

Data Mining, Fuzzy AHP And TOPSIS For Optimizing Taxpayer Supervision

M. Jupri¹, Riyanarto Sarno²

¹Department of Information Technology Management, Faculty of Business and Technology Management

²Department of Informatics, Faculty of Information Technology and Communication

Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

Article Info

Article history:

Received

Revised

Accepted

Keywords:

Data mining

Fuzzy AHP

TOPSIS

Taxpayer supervision

Taxpayer compliance

ABSTRACT

The achievement of accepting optimal tax need effective and efficient tax supervision can be achieved by classifying taxpayer compliance to tax regulations. Considering this issue, this paper proposes the classification of taxpayer compliance using data mining algorithms; i.e. C4.5, Support Vector Machine, K-Nearest Neighbor, Naive Bayes, and Multilayer Perceptron based on the compliance of taxpayer data. The taxpayer compliance can be classified into four classes, which are (1) formal and material compliant taxpayers, (2) formal compliant taxpayers, (3) material compliant taxpayers, and (4) formal and material non-compliant taxpayers. Furthermore, the results of data mining algorithms are compared by using Fuzzy AHP and TOPSIS to determine the best performance classification based on the criteria of Accuracy, F-Score, and Time required. Selection of the taxpayer's priority for more detailed supervision at each level of taxpayer compliance is ranked using Fuzzy AHP and TOPSIS based on criteria of dataset variables. The results show that C4.5 is the best performance classification and achieves preference value of 0.998; whereas the MLP algorithm results from the lowest preference value of 0.131. Alternative taxpayer A233 is the top priority taxpayer with a preference value of 0.433; whereas alternative taxpayer A051 is the lowest priority taxpayer with a preference value of 0.036.

*Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

Riyanarto Sarno,

Department of Informatics, Faculty of Information Technology and Communication, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.

Email: riyanarto@if.its.ac.id

1. INTRODUCTION

Tax is the largest state income to finance government administration, public services, and public infrastructure [1]. Because taxes are very important for the state, tax authorities must know all types of tax evasion, both by individual taxpayers and corporate taxpayers in Indonesia, quickly, effectively and efficiently. The taxation system in Indonesia is a self-assessment system. It means taxpayer calculates, deposits and self-reports the amount of tax he/she should pay based on the tax laws and regulations.

The consequence of self-assessment taxation system in Indonesia is a big data which comes from Periodic Tax Return and Annual Tax Return. It requires an effective and efficient way to be managed based on tax regulations. Effective means the tax authorities supervise taxpayers based on the compliance classification and determine taxpayers priorities based on predetermined variable criteria. Efficient means the tax authorities supervise taxpayers so that taxpayers do tax obligations based on the tax regulations without much time and big effort. Furthermore, taxes are forced by obtaining indirect compensations; thus, taxpayers tend to avoid tax or try to do tax evasion. It will harm the state in the form of reduced tax revenues.

Several studies related researches to detect tax avoidance efforts have been conducted. The researchers [2], conducted a study developing a system that can measure the dimensions of billboards without physically touching them automatically to calculate tax in Indonesia. The researchers [3], conducted a study using a hybrid intelligence system to detect tax evasion from corporate taxpayers, namely taxpayers who do business beverages and textile in Iran. The researchers [4], conducted a study investigating classic tax evasion cases

using several methods aimed at classifying tax evasion behavior based on the network that has been simulated with real data. The researchers [5], conducted a study by applying parallelism techniques that aim to improve the performance of fraud detection algorithms. Another researchers [6], conducted a study to detect fraudulent tax invoices using various types of data mining techniques.

This study proposes taxpayer supervision which variables in this study can be used for all types of taxpayers in Indonesia and can be used for all types of tax in Indonesia. Each state has different tax regulations and administration, so this study has advantages compared to other studies because this study detect tax avoidance in Indonesia. The supervision of taxpayers in this study is to classify taxpayer compliance into four classes with the classification algorithms of data mining. The classification algorithms chosen are C4.5, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes (NB), and Multilayer Perceptron (MLP). Every algorithm is compared to determine the best performance classification using Fuzzy AHP and TOPSIS. The best performance classification is chosen based on the criteria of Accuracy, F-Score and Time required. The priorities of taxpayers compliance supervision are formal and material non-compliant taxpayers, formal compliant taxpayers, material compliant taxpayers, and formal and material compliant taxpayers. Every taxpayer's compliance has a lot of data. Therefore, it has to priority taxpayers to be examined and processed based on tax regulations. In determining the priority of taxpayers, researchers propose fuzzy AHP and TOPSIS methods. The results show that alternative taxpayer A233 is the top priority taxpayer with a preference value of 0.433; whereas alternative taxpayer A051 is the lowest priority taxpayer with a preference value of 0.036.

2. RESEARCH METHOD

This study proposes effective and efficient taxpayer supervision by classifying taxpayer compliance into four classes. In general, the flow of this study is explained in Figure 1.

