Gain ratio in weighting attributes on simple additive weighting

Ade Rizka¹, Syahril Efendi²,* and Pahala Sirait³

¹Student in Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia
²Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia
³Faculty of Master Information Technology, STMIK STIE Mikroskil, Medan, Indonesia

*E-mail:syahril1@usu.ac.id

Abstract. Simple Additive Weighting (SAW) is one of Multi Attribute Decision Making (MADM) method known as simple weighted linear combination and most used. However, based on several studies, it produces lower accuracy values than other MADM methods. Because there is no validation in the weighting system for each attribute so that it affects the decision-making process and for some newly incompatible attributes causing errors in decision making and determining the best alternative. In this study, researchers used gain ratio as the basis of attribute weighting on SAW. Datasets used from UCI machine learning repository, such as cryotherapy, immunotherapy, ILPD and user knowledge modelling. The accuracy result of this research will be compare with the result of SAW method accuracy value based on the weight of the dataset using relative standard deviation. The average value of accuracy obtained by weighting attributes based on the weight of the dataset of 28.1825% and weight gain ratio of 31.6975%. Then on attribute weighting based on the gain ratio has a better accuracy. However, the Cryotherapy dataset value accuracy based on the weight gain ratio more 0.5% lower than the weight of the dataset due to the value in the spread dataset.

1. Introduction
Simple Additive Weighting (SAW) developed by Mac Crimon in 1968, known as a weighted linear combination, assessment method, or weighted summation [14]. SAW uses the principle of weighted average. The Simple Additive Weighting (SAW) method is also call the weighted addition method and is the simplest and widest method used [7]. In research [14] producing Simple Additive Weighting (SAW) cannot provide controlled consistency because they do not have a comparative index as an indicator. Research [15] analyzes several MCDM methods and determines their application in certain situations by evaluating the relative advantages and disadvantages of each method. As a result, the loss of the Simple Additive Weighting (SAW) method on the resulting estimate does not necessarily reflect the true situation, the result may not be logical with the value of one particular attribute that is very different from the others. Some researches compare and analytics two MCDM methods[1]. The result of the SAW method is a good alternative to Analytical Hierarchical Processing (AHP), but if many hierarchical levels with sub attributes then SAW does not qualify for aggregate rank values. Research [12] the results were analyzed using the standard deviation relative from the rank value, the result for SAW reached 12.64% compared while and WPM reached 35.75%. Research [3] results show that evaluation of product ranking results of TOPSIS, SAW and
WPM method is more accurate as it supports alternative products because it uses entropy to evaluate the weight of each attribute. Research [9] the results showed that TOPSIS accuracy reached 50% of at least three matches in order of twenty participants, followed by PROMETHEE II (40%), ELECTRE II (30%), TODIM (25%) and SAW (25%). TOPSIS correctly predicted 15 times the first alternative rank which means 79% accuracy and SAW method 74%. Research [2] results showed that the AHP method (42.42%) was most widely used because the structure was very easy, flexible and easy to understand, the second method TOPSIS (13.33%) and the next method SAW (12.73%). Based on several studies, SAW method were established based on policy analysis, making it easier and less time consuming to calculate than other multi-attribute methods. However, SAW is a method that does not have a clear validation in its weighting system, it can even produce a less accurate [7]. The Simple Additive Weighting (SAW) method that has been combined in the study [6] produces more accurate accuracy and in research [8] modified SAW to obtain a better alternative, Simple Additive Weighting (SAW) can work more optimally with the help of other methods. In research [11] gain ratio is a modification of the information gain that reduces its bias. Therefore, in this research will utilize gain ratio in giving weight of attribute before making process decision. It is hoped this will improve the accuracy of Simple Additive Weighting (SAW).

2. Multi Attribute Decision Making (MADM)
MADM is the method used to find the most optimal alternative of a number of alternatives that exist based on certain attributes. MADM sets the weight value for each attribute, then performs a ranking process to select a number of alternatives. The value of an attribute describes alternate characteristics, quality, and performance. Weight attribute serves to measure the weight value of importance of each attribute. In looking for the weight value of each attribute there are 3 categories of approaches namely, subjective approach, objective approach and hybrid approach [4] [13].

