Integration of the Decision Support System and Machine Learning Hybrid Algorithm Methods to Determine Government Assistance Recipients: A Case Study of Desa Berinovasi Funding Program BRIN

Indra Rusyadi Adiwijaya[1,3*], Sapto Wahyu Indratno[1], Manahan Siallagan[2], Agus Widodo[3]

[1]Department of Computational Science, Faculty of Mathematics and Natural Sciences, Bandung Insitute of Technology, Bandung, West Java, Indonesia
[2]School of Business and Management, Bandung Insitute of Technology, Bandung, West Java, Indonesia
[3]National Research and Innovation Agency, Indonesia

*Corresponding author's email: 20921001@mahasiswa.itb.ac.id

ABSTRACT

To improve the economy and the community's economic recovery, the Government of Indonesia, through the National Innovation Research Agency (BRIN), has pioneered the Innovative Village Program, an incentive to facilitate development for rural communities through legal entities. However, one of the important parts of implementing the village program to innovate is to determine the recipients of incentives who must be right on target and can be accounted for. In the process, the selection and evaluation of incentive recipients is a problem that involves many components or criteria that are assessed or called Multi-Criteria Decision Making (MCDM). The purpose of this research is to help decision-makers determine the right target beneficiaries. In this study, for the MCDM problem, the hybrid Analytical Hierarchy Process (AHP) method with Simple Additive Weighting (SAW) was used and integrated with machine learning modeling using Logistic Regression (LR). The AHP method is used to calculate the weight of each criterion, while the SAW method is used to sort each alternative with the assistance of an expert team assessment. In contrast, the LR method is used for predictive analysis and classification of the resulting data. This study indicates that the value of the consistency ratio (CR) of the AHP is 0.0657 <0.1, which means that the preference for the weight value of the assessment criteria is consistent and feasible to be applied to each criterion. And also the results obtained by the logistic regression modeling equation Y = -17.84 + 1.53X3 + 0.49X4 with 92.11% accuracy value, 92.42% precision value, 98.39% recall value and 95.31% F1 Score.

Keywords: Decision Support System, MCDM, , AHP, SAW, Machine Learning, Logistic Regression

1 Introduction

Currently, the Indonesian government, both central and regional, is committed to implementing a government based on the norms of good governance, including system governance, processes and work procedures that are clear, effective, efficient, measurable and by the principles of good governance. This is stated in the Presidential Regulation of the Republic of Indonesia No. 81 of 2010 concerning the Grand Design of Bureaucratic Reform. The target in 2025 is the realization of good governance with a professional and high-integrity government bureaucracy [1]. Good governance in the public sector encourages more informed and long-term decision-making and the efficient use of resources [2]. Public services are a benchmark for the success of carrying out tasks and measuring government performance through the bureaucracy. Public services as the primary mover are also considered necessary by all actors from the elements of good governance. Therefore, the government needs to implement good governance, supported by an accountability system, adequate and reliable information, and efficiency in resource management and public service delivery (The World Bank, 1992). [3].

On the other hand, the National Research and Innovation Agency (BRIN), one of the government institutions, is committed to implementing the principles of good governance in its institutional control. This is implied in the direction and target of BRIN Head in 2022 [4]. In public services, BRIN has pioneered the innovation ecosystem through the synergy of various parties through its program in the form of an Innovating Village, which is a collaboration between BRIN and the community and local government. This program is an incentive to facilitate development for rural communities through legal entities or those determined by communal/community-based authorized institutions that can be used to increase the added value of innovation-based superior products or services to contribute to economic improvement and community economic recovery [5].

However, an essential part of the village program process to innovate is determining the recipients of incentives who must be well-targeted and accountable. In the process, the selection and evaluation of incentive recipients is a problem that involves many components or criteria being assessed (multi-criteria), so that in its completion, a multi-criteria decision support system is needed or called Multi-Criteria Decision Making (MCDM) [6]. Many popular methods are used for selection and evaluation problems with the MCDM approach, such as Analytical Hierarchy Process (AHP) [7], Analytical Network Process (ANP) [8], Elimination et Choix Traduisant la Realité (ELECTRE) [9], Preference Ranking Organization Method for Enrichment of Evaluations (PROMETHEE) [10], Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [11] and VIseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR) [12]. In a study conducted by Nayli et al. [13] reviewing and surveying the use of these popular methods in various cases, the results of each process have its advantages and disadvantages depending on the complexity and structure of the problem.

In recent years researchers have used the hybrid MCDM method to get maximum results, such as the study conducted by Jicang Xu et al. [14] by combining the AHP and ANP methods to evaluate decision making using sustainable government data. Similarly, Vipul Jain et al. and Chia-NanWang et al. [15-16] combine the Fuzzy AHP and TOPSIS methods to support decision-making in determining suppliers. Furthermore, research conducted by Gülçin Büyüközkan et al. [17] combines AHP and VIKOR to support decision making in hazardous waste treatment. Furthermore, several MCDM methods have been integrated with other approaches, one of which is widely used to incorporate the MCDM method with machine learning. Research conducted by Jianmiao Hu et al. [18] integrates data mining modelling with Support Vector Machine (SVM) classification modelling and regression models with the MCDM AHP method to evaluate and analyze the company's credit risk level. Mazin Abed et al. [19] combined SVM classification modelling with TOPSIS using Entropy criteria weight calculations to assess and classify the COVID-19 diagnosis model. The study conducted by Ratiranjan Jena et al. [20] applied an integrated AHP model with Artificial Neural Network (ANN) to measure and assess earthquake risk areas. In the study, Lanbing Yu et al. [21] applied several machine learning models (logistic regression, decision tree, support vector machines, and random forest) with AHP to calculate the level of vulnerability and map the landslide-prone zone.

Several integration procedures for the hybrid MCDM method have been widely applied in various fields based on previous research. One hybrid method that is popular and interesting to research is the integration of AHP and Simple Additive Weighting (SAW). The integration of AHP and SAW has been widely applied to evaluation and selection problems [22-27]. Therefore, in this study, we combine the AHP method with SAW in a hybrid way and integrate it into machine learning modelling using a logistic regression algorithm for the problem of selecting recipients of the BRIN Innovation Village program funding incentives.

This research has studied the use of the SAW method as a ranking method in multi-attribute decision making supported by the Analytical Hierarchy Process (AHP) method using eigenvalues to determine the weight values for these attributes. The AHP method also calculates the Consistency Ratio, and this can provide the best alternative from various alternatives. The alternative assessment of the SAW method is obtained by adding the assessment contribution from the expert team for each weighted attribute. The resulting data in the form of scoring is used for predictive analysis with machine learning modelling using logistic regression to predict recipients of funding incentives.

2 Research Methodology

This research was conducted using the scheme depicted in Figure 1.