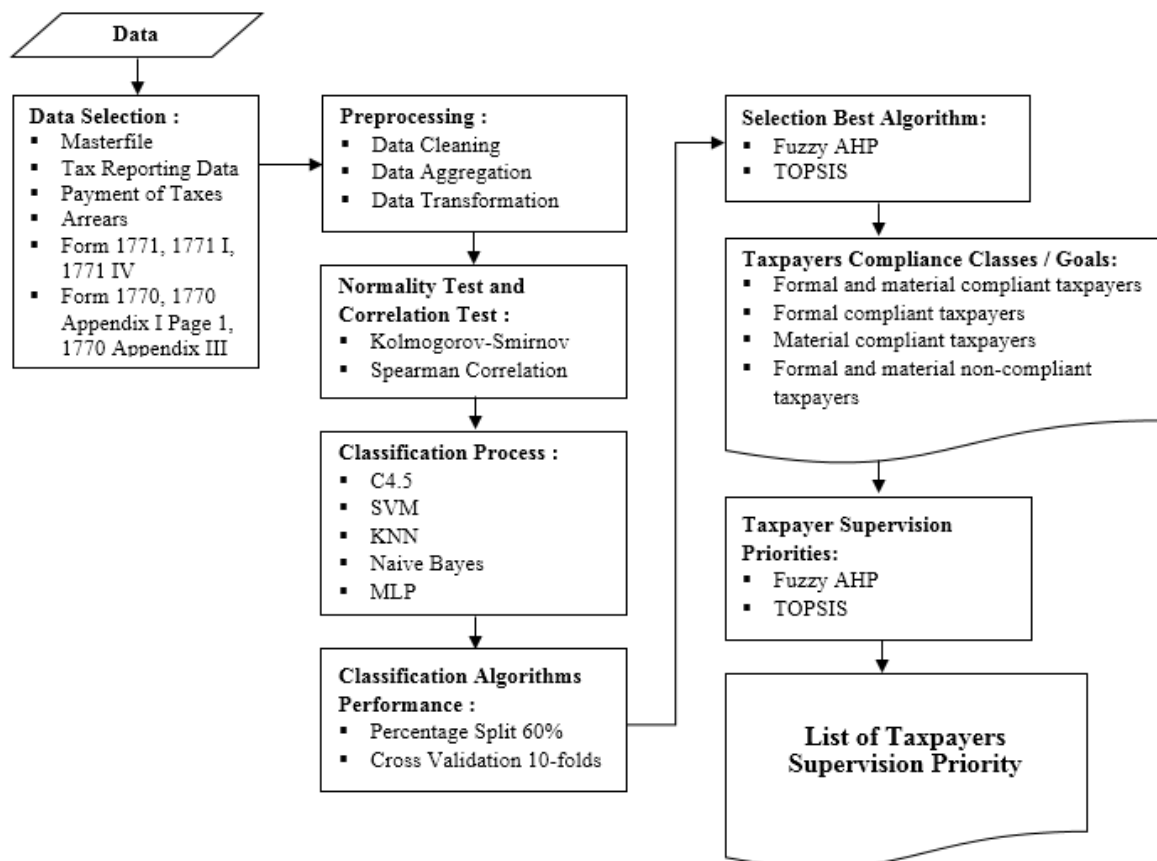


Figure 1. General description of the overall study

2.1. Selection Data

Data mining is the process of finding patterns and trends from large data so that predictions can be made about that data [7], [8]. Selection of variables in data mining is the process of identifying some of the most important variables or attributes that are important in the model for goals prediction [9]. This study uses data sources from the reporting of certain regional taxpayers for the 2015-2017 tax years. Tables from the

database used are Masterfile of taxpayers, Tax Reporting Data, Tax Payments, Arrears, Form 1771, 1771 I, 1771 IV, Form 1770, 1770 Appendix I Page 1, and 1770 Appendix III.

2.2. Preprocessing

Preprocessing is a step doing to make raw data into quality data [10]. The data is cleaned to correct bad data, filtering some incorrect data from the dataset and reducing unnecessary data details [10]. Data aggregation is the preprocessing process to summarize data [11]. Variables from the data aggregation process are Late_Report, Not_Report, Late_To_Pay, GPM, NPM and CTTOR. Late_Report and Not_Report variables are accumulations of late reports and unreported amounts in one tax year derived from tax reporting tables. Gross Profit Margin (GPM), Net Profit Margin (NPM), and Corporate Tax Turn Over Ratio (CTTOR) variables are financial ratio analysis variables. The Taxable Entrepreneur variable (PKP) is the result of transforming the PKP date column in the masterfile of taxpayers table where "Yes" if the PKP date is not null and "No" if the PKP date is null. Data Transformation is the process of consolidating data into forms that are suitable for data mining purposes [11]. The formal and material compliant taxpayers come from the list of compliant taxpayers with the criteria for late reporting within a year no more than three times, not reporting in a year no more than three times and late paying in one year no more than three times. Goals come from data on taxpayers who make corrections to Tax Return or taxpayers are issued Tax Underpayment Assessment Letter (SKPKB).

2.3. Normality Test and Correlation Test

This study uses the Kolmogorov-Smirnov normality test to find the distribution of data normal or not. The normality test determines the correlation test method to select dataset variables. The Kolmogorov Smirnov normality test compares the distribution data with normal distribution standard [12]. The data is not normal if the significance is below 0.05 [13]. The Kolmogorov-Smirnov test in this study used the SPSS Statistic 23 tool and the results showed that the study data had an abnormal distribution so that the correlation test method used the Spearman correlation.

The correlation test serves to select and to ensure the variables correlate with the goals to be achieved. The Spearman correlation test is a nonparametric statistic and it is used in the condition of one or both variables measured are ordinal scale or both variables are quantitative, but normal conditions are not met [14]. The Spearman correlation test in this study uses the SPSS Statistic 23 tool to determine the strength level of the inter-variable relationship guided by the correlation coefficient value. The correlation coefficient value is 0.00-0.25 = little or no relationship, correlation coefficient value is 0.25-0.50 = fair degree of relationship, correlation coefficient value is 0.50-0.75 = moderate to good relationship, and correlation coefficient value is >0.75 = very good to excellent relationship [14].

2.4. Classification Process

The classification process is the process of analyzing data by extracting a model that describes important data classes [11]. The process of classification data is divided into two steps, which consist of learning steps (where the classification model is built) and the classification step (where the model is used to predict goals based on the data given) [11]. The dataset has been created, using the C4.5, SVM, KNN, NB, and MLP classification algorithms to see the best performance classification.

C4.5 is a classification algorithm of data mining processes by forming a decision tree [15]. The first step this algorithm calculates information gain of all the attributes and then selects the root node from attribute having the highest information gain (1). The gain value can be calculated by finding the entropy value first (2). Entropy is diversity.

$$Gain(y, A) = Entropy(y) - \sum_{c \in value(A)} \frac{y_c}{y} Entropy(y_c) \quad (1)$$

where $value(A)$ is all possible values of attribute A, and y_c is a subset of y where A has a value of c

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

where p_1, p_2, \dots, p_n each express the proportions of class 1, class 2, ..., class n in output y

The second step makes a branch based on the possible attribute values, then divide the case into branches, and this process is repeated for each new branch so that the case in the branch has the same class.

Support Vector Machine is a supervised learning model for conducting a linear and non-linear classification of data [16], [17]. The SVM method finds the optimal hyperplane that separates feature points from two different classes with the largest margin possible in the feature space. Classification cases are linearly separable, the separating function sought is a linear function using Equation (3).

$$g(x) = \text{sign}(f(x)) \quad (3)$$

With $f(x) = w^T x + b$, where $x, w, b \in R$. This classification problem can be formulated as follows: we find the set of parameters (w, b) so that $f(x_i) = \langle w, x_i \rangle + b = y_i$ for all i . This technique finds the best separator function between functions that are not limited to separating two objects. If in two dimensions the separator is a line, in three dimensions the separator is a plane, and in high dimensions (more than three) the separator is a hyperplane. The best hyperplane is a hyperplane located in the middle between two sets of objects from two classes. Finding the best hyperplane is equivalent to maximizing the margin, which is the distance between the hyperplane and the support vectors.

K-Nearest Neighbor is a supervised algorithm learning based on the k-nearest neighbor by classifying new instances based on the majority of the k-nearest neighbor categories [18]. This method calculates similarities between samples of unlabeled data and all training data samples [19]. KNN is determined by looking at the shortest distance from the query instance to the training sample data. The Euclidean distance formula is often used to determine the distance between two training and testing objects.

Naive Bayes is a classification method based on the application of the Bayes theorem using knowledge about probability and statistics [20]. The Naive Bayes algorithm predicts future opportunities based on previous experience so that it is known as the Bayes Theorem (4). The main characteristic of Naive Bayes Classifier is a very strong (naive) assumption of independence from each condition/event.

$$P(C|X) = \frac{P(x|c)P(c)}{P(x)}$$

where x is data with an unknown class, c is the data hypothesis is a specific class, $P(c|x)$ is probability of hypothesis based on condition (posterior probability), $P(c)$ is probability of hypothesis (prior probability), $P(x|c)$ is probability based on conditions on the hypothesis, and $P(x)$: Probability c

Multilayer Perceptron adopts the workings of neural networks in living things, namely having an architecture on how to convert two or more inputs into output [21]. The inspiration of ANN comes from the observation that the system of learning from living things, especially humans, from very complex networks consisting of interconnected neurons. ANN appears as an alternative to conventional approaches which are usually less flexible about structural changes in problems. ANN proposes the advantages of being able to get over several problems without making drastic changes to the model.

