IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 23, Issue 6, Ser. III (Nov. –Dec. 2021), PP 41-48
www.iosrjournals.org

Optimization of Simple Additive Weighting Method Based on Information Gain in Decision Support System

Heri Ardiyanto[1], M. Arief Soeleman[2]
[1]*Faculty of Computer Science, Dian Nuswantoro University, Indonesia*
[2]*Faculty of Computer Science, Dian Nuswantoro University, Indonesia*

Abstract:
Decision support systems are an alternative offered in the global era to ease tasks and work by converting the roleof humans in making decisions to machines. However, humans still determine the parameters or criteria as an indicator of decision-making considerations. Feature selection is used to determine the best features in the dataset to be used in the decision-making process. This study proposes the use of feature selection in the decision-making process based on feature ranking using information gain and simple additive weighting (SAW) used in the decision- making process. Private data collection is carried out through instruments containing questions that represent the condition of students to produce objective information. The target of filling out this instrument includes all statehigh school and vocational high school students in the city of Salatiga, while the UCI repository is used as a reference for public resources. Between public data and private data which is influenced by cultural factors belonging to a country. Private data is generated by adjusting the culture and customs that exist in Indonesia. Decision making using the simple additive weighting (SAW) method is considered effective in producing decisions because the resulting decisions are more objective based on ranking data.
***Key Word**: simple additive weighting (SAW), information gain, ranking*

Date of Submission: 10-12-2021 Date of Acceptance: 24-12-2021

I. Introduction

The development of information technology is transforming so rapidly and affects the decision-making system mechanism. Mechanisms that originally relied on human abilities shifted to computer-based mechanisms or ways of working. Decision support systems are an alternative offered inthe global era to ease tasks and work by converting the role of humans in making decisions to machines. However, not all of the considerations are left to the computer, while humans still determine the parameters or criteria as indicators of decision- making considerations. It is revealed in [1] that in theera of big data, it is possible to use data mining as an effective alternative in doing data modeling usingalgorithms. However, the reality shows that researchers use various algorithms according to their wishes. This becomes an obstacle for beginners in determining the right algorithm for data modeling. Soit is proposed to use a decision support system as a tool in determining the most appropriate algorithm to build the model. Decision support systems play a rolein helping provide objective decisions.

In terms of performance measurement, it is necessaryto do it objectively with reliable measuring tools. Various kinds of mathematical methods are offered to get optimal results. One of the tools used is a decision support system. According to [2], a decision support system is an interactive application system by combining data and mathematical models to assist management in making a complex decisions. In the world of education, the decision support system plays an important role in assisting the implementation of academic activities. In addition, the decision support system helps in determining the performance of the academic community more objectively and regardlessof the subjective views of decision makers if it is done manually. The performance of students needs tobe measured to find out the extent of their competence. This can be taken into consideration in determining the award of scholarships or determining other policies that involve the competence of students. What is happening now is that the processof determining and considering is done manually, butit is still being intervened by the subjective views of decision makers. So it is necessary to have an objective tool as the sole decision maker by using mathematical calculations based on predetermined methods. In Portugal, [3] has conducted research on student performance. There is a downward trend in the level of education in Portugal, then a study was conducted using datasets from UCI repository to perform modeling and then the comparison of theeffectiveness of each method used was known.

This study proposes the use of feature selection in the decision-making process based on feature ranking. Feature selection is used to determine the best features in the dataset to be used in the decision-making process. Simple additive weighting (SAW) is used in the decision-making process based on information gain for the feature selection process.The simple additive weighting (SAW) method is also known as the sum method [4]. The ranking method was chosen because of its effectiveness in providing calculation results. The data is clearly presented starting from the highest rank which is assumed to be a weighted feature and has a 100% chance to be used as a decision-making feature to the lowest rank.

The systematics of this paper is divided into (1) an introduction which contains a descriptive description of the problems that occur and a description of the expected results, (2) related research conveys several studies that are relevant to the problems discussed in this paper, (3) the method contains methods or methods. which is used in conducting this research which is divided into two methods, namely data collection methods and problem solving methods, (4) results and discussion explores problem solving steps based on predetermined methods to obtain decision results, (5) closing consists of conclusions from research and suggestions as input or gaps for further research, as well as (6) references containing a list of literature used as a basis for conducting research and writing this paper.

