❑   1257

Eigenvalue of Analytic Hierarchy Process as The Determinant for Class Target on Classification Algorithm

Mustakim[1], Novia Kumala Sari[2], Jasril[3], Ismu Kusumanto[4], Nurul Gayatri Indah Reza[5]
[1,2,5]Puzzle Research Data Technology, Faculty of Science and Technology, UIN Sultan Syarif Kasim Riau,
Pekanbaru-Riau, Indonesia 28293
[1,2,5]Department of Information Systems, Faculty of Science and Technology, UIN Sultan Syarif Kasim Riau,
Pekanbaru-Riau, Indonesia 28293
[3]Department of Informatics Engineering, Faculty of Science and Technology, UIN Sultan Syarif Kasim Riau,
Pekanbaru-Riau, Indonesia 28293
[4]Department of Industrial Engineering, Faculty of Science and Technology, UIN Sultan Syarif Kasim Riau,
Pekanbaru-Riau, Indonesia 28293

Article Info

ABSTRACT

Data mining has two main concepts of data distribution, namely supervised learning and unsupervised learning. The most easily recognizable concepts from data distribution is related to the dataset, with and without target class. Analytic Hierarchy Process (AHP) technique that carries the concept of pairwise comparison able to answer the problem related to the dataset, which is to change unsupervised to be supervised by determining eigenvalue value of each attribute and sub attribute in AHP method. The case study conducted in this issue is related to determining the target classes used to predict the success of a student learning in UIN Suska Riau. The three main attributes are Procrastination, Total Credits (SKS) and Number of Repeated Courses, each having eigenvalues of 0.319; 0.189 and 0.171 which become the feedback in the determination of the Target Timely Graduation (TG) or Possibility of Timely Graduation (PTG). The biggest consistency ratio generated in the AHP case is 9.4% in the GPA attribute. This research recommends that further research should use datasets that have been arranged based on experimental combinations of the three main attributes above, then applied to the classification or prediction algorithm. So that it would obtain a decision of accuracy from data used against the real result on the field.

Corresponding Author:

Mustakim
Puzzle Research Data Technology, Faculty of Science and Technology,
UIN Sultan Syarif Kasim Riau, Pekanbaru-Riau, Indonesia 28293.
Email: mustakim@uin-suska.ac.id

1. INTRODUCTION

Data mining has two main concepts in the study of data distribution, namely supervised learning and unsupervised learning. In the science of data mining, accuracy becomes a major feedback in summing up the results obtained [1]. Apart from the accuracy issues, something that need to be analyzed in data mining is related to the distribution of training data and testing data for supervised learning. Faulty data distribution will be fatal to desired result [2]. The study conducted by Mustakim in 2017 states that the most optimal data distribution is by applying the clustering techniques [3]. Regardless of the accuracy and data distribution, another thing which is directly related to data mining is the process of determining a class. In machine learning, there are two approaches: supervised and unsupervised learning [4], supervised learning is an

approach where data trained already existed, and there are targeted variables so the goal of this approach is to group data into the existing data [5]. This condition requires that a dataset needs to have a class variable to predict the new data [6].

Several types of algorithms included in the frequently used supervised learning groups have been compared are K-Nearest Neighbor (K-NN), Probabilistic Neural Network (PNN) and K-Nearest Centroid Neighbor (KNCN) [7]; Naïve Bayes, Decision Tree and K-NN [8]; Artificial Neural Network (ANN) and Support Vector Regression (SVR) [9]. From these various studies obtained various results from each comparison between the algorithms. Furthermore, the studies which involved these algorithms use training data with the target class attributes, because of the nature which need to do the learning [10].

This research tries to simulate case study based on dataset from field observation, interview and data acquisition process from several experts. The main objective is to collect unsupervised learning data into supervised learning as a prediction model. Based on the results of data collection, variables and observations, there are 7 variables and 38 sub variables used in determining the prediction of Timely Graduation of Students in UIN Sultan Syarif Kasim Riau Indonesia. The problem is that there is no target class variable in the data set which is used to do the prediction. While one of the conditions of supervised learning is the existence of target class [11]. Nevertheless, in the research conducted, there are several assumptions which can be done to determine the target variable such as by using pairwise comparison from all variable used.