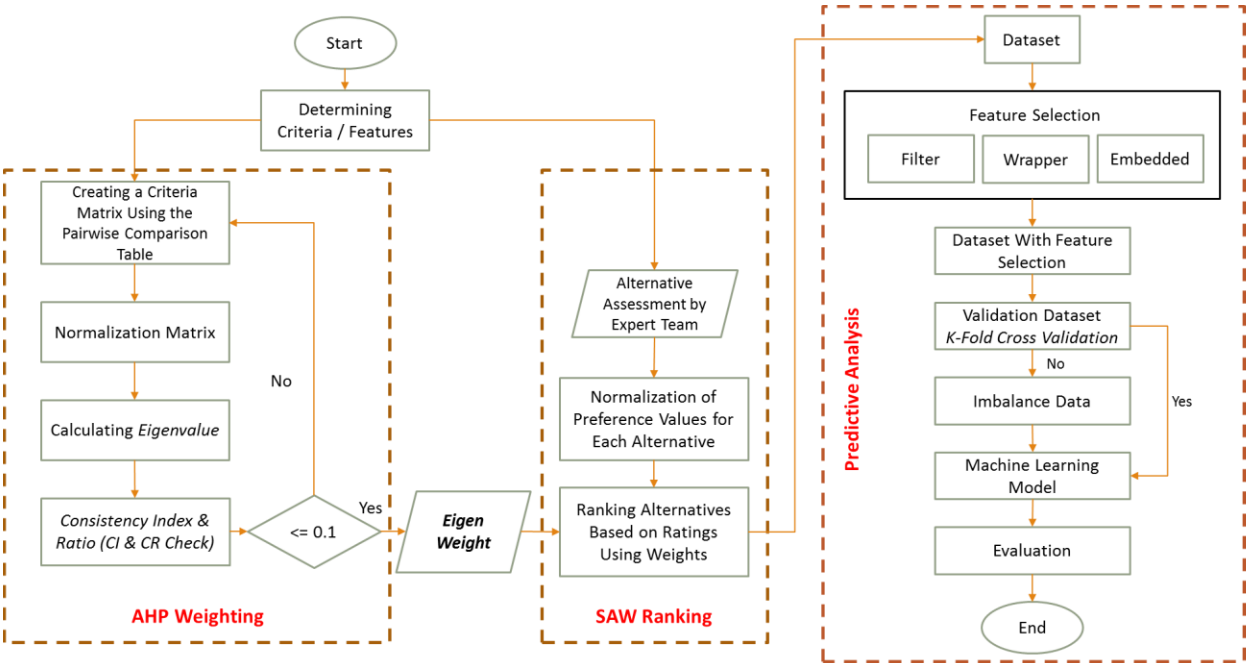

Figure 1: Research Stages

2.1 Area of Study

The research was conducted using data from the Innovated Village Program of the National Research and Innovation Agency in 2021, with a total of 2242 applicants who submitted funding proposals. In particular, there are several stages to determining program funding recipients in the selection process. In the first stage of 2242 applicants, 771 applicants officially submitted proposals, followed by the initial selection stage for the recommendations, which resulted in 138 proposals that passed the selection, then continued to the administrative selection stage by producing 96 proposals. And in the last step of this amount, 80 recipients of funding assistance were made. In this study, the data that will be observed is data on the results of the substance selection assessment of 96 prospective funding recipients. The results consist of 80 alternatives that pass or funded proposals and the remaining 16 recommendations that do not give.

2.2 Analytical Hierarchy Process (AHP) Method

The Analytical Hierarchy Process (AHP) was developed by Thomas L. Saaty, a mathematician from Pittsburg, United States, in the 1970s. This decision support model will describe a complex multi-factor or multi-criteria problem into a hierarchy. A problem is said to be complicated if the structure of the problem is not clear and there is no availability of accurate statistical data and information, so the input used to solve this problem is human intuition. The AHP method breaks a complex, unstructured situation into its parts, arranges the elements or variables in a hierarchical arrangement, assigns a numerical value to subjective judgments about the relative importance of each variable, synthesizes various considerations, and improves reliability. AHP as a decision-making tool.

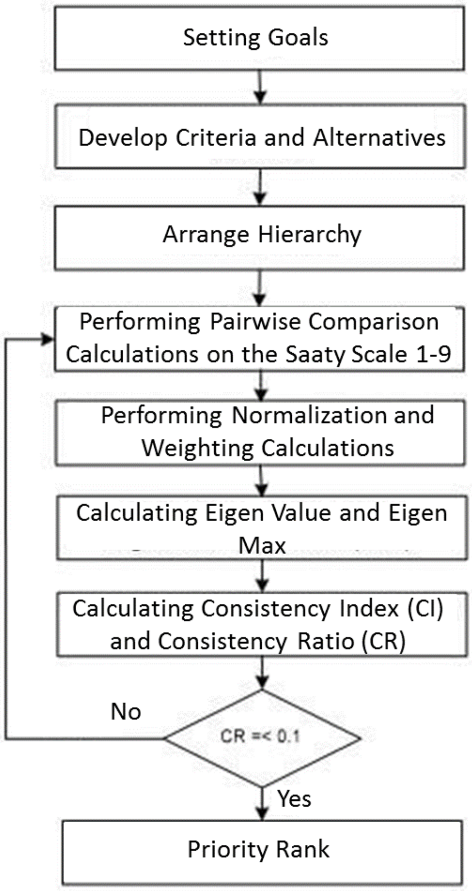

Figure 2. Flow Chart AHP

The steps and procedures in solving problems using the AHP method are as follows:

a. Define the problem and determine the desired solution.
   In setting priorities, the problem of prioritization must be able to be decomposed into the objectives (goals) of an activity, identification of options (alternatives), and formulation of criteria (criteria) for selecting priorities.
b. Develop a hierarchy starting with the primary goal.
   Hierarchy is an abstraction of the structure of a system that studies the function of interactions between components and their impacts on the system. The hierarchy or decision structure arrangement is carried out to describe the system elements or decision alternatives identified. According to Saaty (1993), hierarchy is a representation of a complex problem in a multi-level structure where the first level is the goal, followed by the level of factors, criteria, sub-criteria, and so on down to the last level of alternatives. The first step is to formulate the objectives of a priority-setting activity. After compiling the main goal as the top level, a hierarchy level will be compiled below it, namely suitable criteria for considering or assessing the alternatives given and determining these alternatives, followed by sub-criteria, as shown in Figure 3:

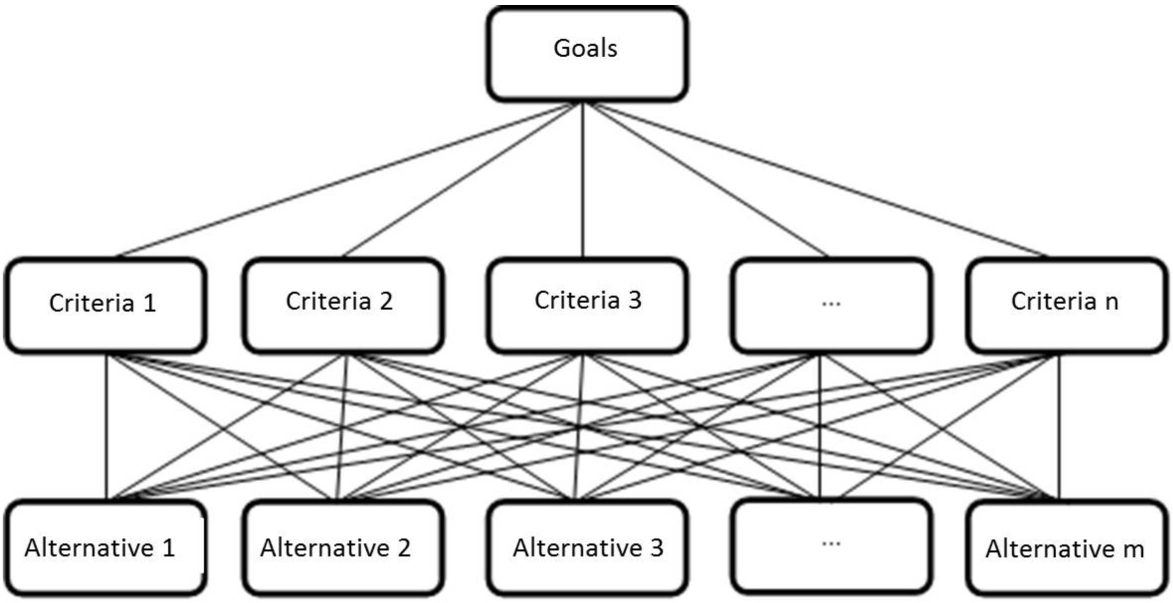

Figure 3: AHP Hierarchy Structure

c. Create a criterion matrix using a pairwise comparison rating scale
After identifying the hierarchy of objectives to alternatives, the next stage is assessing or comparing elements, namely, comparisons between criteria using a paired comparison rating scale. Comparison between criteria is intended to determine the weight of each measure. According to Saaty (1988), a scale of 1 to 9 is the best scale for expressing opinions on various problems. Each pairwise comparison is evaluated on Saaty's scale 1 – 9 as follows:

Table 1: Pairwise Comparison Matrix

Value	Information
1	Criteria/Alternative A is as important as Criteria/Alternative B
3	A is more important than B
5	A is slightly more important than B
7	A is clearly more important than B
9	A is absolutely more important than B
2,4,6,8	When in doubt between two adjacent values
Opposite	If alternative 1 is compared to alternative 2 the value is 3, then alternative 2 is compared to alternative 1 the value is 1/3

d. Matrix Normalization

$$w_i = \sum_{i=1}^{n} a_{ij}/n \qquad (1)$$

Information :
w_i : $weighted\ value$
a_{ij}/n : $matriks\ normalisasi\ baris$

e. Calculating eigen value and eigen value max

$\lambda_i = \sum_{i=1}^{n} a_{ij}/w_i$ (2)

$\lambda_{max} = \sum_{i=1}^{n} (a_{ij}/w_i)/n$ (3)

f. Test consistency by using the Consistency Index (CI)

$CI = \frac{(\lambda_{max}-n)}{(n-1)}$ (4)

Keterangan :

λ_{max} : $eigen\ value\ maximum$

n : $number\ of\ matrices$

g. Calculating Consistency Ratio (CR)

$CR = \frac{CI}{RI}$ (5)

Information :

CR : $Consistency\ Ratio$

RI : $Random\ Consistency\ Index$

Table 2: Random Consistency Index

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0,58	0,90	1,12	1,24	1,32	1,41	1,45	1,49

2.3 Simple Additive Weighting (SAW)

The SAW method is often also known as the weighted addition method. The basic concept of the SAW method is to find the weighted sum of the performance ratings for each alternative on all attributes. The SAW method requires normalizing the decision matrix (X) to a scale that can be compared with all existing alternative ratings. Here is the formula for finding the normalized matrix

$$r_{ij} = \begin{cases} \dfrac{x_{ij}}{\max\limits_i x_{ij}} & If\ j\ is\ a\ benefit\ attribute \\ \dfrac{\min\limits_i x_{ij}}{x_{ij}} & If\ j\ is\ a\ cost\ attribute\ (cost) \end{cases} \quad (6)$$

Information :
rij = normalized performance rating value
xij = attribute value owned by each criterion
Max xij = the largest value of each criterion i
Min xij = the smallest value of each criterion i
benefit = if the biggest value is the best
cost = if the smallest value is the best where rij is the normalized performance rating of alternative Ai on attribute Cj; i=1,2,…,m and j=1,2,…,n.

Steps to solve using the SAW method:

a. Determine the criteria that will be used as a reference in decision making.
b. Determine the weight value of each criterion that has been obtained previously.
c. Determine the suitability rating of each alternative on each criterion.
d. Make a decision matrix based on the criteria, then normalize the matrix based on the equation adjusted to the type of attribute (profit attribute and cost attribute) to obtain a normalized matrix R.
e. Give a preference value for each alternative (Vi) with the formula:

$$V_i = \sum_{j=1}^{n} w_j r_{ij} \qquad (7)$$

Information :

V_i : $rank\ for\ each\ alternative$

w_j : $weight\ value\ of\ each\ criterion$

r_{ij} : $normalized\ performance\ rating\ value$

2.4 Feature Selection

Before a dataset is used to train a machine learning model, a series of steps need to be performed on the data. This series of processes is commonly referred to as data preparation or data preparation. Data preparation is an essential part of improving data quality and minimizing noise because the data generated from this series of processes will determine the efficiency of training and the performance of the resulting model.

Feature selection is one way of preparing data to improve accuracy in a machine learning model. The feature selection process reduces the number of features or input variables by selecting the features that are considered most relevant to the model. There are two types of feature selection, namely supervised and unsupervised. Supervised methods consist of the wrapper, filter, and intrinsic/embedded methods.