2.5. Classification Algorithms Performance

Confusion matrix, Percentage Split Test, and k-fold Cross-validation have a function to determine the performance of the C4.5, SVM, KNN, NB, and MLP classification algorithms in classifying dataset based on goals of this study.

The confusion matrix is the process of evaluating the performance of a system by determining the amount of data that is classified correctly and incorrectly [22], [23]. The table confusion matrix shows the relationship between observed and estimated values for evaluating data classification results [24]. Accuracy is a comparison between data classified correctly with all data tested (5). Precision describes the number of positive category data that are classified correctly divided by the total data classified as positive (6). Recall shows what percentage of positive category data is correctly classified by the system (7). F-score is a combination of Precision and Recall to measure the main quality and can be used as an alternative (7).

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (5)$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (6)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (7)$$

$$\text{F-Score} = \frac{(2 \times \text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (8)$$

where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

Percentage Split Test divides the test method into two, training sets and test sets based on the desired percentage (Figure 2) [25]. The K-fold Cross-validation method divides the dataset consisting of training data as many as $k-1$ and testing data as many as 1. This study uses the value of k 10 (Figure 2) [26]. This study, the results of value of accuracy, F-Score, and Time required (Time) from Percentage Split Test method and K-fold Cross-validation method are searched the average value of each classification algorithm where is the average results as a basis for calculating the Alternative Data Normalization Matrix in TOPSIS method for determining algorithm with the best performance classification.

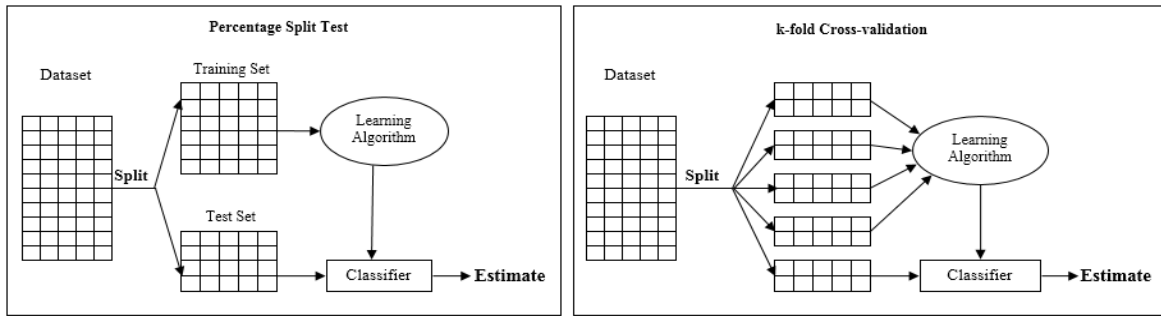


Figure 2. Algorithms Testing Methods

2.6. Selection Best Algorithm

Accuracy, F-Score, and Time required are the criteria for determining the best performance classification algorithm. The criteria were given weights with Fuzzy AHP and were ranked using TOPSIS. Fuzzy AHP functions to determine the weight of the specified criteria and TOPSIS to rank the selected alternatives [27]. This study uses Fuzzy AHP and TOPSIS because this method can provide good results and this method is suitable for solving complex problems that are not too subjective [28]. Figure 3 is the hierarchy to determine the best performance classification algorithm in this study.

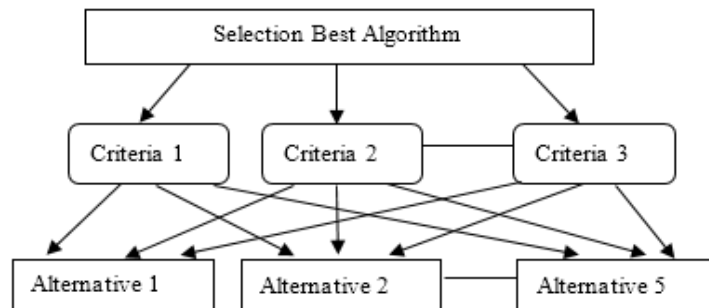


Figure 3. Hierarchy of purpose selection best algorithm

Fuzzy AHP method combines Analytic Hierarchy Process (AHP) and Fuzzy functions to minimize the obscurity that is formed in the fuzzy ratios [29]. Fuzzy AHP combines the AHP ranking with a fuzzy concept approach [30]. Fuzzy AHP is used to improve the deficiency of the AHP method, namely to solve problems that are not too subjective. AHP method is used to help solve multi-criteria problems from several alternative decisions [31], [32], [33]. AHP method solves the problem by breaking the problem into several parts so that it forms a hierarchy of three parts of objectives, criteria and alternatives [34], [35]. To get better results, AHP is combined with other techniques such as Fuzzy logic [36].

The Fuzzy set approach in AHP has the purpose of solving the problem of obscurities of human thought that was first used by Zadeh [37], [38]. Fuzzy numbers make it possible to solve the problems where criteria are not precisely defined [39]. To solution this, a number of special Triangular Fuzzy Numbers (TFN) were formed into AHP values which are divided into three parts, namely l = lowest value, m = middle value and u = highest value, using Equation (9) [40], [41]:

$$\mu_T(x) = \begin{cases} \frac{x-l}{m-l}, & \text{if } x \in [l, m] \\ \frac{u-x}{u-m}, & \text{if } x \in [m, u] \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

The first steps of Fuzzy AHP create the hierarchical structure of the problem that must be solved and determine the comparison of the paired matrices between the criteria and the scale of the TFN. The next steps determine the value of fuzzy Synthesis (Si) using Equation (10).

$$S_i = \sum_{j=1}^m M^j g_i \otimes \left[\sum_{i=1}^n \sum_{j=1}^n M^j g_i \right]^{-1} \tag{10}$$

where : $\sum_{j=1}^m M^j g_i$ is the sum of each TFN value in a paired matrix (11)

$$\sum_{j=1}^m M^j g_i = \left(\sum_{j=1}^m l_j, \sum_{j=1}^m m_j, \sum_{j=1}^m u_j \right) \tag{11}$$

While $[\sum_{i=1}^n \sum_{j=1}^n M^j g_i]^{-1}$ is the inverse of the operation of the number of TFN (12)

$$\begin{aligned} & \left[\sum_{i=1}^n \sum_{j=1}^n M^j g_i \right]^{-1} \\ &= \left(\frac{1}{\sum_{i=1}^n u_i}, \frac{1}{\sum_{i=1}^n m_i}, \frac{1}{\sum_{i=1}^n l_i} \right) \end{aligned} \quad (12)$$

After the operational fuzzy continues, then the search process for the possible level of operational fuzzy continues. The probability level is assumed to be $M_2 \geq M_1$ where $M_1 = (l_1, m_1, u_1)$ and $M_2 = (l_2, m_2, u_2)$ for more details see Equation (13).

$$V(M_2 \geq M_1) = \text{SUP}_{y \geq x} [\min(\mu_{M_1}(x), \mu_{M_2}(y))] \quad (13)$$