3. Simple Additive Weighting (SAW)
Simple Additive Weighting (SAW) based on a weighted average. The preference value is calculated for each alternative of the attribute, by multiplying the given scale value to an alternative of the attribute with the relative importance weight value provided by the decision maker and summing the product for all criteria [12]. SAW is able to balance between each criterion or attribute, intuitive in making decisions, and simple calculation process so it does not need to use complicated computer programs [15]. In the Simple Additive Weighting Method (SAW) there are two attributes: attribute of benefits or benefits and cost attributes. Both attributes have a fundamental difference in the selection of attributes when making decisions [4]. The preference value of each alternative calculated as follows [13]:

$$V_i = \sum_{j=1}^{n} W_j R_{ij}$$  \hspace{1cm} (1)

Where, \(V_i\) is the preference value for each alternative, \(W_j\) is the weight value of each attribute and \(R_{ij}\) is normalized work rating value

4. Gain Ratio
The C4.5 algorithm is a decision tree method in the selection of its attributes based on the gain ratio. Gain ratio is a modification of the information gain that reduces its bias. Gain Ratio by taking intrinsic information from each attribute can improve information gain [11].
Entropy dataset is the number of values needed to declare an attribute. The more information gain is obtained, the stronger the attribute. Thus, the partition will occur in the attribute that has the highest information gain [5]. The gain ratio found in the C4.5 algorithm, used to calculate the attribute's influence on the target of a data [11]. Gain ratio is the development of information gain.

5. Relative Standard Deviation (RSD)
Relative Standard Deviation (RSD) used to compare how well an accuracy value is. The relative standard deviation expressed in percent and obtained by multiplying the standard deviation (σ) by 100 and dividing the dataset by mean [12]. Standard deviation is a measure of how exactly the average is, that is how well individual numbers fit into each other. Standard deviation is also an index that represents deployment in a set of values. If the values grouped close to the average, the standard deviation will be small, and the standard deviation will be large [12].

6. Methodology
The method used in this study consists of several stages, namely the pre-processing of data by determining the attributes, performing weight calculations using the gain ratio and decision making using Simple Additive Weighting (SAW) method then calculate the relative standard deviation of each dataset based on the weighting of the gain ratio and the weight of the dataset.

6.1 Pre-processing Data
Pre-processing stages of data are the exact data selection stage to use. This study using 4 datasets derived from UCI machine learning repository. The dataset is as follows 90 cryotherapy patient data, 90 immunotherapy patient data, 583 Indian liver patient dataset (ILPD) data and 403 user knowledge modelling data.

6.1.1 Types of attributes.
The attributes used in the decision-making process and weighting are as follows cryotherapy dataset with 5 attributes as follows: sex, age, time, number of warts, type, area and result of treatment with ratings 0 and 1. Dataset immunotherapy with 6 attributes as follows: sex, age, time, number of warts, type, area, induration diameter and result of treatment with ratings 0 and 1. Dataset of Indian liver patient dataset (ILPD) with 10 attributes as follows: age, gender, TB, DB, the alkphos, sgpt, sgot, TP, ALB, A / G and the results are selectors with ratings 1 and 2. And the last data user knowledge modelling with 5 attributes as follows: STG, SCG, STR, LPR, PEG and UNS results with very low valuation, low, middle and high.

6.2 Data Processing
The specified dataset will processed by several stages. The weighting phase of each attribute based on the weight of the dataset, the weight gain ratio, the calculation phase of the Simple Additive Weighting (SAW) method to generate the preference value of each attribute and then calculate the value of the relative standard deviation dataset based on the preference values of the alternatives.

6.2.1 Weighted Attributes.
Determining the weight of the attribute used the weight of the dataset and the weight gain ratio. The weights of the dataset are deriving from a predefined dataset source. In determining the weight Gain Ratio there are several processes, are as follows [10]:

3
Step 1: Calculates the entropy value of all data based on the sum of the entropy value of the result of the dataset

\[
\text{Entropy } (S) = \sum_{i=1}^{n} - p_i \times \log_2 p_i
\]  

Where, \( S \) is the set of cases, \( n \) is number of partitions \( S \) and \( p_i \) is the proportion of \( S_i \) to \( S \).