The linkage of comparisons between one variable with another is often used by the term of Decision Making modeling. The most famous algorithms are Analytic Hierarchy Process (AHP) and Analytic Network Process (ANP) [12, 13]. AHP and ANP have several advantages such as being able to solve complex problems in decision making [14], has good validation because it uses consistency value [15] and is able to represent human perception into matrix [16], [18]. Related with AHP and ANP there is a reference in determining the decision, that is based on the eigenvalue [17], [19]. Eigenvalue is represented from each of the variable's criteria and the alternative variable with reference to the value of consistency ratio [20].

In this case, the pairwise combination experiments of AHP Methods will be implemented for the determination of class variables in Data Mining algorithm especially classification and prediction. In some studies, AHP was only combined with the same type of Multi Attribute Decision Making (MADM) algorithm such as Technique For Others Reference by Similarity to Ideal Solution (TOPSIS) and Simple Additive Weighting (SAW) with the application of decision making only [21-23]. Therefore, the final result of this research is how to obtain the best variable from a series of variables used and get the target class variable used in the process of predicting the timely graduation of the Student.

2.   RESEARCH METHOD

This activity starts by searching for literature review and discuss about the timely graduation of Students in UIN Suska Riau with thesis advisor, lecturer in the field of student psychology, Deputy Dean of Faculty and Student. Some of the activities are collecting primary data, observation and interview and then then pairwise comparison from experts to make knowledge acquisition into knowledge base. Generally the research methodology can be seen in Figure 1.

2.1. Supervised and Unsupervised Learning

Supervised learning is an approach of machine learning techniques where data trained already existed, and there are determined variables, so the purpose of this approach is to group data into existing data [24-27]. While unsupervised learning does not have training data, so existing data can be grouped into several sections according to the needs [25], [28].

2.2. Classification

Classification is one technique in data mining used to classify data into predetermined classes [29]. In the classification there is a target variable called class. This model will test data sets containing variables or attributes information based on the input or predictor variables [30, 31]. In addition to classification techniques, the classification also serves to make predictions continuously, the modeling can be regarded as a predictor [32].
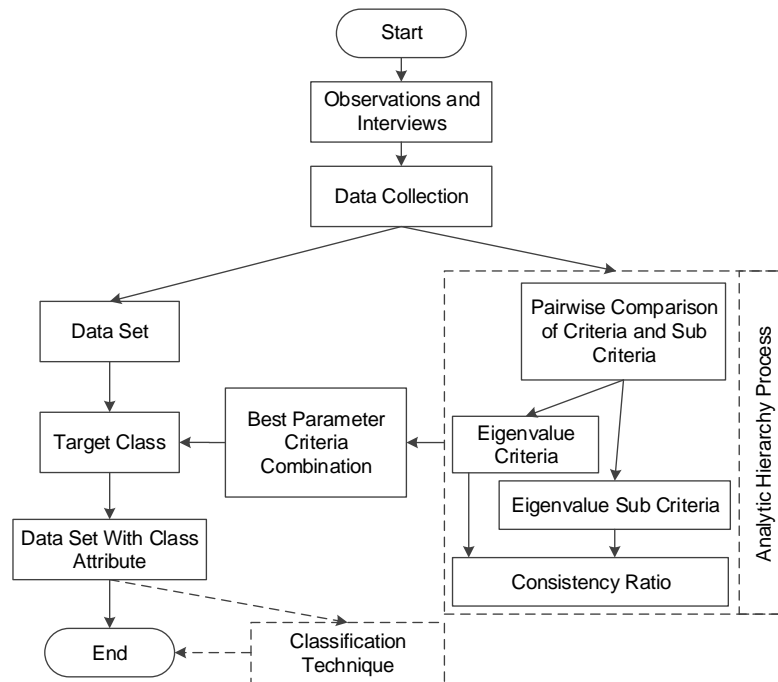
Figure 1. Research Methodology