II. Related Research

Feature selection has been used in several studies related to data management. It was mentioned in [5] that high data duplication or redundancy can lead to a non-optimal selection process, so an efficient framework model is proposed for feature selection using an unsupervised method. This is in line with that conveyed by [6], where high data dimensions in some machine learning applications require complex computational analysis. It is necessary to decide that the right feature is chosen as a decision-making indicator, so deep feature selection is proposed in this case the relationship between teacher and student. In [7], the ensemble classification method is considered to have better performance than the use of a single classifier method. In this study, a formal comparison of different ensemble methods was conducted in the feature selection domain. This comparison involves five machine learning techniques, namely logistic regression, support vector machine (SVM), extreme learning machine, naive bayes and decision tree. Most feature selection is applied in research to reduce high data dimensionality and reduce duplication or data redundancy, such as the application of feature selection in [8]. Feature selection for text grouping has been carried out by [9] in Improved information gain. Similarly, [10] combines information gain with feature selection for web filtering. The implementation of feature selection in statistical calculations was proposed through research [11]. The application of feature selection was developed to analyze the development of big data, as in the research described in [12] which carried out an online feature selection technique to handle classification in big data. Several studies were conducted to seek novelty regarding decision support systems. Some of them are combined with data mining modeling. [13] suggested that the decision support system is a combination of information systems and decision- making technology. Visual interaction between humans and computers in decision making is a key decision support system technique. His research proposes a decision support system that can be applied in industry. In the study [14] identified a decision model that supports decisions in choosing an information technology system in a company. The study was conducted based on the literature with reference to five basic criteria that influence the selection of an information technology system by an entrepreneur. [15] analyzed the development of spatially based decision support systems over a three- decade time period. It is an attraction and a challenge to conduct research related to decision support systems. The future development of decision support systems is an important point in terms of data mining management. In a study conducted by [16], several decision-making methods were carried out to determine the performance of each method. In this study, the SAW, TOPSIS, GRA methods were used and implemented in Multiple criteria decision- making (MCDM). MCDM is implemented by [17] in a decision support system in providing support tools to tourists in determining hotel choices. The method used in this study using TOPSIS. Comparison between SAW and TOPSIS in the analysis through research conducted [18]. SAW uses several criteria to do calculations using the weights of each criterion with the results in the form of a ranking with advantages that are simple and easy to understand, while TOPSIS uses positive and negative value indicators which indicate the best solution marked positive and negative to indicate discrepancies. The development of a decision support system is very important. [19] develop a decision support system for emergencies.

III. Proposed Method

The proposed method of this paper is divided into two, namely data collection methods and problem solving methods. This study uses a data set or dataset which is divided into private data and public data. The UCI repository is used as a public resource reference. In addition, this research involves partners, in this case students, to generate private data through filling out questionnaires. The data used is about student data where in this study measures the performance of each student in order to determine the achievement rating of students. There are differences in data features between public data and private data used in this study, where these differences are influenced by cultural factors belonging to a country. Private data is generated by adjusting the culture and customs that exist in Indonesia.

In collecting private data, it is done by distributing instruments containing questions that represent the condition of students. Instruments are arranged in as much detail as possible to produce objective information. The target of filling out this instrument includes all state high school students in the city of Salatiga. The information contained in the data shows indicators of student performance. There are differences in the features contained in private and public datasets. This is influenced by differences in culture and habits of the dataset producing country. The differences in the features of private and public datasets are shown in table 1.

Table no 1: Differences between private and public dataset features

No	Field	
	Private	**Public**
1	Name	School – student's school
2	School	Sex – student's sex
3	Sex	Age – student'sage
4	Address	Address – student's home address type
5	Age	Famsize – family size
6	Family_number	Pstatus – parent's cohabitation status
7	residence	Medu – mother's education
8	Father_education	Fedu – father's education
9	Mother_education	Mjob – mother's job
10	Father_job	Fjob – father's job
11	Mother_job	Reason – reason tochoose this school
12	choosin_private_school	Guardian – student's guardian
13	Travel_time_to_school	Traveltime – home to school travel time
14	Study_time	Studytime – weekly study time
15	Stay_class	Failures – number of past class failures
16	Family_education_support	Schoolsup – extra educational support
17	Extra_lesson	Famsup – family educational support
18	extracurricular	Paid – extra paid classes within the course subject
19	Capacity_building_enthusiasm	Activities – extra-curricular activities
20	Higher_education_motivation	Nursery – attended nursery school
21	Internet_facility_at_home	Higher – wants to take highereducation
22	Learning_resource_facility	Internet – Internet access at home
23	Love_relationship	Romantic – with aromantic relationship
24	Family_relationship	Famrel – quality offamily relationships
25	Free_time	Freetime – free time after school
26	Hang_out_time	Goout – going out with friends
27	Healthy	Dalc – work dayalcohol consumption
28	Absence	Walc – weekend alcohol consumption
29	Worship_intensity	Health – current health status
30	-	Absences – number of school absences
31	-	G1 – first period grade
32	-	G2 – second period grade
33	-	G3 – final grade