So that obtaining the possibilities (14):

$$V(M_2 \geq M_1) = \begin{cases} 1 & \text{if } m_2 \geq m_1 \\ 0 & \text{if } l_1 \geq l_2 \\ \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)} & \end{cases} \quad (14)$$

The next step is to calculate the defuzzification of the ordinate (d^i) and the weight of the vector (V). Before calculating vector weight values, the first thing to do is calculate the value of defuzzification ordination. To calculate the ordinate defuzzification can use Equation (15).

$$d^n(A_i) = \min V(S_i \geq S_k) \quad (15)$$

The next step after obtaining the defuzzification ordinate is to calculate the vector weight value with Equation (16).

$$W^n = (d^n(A_1), d^n(A_2), d^n(A_3), \dots, d^n(A_n))^T \quad (16)$$

Where $A_i = 1, 2, \dots, n$ is a fuzzy vector

After the vector weight value is known, the next step is to normalize the value of fuzzy vector weight (W) using Equation (17).

$$w = (d(A_1), d(A_2), d(A_3), \dots, d(A_n))^T \quad (17)$$

The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is a technical method for order preference with similarity to an ideal solution algorithm [36]. TOPSIS was first introduced by C.LHwang and K.Yoon in 1981 [42], [43]. In the case of the Multi-Criteria Decision Making (MCDM) many ranking methods can be used, one of which is a TOPSIS method [44]. MCDM was developed to provide solutions in the decision making the process [45]. There are five steps in the TOPSIS method [46], [47], namely:

Step 1: Calculate the normalized decision matrix (18).

$$r_{ab} = \frac{x_{ab}}{\sqrt{\sum_{i=1}^m x_{ab}^2}}, a = 1, 2, 3, \dots, m, b = 1, 2, 3, \dots, n \quad (18)$$

Step 2: Calculate the weighted normalized decision matrix (19).

$$v_{ab} = w_b \cdot r_{ab}, a = 1, 2, 3, \dots, m, b = 1, 2, 3, \dots, n \quad (19)$$

w_b is the weight of the b th criterion.

Step 3: Calculate the positive ideal solutions A^+ and negative ideal solutions A^- (20).

$$\begin{aligned} A^+ &= \{v_1^+, v_2^+ \dots v_n^+\} \\ A^- &= \{v_1^-, v_2^- \dots v_n^-\} \end{aligned} \quad (20)$$

Step 4: Using the Euclidean n -dimensional distance method do the calculation of the size of the separation of each alternative from the ideal solution (D_a^+) and separation of each alternative from the negative-ideal solution (D_a^-) (21).

$$D_a^- = \sqrt{\sum_{b=1}^n (v_{ab} - v_b^-)^2} \quad (21)$$

$$D_a^+ = \sqrt{\sum_{b=1}^n (v_{ab} - v_b^+)^2}$$

Step 5: Determine the preference value with Equation (22).

$$C_a = \frac{D_a^-}{(D_a^+ + D_a^-)} \tag{22}$$

Step 6 : Ranking the alternatives from largest to lowest preference value.

2.7. Taxpayer Supervision Priorities

After finding the algorithm with the best performance classification, the taxpayers are chosen as the priority of supervision. Based on the level of taxpayer compliance, the priority of taxpayers supervision is taxpayers who are at the level of compliance with the order of formal and material non-compliant taxpayers, formal compliant taxpayers, material compliant taxpayers, and formal and material compliant taxpayers.

This study uses the results of taxpayer data from the best algorithm classification with a formal and material non-compliant taxpayers goal as a role model because the priority taxpayers supervision has the same method for other levels of compliance. Twenty-five dataset variables are criteria for determining taxpayers supervision priorities. The results of taxpayer data from the best classification algorithm are the alternative to rank based on these Twenty-five criteria. Fuzzy AHP is used to give weight criteria and TOPSIS is used to rank alternatives. The flow of taxpayers supervision priorities determine the priority of taxpayers be supervised in this study as shown in Figure 4.

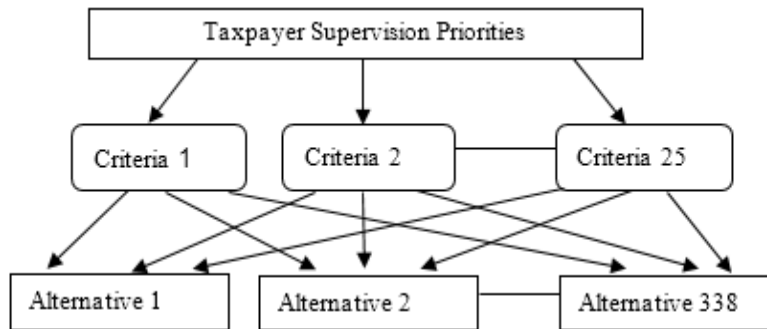


Figure 4. Hierarchy of purpose taxpayers supervision priorities

3. RESULTS AND DISCUSSION

The results of the preprocessing process, normality test, and correlation test in this study produce dataset variables as shown in Table 1.

Table 1. The Variables of The Research dataset

No	Variable	Description	Type
1	PKP_Status	Yes = Taxpayers are taxable entrepreneurs, No = Non-taxable entrepreneurs	Nominal
2	Type_of_taxpayer	Badan = Corporate taxpayer, OP = Individual taxpayer	Nominal
3	Sector	The sector code of Taxpayer business	Nominal
4	Late_Report	Late Amount of Report in one tax year	Numeric
5	Not_Report	Amount Not Reported in one tax year	Numeric
6	Late_to_pay	Late Pay Amount in one tax year	Numeric
7	Arrears	Arrears Value of SKPKB for annual tax returns that have not been paid off	Numeric
8	Turnover	Turnover	Numeric
9	VAT	Value Added Tax (VAT) with MAP Code = 411211	Numeric
10	Fiscal_Net_Income	Value of Fiscal Net Income	Numeric
11	PKP	Value of Taxable Income	Numeric
12	Compensation	Value of Compensation for Losses	Numeric
13	Income_Tax_payable	Value of Income Tax payable	Numeric
14	PPh29_payment	Value of Income Tax Article 29 payment with MAP Code = 411125 or 411126 with payment type code = 200	Numeric
15	Installment_Psl25	The instalment amount of Article 25 Income Tax per month	Numeric
16	Psl25_Payment	Value of Income Tax Article 29 payment with MAP Code = 411125 or 411126 with payment type code = 100	Numeric
17	HPP	Cost of goods sold	Numeric
18	DPP_Peng_TB	Basic Imposition of Taxes Transfer of Rights to Land and/or Buildings	Numeric
19	42_TB_Payment	Payment of Final PPh Article 4 paragraph (2) for the transfer of rights to land and/or buildings with MAP Code = 411128 with payment type code = 402	Numeric
20	DPP_Construction	Tax Base for construction business	Numeric
21	DPP_PP46	Basic Imposition of Certain Income Taxes (PP 46)	Numeric

22	PP46_Payment	Payment of PP 46 with MAP Code = 411128 and payment type code = 420	Numeric
23	GPM	Gross Profit Margin	Numeric
24	NPM	Net Profit Margin	Numeric
25	CTTOR	Corporate Tax Turn Over Ratio	Numeric
26	Tax_Compliance	Goals of this study	Nominal

This study uses the Weka version 3.8.1 tool to test dataset. The dataset in this study has 2424 data. The C4.5, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes (NB), and Multilayer Perceptron (MLP) classification algorithms were tested and analyzed with a confusion matrix to show the performance classification of the algorithms. The test models used are 10-folds Cross-validation (M1) and Split Mode Percentage of 60% (M2).

