qualifications and dedication as teachers who guide students into qualified human resources. The achievement of a teacher makes the teacher an example to the school and its students. According to the Kompasiana newspaper quoted on 15 September 2016, the selection of exemplary teachers aims to 1) appoint teachers as honorable, noble, dignified, and...
protected professions, 2) improve teacher's motivation and professionalism in performing their professional duties, 3) build teachers' and improving the quality of education and learning more equitably, 4) assessing achievements achieved by teachers and 5) improving teachers' motivation and professionalism in the execution of tasks. In reality, in making decisions model teacher in SMA Negeri 2 Pematangsiantar still not done maximally and is subjective. Efficiently the teacher's assessment files are not well organized and have low security so that they can be influenced by internal and external factors in their assessment. So it takes a decision-making system that is able to assist in determining teacher model in SMK Negeri 2 Pematangsiantar objectively.

Based on the above problems researchers using data mining classification with C4.5 method in the selection of the best teachers in SMA Negeri 2 Pematangsiantar. Some previous research that can solve the problem using C4.5 method.

2. Literature Review
2.1. Artificial intelligence
Artificial Neural Network (ANN) is one of the artificial representations of the human brain that always tries to simulate the learning process in the human brain [1][2][3]. To achieve this, however, two major issues need to be addressed. The first issue is knowledge representation, and the second is knowledge manipulation [4][5]. Salah satu contoh kecerdasan buatan adalah data mining.

2.2. Classification
Classification is the operation of separating various entities into several classes. These classes can be defined by business rules, class boundaries, or some mathematical function [6]. A classifier should first get that knowledge by learning the representation of classes using a given set of pre-classified samples [7]. A classification approach that aims to predict unseen objects to one of their predefined classes using association rules. Two well-known data mining tasks, association rules mining and classification are integrated where the association rule discovers relationships between attribute values in training datasets whereas the goal of classification is to guess the class values of the test datasets [8].

2.3. Data Mining C4.5
C4.5 is an efficient and effective approach for real-time evaluation and classification classifier design. [9][10]. The advantages of the C4.5 is that it models can be easily interpreted and implemented with both continuous values and discrete values. It also work efficiently work with signals electrocardiogram. The C4.5 algorithm is an upgraded ID3 algorithm [11].

3. Research Methodology
Stages or descriptions done in conducting research that aims to facilitate researchers in conducting research. The research framework as described in (Figure 1).
4. Results And Discussion

4.1 Data processing
The study used primary data from SMA Negeri 2 Pematangsiantar, with the total of 72 teacher data with NUPLEV attribute, name, age, education, status, Appointment Letter, competence, award, work, teaching load, personality, Position, and label. From the data taken 16 sample data used as an example of calculation of entropy and gain manuals, on the data sample there are 7 data model teacher (appropriate) and 9 data of ordinary teacher (not appropriate as) shown as Table 1.

| Attribute          | explanation                                           |
|--------------------|-------------------------------------------------------|
| Age                | It is the age attribute of the teacher who is a potential teacher |
| Education          | It is the last educational attribute of the teacher who is an aspiring teacher |
| Status             | It is a status attribute of the teacher who is a potential teacher |
| Appointment Letter | It is an attribute of the appointment letter as a civil servant teacher or as a teacher of honor from a teacher who is a potential teacher |
| Competence         | It is an attribute of Competence expertise possessed by the teacher who is the best teacher participant candidate |
| appreciation       | It is an attribute of the competence of expertise possessed by the teacher of the teacher who is a candidate for the role model teacher |
| Creation           | Is an attribute of the scientific work ever created or published from the teacher who is a candidate for the role model teacher |
| Teaching Expenses  | It is the attribute of the teacher's teaching load in one week per semester from the teacher who is the ideal teacher |
| Personality        | Is an attribute never problematic or not with the law indicated by a certificate from the authorized institution of the teacher who is a candidate for the role model teacher |
| Position           | It is an attribute of a Position that has ever been assumed by a teacher who is a prospective teacher |
| Label              | The attributes used to describe the outcome of an exemplary and inappropriate assessment of a model teacher |