In solving the existing problems, a mathematical method is needed so that the data obtained from the calculation results are quantitative. This study proposes the Simple Additive Weighting (SAW) method as a decision-making method by combining the feature selection process using information gain and entropy calculation. The ranking method was chosen because it is considered more effective in determining quality based on ranking. Diagram 1 shows the flow of the data processing process from the start of data collection until the data produces useful information and knowledge for users. Supporting devices for data processing using tools with the following specifications:

a. Software: Microsoft Excel 2010, XAMP server, Google Form.
b. Hardware: Laptop Lenovo S10-3, Intel Atom 1,66 GHz, RAM 2 GB, 64-bit OS Windows 7 Ultimate SP 1, HD SATA 300 GB

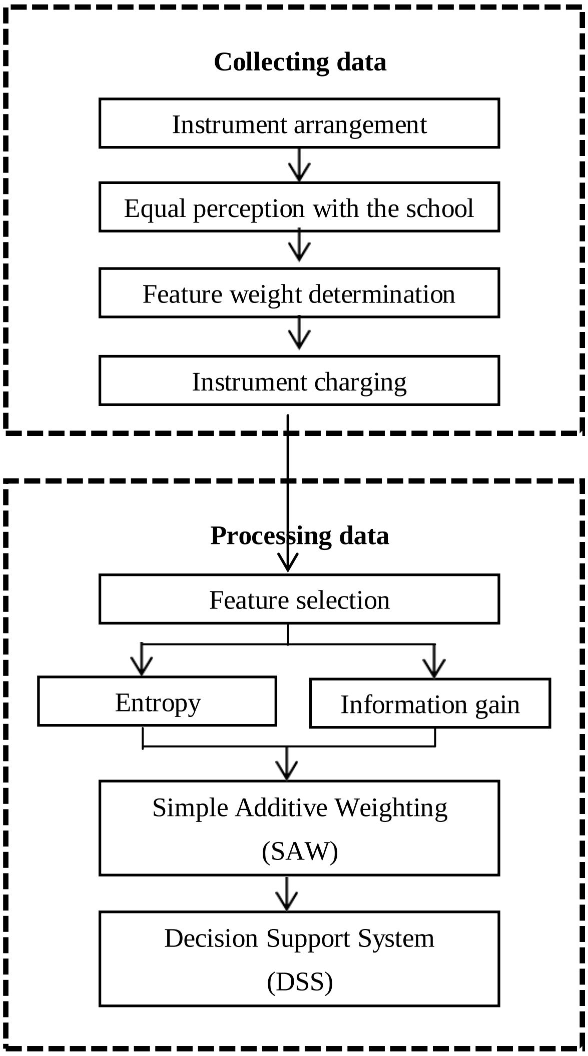

Diagram 1 Data processing flow

IV. Results and Discussion

Feature selection is used to simplify features so that the processed data is of high quality and provides optimal results. Quoted in [20], the main function of feature selection is to select the features that are most relevant to the classification problem. The public dataset of student performance used in this study before feature selection has a total of 33 features and it is necessary to carry out a preprocessing stage with feature selection to produce quality features. The feature selection process using information gain is shown in equation (ii) with entropy calculation shown in equation (i).

$$Entropy(S) = \sum_{i=1}^{n} -pi * log_2\, pi \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (i)$$

Description:
S = Case set
n = Number of partitions S
P_i = Proportion S_i to S

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} - \frac{|S_i|}{|S|} * Entropy(S_i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (ii)$$

Description:
S = Case set
A = Feature
n = Number of attribute partitions A
$|S_i|$ *= Proportion* S_i *to S*
$|S|$ *= Number of case in S*

The calculation in equation (i) will produce an entropy value of 0.998661155289733, while the formula in equation (ii) is used to obtain the most optimal feature based on the weight of each feature that has been determined, so that the resulting data is shown in table 2.