2.3. Analytic Hirarchy Process (AHP)

AHP is a method that combines quantitative and qualitative, proposed by Thomas L. Saaty of American Operation Research in 1970 [33], [36]. The AHP method can solve complex problems in the case of decision making. The elements of a decision are divided into sections which consist of targets, attributes or criteria and solutions, often referred as problem solving based on certain sub-levels [14], [33, 34]. The scoring scale on AHP can be seen in Table 1 [20], [35]:

Table 1. The fundamental scale of absolute numbers Saaty [18]

Intensity	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Moderate importance	Experience and judgement slightly favour one activity over another
5	Strong importance	Experience and judgement strongly favour one activity over another
7	Very strong or demonstrated importance	An activity is favoured very strongly over another; its dominance demonstrated in practice
9	Extreme importance	The evidence favouring one activity over another is of the highest possible order of affirmation
2,4,6,8	When in doubt between two adjacent values	This value is given when there are two compromises between two options
Reciprocals of above	If activity i has one of the above non-zero numbers assigned to it when compared with activity j, then j has the reciprocal value when compared with i	

3.   RESULTS AND ANALYSIS

Direct data collection and field observation to produce some data such as obtaining the attributes that can delay the timely graduation of Students based on factors mentioned by Experts. Associated with the Students obtain an information related to the cause of the delay in the study process. And furthermore, obtain information associated to the attributes related to the topic of case study. This was done by conducting interviews with experts namely Vice Dean 1 Faculty of Science and Technology, Lecturer of Psychology Educational UIN Suska Riau, Head of Information System Department, some Lecturer of Information System, some Students who are working on the final project and some Alumni who have graduated from the Information Systems to find out why Students are late in completing their final project. Thus, an attribute and sub-attribute that are directly related to the case study can be seen in Figure 2.
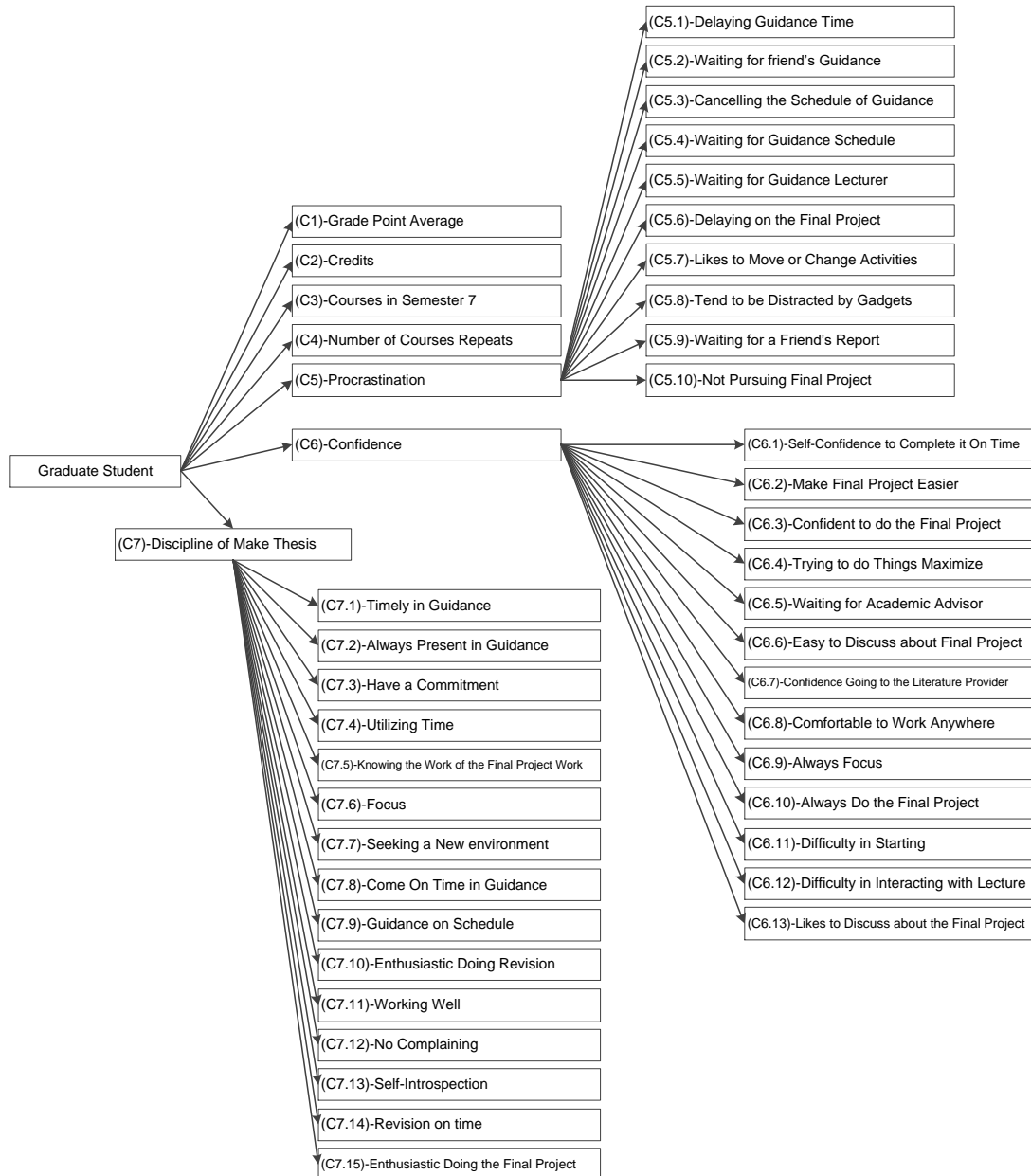
Figure 2. Attribute and Sub Attribute of Student Timely Graduation