Next is the selection of data attributes used. Changed attributes are also used because of the use of attributes that are related (shown as Table 2).
After the classification of rules using the algorithm C4.5, then formed rules as many as 37 rules. There are 13 rules that are successfully classified with appropriate values and 24 rules with no Appropriate value. From the whole rule used is as follows (Table 5):

### Table 5. Fuzzy Rule Using C4.5 Algorithm

| No | Rules                                      | Decision  |
|----|--------------------------------------------|-----------|
| 1  | IF Position = Teacher Adult Then Value_Description = Appropriate | Appropriate |
| 2  | IF Position = Teacher Madya Lev.1 Then Value_Description = No Appropriate | Disagree |
| 3  | IF Position = Teacher Adult Lev.1 Then Value_Description = No Appropriate | Disagree |
| 4  | IF Position = Teacher Builder And Competence = Adequate Then Value_Description = Appropriate | Appropriate |
| 5  | IF Position = Teacher Builder And Competence = Incompetent Then Value_Description = No Appropriate | Disagree |
| 6  | IF Position = Teacher Builder Lev.1 And Competence = Incompetent Then Value_Description = No Appropriate | Disagree |
| 7  | IF Position = Teacher Builder Lev.1 And Competence = Adequate Then Value_Description = Appropriate | Appropriate |

The attributes used to support research to get maximum results. Data processing using RapidMiner software tools.

### 4.2. Testing Data

This study uses a sample of 16 teachers SMA Negeri 2 Pematangsiantar to determine the accuracy of classification. There are 16 teacher data of SMA Negeri 2 Pematangsiantar used as sample data for entropy and gain calculation (shown as Table 3 and Table 4).

### Table 3. Testing Data

| NUPTK                  | Age  | Education | Status          | Appointment Letter | Competence | Appreciation |
|------------------------|------|-----------|-----------------|--------------------|------------|--------------|
| 2133739640200003       | Old  | S1/DIV    | Married         | Central government | Competent  | Teacher Certification |
| 3446763664200023       | Adult| S2        | Single          | District / City Gov| Adequate   | Not yet       |
| 7146759660110043       | Old  | S1/DIV    | Married         | District / City Gov| Incompetent| Not yet       |
| 093873839300102        | Old  | S1/DIV    | Married         | Central government | Competent  | Teacher Certification |
| ...                    | ...  | ...       | ...             | ...                | ...        | ...           |
| 4544740641300112       | Old  | S2        | Married         | Central government | Adequate   | Teacher Certification |
| 1952747648300002       | Old  | S2        | Married         | Central government | Competent  | Teacher Certification |
| 5257751653300003       | Old  | S1/DIV    | Married         | Central government | Competent  | Teacher Certification |
| 1836744646300072       | Old  | S2        | Married         | Central government | Adequate   | Not yet       |

### Table 4. Advanced Data Testing

| NUPTK                  | Creation | Teach/Week | Personality | Position       | Label            |
|------------------------|----------|------------|-------------|----------------|------------------|
| 2133739640200003       | Scientific Publications | Good       | Good        | Teacher Builder | Appropriate      |
| 3446763664200023       | Book     | Good       | Good        | Teacher Adult  | Appropriate      |
| 7146759660110043       | Book     | Good       | Not Good    | Teacher Madya Lev.1 | No Appropriate |
| 093873839300102        | Scientific Publications | Enough     | Enough      | Teacher Builder Lev.1 | No Appropriate |
| ...                    | ...      | ...        | ...         | ...             | ...              |
| 4544740641300112       | Scientific Publications | Not Good   | Good        | Teacher Builder | Appropriate      |
| 1952747648300002       | There is no | Good       | Very Good   | Teacher Adult  | Appropriate      |
| 5257751653300003       | Scientific Publications | Not Good   | Enough      | Teacher Adult Lev.1 | No Appropriate |
| 1836744646300072       | Scientific Publications | Not Good   | Not Good    | Teacher Builder Lev.1 | No Appropriate |