Table 2 Information gain of students performance dataset

No	Features	Weight
1.	Mjob	0,597501148
2.	absences	0,051268248
3.	G1	0,02035119
4.	G2	0,018395952
5.	G3	0,015893431
6.	Fedu	0,007193007
7.	age	0,006731174
8.	Fjob	0,006683177
9.	sex	0,004526564
10.	reason	0,003495889
11.	famrel	0,003214155
12.	Dalc	0,002175279
13.	studytime	0,002060688
14.	Medu	0,001903606
15.	famsup	0,001710373
16.	goout	0,001502321
17.	Walc	0,001060899
18.	failures	0,000942961
19.	traveltime	0,000911323
20.	health	0,000855747
21.	freetime	0,000789435
22.	higher	0,000369799
23.	schoolsup	0,000310738
24.	activities	0,000127759
25.	internet	0,000101565
26.	guardian	8,80082E-05
27.	school	7,79496E-05
28.	address	7,75196E-05
29.	nursery	6,00963E-05
30.	paid	5,79397E-05
31.	Pstatus	4,79299E-05
32.	famsize	7,56744E-06

33.	romantic	8,13317E-07

Table 2 shows the weights of each feature of the student performance dataset sorted from the highest to the lowest weight. Based on table 2, the mother's job (job) has the highest weight, which means that the mother's job has a big influence on student performance in learning and is followed by the number of student attendance (absence) which is the next influencing factor.

The results obtained from feature selection using information gain are features with the highest to lowest weights. From these features, the most optimal feature is chosen, namely the feature that has the highest weight. In table 3, the most optimal features have been determined which will later be used for the decision-making process.

Table 3 Optimal features from feature selection

No	Features	Description
1	Mjob	Mother job
2	G1	Grades on the first level
3	G2	Grades on the second level
4	G3	Grades on the third level
5	Fedu	Father education
6	age	Age
7	Fjob	Father job
8	famrel	Family relationship
9	studytime	Study time
10	sex	Sex
11	reason	Reason for choosing school

The next step is the decision-making process. In this study using the Simple Additive Weighting (SAW) method. The SAW method is a method used in the decision-making process based on the ranking and weighting of each feature used. In the calculation process using the SAW method, there are several stages that are correlated with each other. The data to be used for the ranking process needs to be normalized first, as shown in equation (iii). Normalization is done to minimize the occurrence of data redundancy and ensure the data is in the right table with values that comply with data processing standards. In addition, normalization is carried out to handle the occurrence of anomalies or data deviations and data inconsistencies.

$$r_{ij} = \begin{cases} \frac{x_{ij}}{Max\ i\ x_{ij}} & if\ j\ is\ a\ benefit \\ \frac{Min\ i\ x_{ij}}{x_{ij}} & if\ \ i\ is\ a\ cost \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(iii)$$

Description:
r = Data normalization
xij = Case set
Maxi xij = Highest value
Mini xij = Lowest value

Benefit is a value or criterion that is beneficial in the calculation process, while cost is the opposite. This means that the data is considered a benefit if the value is higher, it will have a positive impact, while the data is considered a cost if the value is higher, it will have a negative or bad impact. Based on the features shown in table 3, Job, G1, G2, G3, Fedu, age, Fjob, Famrel, studytime are classified into benefits. While absences, sex, reason are classified into costs. Each data is divided by the largest value of the total data on each feature for benefits. Cost is calculated by dividing the smallest value of each feature by all data on each feature. The ranking stage is the core of the SAW method, in which the data that has been normalized is carried out a ranking process. The calculation uses the formula shown in equation (iv).

$$V_i = \sum_{j=1}^{n} w_j\ r_{ij} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(iv)$$

Description:
Vi = Ranking
wi = Weight
rij = Data normalization

The normalized data is multiplied by the weight of each feature, so that a number is obtained which will be used for the ranking process. Table 4 shows the sample data from the calculation process using the SAW method. The data is sorted by rank from highest to lowest value. The resulting value represents the performance of students.