3.1. Pairwise Comparison of AHP Attributes

Based on Table 1 the paired comparison scale proposed by Saaty is related to the comparison of each attribute/criteria which is the basic reference in the assessment based on human perception. In the assessment of this study simulated by the First Expert with the eigenvalue shown in Table 2:

Table 2. Eigenvalue of pairwise comparison for Main Attribute of Expert 1

Matrix							Summary of Matrix	Summary of Matrix
6.97	0.86	4.44	1.96	2.41	1.75	3.90	22.29	0.035
67.00	6.95	43.65	15.63	20.98	14.98	29.64	198.83	0.317
12.93	1.70	6.94	3.41	3.68	3.02	4.95	36.62	0.058
39.65	3.90	22.98	6.96	9.91	7.91	16.97	108.28	0.172
35.64	4.08	21.63	8.29	6.96	6.30	12.96	95.86	0.153
37.65	4.30	23.64	8.96	8.97	6.97	13.63	104.12	0.166
22.95	2.80	12.96	5.41	6.57	4.57	6.95	62.21	0.099
Total Summary of Matrix							628.21	1.000

For the overall pairwise comparison measurement conducted by 4 experts, the average eigenvalues of the main attributes are GPA (0.054), Credits Total (0.189), Taking Final Project Subject in 7th Semester (0.050), Number of Repeated Courses (0.171), Procrastination (0.319), Self Confidence (0.125) and Discipline (0.092). In detail can be shown in Table 3:

Table 3. The main attribute eigenvalue for the whole measurement

Attribute	Assessor 1	Assessor 2	Assessor 3	Assessor 4
C1	0.035	0.082	0.066	0.032
C2	0.317	0.115	0.031	0.293
C3	0.058	0.046	0.042	0.055
C4	0.172	0.167	0.223	0.121
C5	0.153	0.448	0.376	0.298
C6	0.166	0.056	0.181	0.098
C7	0.099	0.086	0.081	0.103
Total	1,000	1,000	1,000	1,000

Based on Table 3 can be seen that Procrastination attribute has a very high potential value as the main factor of student success, it is found that in the determination of the timely study could potentially be influenced by attribute Procrastination (C5), then followed by attribute Credits (C2) and Number of Repeated Courses (C4). The Procrastination attribute consists of 10 sub attributes with the following average of eigenvalue:
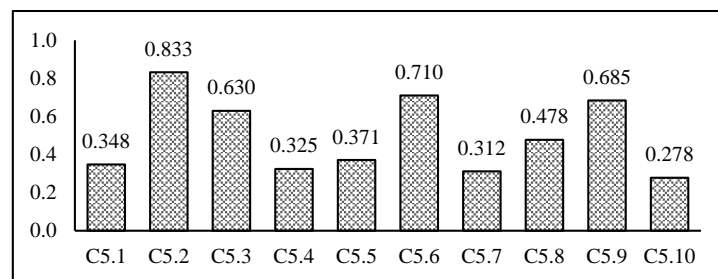

Figure 3. Average eigenvalues on Procrastination attributes

If seen from Figure 3, it can be concluded that sub attributes C5.2, C5.6 and C5.9 greatly affect the Procrastination attribute where sub attributes are Waiting for Friends, Delaying on Final Project and Waiting for Friend's Report. This greatly affects the tardiness of students in completing their studies.