### 4.3. If-Then Rule Formation Based on Decision Tree

After the classification of rules using the algorithm C4.5, then formed rules as many as 37 rules. There are 13 rules that are successfully classified with appropriate values and 24 rules with no Appropriate value. From the whole rule used is as follows (Table 5):
The 3rd International Conference on Computing and Applied Informatics 2018
IOP Conf. Series: Journal of Physics: Conf. Series 1235 (2019) 012005 DOI:10.1088/1742-6596/1235/1/012005

4.4. Validation and Accuracy Testing of Decision Tree Algorithm C4.5

Decision tree validation is used to see the accuracy of the algorithm model C 4.5 in determining the exemplary Teacher using software RapidMiner. In validating the rule decision tree, the author uses the split validation utility used to divide the two training areas and testing the data. In testing the data the

| No | Rules | Decision |
|----|-------|----------|
| 8  | IF position = Teacher Builder Lev 1 And Competence = Occupied Then Value_Description = No Appropriate | Disagree |
| 9  | IF Position = Teacher Builder Lev 1 And Personality= Very Good Then Value_Description = Appropriate | Appropriate |
| 10 | IF Position = Teacher Builder Lev 1 And Personality = Good Then Value_Description = No Appropriate | Disagree |
| 11 | IF Position = Teacher Builder Lev 1 And Personality = Enough Then Value_Description = No Appropriate | Disagree |
| 12 | IF Position = Teacher Builder Lev 1 And Appreciation = Certification Then Value_Description = No Appropriate | Disagree |
| 13 | IF Position = Teacher Builder And Competence = Occupied And Teach= Enough Then Value_Description = No Appropriate | Disagree |
| 14 | IF Position = Teacher Builder And Competence = Occupied And Teach= Incompetent Then Value_Description = No Appropriate | Disagree |
| 15 | IF Position = Teacher Builder Lev 1 And Appreciation= Not yet And Teach= Good THEN Value_Description = No Appropriate | Disagree |
| 16 | IF Position = Teacher Builder Lev 1 And Appreciation= Not yet And Teach= Enough THEN Value_Description = Appropriate | Appropriate |
| 17 | IF Position = Teacher Builder Lev 1 And Appreciation= Not yet And Teach= Incompetent Then Value_Description = No Appropriate | Disagree |
| 18 | IF Position = Teacher Builder And Competence = Occupied And Teach= Good And Creation= Book Then Value_Description = No Appropriate | Disagree |
| 19 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Creation= There is no THEN Value_Description = No Appropriate | Disagree |
| 20 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Not yet Then Value_Description = No Appropriate | Disagree |
| 21 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Enough Then Value_Description = No Appropriate | Disagree |
| 22 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Creation= Publication And Personality= Enough Then Value_Description = No Appropriate | Disagree |
| 23 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Creation= Publication And Personality= Very Good Then Value_Description = Appropriate | Appropriate |
| 24 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Creation= Publication And Personality= Good And Education= S2 Then Value_Description = Appropriate | Appropriate |
| 25 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Creation= Publication And Personality= Good And Education= S1/DIV And Appointment Letter= District / City Government THEN Value_Description = No Appropriate | Disagree |
| 26 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Creation= Publication And Personality= Good And Education= S1/DIV And Appointment Letter= Central government Then Value_Description = Appropriate | Appropriate |
| 27 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Very Good And Creation= Publication Then Value_Description = Appropriate | Appropriate |
| 28 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Very Good And Creation= There is no THEN Value_Description = No Appropriate | Appropriate |
| 29 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Very Good And Education= S1/DIV Then Value_Description = Appropriate | Appropriate |
| 30 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Very Good And Education= S2 Then Value_Description = No Appropriate | Disagree |
| 31 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Good And Education= S2 Then Value_Description = Appropriate | Appropriate |
| 32 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Good And Education= S1 And Creation= There is no THEN Value_Description = No Appropriate | Disagree |
| 33 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Good And Education= S1 And Creation= Book And Appointment Letter= District / City Government Then Value_Description = No Appropriate | Disagree |
| 34 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Good And Education= S1 And Creation= Book And Appointment Letter= Central government Then Value_Description = Appropriate | Appropriate |
| 35 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Good And Education= S1 And Creation= Publication And Appointment Letter= Central government Then Value_Description = No Appropriate | Disagree |
| 36 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Good And Education= S1 And Creation= Publication And Age= Adult Then Value_Description = No Appropriate | Disagree |
| 37 | IF Position = Teacher Builder And Competence= Occupied And Teach= Good And Appreciation= Certification And Personality= Good And Education= S1 And Creation= Publication And Status= Pemerintah Single Then Value_Description = No Appropriate | Disagree |
author uses utility apply model and % performance. For AUC and ROC pessimistic as Figure 2, Accuracy percentage as Figure 3, Precision value as Figure 4 and Recall percentage as Figure 5.