Step 2: Followed by calculating the Information Gain as one of the selection attributes used to select the attribute test of each alternative to the dataset. The Information Gain of each attribute could be calculated based on the reduction of the overall entropy value of the selector by the amount of data value of each alternative multiplied by the entropy value of each result.

\[
\text{Information Gain } (S,A) = \text{Entropy}(S) - \sum_{i=1}^{n} \left| S_i \right| \times \text{Entropy}(S_i)
\]  

Where, \( S \) is all dataset, \( A \) is subset attribute, \( n \) is Number of Partition Attributes \( A \), \( \left| S_i \right| \) is Size of a subset of the dataset that belongs to the attribute on \( A \) partition \(-I \) and \( \left| S \right| \) is Number of Case Size in Dataset.

Step 3: Then, Split Info of each attribute is calculated based on the number of values of each alternate one attribute.

\[
\text{Split Info }_A(D) = -\sum_{j=1}^{V} \frac{\left| D_j \right|}{\left| D \right|} \times \log_2 \left( \frac{\left| D_j \right|}{\left| D \right|} \right)
\]  

Where, \( D \) is all dataset, \( A \) is subset attribute, \( V \) is number of partition attributes \( A \), \( \left| D_j \right| \) is size of a subset of the dataset belonging to the attribute on \( A \) partition \(-j \) and \( \left| D \right| \) is number of case size in dataset.

Step 4: And then calculated the value of Gain Ratio for each attribute by dividing the value of Information Gain with Split Info.

\[
\text{Gain Ratio } (A) = \frac{\text{Gain } (A)}{\text{Split Info}_A(A)}
\]  

Weighted attribute based on the weight of the dataset and the weight Gain Ratio is used in the process of calculating the Simple Additive Weighting (SAW) method to determine the value of the preference value of the two weightings of different attributes.

6.2.2 Calculation of Simple Additive Weighting (SAW). In the process of calculating the Simple Additive Weighting (SAW) method, the weight of the dataset and the weight Gain Ratio are calculate using the Simple Additive Weighting (SAW) method, thus generating the preference value of each alternative dataset. The process of calculating the Simple Additive Weighting (SAW) method uses attribute weighting based on the weight of the dataset and the weight Gain Ratio is as follows [13].

Step 1: Determine the decision Z size \( m \times n \), where \( m \) = alternative \( (Ai) \) is selected and \( n \) = criterion or attribute \( (Cj) \).

Step 2: Gives the value of \( x \) each alternative \( (i) \) on any specified criterion or attribute \( (j) \), where \( i = 1,2,...,m \) and \( j = 1,2,...,n \) on the decision matrix Z.

Step 3: Input weight preference value \( W \) based on the weight of the dataset and the weight Gain Ratio for each attribute.

Step 4: Then, normalize the Z decision matrix according to the type of criterion or attribute (advantage and cost) to obtain a normalized matrix \( (R) \).

If \( (j) \) is an advantage:

\[
R_{ij} = \frac{x_{ij}}{\max x_{ij}}
\]  

If \( (j) \) is cost:

\[
R_{ij} = \frac{\min x_{ij}}{x_{ij}}
\]
Step 5: The result of a normalized work rating value \( (rij) \) forms a normalized matrix \( (R) \).

Step 6: And then, multiply the normalized matrix \( (R) \) by the preference weight value \( (W) \) based on the weight of the dataset and the weight Gain Ratio. So it generates the preference value of each alternative.

Step 7: The preference value for each alternative \( (Vi) \) generated can determine the best alternative and accuracy of the Simple Additive Weighting (SAW) method using equation (1).

### 6.2.3 Calculate Relative Standard Deviation.

The result of preference values of all alternatives based on the weight of the dataset and the weight Gain Ratio will determine the accuracy of the SAW method. In determining the accuracy, the root result of the sum of squares of the preferences value of all alternatives minus the sum of preference values of all alternatives divided by all alternatives and divided by the number of alternatives in the dataset minus 1, will result in the standard deviation value \( [12] \).

\[
\sqrt{\frac{\sum x^2 - (\sum x)^2}{n-1}}
\]  
(8)

Where, \( S \) is standard deviation, \( \sum x^2 \) is the sum of the squares of individual measurements, \( \sum x \) is number of individual measurements and \( n \) is number of samples of data analyzed.