From the assessment of all attributes and sub attributes of AHP pairwise comparison obtain the smallest Consistency Ratio value of 0.023 or 2.3% for the 2nd attributes of Self Confidence and the highest Consistency Ratio is in the attribute of GPA with value of 0.094 or 9.4% in the 4th attribute. However, all assessment is still below 10% which means the assessment is stated as Consistent.

3.2. The combination of Class Attribute Determination

There are 2 targets set for the Student's graduation estimation such as the Possibility of Timely Graduation (PTG) or Timely Graduation (TG). Each data will be grouped to one of these target categories. For target determination refers to the result of AHP method calculation with the 3 highest eigenvalue as the most priority attribute namely Procrastination, Total Credits and Number of Repeated Courses.

From 3 attributes which become the priority and the result of discussion with Expert 2, the Lecturer of Psychology UIN Suska Riau, for determination of target can be determined with condition from 3 influential criteria. For the first is, if the procrastination score in the "low" category then the student can be categorized as on time. Second, if the student has taken the entire course or total credit is more than 136 in 7th semester then the student can be categorized as on time. And last, if the student repeats less than 3 subjects then the student can be categorized as on time. If the 3 conditions are met then the Student can reach the Timely Graduation (TG) and if one or all conditions is not fulfilled, then the student only reaches the Target of Possible Timely Graduation (PTG). Generally the resulting combination can be illustrated in Table 4:

Table 4. Reference of Class Determination Combination

Attribute	C4	C2	C5	Class
Combination 1	Yes	Yes	Yes	TG
Combination 2	Yes	Yes	No	PTG
Combination 3	Yes	No	Yes	PTG
Combination 4	No	Yes	Yes	PTG

3.3. Dataset for Classification

The next process is to do a combination of class determination based on Table 5. A total of 112 data was obtained from the distribution of questionnaires to students related to the problem of delaying in study period, complete data with predetermined Target Class was obtained, and shown in the following Table 5:

Table 5. Dataset with predetermined class target

No	Code	C1	C2	C3	C4	C5	C6	C7	Target Class
1	SI026	2.99	149	9	0	2	3	3	PTG
2	SI028	2.93	140	6	1	2	3	3	PTG
3	SI029	3.53	146	0	1	1	3	3	TG
4	SI030	3.05	150	3	1	2	3	3	PTG
5	SI037	3.18	149	5	0	2	3	3	PTG
6	SI042	3.33	142	0	1	1	3	3	TG
7	SI043	3.30	146	0	1	1	3	3	TG
8	SI051	3.00	146	0	1	2	3	3	TG
9	SI055	3.02	146	2	1	1	3	3	TG
10	SI057	3.40	146	0	1	1	3	3	TG
…	…	…	…	…	…	…	…	…	…
112	SI034	3.32	149	0	1	2	2	2	TG

Datasets which do not initially have a target class or better known as unsupervised learning, with a pairing combination process and AHP pairwise comparison can have a target class. The dataset can then be used to classify and predict the success of the Student study period for new data. Classification can be done by implementing Back Propagation Neural Network (BPNN), Probabilistic Neural Network (PNN), Learning Vector Quantification (LVQ) and other Classification Algorithms.

In general, the success of Students to achieve timely graduation is strongly influenced by three main aspects namely Procrastination, Total credits and Number of Repeated Courses. While other attributes such as GPA, Subjects in 7th Semester, Self Confidence and Discipline are supporting attributes, but can't be ignored. Therefore, the novelty of this research is the presence of key attributes and supporting attributes to predict student success. Furthermore, unsupervised learning techniques can be used for processes closely related to supervised learning datasets, such as predictions and classifications. The weakness of this study is in determining the pairwise comparisons, where the human perceptions sometimes have a great different value between one assessor with another, so that the number of experts affects the generated eigenvalues. While linked to the dataset, this research has not been validated against the real predictions on the field, which allows this data to be applied first using predictive algorithms such as BPNN or the like to test the accuracy.