Based on the figure above which of 16 data used to determine the exemplary teacher by using the C4.5 algorithm with help of rapidminer tools, the accuracy value was 87.50%, 77.78% Precision, 100% recall and AUC (Area Under Curve) of 0.841. Based on these results it can be concluded that the determination of the exemplary teacher uses the C4.5 decision tree model into the excellent category and can be used to determine the next model teacher based on the rules generated.

5. Conclusion

Based on the results of the study can be concluded in the case of determining the best teacher using the C4.5 decision tree algorithm method as follows:

a. There is a rule model that can show the rules in determining exemplary Teacher in SMA Negeri 2 Pematangsiantar through NUPLEV attributes, name, age, Education, status, Appointment, Competence, Appreciation, Creation, Teach, Personality, Position, and Lable.

b. In a case study at SMA Negeri 2 in determining the Exemplary Teacher that most Teachers are not included in the category of 24 rules with No Appropriate value

c. This study has shown that four of the ten variables have a very significant correlation with the rules of connectedness model in determining the role model as proposed (Position, Competence, Education and Personality). All four predictors of the above variables contributed 82.8%.

References

[1] Wanto A, Zarlis M, Sawaluddin, and Hartama D 2017 Analysis of Artificial Neural Network Backpropagation Using Conjugate Gradient Fletcher Reeves in the Predicting Process Journal of Physics: Conference Series 930 1 pp 1–7.

[2] Wanto, A. Windarto P, Hartama D, and Parlina I 2017 Use of Binary Sigmoid Function And Linear Identity In Artificial Neural Networks For Forecasting Population Density International Journal Of Information System & Technology 11 1 pp. 43–54.

[3] Huang D and Wu Z 2017 Forecasting Outpatient Visits Using Empirical Mode Decomposition Coupled With Backpropagation Artificial Neural Networks Optimized By Particle Swarm Optimization PLoS ON 12 2 pp. 1–18.
[4] Sumijan A, Windarto P, Muhammad A, and Budiharjo 2016 Implementation of Neural Networks in Predicting the Understanding Level of Students Subject International Journal of Software Engineering and Its Applications 10 10 pp. 189–204.

[5] Siregar S P and Wanto A 2017 Analysis Accuracy of Artificial Neural Network Using Backpropagation Algorithm In Predicting Process (Forecasting) International Journal Of Information System & Technology 11 pp. 34–42.

[6] Nisbet R, Miner G, and Yale K 2018 Handbook of Statistical Analysis and Data Mining Applications Second Edition, ed Nisbet R, Miner G (Elsevier).

[7] Aziz A S A, Hanafi S E O, and Hassanien A E 2017 Comparison of classification techniques applied for network intrusion detection and classification Journal of Applied Logic 24 pp. 109–118.

[8] Hadi W, Al-Radaideh Q A, and Alhawari S 2018 Integrating Associative Rule-based Classification with Naïve Bayes for Text Classification Applied Soft Computing 69 pp. 344–356.

[9] Kudla P and Pawlak T P 2018 One-class synthesis of constraints for Mixed-Integer Linear Programming with C4.5 decision trees Applied Soft Computing 68 pp. 1–12.

[10] Saeh I S, Mustafa M W, Mohammed Y S and Almaktar M 2016 Static Security classification and Evaluation classifier design in electric power grid with presence of PV power plants using C-4.5 Renewable and Sustainable Energy Reviews 56 pp. 283–290.

[11] 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.