Table 4 Rangking using the Simple Additive Weghting (SAW)

Mjob	G1	G2	G3	Fedu	age	Fjob	famrel	studytime	sex	reason	SAW
0.05	0.13	0.12	0.17	0.06	0.04	0.05	0.08	0.14	0.08	0.20	**1.11**
0.06	0.13	0.13	0.18	0.04	0.04	0.06	0.08	0.14	0.08	0.15	**1.11**
0.06	0.15	0.15	0.20	0.06	0.04	0.05	0.06	0.14	0.04	0.15	**1.10**
0.05	0.14	0.14	0.18	0.08	0.04	0.08	0.06	0.04	0.08	0.20	**1.09**
0.08	0.12	0.13	0.18	0.06	0.03	0.05	0.08	0.11	0.04	0.20	**1.07**
0.08	0.10	0.10	0.14	0.08	0.04	0.05	0.08	0.11	0.08	0.20	**1.05**
0.02	0.15	0.14	0.19	0.08	0.04	0.02	0.08	0.11	0.08	0.15	**1.05**
0.05	0.13	0.14	0.18	0.06	0.04	0.02	0.08	0.11	0.04	0.20	**1.03**
0.05	0.13	0.13	0.17	0.08	0.04	0.05	0.08	0.07	0.08	0.15	**1.03**
0.05	0.13	0.14	0.18	0.06	0.04	0.02	0.05	0.07	0.08	0.20	**1.02**
0.05	0.13	0.11	0.15	0.04	0.03	0.05	0.08	0.14	0.04	0.20	**1.02**
0.05	0.13	0.12	0.16	0.08	0.04	0.08	0.08	0.04	0.04	0.20	**1.01**
0.08	0.14	0.15	0.19	0.08	0.03	0.08	0.08	0.04	0.04	0.10	**1.01**
0.08	0.12	0.11	0.14	0.08	0.04	0.08	0.05	0.07	0.04	0.20	**1.01**
0.05	0.14	0.14	0.18	0.02	0.04	0.02	0.05	0.14	0.08	0.15	**1.00**
0.06	0.13	0.13	0.16	0.06	0.04	0.02	0.06	0.07	0.08	0.20	**1.00**
0.06	0.12	0.11	0.15	0.04	0.03	0.05	0.05	0.11	0.08	0.20	**1.00**
0.06	0.12	0.12	0.15	0.06	0.04	0.05	0.06	0.11	0.08	0.15	**1.00**
0.08	0.12	0.10	0.15	0.08	0.04	0.08	0.08	0.07	0.04	0.15	**0.99**
0.06	0.11	0.12	0.16	0.04	0.04	0.05	0.06	0.07	0.08	0.20	**0.99**

Ranking is useful for decision making in terms of finding the best from a set of objects. An example of its application is in determining the award of scholarships and selecting outstanding student ambassadors. The attributes possessed by the candidate determine the decisions taken. The objectivity and validity of the data at the data collection stage greatly affect the value generated in the ranking process.

V. Results and Discussion

Decision support systems play an important role in assisting the process of making a decision where with the existence of a systematic decision-making mechanism it can produce decisions that are more objective and easy to make considerations. The feature selection provides a better alternative in determining the optimal features to be used as data to support decision making. The ranking-based decision-making method is considered effective in carrying out the decision-making process. In terms of ranking, this research uses information gain for the feature selection process and simple additive weighting (SAW) for the decision support system method.

The cultural differences of a country determine the type of data used for the decision-making process. So in this study two types of data were used, namely public data sourced from other countries and private data collected through filling out instruments for SMA N 2 students in the city of Salatiga.

VI. Suggestion

The application of the ranking-based Simple Additive Weighting (SAW) method is very effective as a decision-making method. Furthermore, this method can be developed for measuring the performance of teachers or instructors as a basis for consideration in providing performance rewards or job competency assessments.

References

[1] T. Man, N. A. Zhukova, A. M. Thaw, and S. A. Abbas, "A decision support system for DM algorithm selection based on module extraction,"*Procedia Comput. Sci.*, vol. 186, pp. 529–537, 2021, doi: 10.1016/j.procs.2021.04.173.

[2] D. Galar and U. Kumar, "Maintenance Decision Support Systems," in *eMaintenance*, Elsevier, 2017, pp. 371–474. doi: 10.1016/B978-0-12- 811153-6.00007-5.

[3] P. Cortez and A. Silva, "USING DATA MINING TO PREDICT SECONDARY SCHOOL STUDENT PERFORMANCE," p. 9.

[4] S.-Y. Chou, Y.-H. Chang, and C.-Y. Shen, "A fuzzy simple additive weighting system under group decision-making for facility location selection with objective/subjective attributes," *Eur. J. Oper. Res.*, vol. 189, no. 1, pp. 132–145, Aug. 2008, doi: 10.1016/j.ejor.2007.05.006.