The standard deviation value multiplied by 100 and divided by the mean of the total preference value of all alternatives will result in the value of relative standard deviation.

\[
RSD = \frac{S}{\bar{x}} \times 100
\]  
(9)

Where, RSD is relative standard deviation, \( S \) is standard deviation and \( \bar{x} \) is the mean of the data.

### 7. Result and Discussion

In this chapter will discuss the results of testing the methods of Simple Additive Weighting (SAW) obtained such as attribute weighting, the value of each alternative preference and the value of performance accuracy.

#### 7.1 Attribute Weighting

Weights based on the weight of the dataset are determined using the following 3 categories of approaches, subjective approaches, objective approaches and hybrid approaches. The weights on the 4 tested datasets are determined from the source dataset. When the weighting based on the Gain Ratio was determined based on the amount of data and data groups corresponding to each other, the data value of each attribute of all alternatives has a great influence in the determination of the weights because the gain ratio determines the weight by reducing the bias of the data. Weights for 4 datasets that have been tested according to the number of attributes can be seen in table 1.

| Dataset          | Weight by Dataset | Weight Gain Ratio |
|------------------|-------------------|-------------------|
| Cryotherapy      | 0.03; 1; 0.917; 0.107; 0.841; 0.653 | 0.005; 0280; 0.416; 0.042; 0.223; 0.178 |
| Immunotherapy    | 0.052; 0.596; 1; 0.032; 0.579; 0.462; 0.769 | 2.568; 0.286; 0.370; 0.050; 0.008; 0.080; 0.097 |
| ILPD             | 7; 6; 1; 2; 10; 5; 4; 8; 3; 9 | 0.881; 0; 0.486; 0.496; 0.271; 1; 0.743; 0.249; 0.218; 0.240 |
| User Knowledge   | 0.19251; 0.03063; 0.13285; 0.22119; | 0.102; 0; 0.194; 0.194; 1 |
| Modeling         | 0.84544 | |
7.2 Alternative Preference Value
The result of testing the Simple Additive Weighting (SAW) method by using weighting based on the weight of the dataset and the weight gain ratio is the preferred value of each alternative dataset. The preference value determines the best alternative of the dataset. The best alternative of 4 datasets can be seen in table 2.

| Dataset             | Alternative | Preference value       |
|---------------------|-------------|------------------------|
|                     |             | Weight by Dataset | Weight Gain Ratio |
| Cryotherapy         | 64          | 3.0256                | 0.9992            |
| Immunotherapy       | 59          | 2.3972                | 0.7429            |
| ILPD                | 136         | 32.9878               | 2.6906            |
| User Knowledge Modeling | 242     | 1.2426                | 1.3460            |

7.3 Value of method performance accuracy
The result of performance accuracy of Simple Additive Weighting (SAW) method based on the weight of the dataset and the weight gain ratio is generated from the preference value of the overall alternative data calculated using equation (8) and (9) in decision making and determining the best alternative. The result of performance accuracy can be seen in table 3.

| Dataset             | Accuracy     |
|---------------------|--------------|
|                     | Weight by Dataset | Weight Gain Ratio |
| Cryotherapy         | 34.54 %       | 34.05 %          |
| Immunotherapy       | 23.77 %       | 28.38 %          |
| ILPD                | 17.5 %        | 24.79 %          |
| User Knowledge Modeling | 36.92 %   | 39.57 %          |
| Average             | 28.1825 %     | 31.6975 %        |

Based on table 2 test results on the 4 data sets, the average accuracy value obtained based on two different attribute weights that is, weighting attribute based on the weight of data set of 28.1825% and weighting attribute based on the weight gain ratio of 31.6975%. Then it can be see that the accuracy on attribute weighting based on gain ratio has better accuracy. However, in the cryotherapy data set the attribute weighted accuracy value based on the lower gain ratio is 0.5% compared to the weighting of attributes based on the weight of the data set because the values in the data sets are scattered thus affecting the results of the alternative preference values.