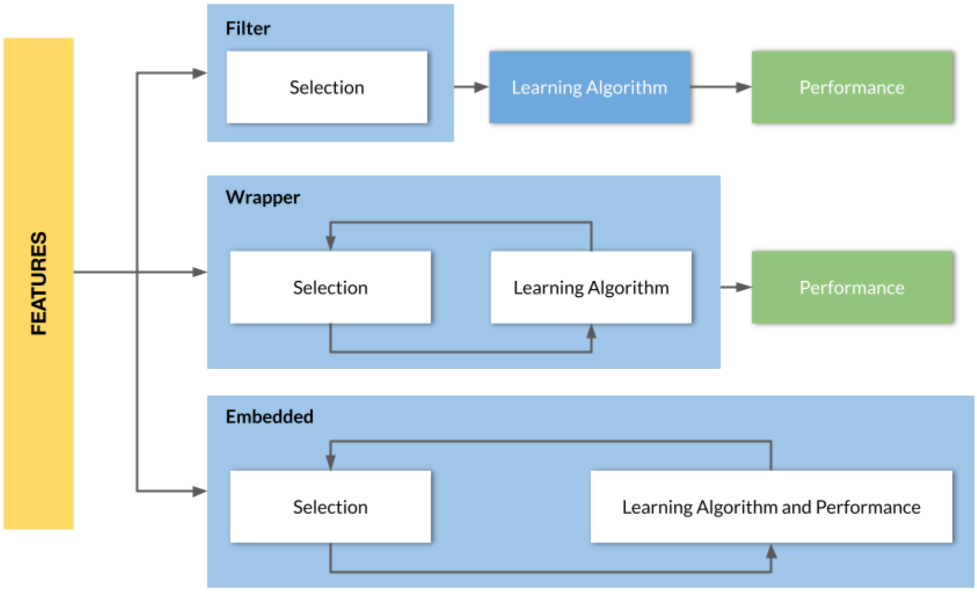

Figure 4: Feature Selection Method

2.5 Imbalanced Data Handling

Unbalanced classes are a common problem in machine learning classification. Where imbalance class is a disproportionate distribution between classes in a dataset, one class has a considerable amount of data (majority class) compared to other classes (minority class). The significant difference in the amount of data between classes can result in the classification model being often unable to predict the minority class accurately. A lot of test data that should be in the minority class is mispredicted by the classification model. Some solutions to overcome this include using the right evaluation metrics, resampling Over-Sampling or Under-Sampling, and K-fold Cross-Validation.

2.6 Regresi Logistik

Logistics regression is a data analysis technique in statistics that determines the relationship between several variables. The response variable is categorical, either nominal or ordinal, with the explanatory variable being categorical or continuous. Binary logistic regression is a mathematical model approach used to analyze the relationship between several factors and a binary variable. In logistic regression, if the response variable consists of two categories, Y=1 indicates the results obtained are "successful," and Y=0 shows the results obtained are "failed." The logistic regression uses binary logistic regression.
The logistic regression method has techniques and procedures that are not much different from the linear regression method. To find the logistic equation, the model used is:

$$\pi(\mathrm{x}) = \frac{e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}}{1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}} \quad (8)$$

From equation (8) obtained 1 – π(x) as follows:

$$1-\pi(\mathrm{x}) = 1-\frac{e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}}{1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}}$$

$$1-\pi(\mathrm{x}) = \frac{1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j} - e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}}{1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}}$$

$1-\pi(\mathrm{x}) = \frac{1}{1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}}$, Sehingga $\frac{\pi(\mathrm{x})}{1-\pi(\mathrm{x})}$ sebagai berikut : $\frac{\pi(\mathrm{x})}{1-\pi(\mathrm{x})} = e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}$

So the logistic equation is :

$$\begin{aligned}\hat{\boldsymbol{g}}(\boldsymbol{x}) &= \mathbf{ln}\left(\frac{\pi(\mathrm{x})}{1-\pi(\mathrm{x})}\right)\\ &= \boldsymbol{ln}\left(e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}\right)\\ &= \beta_0+\textstyle\sum_{j=1}^{p}\beta_j x_j \qquad (9)\end{aligned}$$

3 Results and Discussion

In the results and discussion, the results and application of the integration of the decision support system algorithm using the hybrid AHP and SAW methods with predictive analysis using machine learning will be explained to assist the best decision-making process in determining the recipients of funding assistance for the Desa Berinovasi program at BRIN. The AHP method is used to calculate the weight of each criterion, followed by the SAW method using the AHP weight to calculate the rating value of each alternative from all requirements with the results of the assessment team's assessment. In contrast, machine learning methods are applied for predictive analysis and classification of the resulting data. Processing data using Microsoft Excel and Python programming language

3.1 Weighting Criteria Using AHP Method

The weight of each criterion was determined using the AHP method. The process of weighting the criteria is processed through brainstorming with the chief executive of the program as the respondent and giving the weight values in table 3 of the pairwise comparison matrix of the AHP calculations.

Table 3: Assessment Criteria

Code	Assessment Aspect
K1	Locus Profile
K2	Regional Featured Product Profile
K3	Technology and Innovation Profile
K4	Stakeholder/Stakeholders Support
K5	Human Resources
K6	Management Agency
K7	Activity Implementation Method
K8	Work Plan and Implementation Strategy
K9	Detailed Budget

The weighting for the criteria begins with defining a pairwise comparison matrix, which can be seen in Table 4. Then the matrix will be normalized, and the normalization results can be seen in Table 5. The calculation of the eigenvectors, which are the weighted values, can be seen in Table 6.

Table 4: Criteria Comparison Matrix Value

	K1	K2	K3	K4	K5	K6	K7	K8	K9
K1	1.0000	0.1429	0.2000	0.1667	0.3333	0.3333	1.0000	0.3333	1.0000
K2	7.0000	1.0000	5.0000	3.0000	3.0000	3.0000	7.0000	3.0000	7.0000
K3	5.0000	0.2000	1.0000	0.2000	1.0000	1.0000	3.0000	1.0000	3.0000
K4	6.0000	0.3333	5.0000	1.0000	5.0000	5.0000	7.0000	5.0000	7.0000
K5	3.0000	0.3333	1.0000	0.2000	1.0000	1.0000	5.0000	1.0000	5.0000
K6	3.0000	0.3333	1.0000	0.2000	1.0000	1.0000	5.0000	1.0000	5.0000
K7	1.0000	0.1429	0.3333	0.1429	0.2000	0.2000	1.0000	0.2000	1.0000
K8	3.0000	0.3333	1.0000	0.2000	1.0000	1.0000	5.0000	1.0000	5.0000
K9	1.0000	0.1429	0.3333	0.1429	0.2000	0.2000	1.0000	0.2000	1.0000
Total	**30.0000**	**2.9619**	**14.8667**	**5.2524**	**12.7333**	**12.7333**	**35.0000**	**12.7333**	**35.0000**

The initial step with the comparison matrix criteria in the previous table is used to get the eigenvalues of each row after the matrix is normalized. An example of the process of calculating the normalization of pairwise comparisons based on equation (1) is as follows:

$$X_{1,1} = \frac{1}{1+7+5+6+3+3+1+3+1} = 0.0333$$

$$X_{2,1} = \frac{7}{1+7+5+6+3+3+1+3+1} = 0.2333$$

The example calculation above produces the normalized value for the first column of the pairwise comparison matrix. The resulting values are shown in Table 5.

Table 5: Normalization of Comparison Matrix

	K1	K2	K3	K4	K5	K6	K7	K8	K9	Total
K1	0.0333	0.0482	0.0135	0.0317	0.0262	0.0262	0.0286	0.0262	0.0286	0.2624
K2	0.2333	0.3376	0.3363	0.5712	0.2356	0.2356	0.2000	0.2356	0.2000	2.5853
K3	0.1667	0.0675	0.0673	0.0381	0.0785	0.0785	0.0857	0.0785	0.0857	0.7466
K4	0.2000	0.1125	0.3363	0.1904	0.3927	0.3927	0.2000	0.3927	0.2000	2.4173
K5	0.1000	0.1125	0.0673	0.0381	0.0785	0.0785	0.1429	0.0785	0.1429	0.8392
K6	0.1000	0.1125	0.0673	0.0381	0.0785	0.0785	0.1429	0.0785	0.1429	0.8392
K7	0.0333	0.0482	0.0224	0.0272	0.0157	0.0157	0.0286	0.0157	0.0286	0.2354
K8	0.1000	0.1125	0.0673	0.0381	0.0785	0.0785	0.1429	0.0785	0.1429	0.8392
K9	0.0333	0.0482	0.0224	0.0272	0.0157	0.0157	0.0286	0.0157	0.0286	0.2354

An example of the process of calculating the eigenvalues, which are the weight values of each criterion based on equation (2) as follows:

$$\lambda_1 = \frac{0.2624}{30.000} = 0.0292$$

$$\lambda_2 = \frac{2.5853}{2.9619} = 0.2873$$

The example calculation above produces eigenvalues which are weight values for criteria 1 and 2. The results of calculating eigenvalues for all requirements are shown in Table 6.