[5] H. Zhang, R. Zhang, F. Nie, and X. Li, "An efficient framework for unsupervised feature selection," *Neurocomputing*, vol. 366, pp. 194– 207, Nov. 2019, doi:10.1016/j.neucom.2019.07.020.

[6] A. Mirzaei, V. Pourahmadi, M. Soltani, and H. Sheikhzadeh, "Deep feature selection using a teacher-student network," *Neurocomputing*, vol. 383, pp. 396–408, Mar. 2020, doi: 10.1016/j.neucom.2019.12.017.

[7] H. E. Kiziloz, "Classifier ensemble methods in feature selection," *Neurocomputing*, vol. 419, pp. 97–107, Jan. 2021, doi: 10.1016/j.neucom.2020.07.113.

[8] E. S. Hosseini and M. H. Moattar, "Evolutionary feature subsets selection based on interaction information for high dimensional imbalanced data classification," *Appl. Soft Comput.*, vol. 82,p. 105581, Sep. 2019, doi:10.1016/j.asoc.2019.105581.

[9] Z. Gao, Y. Xu, F. Meng, F. Qi, and Z. Lin,"Improved information gain-based feature selection for text categorization," in *2014 4th International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems (VITAE)*, Aalborg, Denmark, May 2014,pp. 1–5. doi: 10.1109/VITAE.2014.6934421.

[10] Lifang Wu, Dan Liu, Xiu Li, and Hongbo Li, "An information gain feature extraction model for web filtering," in *2012 IEEE International Conference on Computer Science and Automation Engineering*, Beijing, China, Jun. 2012, pp. 169–172. doi: 10.1109/ICSESS.2012.6269432.

[11] N. Pokhriyal and S. K. Verma, "Statistical feature extraction/selection for small infrared target," in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Jaipur, India, Sep. 2016, pp. 2064–2070. doi: 10.1109/ICACCI.2016.7732355.

[12] S. G. Devi and M. Sabrigiriraj, "Feature Selection, Online Feature Selection Techniques for Big Data Classification: - A Review," in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Coimbatore, Mar. 2018, pp. 1–9. doi:10.1109/ICCTCT.2018.8550928.

[13] Y. Yun, D. Ma, and M. Yang, "Human–computer interaction-based Decision SupportSystem with Applications in Data Mining," *Future Gener. Comput. Syst.*, vol. 114, pp. 285– 289, Jan. 2021, doi:10.1016/j.future.2020.07.048.

[14] B. Paradowski and Z. Drążek, "Identification of the decision-making model for selecting an information system," *Procedia Comput. Sci.*, vol. 176, pp. 3802–3809, 2020, doi: 10.1016/j.procs.2020.09.007.

[15] P. B. Keenan and P. Jankowski, "Spatial Decision Support Systems: Three decades on," *Decis. Support Syst.*, vol. 116, pp. 64–76, Jan. 2019, doi: 10.1016/j.dss.2018.10.010.

[16] P. Wang, Z. Zhu, and Y. Wang, "A novel hybrid MCDM model combining the SAW, TOPSIS and GRA methods based on experimental design," *Inf. Sci.*, vol. 345, pp. 27–45, Jun. 2016,doi: 10.1016/j.ins.2016.01.076.

[17] P. K. Kwok and H. Y. K. Lau, "Hotel selection using a modified TOPSIS-based decisionsupport algorithm," *Decis. Support Syst.*, vol. 120, pp. 95–105, May 2019, doi: 10.1016/j.dss.2019.02.004.

[18] T.-Y. Chen, "Comparative analysis of SAW and TOPSIS based on interval-valued fuzzy sets: Discussions on score functions and weightconstraints," *Expert Syst. Appl.*, vol. 39, no. 2, pp. 1848–1861, Feb. 2012, doi:10.1016/j.eswa.2011.08.065.

[19] X. Ye, Y. Wang, H. Li, and Z. Dai, "An Emergency Decision Support System Based on the General Decision Process," in *2008*

[20] *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Sydney, Australia, Dec. 2008, pp. 215–218. doi: 10.1109/WIIAT.2008.173.

[21] A. Nandi and H. Ahmed, *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machine*, 1st ed. Wiley, 2019. doi:10.1002/9781119544678.

Heri Ardiyanto, et. al. "Optimization of Simple Additive Weighting Method Based on Information Gain in Decision Support System." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 23(6), 2021, pp. 41-48.