4.   CONCLUSION

Based on the research conducted on the case study to predict the success of students study in UIN Suska Riau, it can be concluded that the main attribute that becomes the feedback to determine the success rate of students is procrastination with the largest eigenvalue compared with others. While the sub-attributes that is very influential on procrastination are the attribute of waiting for friends and postponing the final project. Four combinations in determining the target class for supervised learning dataset in this case are procrastination, total of credits and Number of Repeated Courses. These three attributes are combined with four other attributes into a single unit in the datasets for the classification and prediction process. The consistency ratio of attributes and sub-attributes on average shows a small percentage or less than 10%, thus the assessment is considered to be very consistent.

REFERENCES

[1]   Gupta N, Rawal A, Narasimhan VL, Shiwani S. Accuracy, Sensitivity and Specificity Measurement of Various Classification Techniques on Healthcare Data. *IOSR Journal of Computer Engineering (IOSR-JCE).* 2013; 11(5): 70-73.

[2]   Wei Q, Dunbrack RL. The Role of Balanced Training and Testing Data Sets for Binary Classifiers in Bioinformatics. *PLOS-ONE Journal.* 2013; 8(7): 1-12.

[3]   Mustakim, Effectiveness of K-means Clustering to Distribute Training Data and Testing Data on K-Nearest Neighbor Classification. *Journal of Theoretical and Applied Information Technology (JATIT),* 2017; 95(21): 5693-5700.

[4]   Isfahani ZB, Jafari S, Akbarian R. Comparison of Supervised and Unsupervised Learning Classifiers for Travel Recommendations. *Journal of Global Research in Computer Science.* 2012; 3(8):51-55.

[5]   Sathya R and Abraham A. Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence (IJARAI).* 2013; 2(2): 34-38.

[6]   Sharareh R, Kalhori N, Zeng XJ. Improvement the Accuracy of Six Applied Classification Algorithms through Integrated Supervised and Unsupervised Learning Approach. *Journal of Computer and Communications.* 2014; 2(1): 201-209.

[7]   Tamouk J, Allahakbari F. A comparison among accuracy of KNN, PNN, KNCN, DANN and NFL. *International Journal of Computer Science Issues (IJCSI).* 2012; 9(3): 319-322.

[8]   Ashari A, Paryudi I, Tjoa AM. Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool. *International Journal of Advanced Computer Science and Applications (IJACSA).* 2013; 4(11): 33-39.

[9]   Mustakim, Buono A, Hermadi I. Performance Comparison Between Support Vector Regression and Artificial Neural Network for Prediction of Oil Palm Production. *Journal of Computer Science and Information.* 2016; 9(1): 1-8.

[10] Paulin F and Santhakumaran A. Classification of Breast cancer by comparing Back propagation training algorithms. *International Journal on Computer Science and Engineering (IJCSE).* 2011; 3(1): 327-332.

[11] Kaura P, Singhb M, Josanc GS. *Classification and prediction based data mining algorithms to predict slow learners in education sector.* 3rd International Conference on Recent Trends in Computing 2015(ICRTC-2015). Procedia Computer Science. 2015; 57: 500-508.

[12] Triantaphyllou E, Mann SH. Using The Analytic Hierarchy Process for Decision Making In Engineering Applications: Some Challenges. *Inter'l Journal of Industrial Engineering: Applications and Practice.* 1995; 2(1): 35-44.

[13] Khademi N, Behnia K, Saedi R. Using Analytic Hierarchy/ Network Process (AHP/ANP) in Developing Countries: Shortcomings and Suggestions. *A Journal Devoted to the Problems of Capital Investment.* 2014; 59(1): 2-29.

[14] Balubaid M, Alamoudi R. Application of the Analytical Hierarchy Process (AHP) to Multi-Criteria Analysis for Contractor Selection. *American Journal of Industrial and Business Management.* 2015; 5(1): 581-589.