Table 2. Confusion Matrix on M1 and on M2

	C4.5		SVM		KNN		NB		MLP	
	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
Correctly Classified	2371	950	2126	861	2081	841	1871	749	2141	875
Incorrectly Classified	53	23	298	112	343	132	553	224	283	98

Table 2 shows a confusion matrix resulting from the 10-folds Cross-validation (M1) testing method and Percentage Split Mode 60% (M2) method from each classification algorithm. Comparison of the classification results with the average value (A) of M1 and M2 is shown in Table 3.

Table 3. Comparison on M1 and on M2

	C4.5			SVM			KNN			NB			MLP		
	M1	M2	A	M1	M2	A	M1	M2	A	M1	M2	A	M1	M2	A
Precision	0.98	0.98	0.98	0.88	0.89	0.89	0.86	0.87	0.86	0.74	0.72	0.73	0.88	0.90	0.89
Recall	0.98	0.98	0.98	0.88	0.89	0.88	0.86	0.86	0.86	0.77	0.77	0.77	0.88	0.90	0.89
F-Score	0.98	0.98	0.98	0.88	0.89	0.88	0.86	0.86	0.86	0.75	0.74	0.75	0.88	0.90	0.89
Accuracy	97.81	97.64	97.73	87.71	88.49	88.10	85.85	86.43	86.14	77.19	76.98	77.09	88.33	89.93	89.13
Time (seconds)	0.11	0.13	0.12	3.34	3.04	3.19	0.00	0.00	0.00	0.05	0.02	0.04	72.67	69.37	71.02

There are two phases, that are Fuzzy AHP and TOPSIS. Fuzzy AHP is used to obtain fuzzy vector weights based on the criteria of Accuracy (C1), F-Score (C2), and Time required (C3) whereas TOPSIS is used to determine the rank of all alternatives based on the largest preference value. The alternatives to determine the best algorithm are C4.5 (A1), SVM (A2), KNN (A3), NB (A4) and MLP (A5). The initial phases make a comparison of AHP criteria matrix from each criterion and make a comparison matrix of fuzzy AHP criteria and calculate the value of fuzzy Synthesis (Si) using Equation (10) as Table 4.

Table 4. Comparison matrix of each criterion and Value of Si

C	C1			C2			C3			Synthesis Value		
	L	M	U	L	M	U	L	M	U	L	M	U
C1	1.0	1.0	1.0	0.5	1.0	1.5	1.0	1.5	2.0	0.20	0.38	0.63
C2	0.7	1.0	2.0	1.0	1.0	1.0	1.0	1.5	2.0	0.21	0.38	0.70
C3	0.5	0.7	1.0	0.5	0.7	1.0	1.0	1.0	1.0	0.16	0.25	0.42

The next steps calculate the degree of possibility by using Equation (13) and calculate the ordinate defuzzification value (d') using Equation (15). The results are shown in Table 5.

Table 5. Degree of Possibility and Defuzzification Ordinate

Criteria	Degree of possibility	Summary Of Degree	d'
C1	C1 >= C2	1	1
	C1 >= C3	1.365	
C2	C2 >= C1	1	1
	C2 >= C3	1.303	
C3	C3 >= C1	0.636	0.622
	C3 >= C2	0.622	

The final steps of Fuzzy AHP calculate Weight Vector values using Equation (16) and normalize the value of the fuzzy Weight Vector using Equation (17). The results are shown in Table 6.

Table 6. Value of The Weight Vector

Criteria	W'	W
C1	1	0.381
C2	1	0.381
C3	0.622	0.237

After obtaining the weight of each criterion as in Table 6, the second phases rank alternative data to determine the best algorithm using the TOPSIS method. The first steps of the TOPSIS method create Alternative Data Normalization Matrix using Equation (18) and create the Alternative Data Normalization

Matrix using Equation (19) based on the weight vector as shown in Table 6. The results of the Alternative Data Normalization Matrix and The Weighted Alternative Data Normalization Matrix are shown in Table 7.

Table 7. Alternative Data Normalization Matrix and The Weighted Alternative Data Normalization Matrix

Alternative	Normalization Matrix			Weighted Normalization Matrix		
	C1	C2	C3	C1	C2	C3
A1	0.497	0.501	0.002	0.19	0.191	0
A2	0.448	0.45	0.045	0.171	0.172	0.011
A3	0.438	0.439	0	0.167	0.168	0
A4	0.392	0.383	0.001	0.15	0.146	0
A5	0.454	0.455	0.999	0.173	0.173	0.237

The next steps calculate the value of Positive Ideal Solutions Matrix and Negative Ideal Solutions Matrix using Equation (20). The results of Positive Ideal Solutions Matrix and Negative Ideal Solutions Matrix are shown in Table 8.

Table 8. Positive Ideal Solutions Matrix and Negative Ideal Solutions Matrix

Alternative	C1	C2	C3
A+	0.19	0.191	0
A-	0.15	0.146	0.237

The value of Positive Ideal Solutions Matrix and Negative Ideal Solutions Matrix as shown in Table 8 are used to calculate Alternative Distance values using Equation (21) and to calculate the preference values of each alternative using Equation (22) as shown in Table 9.

Table 9. Alternative Distance and Value of Preference

Alternative	Distance + (D+)	Distance - (D-)	Preference
A1	0.0000	0.2449	0.998
A2	0.0000	0.2280	0.887
A3	0.0447	0.2366	0.880
A4	0.0632	0.2366	0.797
A5	0.2366	0.0447	0.131

The final results of TOPSIS method rank preference values based on the largest preference value as shown in Table 10.

Table 10. Ranking alternative

Ranking	Alternative	Preference
1	A1	0.998
2	A2	0.887
3	A3	0.880
4	A4	0.797
5	A5	0.131

Table 10 shows that the C4.5 (A1) is the best performance classification because it has the largest preference. After finding the best algorithm, the next steps determine the priority of the taxpayer be supervised. The priority of taxpayers in this study is divided into two phases, namely Fuzzy AHP to obtain fuzzy vector weights and TOPSIS to rank all alternatives based on the largest preference value. The criterias for obtaining fuzzy vector weights are PKP_Status (C1), Type_of_taxpayer (C2), Sector (C3), Late_Report (C4), Not_Report (C5), Late_to_pay (C6), Arrears (C7), Turnover (C8), VAT (C9), Fiscal_Net_Income (C10), PKP (C11), Compensation (C12), Income_Tax_payable (C13), PPh29_payment (C14), Installment_Psl25 (C15), Psl25_Payment (C16), HPP (C17), DPP_Peng_TB (C18), 42_TB_Payment (C19), DPP_Construction (C20), DPP_PP46 (C21), PP46_Payment (C22), GPM (C23), NPM (C24), CTTOR (C25). This study ranks the priority of taxpayer's from class/goal formal and material non-compliant taxpayers data. It is the results of the best algorithm. The formal and material non-compliant taxpayers are 338 data. The first steps determining the priority of taxpayers to be supervised make a comparison matrix of AHP criteria as shown in Table 11.