Table 6: Eigen Value

	Eigen Value	Weight Value (%)
K1	0.0292	2.92
K2	0.2873	28.73
K3	0.0830	8.30
K4	0.2686	26.86
K5	0.0932	9.32
K6	0.0932	9.32
K7	0.0262	2.62
K8	0.0932	9.32
K9	0.0262	2.62

Then calculate the index consistency value (CI) and the consistency ratio value (CR) using the following formula:

$$CI = \frac{\lambda maks - n}{n-1}$$

$$CR = \frac{CI}{RI}$$

Then the index consistency value for all criteria is,

$$CI = \frac{9.7627 - 9}{9 - 1} = 0.0953$$

Based on the Random Consistency Index table for n = 9, the value of RI = 1.45, then the value of the consistency ratio is,

$$CR = \frac{0.0953}{1.45} = 0.0657$$

Because the value of CR = 0.0657 < 0.1, the preference value of the assessment criteria is consistent and does not require a revision of the assessment. The weight value, the eigenvalue, can be used using the SAW method in the following calculation process.

3.2 Calculation of SAW Method

The SAW method is used to calculate the final alternative value by sorting the importance of all alternatives using the weights of the AHP method. The output produced is a sequence of other options from the highest to the choice with the lowest value. The alternatives referred to are 96 potential recipients of funding assistance. The basic concept in the SAW method is to find the weighted sum of the performance ratings on each alternative in all attributes. The SAW method requires normalizing the decision matrix to a scale that can be compared with all existing alternative ratings. Of the nine assessment criteria used to determine the prospective recipients of funding assistance, table 7 identifies the classification of measures that benefit and cost the matrix normalization process.

Table 7: Normalization of Benefit & Cost

Criteria	Attribute	Value Weight (%)	Rounding Value Weight (%)
K1	Benefit	2.92	3
K2	Benefit	28.73	29
K3	Benefit	8.30	8
K4	Benefit	26.86	27
K5	Benefit	9.32	9
K6	Benefit	9.32	9
K7	Benefit	2.62	3
K8	Benefit	9.32	9
K9	Benefit	2.62	3

From the table of criteria classification, all criteria are included in the benefit attribute, which means that the most excellent value is the best value. Therefore, at this stage, the matrix normalization of all alternative values is converted to a percentage of the full value scale using the initial weight before calculating the sum using the AHP weight.

Table 8: Assessment Results Before Normalization

Alternative	Recapitulation of the Reviewer Team's Assessment Results								
	K1 (5%)	K2 (25%)	K3 (10%)	K4 (20%)	K5 (10%)	K6 (10%)	K7 (5%)	K8 (10%)	K9 (5%)
1	45	150	85	150	80	90	50	90	45
2	45	150	90	140	80	90	40	80	40
…	…	…	…	…	…	…	…	…	…
96	35	160	80	150	80	80	40	80	35

Tabel 9: Assessment Results After Normalization

Alternative	Recapitulation of the Reviewer Team's Assessment Results								
	K1 (5%)	K2 (25%)	K3 (10%)	K4 (20%)	K5 (10%)	K6 (10%)	K7 (5%)	K8 (10%)	K9 (5%)
1	90	60	85	75	80	90	100	90	90
2	90	60	90	70	80	90	80	80	80
…	…	…	…	…	…	…	…	…	…
96	70	64	80	75	80	80	80	80	70

The final result is obtained from the ranking process, namely the addition of the normalized matrix multiplication with the AHP weight vector. The most considerable value is chosen as the best alternative as a candidate solution for funding assistance recipients.

Table 10: Multiply Normalized Matrix with AHP Weights

Alternative	Recapitulation of the Reviewer Team's Assessment Results									Total
	K1 (3%)	K2 (29%)	K3 (8%)	K4 (27%)	K5 (9%)	K6 (9%)	K7 (3%)	K8 (9%)	K9 (3%)	
1	2.7	17.4	6.8	20.25	7.2	8.1	3	8.1	2.7	76.25
2	2.7	17.4	7.2	18.9	7.2	8.1	2.4	7.2	2.4	73.50
…	…	…	…	…	…	…	…	…	…	…
96	2.1	18.56	6.4	20.25	7.2	7.2	2.4	7.2	2.1	73.41

Table 11: Ranking of Values from Largest to Lowest

Alternative	Recapitulation of the Reviewer Team's Assessment Results									Total
	K1 (3%)	K2 (29%)	K3 (8%)	K4 (27%)	K5 (9%)	K6 (9%)	K7 (3%)	K8 (9%)	K9 (3%)	
27	3	29	8	27	9	9	3	7.2	2.4	97.60
30	3	29	8	23.625	9	9	3	9	2.4	96.03
48	2.7	24.36	7.2	26.325	8.55	8.55	2.7	8.55	2.7	91.64
23	3	26.1	7.2	27	9	9	3	3.6	2.4	90.30
…	…	…	…	…	…	…	…	…	…	…
87	1.8	20.3	5.2	16.2	5.4	6.3	2.1	5.85	1.8	64.95
88	1.8	20.3	5.2	16.2	5.4	6.3	2.1	5.85	1.8	64.95

3.3 Predictive Analysis

3.3.1 Feature Selection

There are two types of feature selection, namely supervised and unsupervised. In this study, feature selection with a supervised method is used, which consists of a filter, wrapper, and intrinsic/embedded plans to reduce the number of features or input variables by selecting the features that are considered the most relevant and affect the model to be made. The following table of feature selection results

Table 12: Feature Selection Results

No	Method	Description	Selection Result
1	Filter Feature Selection	Using Pearson Correlation and Variance Inflation Factor (VIF)	K4, K3, K1, K6
2	Wrapper Feature Selection	Using Recursive Feature Elimination (RFE) mechanism with Learning Random Forest Algorithm	K4, K3, K1, K9
3	Embedded Feature Selection	Using the Learning Logistic Regression Algorithm and using the L2 regularization feature as a penalty function to eliminate feature	K4, K3, K2, K6

Based on the feature selection table above, of the nine criteria or input variables used, the criteria that are considered the most relevant and influential on the model to be made are K3 (Profile of Technology and Innovation) and K4 (Stakeholder Support). This also proves that in predictive analysis, the weight value of each criterion or variable obtained from the calculation of the AHP method does not automatically represent the value of relevance to the input variable. It is proven that the K2 criteria (Profile of Regional Leading Products) are the most considerable weight value (29%) in the calculation of the AHP method, in contrast to the feature selection process of predictive analysis of the most relevant and influential K3 criteria.