[15] Whitaker R. *Alidation examples of the Analytic Hierarchy Process and Analytic Network Process.* Mathematical and Computer Modelling. 2007. 56: 840–859.

[16] Wedley WC. Consistency Prediction for Incomplete Ahp Matrices. *Mathl. Comput. Modelling.* 1993; 17(415): 151-161.

[17] Xuli H. The Eigenvalue Method on Wei ght Matrix in AHP. *Journal of Systems Science and Systems Engineering.* 1997; 6(3): 293-296.

[18] Saaty TL. How to Make a Decision: The Analytic Hierarcy Process. *European Journal of Operational Research.* 1990; 48: 9-26.

[19] Saaty TL. Decision-making with the AHP: Why is the principal eigenvector necessary. *European Journal of Operational Research.* 2003; 145: 85–91.

[20] Saaty TL. Decision making with the analytic hierarchy process. *Int. J. Services Sciences.* 2008; 1(1): 83-98.

[21] Kusumawardani RP, Agintiara M. *Application of Fuzzy AHP-TOPSIS Method for Decision Making in Human Resource Manager Selection Process.* The Third Information Systems International Conference. Procedia Computer Science. 2015; 72: 638–646.

[22] Fox WP, Ormond B, Williams A. Ranking terrorist targets using a hybrid AHP–TOPSIS methodology. *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology.* 2016; 13(1): 77–93.

[23] Afshari AR, Nikolić M, Akbari Z. Personnel Selection Using Group Fuzzy AHP and SAW Methods. *Journal of Engineering Management and Competitiveness (JEMC).* 2017; 7(1): 3-10.

[24] Blume C, Matthes K. Supervised Learning Approaches to Classify Sudden Stratospheric Warming Events. *Journal of The Atmospheric Sciences - American Meteorogical Society.* 2012; 69(1):1824-1840.

[25] Tarassenko I, Roberts S. *Supervised and unsupervised learning in radial basis function classifiers.* IEE Proceedings - Vision, Image and Signal Processing. 1994; 141(4): 210-216.

[26] Carneiro G, Chan AB, Moreno PJ, Vasconcelos N. Supervised Learning of Semantic Classes for Image Annotation and Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2007; 29(3): 394-410.

[27] Wei S, Kosorok MR. Latent Supervised Learning. *Journal of the American Statistical Association*. 2013. 108(503): 957-970.
[28] Dy JG, Brodley CE. Feature Selection for Unsupervised Learning. *The Journal of Machine Learning Research archive*. 2004; 5(1): 845-889.
[29] Agrawal R. K-Nearest Neighborn for Uncertain Data. *International Journal of Computer Applications*. 2014; 105(11): 13-16.
[30] Okfalisa, Gazalba I, Mustakim, Reza NGI. *Comparative Analysis of K-Nearest Neighbor and Modified K-Nearest Neighbor Algorithm for Data Classification.* IEEE Conferences: 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE). 2017; 2: 294-298.
[31] Larose DT. Discovering Knowledge in Data An Introduction to Data Mining. New York. Wiley Interscience. 2005: 90-106.
[32] Han JK, Pei MJ. Data Mining: Concepts and Techniques. Third Edition, British Library Cataloguing-in-Publication. New York. Morgan Kaufmann. 2011: 235-236.
[33] He M, An X. Information Security Risk Assessment Based on Analytic Hierarchy Process. *Indonesian Journal of Electrical Engineering and Computer Science*. 2016; 1(3): 656-664.
[34] Wei Z, Li M. *Information Security Risk Assessment Model Base on FSA and AHP*. International Conference on Machine Learning and Cybernetics (ICMLC). Qingdao. 2010; 5: 2252-2255.
[35] Bhattacharya S, Raju V. *A Condorcet Voting theory based AHP approach for MCDM problems*. Indonesian Journal of Electrical Engineering and Computer Science. 2017. 7(1):276-286.
[36] Sheng W, Zhang L, Tang W, Wang J, Fang H. *Optimal Multi Distributed Generators Planning Under Uncertainty using AHP and GA. TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2014. 12(4): 2582-2591.