Table 11. Comparison matrix of each criterion

	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25
C01	1.0	5.0	3.0	3.0	3.0	4.0	0.1	0.3	0.2	0.3	0.3	0.3	0.3	0.2	3.0	0.2	0.3	3.0	0.2	3.0	0.3	0.2	0.2	0.2	0.3
C02	0.2	1.0	0.3	3.0	3.0	3.0	0.1	0.3	0.2	3.0	0.3	0.5	0.3	0.2	5.0	0.2	0.3	2.0	0.2	2.0	0.3	0.2	0.2	0.2	0.2
C03	0.3	3.0	1.0	4.0	4.0	4.0	0.1	0.3	0.2	4.0	0.3	0.3	0.3	0.2	3.0	0.2	0.3	0.3	0.2	0.3	0.2	0.2	0.2	0.2	0.3
C04	0.3	0.3	0.3	1.0	0.5	1.0	0.1	0.3	0.3	0.3	0.3	0.5	0.3	0.3	0.3	0.3	0.5	0.3	0.5	0.5	0.3	0.3	0.3	0.3	0.3
C05	0.3	0.3	0.3	2.0	1.0	2.0	0.1	0.3	0.3	0.3	0.3	0.5	0.3	0.3	0.3	0.3	0.5	0.3	0.5	0.5	0.3	0.3	0.3	0.3	0.3
C06	0.3	0.3	0.3	1.0	0.5	1.0	0.1	0.3	0.3	0.3	0.3	0.5	0.3	0.3	0.3	0.3	0.5	0.3	0.5	0.5	0.3	0.3	0.3	0.3	0.3
C07	7.0	7.0	7.0	7.0	7.0	7.0	1.0	5.0	3.0	4.0	5.0	5.0	5.0	3.0	4.0	3.0	4.0	4.0	3.0	5.0	5.0	3.0	3.0	3.0	2.0
C08	4.0	4.0	4.0	3.0	3.0	3.0	0.2	1.0	0.3	3.0	3.0	5.0	4.0	0.3	3.0	0.3	3.0	4.0	0.3	3.0	3.0	0.3	0.3	0.3	0.3
C09	5.0	5.0	5.0	4.0	4.0	4.0	0.3	3.0	1.0	3.0	5.0	5.0	5.0	1.0	5.0	1.0	3.0	3.0	1.0	3.0	3.0	1.0	3.0	3.0	2.0
C10	4.0	0.3	0.3	3.0	3.0	3.0	0.3	0.3	0.3	1.0	1.0	3.0	2.0	0.3	3.0	0.3	3.0	3.0	0.3	3.0	0.3	0.3	0.3	0.3	0.3
C11	4.0	4.0	4.0	3.0	3.0	3.0	0.2	0.3	0.2	1.0	1.0	3.0	2.0	0.3	3.0	0.3	3.0	3.0	0.3	3.0	0.3	0.3	0.3	0.3	0.3
C12	3.0	2.0	3.0	2.0	2.0	2.0	0.2	0.2	0.2	0.3	0.3	1.0	0.3	0.2	3.0	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
C13	4.0	4.0	4.0	3.0	3.0	3.0	0.2	0.3	0.2	0.5	0.5	3.0	1.0	0.3	0.5	0.3	0.5	0.3	0.5	0.3	0.3	0.3	0.3	0.3	0.3
C14	5.0	5.0	5.0	3.0	3.0	3.0	0.3	3.0	1.0	3.0	3.0	5.0	3.0	1.0	3.0	1.0	3.0	3.0	1.0	3.0	3.0	1.0	3.0	3.0	2.0
C15	0.3	0.2	0.3	3.0	3.0	3.0	0.3	0.3	0.2	0.3	0.3	0.3	2.0	0.3	1.0	0.3	0.3	2.0	0.3	3.0	0.3	0.3	0.3	0.3	0.3

C16	5.0	5.0	5.0	3.0	3.0	3.0	0.3	4.0	1.0	3.0	3.0	3.0	3.0	1.0	3.0	1.0	3.0	3.0	1.0	3.0	3.0	1.0	3.0	3.0	2.0	
C17	4.0	4.0	4.0	3.0	3.0	3.0	0.3	0.3	0.3	0.3	3.0	4.0	3.0	0.3	3.0	0.3	1.0	3.0	0.3	3.0	0.3	0.3	0.3	0.3	0.3	
C18	0.3	0.5	3.0	2.0	2.0	2.0	0.3	0.3	0.3	0.3	0.3	3.0	2.0	0.3	0.5	0.3	0.3	1.0	0.3	1.0	1.0	0.3	0.3	0.3	0.3	
C19	5.0	5.0	5.0	3.0	3.0	3.0	0.3	3.0	1.0	4.0	4.0	4.0	3.0	1.0	3.0	1.0	3.0	3.0	1.0	3.0	3.0	1.0	3.0	3.0	2.0	
C20	0.3	0.5	3.0	2.0	2.0	2.0	0.2	0.3	0.3	0.3	0.3	3.0	2.0	0.3	0.3	0.3	1.0	0.3	1.0	2.0	0.3	0.3	0.3	0.3	0.3	
C21	3.0	3.0	3.0	2.0	2.0	2.0	0.2	0.3	0.3	0.3	0.3	4.0	3.0	0.3	0.3	0.3	1.0	0.3	0.5	1.0	0.3	0.3	0.3	0.3	0.3	
C22	5.0	5.0	5.0	3.0	3.0	3.0	0.3	3.0	1.0	3.0	3.0	4.0	3.0	1.0	3.0	1.0	3.0	3.0	1.0	3.0	3.0	1.0	3.0	3.0	2.0	
C23	5.0	5.0	5.0	3.0	3.0	4.0	0.3	3.0	0.3	3.0	3.0	3.0	3.0	3.0	0.3	3.0	0.3	3.0	0.3	3.0	0.3	3.0	0.3	1.0	0.5	0.3
C24	5.0	5.0	5.0	4.0	4.0	4.0	0.3	3.0	0.3	3.0	3.0	3.0	3.0	3.0	0.3	3.0	0.3	3.0	0.3	3.0	0.3	3.0	0.3	2.0	1.0	0.5
C25	4.0	5.0	4.0	3.0	3.0	3.0	0.5	3.0	0.5	3.0	3.0	3.0	3.0	0.5	3.0	0.5	3.0	0.5	3.0	0.5	3.0	0.5	3.0	2.0	1.0	0.10

The next steps make a comparison matrix of fuzzy AHP criteria with TFN scale and calculate fuzzy Synthesis (Si) using Equation (11) based on Table 11. The values of fuzzy Synthesis are shown in Table 12.