3.3.2 Data Validation and Machine Learning Modeling

Imbalanced Data Test

In the classification process, we usually have various problems with data, both from data preprocessing, modeling, evaluation, and others. Sometimes, one thing that is not realized from the classification process is the number or proportion of existing labels/classes. It could be that the data we are dealing with is an Imbalanced Dataset. So in this study, dataset validation was also carried out to help form the desired model. The expected model is a model that can distinguish between proposals that are accepted/funded (rare class) and those that are not (abundant class). The results of plotting the dataset using the feature selection variable can be seen in Figure 5.

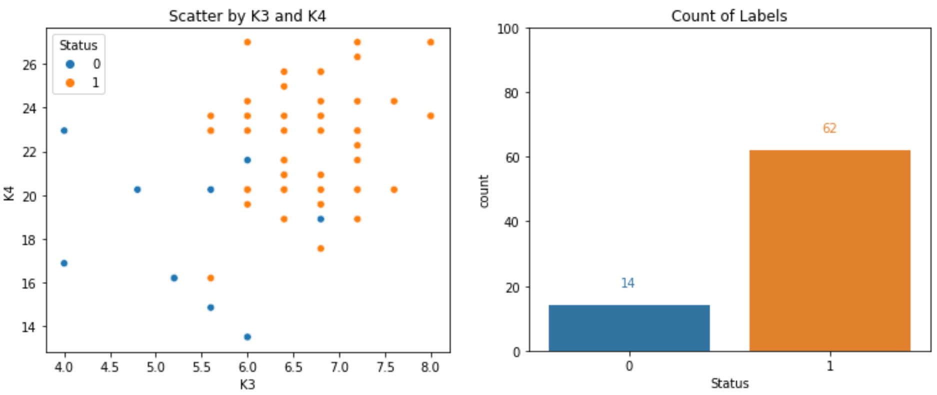

Figure 5: Dataset Plotting Results Using Feature Selection Results Variables Without Imbalanced Data

Based on the results of plotting the dataset, it can be seen that one of the classes/labels (not passed) has a value that looks much different in number from the graduated label class. So dataset validation is essential here to help form the desired model. Figures 6 and 7 are the results of resampling the dataset using Over-Sampling and Under-Sampling.

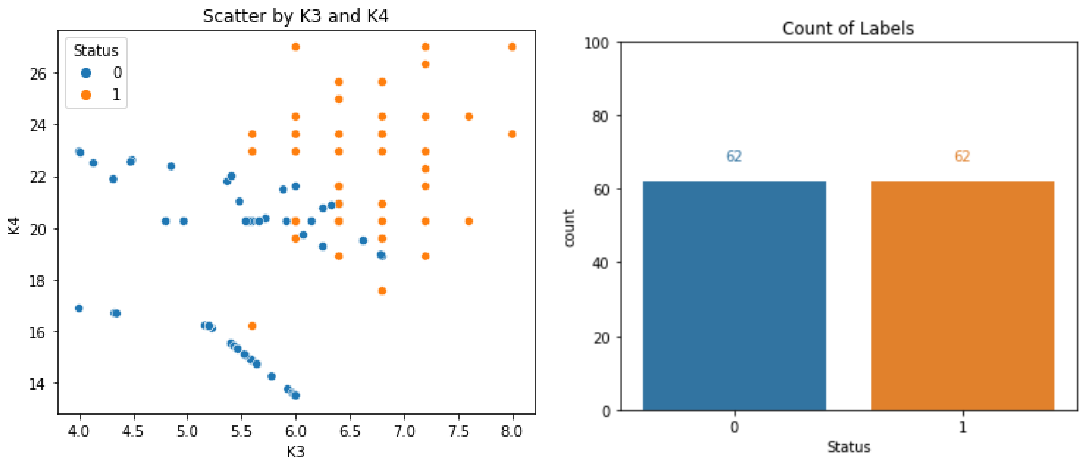

Figure 6: SMOTE (Oversampling) Data Imbalance Plotting Results

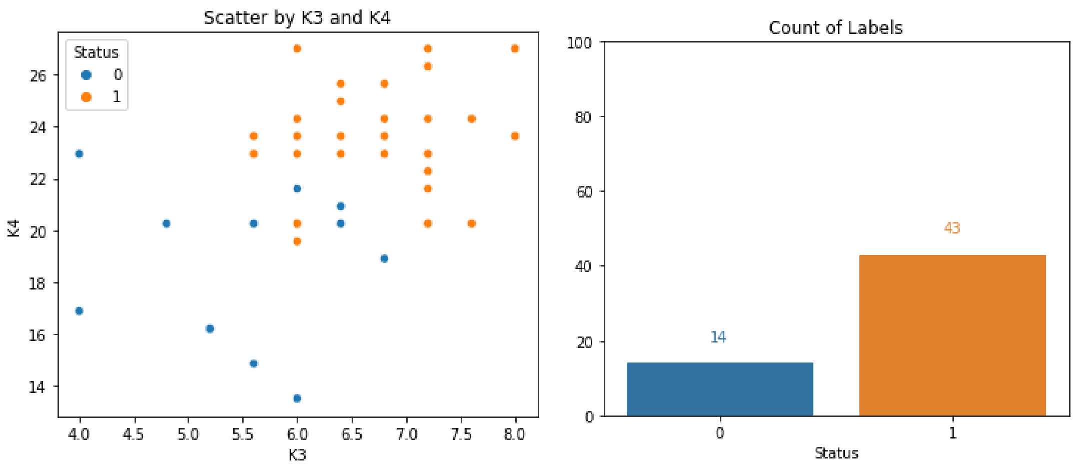

Figure 7: ENN (Under Sampling) Data Imbalance Plotting Results

Table 13: Dataset Evaluation & Validation Table

Evaluation Model	Base Logistic Regression (%)		
	Without Imbalance Data	**Imbalance Data SMOTE (Over Sampling)**	**Imbalance Data ENN (Under Sampling)**
Accuracy	92.11	67.11	89.47
Precision	92.42	95.24	90.91
Recall	98.39	63.49	96.77
F1 Score	95.31	76.19	93.75

Based on the results of table 13, the dataset without imbalanced data handling has the best model evaluation score compared to the dataset with imbalanced data handling using SMOTE (Over-Sampling) and ENN (Under-Sampling).

K-Fold Cross Validation

K-fold cross-validation is a validation technique to assess how the results of a statistic analysis generalize a data set. This technique is used as a predictive model, accommodating estimates from a model when it is run. One cross-validation technique is k-fold cross-validation, which breaks down data into k datasets of equal size. Use of k-fold cross-validation to eliminate bias in the data. Figure 8 is the result of the k-fold cross-validation test on the dataset with the previous three scenarios.