8. Conclusion
Based on the results of research and testing of 4 datasets that the average accuracy of the four data sets increased by 3.5150%. Then it can be see the accuracy of attribute weighting based on Gain Ratio has better accuracy. The higher the weight of attribute to eat the higher the level of importance of the attribute and have a considerable influence in decision making and determine the best alternative. However, if the weights between attributes are much different then it affects the preference value of each alternative and the performance of the Simple Additive Weighting SAW method.
References

[1] Annette, J. R., Banu, A. & Chandran, P. S. 2015. Comparison of Multi Criteria Decision Making Algorithms for Ranking Cloud Renderfarm Services. *Indian Journal of Science and Technology* 9(31): 1-5.

[2] Brito, M. M. de & Evers, M. 2016. Multi-criteria decision-making for flood risk management: a survey of the current state of the art. *Natural Hazards and Earth System Sciences* 16: 1019-1033.

[3] Dashore, K., Pawar, S. S., Sohani, N. & Verma, D. S. 2013. Product Evaluation Using Entropy and Multi Criteria Decision Making Methods. *International Journal of Engineering Trends and Technology (IJET)* 4(5): 2183-2187.

[4] Ding, T., Liang L., Yang, M. & Wu, H. 2016. Multiple Attribute Decision Making Based on Cross-Evaluation with Uncertain Decision Parameters. *Hindawi Publishing Corporation Mathematical Problems in Engineering* 1-10.

[5] Dujena, A., Puyalnithi, T. 2017. Enhancing Classification Accuracy of K-Nearest Neighbours Algorithm Using Gain Ratio. *International Research Journal of Engineering and Technology (IRJET)* 4(9): 1385-1388.

[6] Dogan, M., Aktar, T., Toker, O. S. & Tatlisu, N. B. 2015. Combination of the Simple Additive (SAW) Approach and Mixture Design to Determine Optimum Cocoa Combination of the Hot Chocolate Beverage. *International Journal of Food Properties* 18: 1677-1692.

[7] Kittur, J. 2015. Optimal Generation Evaluation using SAW, WP, AHP and PROMETHEE Multi - Criteria Decision Making Techniques. *IEEE International Conference on Technological Advancements in Power & Energy* 15: 304-309.

[8] Korsemov, D. & Borissova, D. 2018. Modifications of Simple Additive Weighting and Weighted Product Models for Group Decision Making. *AMO – Advanced Modeling and Optimization* 20(1): 101-112.

[9] Leoneti, A. B. 2016. Considerations Regarding the Choice of Ranking Multiple Criteria Decision Making Methods. *Sociedade Brasileira de Pesquisa Operacional* 36(2): 259-277.

[10] Mitchell, T. M. 1997. *Machine Learning*. McGraw-Hill Science: New York.

[11] Priyadarsini, R. P., Valarmathia, M. L., & Sivakumari, S. 2011. Gain Ratio Based Feature Selection Method for Privacy Preservation. *Ictact Journal On Soft Computing* 1(4): 202-205.

[12] Savitha, K. & C. Chandrasekar, Dr. 2011. Vertical Handover decision schemes using SAW and WPM for Network selection in Heterogeneous Wireless Networks. *Global Journal of Computer Science and Technology* 11(9): 18–24.

[13] Sembiring, A. P., Tulus, Sembiring, W. R., & Maulana, H. 2018. Rule Model With Fuzzy Simple Additive Weighting Approach And Weighted Product On Determination Of Position In High Education Institution. *International Journal of Latest Trends in Engineering and Technology* 10(1): 054-062.

[14] Thor, J., Ding, S-H. & Kamaruddin, S. 2013. Comparison of Multi Criteria Decision Making Methods From The Maintenance Alternative Selection Perspective. *The International Journal Of Engineering And Science (IJES)* 2(6): 27-34.

[15] Velasquez, M. & Hester, P. T. 2013. An Analysis of Multi-Criteria Decision Making Methods. *International Journal of Operations Research* 10(2): 56-66.