Table 12. Value of Si

	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	
L	0.02	0.02	0.02	0.01	0.01	0.01	0.05	0.03	0.03	0.02	0.02	0.01	0.02	0.03	0.02	0.03	0.02	0.02	0.03	0.02	0.02	0.03	0.03	0.03	0.03	0.03
M	0.03	0.03	0.03	0.03	0.03	0.03	0.08	0.05	0.06	0.04	0.04	0.03	0.04	0.05	0.03	0.05	0.04	0.03	0.05	0.03	0.03	0.05	0.05	0.05	0.05	0.05
U	0.05	0.05	0.05	0.05	0.05	0.05	0.13	0.08	0.10	0.07	0.07	0.05	0.07	0.09	0.06	0.09	0.08	0.06	0.09	0.06	0.06	0.09	0.09	0.09	0.09	0.10

The next steps calculate the degree of possibility using Equation (13) and calculate the ordinate defuzzification values (d') using Equation (15). The next steps calculate the vector weight values using Equation (16) and normalize the values of the fuzzy Weight Vector using Equation (17) as shown in Table 13.

Table 13. Value of The Weight Vector

	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25
W'	0.13	0.07	0.09	0.12	0.12	0.10	1.00	0.51	0.72	0.32	0.36	0.07	0.36	0.62	0.17	0.61	0.44	0.18	0.64	0.17	0.25	0.61	0.56	0.59	0.63
W	0.01	0.01	0.01	0.01	0.01	0.01	0.11	0.05	0.08	0.03	0.04	0.01	0.04	0.07	0.02	0.07	0.05	0.02	0.07	0.02	0.03	0.07	0.06	0.06	0.07

After obtaining the weights of each criterion as shown in Table 13, the second phases rank alternative data using the TOPSIS method. The initial steps of the TOPSIS method create the alternative data normalization matrix using Equation (18) and create the Weighted Alternative Data Normalization Matrix using Equation (19). The results of the Weighted Alternative Data Normalization Matrix are used to calculate the values of Positive Ideal Solutions Matrix and Negative Ideal Solutions Matrix using Equation (20) as shown in Table 14.

Table 14. Positive Ideal Solutions Matrix and Negative Ideal Solutions Matrix

	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	
A+	0.00	0.00	0.00	0.00	0.01	0.00	0.08	0.02	0.05	0.02	0.02	0.01	0.02	0.05	0.01	0.03	0.01	0.01	0.06	0.01	0.02	0.04	0.01	0.02	0.03	
A-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.00

The values of Positive Ideal Solutions Matrix and Negative Ideal Solutions Matrix as shown in Table 14 are used to calculate Alternative Distance values using Equation (21). After that, the preference values of each alternative are calculated using Equation (22). The final results determine the priority of taxpayers be supervised as shown in Table 15 which rank alternatives based on the largest reference values. A taxpayer with Alternative A233 is priority number one taxpayer to be supervised while the taxpayer with alternative A051 is the last priority taxpayer.

Table 15. Ranking alternative

Ranking	Alternative	Preference
1	A233	0.433
2	A163	0.362
3	A240	0.330
4	A316	0.326
5	A151	0.304
.	.	.
.	.	.
.	.	.
337	A035	0.044
338	A051	0.036

4. CONCLUSION

Table 10 shows that the C4.5 (A1) algorithm is the best performance classification to classify taxpayer compliance based on this study dataset with a preference value of 0.998 whereas the MLP (A5) algorithm is the lowest classification algorithm with a preference value of 0.131. Table 15 shows that fuzzy AHP and TOPSIS methods rank taxpayer priorities for supervision at each level of compliance from the best algorithm classification results. Alternative taxpayer A233 is the top priority taxpayer with preference value 0.433 whereas alternative taxpayer A051 is the lowest priority taxpayer with preference value 0.036.

Further studies can be developed by expanding data sources for data mining selection. Tax avoidance can be known and detected as early as possible if the criteria used are complex and current with tax avoidance behavior.

ACKNOWLEDGEMENTS



Authors give a deep thanks to Institut Teknologi Sepuluh Nopember for supporting the study.

REFERENCES

- [1] S. Kafkalas, P. Kalaitzidakis, and V. Tzouvelekas, "Tax evasion and public expenditures on tax revenue services in an endogenous growth model," *Eur. Econ. Rev.*, 2014.
- [2] R. F. Rahmat, S. Purnamawati, H. Saito, M. F. Ichwan, and T. M. Lubis, "Android-based automatic detection and measurement system of highway billboard for tax calculation in Indonesia," *IJECS*, vol. 14, no. 2, pp. 877–886, 2019.
- [3] E. Rahimikia, S. Mohammadi, T. Rahmani, and M. Ghazanfari, "Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran," *Int. J. Account. Inf. Syst.*, vol. 25, pp. 1–17, 2017.
- [4] F. Tian *et al.*, "Mining suspicious tax evasion groups in big data," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 10, pp. 2651–2664, 2016.
- [5] M. S. Rad and A. Shahbahrami, "High performance implementation of tax fraud detection algorithm," *2015 Signal Process. Intell. Syst. Conf. Sp. 2015*, pp. 6–9, 2016.
- [6] P. Castellón González and J. D. Velásquez, "Characterization and detection of taxpayers with false invoices using data mining techniques," *Expert Syst. Appl.*, vol. 40, no. 5, pp. 1427–1436, 2013.
- [7] S. Mohamed and A. Ezzati, "A data mining process using classification techniques for employability prediction," *IJECS*, vol. 14, no. 2, pp. 1025–1029, 2019.
- [8] Arisman, "Nitric Oxide Chemistry and Velocity Slip Effects in Hypersonic Boundary Layers," *Thesis*, vol. 25, no. 1, pp. 1–64, 2014.
- [9] V. Kotu and B. Deshpande, *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. 2014.
- [10] S. García, J. Luengo, and F. Herrera, "Preface," *Intell. Syst. Ref. Libr.*, vol. 72, 2015.
- [11] H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [12] Z. Hanusz and J. Tarasińska, "Normalization of the Kolmogorov–Smirnov and Shapiro–Wilk tests of normality," *Biometrical Lett.*, vol. 52, no. 2, pp. 85–93, 2015.
- [13] C. Dissanayaka, H. Abdullah, B. Ahmed, T. Penzel, and D. Cvetkovic, "Classification of healthy subjects and insomniac patients based on automated sleep onset detection," in *IFMBE Proceedings*, 2016, vol. 56, pp. 188–192.
- [14] L. Liu, "Biostatistical Basis of Inference in Heart Failure Study," *Hear. Fail. Epidemiol. Res. Methods*, pp. 43–82, 2017.
- [15] A. Priya and E. Meenakshi, "Detection of phishing websites using C4.5 data mining algorithm," *RTEICT 2017 - 2nd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. Proc.*, vol. 2018-Janua, pp. 1468–1472, 2018.
- [16] M. Fikri and R. Sarno, "A comparative study of sentiment analysis using SVM and SentiWordNet," *IJECS*, vol. 13, no. 3, pp. 902–909, 2019.
- [17] P. A. Abhang, B. W. Gawali, and S. C. Mehrotra, *Introduction to EEG- and Speech-Based Emotion Recognition*. 2016.
- [18] T. Cohen and D. Widdows, "Geometric Representations in Biomedical Informatics: Applications in Automated Text Analysis," in *Methods in Biomedical Informatics: A Pragmatic Approach*, 2013, pp. 99–139.
- [19] F. Chen, Z. Ye, C. Wang, L. Yan, and R. Wang, "A feature selection approach for network intrusion detection based on tree-seed algorithm and k-nearest neighbor," *Proc. 2018 IEEE 4th Int. Symp. Wirel. Syst. within Int. Conf. Intell. Data Acquis. Adv. Comput. Syst. IDAACS-SWS 2018*, pp. 68–72, 2018.
- [20] R. Dorajoo, J. Liu, and B. O. Boehm, "Genetics of type 2 diabetes and clinical utility," *Genes (Basel)*, vol. 6, no. 2, pp. 372–384, 2015.
- [21] N. E. Karkalos and A. P. Markopoulos, *Modeling of hard machining*, vol. 1. Elsevier Ltd., 2017.
- [22] S. I. Sabilla and R. Sarno, "Development of wavelet transforms to predict methane in chili using the electronic nose," *Proceeding - ICAMIMIA 2017 Int. Conf. Adv. Mechatronics, Intell. Manuf. Ind. Autom.*, pp. 271–276, 2018.
- [23] A. Meyer-Baese and J. J. Schmid Volker, "Pattern Recognition and Signal Analysis in Medical Imaging: Second Edition," *Pattern Recognit. Signal Anal. Med. Imaging Second Ed.*, pp. 1–444, 2014.
- [24] R. Nisbet, G. Miner, and K. Yale, "Model Evaluation and Enhancement," *Handb. Stat. Anal. Data Min. Appl.*, pp. 215–233, 2018.
- [25] M. Bramer, *Principles of Data Mining*. 2007.
- [26] R. Sarno, D. R. Wijaya, and M. N. Mahardika, "Music fingerprinting based on bhattacharya distance for song and cover song recognition," *IJECE*, vol. 9, no. 2, pp. 1036–1044, 2019.
- [27] U. Yudatama and R. Sarno, "Evaluation maturity index and risk management for it governance using Fuzzy AHP and Fuzzy TOPSIS (case Study Bank XYZ)," *2015 Int. Semin. Intell. Technol. Its Appl. ISITIA 2015 - Proceeding*, pp. 323–327, 2015.
- [28] C. C. Sun, "A performance evaluation model by integrating fuzzy AHP and fuzzy TOPSIS methods," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7745–7754, 2010.
- [29] I. Ertuğrul and N. Karakaşoğlu, "Comparison of fuzzy AHP and fuzzy TOPSIS methods for facility location selection," *Int. J. Adv. Manuf. Technol.*, vol. 39, no. 7–8, pp. 783–795, 2008.