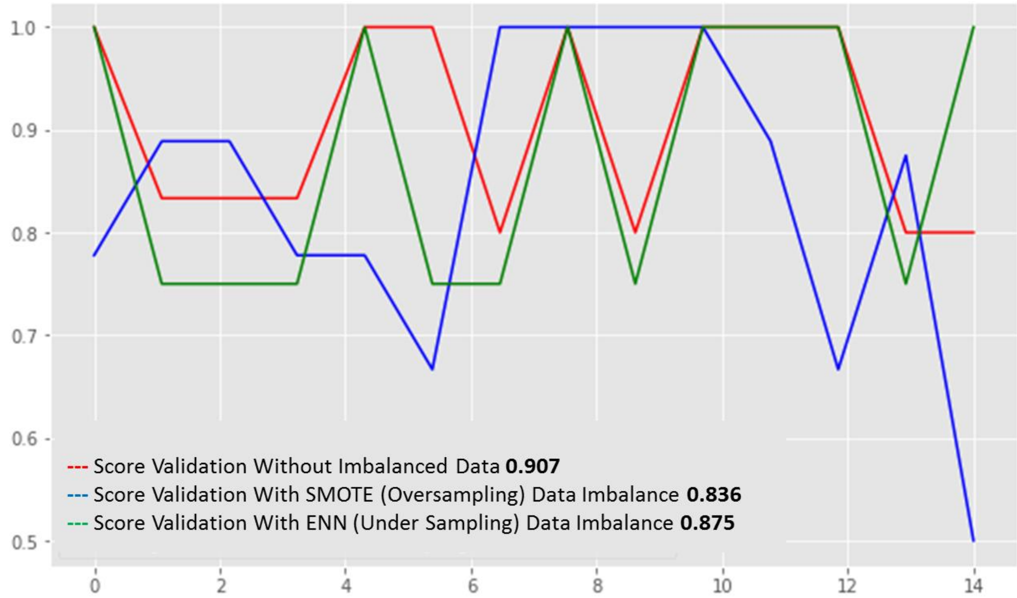

Figure 8: K-Fold Cross Validation Test Results

Similar to the validation with imbalanced data, the results of the k-fold cross-validation test resulted in the dataset without imbalanced data handling having the best score with a value of 0.907 compared to the dataset with imbalanced data handling using SMOTE (Over-Sampling) and ENN (Under-Sampling).

Logistics Regression Equation Model

Based on the results of feature selection and data validation, it is known that for machine learning modeling, the dataset that will be used is a dataset without resampling with two explanatory variables, K3 (Profile Technology and Innovation) and K4 (Stakeholder Support) which have a significant influence on the response variable so that These two variables are included in the logistic regression equation, so that the logistic regression model obtained with the response variable Y = (x) is the status of the recipient of funding with the explanatory variables K3 and K4. The following is the equation of the logistic regression model:

$$\hat{g}(x) = -17.84 + 1.53K3 + 0.49K4$$

Furthermore, an evaluation of the classification of the logistic regression equation model is carried out using the confusion matrix table with the classification results carried out by the system (model) with the actual classification results. The confusion matrix is in the form of a matrix table that describes the performance of the classification model on a series of test data whose actual values are known. The picture below is a confusion matrix with four different combinations of predicted values and actual values.

Table 14: Confution Matrix

Confution Matriks N=96		Prediction	
		Accepted	Not Accepted
Actual	Accepted	78 (TP)	2 (FP)
	Not Accepted	7 (FN)	9 (TN)

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{87}{96} = 0.90625 = \mathbf{90.62}\%$$

$$\boldsymbol{Precission} = \frac{TP}{TP + FP} = \frac{78}{80} = 0.975 = \mathbf{97.50}\%$$

$$\boldsymbol{Recall} = \frac{TP}{TP + FN} = \frac{78}{85} = 0.9176 = \mathbf{91.76}\%$$

4 Conclusions

Based on the results of the study, it can be concluded that the weight value of each assessment criterion is obtained from the AHP calculation with a value of K1 3%, K2 29%, K3 8%, K4 27%, K5 9%, K6 9%, K7 3%, K8 9% and K9 3%, and the consistency ratio value is 0.065749 < 0.1, then the preference value of the assessment criteria is consistent and does not require a revision of the assessment. The weight value, the eigenvalue, can be used in the following calculation process using the SAW method. The final result is obtained from the ranking process, namely the addition of the normalized matrix multiplication with the AHP weight vector. The most considerable value is chosen as the best alternative as a candidate solution for funding assistance recipients.

In the predictive analysis of the feature selection process, which aims to select variables relevant to the grantee's response variables, the relevant variables are K3 (Profile Technology and Innovation) and K4 (Stakeholder Support). These results prove that in predictive analysis, the weight value of each of the criteria or variables obtained from the calculation of the AHP method does not automatically represent the value of relevance to the input variable. It is proven that the K2 criteria (Profile of Regional Leading Products) are the most considerable weight value (29%) in the calculation of the AHP method. In contrast to the feature selection process, the K3 criteria are the most relevant and influential variables.

In the data validation process, the best model is without imbalanced data handling compared to imbalanced data handling using Over Sampling and Under Sampling, resulting in 92.11% Accuracy, 92.42% Precision, 98.39% Recall, and F1 Score. 95.31%. Likewise, validation testing using K-Fold Cross Validation with a scoring accuracy value of 0.907 or 91%. Therefore, a logistic regression equation is obtained to predict the response variable y (recipient of funding) Y = -17.84 + 1.53K3 + 0.49K4 with evaluation values using a confusion matrix, namely the Accuracy value of 90.62%, Precision 97.50% and Recall 91.76%. Judging from the regression equation, the K3 value is greater than the K4 value. The K3 value indicates the slope of X (Profile of Technology and Innovation), and K4 indicates the slope of X (Stakeholders Support). In this case, it can be concluded that the Technology and Innovation Profile percentage is more influential than Stakeholder Support.

It is hoped that the following research can use serial data to enrich the dataset used and be tested to find weights using AHP with group preferences rather than individuals. The subsequent research can compare several machine learning algorithms for predictive analysis to get the best model from several algorithms used. It is also hoped that the following output can be application-based.

5 References

[1] Pemerintah Indonesia, "Peraturan Presiden Republik Indonesia Nomor 81 Tahun 2010 Tentang Grand Design Reformasi Birokrasi 2010-2025," Sekretariat Negara, Jakarta, 2010.

[2] IFAC & CIPFA, "International Framework: Good Governance in the Public Sector," IFAC and CIPFA, New York, 2014.

[3] The World Bank, "Governance and Development," The World Bank, Washington, D.C., 1992.

[4] Kementerian Riset, Teknologi dan Pendidikan Tinggi / Badan Riset Inovasi Nasional, "Peraturan Menteri Riset, Teknologi dan Pendidikan Tinggi Republik Indonesia Nomor 38 Tahun 2019 Tentang Prioritas Riset Nasional Tahun 2020-2024," Kementerian Riset, Teknologi dan Pendidikan Tinggi / Badan Riset Inovasi Nasional, Jakarta, 2019.

[5] Kementerian Riset, Teknologi dan Pendidikan Tinggi / Badan Riset Inovasi Nasional, "Petunjuk Teknis Bantuan Pemerintah Berupa Desa Berinovasi," Kementerian Riset, Teknologi dan Pendidikan Tinggi / Badan Riset Inovasi Nasional, Jakarta, 2021.

[6] C. N. Wang, H. T. Tsai, T. P. Ho, V. T. Nguyen and Y. F. Huang, "Multi-criteria decision making (MCDM) model for supplier evaluation and selection for oil production projects in vietnam," *Processes MDPI,* vol. 8, no. 2, 2020.

[7] K. Zhao, Y. Dai, Y. Ji and Z. Jia, "Decision-Making Model to Portfolio Selection Using Analytic Hierarchy Process (AHP) with Expert Knowledge," *IEEE Access,* vol. 9, pp. 76875-76893, 2021.

[8] D. Rodrigues, R. Godina and P. E. da Cruz, "Key performance indicators selection through an analytic network process model for tooling and die industry," *Sustainability (Switzerland) MDPI,* vol. 13, no. 24, 2021.

[9] D. Azzeddine, S. Jabri, B. Yousse and G. Taoufiq, "The selection of the relevant association rules using the electre method with multiple criteria," *IAES International Journal of Artificial Intelligence,* vol. 9, no. 4, pp. 638-645, 2020.

[10] A. P. Lopes and N. Rodriguez-Lopez, "A decision support tool for supplier evaluation and selection," *Sustainability (Switzerland) MDPI,* vol. 13, no. 22, 2021.

[11] M. Dursun and O. Ogunclu, "Agile Supplier Evaluation Using Hierarchical TOPSIS Method," *WSEAS Transactions on Information Science and Applications,* vol. 18, pp. 12-19 , 2021.

[12] J. H. Kim and B. S. Ahn, "The hierarchical VIKOR method with incomplete information: Supplier selection problem," *Sustainability (Switzerland) MDPI,* vol. 12, no. 22, pp. 1-4, 2020.

[13] N. A. Azhar, N. A. M. Radzi and W. S. H. M. Wan Ahmad, "Multi-criteria Decision Making: A Systematic Review," *Electrical & Electronic Engineering Bentham Science,* vol. 14, no. 8, pp. 779-801, 2021.

[14] J. Xu, L. Li and M. Ren, "A Hybrid ANP Method for Evaluation of Government Data Sustainability," *Sustainability (Switzerland) MDPI,* vol. 14, no. 2, 2022.

[15] V. Jain, A. K. Sangaiah, S. Sakhuja, N. Thoduka and R. Aggarwal, "Supplier selection using fuzzy AHP and TOPSIS: a case study in the Indian automotive industry," *Neural Computing and Applications Springer,* vol. 29, no. 7, pp. 555-564, 2018.

[16] C. N. Wang, C. F. Pan, V. T. Nguyen and S. T. Husain, "Sustainable supplier selection model in supply chains during the COVID-19 pandemic," *Computers, Materials and Continua - Tech Science Press,* vol. 70, no. 2, pp. 3005-3019, 2022.

[17] G. Büyüközkan, F. Göçer and Y. Karabulut, "A new group decision making approach with IF AHP and IF VIKOR for selecting hazardous waste carriers," *Journal of the International Measurement Confederation Elsevier,* vol. 134, pp. 66-82, 2019.

[18] J. Hu, C. Chen and K. Zhu, "Application of Data Mining Combined with Fuzzy Multicriteria Decision-Making in Credit Risk Assessment from Legal Service Companies," *Mathematical Problems in Engineering Hindawi,* vol. 2021, 2021.

[19] M. A. Mohammed, K. H. Abdulkareem, A. S. Al-Waisy, S. A. Mostafa, S. Al-Fahdawi, A. M. Dinar, W. Alhakami, A. Baz, M. N. Al-Mhiqani, H. Alhakami, N. Arbaiy, M. S. Maashi and A. Mutlag, "Benchmarking Methodology for Selection of Optimal COVID-19 Diagnostic Model Based on Entropy and TOPSIS Methods," *IEEE Access,* vol. 8, pp. 99115-99131, 2020.

[20] R. Jena, B. Pradhan, G. Beydoun, Nizamuddin, Ardiansyah, H. Sofyan and M. Affan, "Integrated model for earthquake risk assessment using neural network and analytic hierarchy process: Aceh province, Indonesia," *Geoscience Frontiers Elsevier,* vol. 11, no. 2, pp. 613-634, 2020.

[21] L. Yu, C. Zhou, Y. Wang, Y. Cao and D. J. Peres, "Coupling Data-and Knowledge-Driven Methods for Landslide Susceptibility Mapping in Human-Modified Environments: A Case Study from Wanzhou County, Three Gorges Reservoir Area, China," *Remote Sensing MDPI,* vol. 14, no. 3, 2022.

[22] Angelina, R. D. Surbakti, R. S. Simamora, E. Cendana, D. Sitanggang, J. Banjarnahor and M. Turnip, "Application Selection Lending Houses Subsidized by the Method of AHP and SAW," *Journal of Physics: Conference Series IOP Publishing,* vol. 1230, no. 1, 2019.

[23] S. Wijayanto, D. Napitupulu, K. Adiyarta and A. P. Windarto, "Decision Support System of New Student Admission Using Analytical Hierarchy Process and Simple Additive Weighting Methods," *Journal of Physics: Conference Series IOP Publishing,* vol. 1255, no. 1, 2019.

[24] M. Macieira, P. Mendonça, J. Miranda Guedes and A. Tereso, "Evaluating the efficiency of membrane's refurbishment solutions to perform vertical extensions in old buildings using a multicriteria decision-support model," *Architectural Engineering and Design Management Taylor & Francis,* vol. 18, no. 1, pp. 1-25, 2022.

[25] Y. Chen, "Application of analytic hierarchy process (AHP) and simple additive weighting (SAW) methods in mapping flood-prone areas," *Frontiers in Artificial Intelligence and Applications,* vol. 341, pp. 435-441, 2021.

[26] Melvin, T. Sutrisno and D. E. Herwindiati, "Decision support system for election and evaluation of assistant lecturer using analytical hierarchy process and simple additive weighting: Case study faculty of information technology Tarumanagara university," *IOP Conference Series: Materials Science and Engineering,* vol. 1007, no. 1, 2020.

[27] M. Saputra, O. S. Sitompul and P. Sihombing, "Comparison AHP and SAW to promotion of head major department SMK Muhammadiyah 04 Medan," *Journal of Physics: IOP Conference Series,* vol. 1007, no. 1, 2018.

[28] A. Cahyapratama and R. Sarno, "Application of Analytic Hierarchy Process (AHP) and Simple Additive Weighting (SAW) methods in singer selection process," *International Conference on Information and Communications Technology, ICOIACT 2018,* Vols. 2018-January, pp. 234-239, 2018.

[29] F. Noviyanto, A. Tarmuji and H. Hardianto, "Food Crops Planting Recommendation Using Analytic Hierarchy Process (AHP) And Simple Additive Weighting (SAW) Methods," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH,* vol. 9, p. 2, 2020.