- [30] K. A. Wiguna, R. Sarno, and N. F. Ariyani, "10.1109@ICTS.2016.7910305.pdf," pp. 237–243, 2016.
- [31] T. L. Saaty, *The Analytic Hierarchy Process*. 1980.
- [32] N. Noradachanon and T. Senivongse, "Decision Model for Identity Management Product Selection Using Fuzzy AHP," pp. 269–275, 2017.
- [33] T. L. Saaty, "Decision making with the analytic hierarchy process," *Int. J. Serv. Sci.*, vol. 1, no. 1, p. 83, 2008.
- [34] N. Saaludin, M. H. Ismail, B. C. Mat, and S. Harun, "Improving lecturers' evaluation score by using analytic hierarchy process (AHP): A case at Universiti Kuala Lumpur," *IJECS*, vol. 15, no. 1, pp. 391–398, 2019.
- [35] S. Pavani, L. K. Sharma, and H. S. Hota, "A group expert evaluation for teachers by integrating fuzzy AHP and TOPSIS models," in *Proceedings of the 2013 IEEE International Conference in MOOC, Innovation and Technology in Education, MITE 2013*, 2013, pp. 85–90.
- [36] C. Science, E. M. Mohamed, B. Bouikhalene, F. Ouatik, S. Safi, and A. Info, "AHP and TOPSIS methods applied in the field of scientific," *IJECS*, vol. 14, no. 3, pp. 1382–1390, 2019.
- [37] S. Grazioso, M. Selvaggio, D. Marzullo, G. Di Gironimo, and M. Gospodarczyk, "ELIGERE: A fuzzy AHP distributed software platform for group decision making in engineering design," in *IEEE International Conference on Fuzzy Systems*, 2017.
- [38] L. a. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [39] H. T. X. Chi and V. F. Yu, "Ranking generalized fuzzy numbers based on centroid and rank index," *Appl. Soft Comput. J.*, vol. 68, pp. 283–292, 2018.
- [40] S. A. Halim, S. Deris, and M. Z. M. Zaki, "Fuzzy AHP based design decision for Product Line architecture," *2014 8th. Malaysian Softw. Eng. Conf.*, pp. 119–124, 2014.
- [41] Z. Yan, "Estimating Quality of the Construction Engineering Project Using Fuzzy AHP Approach," in *2016 International Conference on Smart Grid and Electrical Automation (ICSGEA)*, 2016, pp. 98–101.
- [42] G. Sophia and R. Sarno, "AHP-TOPSIS for analyzing job performance with factor evaluation system and process mining," *TELKOMNIKA*, vol. 17, no. 3, p. 1344, 2019.
- [43] X. Li, D. Zhou, Z. Yang, J. Huang, K. Zhang, and Q. Pan, "UAV route evaluation algorithm based on CSA-AHP and TOPSIS," *2017 IEEE Int. Conf. Inf. Autom. ICIA 2017*, no. July, pp. 915–920, 2017.
- [44] S. A. A. Alrababah, K. H. Gan, and T. P. Tan, "Comparative analysis of MCDM methods for product aspect ranking: TOPSIS and VIKOR," in *2017 8th International Conference on Information and Communication Systems, ICICS 2017*, 2017, pp. 76–81.
- [45] A. Cahyapratama and R. Sarno, "Application of Analytic Hierarchy Process (AHP) and Simple Additive Weighting (SAW) methods in singer selection process," *2018 Int. Conf. Inf. Commun. Technol. ICOIACT 2018*, vol. 2018-Janua, no. Mcdm, pp. 234–239, 2018.
- [46] M. R. Sanjay, M. Jawaid, N. V. R. Naidu, and B. Yogesha, *TOPSIS method for selection of best composite laminate*. Elsevier, 2018.
- [47] N. Ploskas and J. Papathanasiou, "A decision support system for multiple criteria alternative ranking using TOPSIS and VIKOR in fuzzy and nonfuzzy environments," *Fuzzy Sets Syst.*, vol. 1, pp. 1–30, 2019.

BIOGRAPHIES OF AUTHORS

	<p>M. Jupri is now a second-year student of Department of Information Technology Management, Faculty of Business and Technology Management, Institut Teknologi Sepuluh Nopember. His current interests are in Data Mining, Decision Making, and Machine Learning E-mail: m.jupri@gmail.com</p>
	<p>Riyanarto Sarno received M.Sc and Ph.D in Computer Science from the University of Brunswick Canada in 1988 and 1992. In 2003 he was promoted to a Full Professor. His teaching and research interests includes Internet of Things, Process Aware Information Systems, Intelligent Systems and Business Process Management. E-mail: riyanarto@if.its.ac.id</p>