Performance Evaluation of Supervised Machine Learning Algorithms Using Multi-Criteria Decision Making Techniques

Akinsola J. E. T [*1]
Department of Computer Science,
Babcock University (BU),
Ilishan-Remo, Ogun State, Nigeria
akinsolajet@gmail.com

Awodele, O.[*2]
Department of Computer Science,
Babcock University (BU),
Ilishan-Remo, Ogun State, Nigeria
awodeleo@babcock.edu.ng

Kuyoro S. O.[*3]
Department of Computer Science,
Babcock University (BU),
Ilishan-Remo, Ogun State, Nigeria
afolashadeng@gmail.com

Kasali, F, A.[4]**
Department of Computer Science
Mountain Top University.
Prayer City, Ogun State, Nigeria
kasalifunmilayo@gmail.com

Abstract - *The choice of classification algorithm in Machine Learning (ML) is a major issue cutting across several disciplines due to the uncertainty in human judgment in the ranking of performance metrics. The process of algorithm selection can be modelled as Multi-Criteria Decision Making (MCDM) problem which involves more than one criterion. In this work, seven classification algorithms, and ten performance criteria were considered to test the proposed Fuzzy Analytical Hierarchical Process (FAHP) and Technique or Order of Preference by Similarity to Ideal Solution (TOPSIS) model. The model was developed using respective priority weights based on AHP and fuzzy logic principle. Pairwise comparison matrix was formulated based on decision makers' judgments that were aggregated and normalized. The study applied FAHP in assigning weights to the criteria and ranking the performance criteria, while Simple Additive Weighting (SAW) and TOPSIS were implemented in MATLAB to rank the classifiers for comparison. Fuzzification was done using Triangular Fuzzy Numbers (TFNs) and defuzzification was done using Graded Mean Integration (GMI) approach. Consistency of the decision makers' judgments were obtained using Saaty's Eigen value and Eigen vector approach. Unlike the usual practice, in addition to Accuracy as the benchmark for selecting an algorithm, the Kappa Statistic measure was also considered. The result of algorithm performance evaluation shows that Logistic Regression (LRN) from Waikato Environment for Knowledge Analysis (WEKA) has the highest Kappa Statistic. Also, FAHP result for criteria weights determination shows that Kappa Statistic has the highest priority weight then Accuracy based on decision makers' judgments. FAHP Consistency Ratio (CR) has a value of 0.017, which is less than 10%. Hence, criteria weights results are reliable. The TOPSIS ranking result of ML algorithms shows that LRN has the highest ranking. The study concluded that LRN being the algorithm with the highest ranking is considered as the best classifier. Therefore, MCDM techniques can be used in selecting the best Supervised Machine Learning Algorithm for classification and regression.*

Keywords— Machine Learning Algorithms, Multi-Criteria Decision Making (MCDM), Fuzzy Analytical Hierarchy Process (FAHP), Triangular Fuzzy Numbers (TFN), Technique or Order of Preference by Similarity to Ideal Solution (TOPSIS))

I. INTRODUCTION

Despite the advancements in machine learning classification approaches, lately, the attention of academics, scholars and scientist have moved to comparative criteria weighting, using multi- criteria. When it comes to medical knowledge attainment, relative criteria weighting has been proposed [34] which makes use of AHP method [61]. The limitation of AHP is that it cannot unravel non straight models, that is, one whose yield is not straightforwardly corresponding to its information. Hence, the need for Fuzzy Analytical Hierarchical Process (FAHP). The researchers utilized average training time, accuracy and memory usage as the criteria. Training time could give biased estimate of the error rate of the classifier and Accuracy has inability to find the relevant cases within a dataset.

Five multi-criteria decision making methods, including TOPSIS [76], ELimination Et Choix Traduisant la Realité III (ELECTRE III) [16], Grey Relational Analysis (GRA), Vlse kri-terijumska Optimizacija I KOmpromisno Resenje (VIKOR), and Preference Ranking Organization Method for Enrichment of Evaluations II (PROMETHEE II) have been discussed in article [38]. Therefore, there is a need for hybrid MCDM techniques that are closely applicable in the type of problem such as performance evaluation of supervised machine learning algorithms which ensures that the disadvantage of one technique will be compensated for in the other. The development of ultra-modern approaches in solving classification ranking problems in relation to Multi Criteria Decision Making (MCDM) are noteworthy.

Machine Learning (ML), a sub-field of Artificial Intelligence (AI), which focuses on the task of enabling computational systems to learn from data about how to perform a desired task automatically. Machine learning has many applications including decision making, forecasting or predicting and it is a key enabling technology in the deployment of data mining and big data techniques in the diverse fields of healthcare, science, engineering, business and finance [5]. The goal of machine learning is to enable a system to learn from the past or present and use that knowledge to make predictions or decisions regarding unknown future events. Machine learning has been applied in several areas of research and combination of ideas from many disciplines including Artificial Intelligence, Probability and Statistics, Computational Complexity, Information Theory, Psychology and Neurobiology, Control Theory, Evolutionary Models and Philosophy.

ML has achieved success due to its strong theoretical foundations and its multi-disciplinary approach by integrating aspects of Computer Science, Applied Mathematics and Statistics, among others. According to [62], Machine Learning has proven to be of great value in data mining problems especially where large databases contain valuable implicit regularities that can only be discovered automatically. The field of machine learning is one of the fastest growing areas of Computer Science with far reaching applications [64]. Machine learning approach when utilized in solving problems involves a number of choices such as choosing the type of training experience, the target function to be learned, a representation for this target function, and an algorithm for learning the target function from training examples. Consequently, choosing that algorithm for learning is a herculean task. ML is known to have three major learning approaches namely Supervised, Unsupervised and Semi-supervised. Machine learning classification algorithms selection can be modeled as a Multi-Criteria Decision Making (MCDM) problem. Choosing a best fit classifier in any machine learning problem is always confronted my biasness and subjectivity, hence, the need for Multi-Criteria Decision making approaches and techniques for effective decision making.

Multiple Criteria Decision Making or Multi-Criteria Decision Making (MCDM) refers to making decisions in the presence of multiple, usually conflicting criteria. MCDM problems are common in everyday life. In business context, MCDM problems are more complicated and usually of large scale. In general, there exist two distinctive types of MCDM problems due to the different problems settings: *one type having a finite number of alternative solutions and the other an infinite number of solutions*. The development of the MCDM discipline is closely related to the advancement of computer technology. Though MCDM problems are widespread all the time, MCDM as a discipline only has a relatively short history of about 30 years [78]. The MCDM techniques are remarkable to judge different alternatives on various criteria [31]. MCDM can be effectively used in decision making among multiple criteria [46] and it is an important issue because it involves many criteria. The choice of the decision to be made has tremendous implication on the model that will finally be built from the chosen ML algorithm.

Making decisions can be described as a cognitive action that entails judging numerous available alternatives or options so as to choose the preferred alternative that will be effective as well as acceptable by decision makers [69]. Effective decision making is an action that involves a careful and systematic process; in other to achieve the main aim which is to select the best choice among numerous alternatives. Different MCDM techniques evaluate classifiers and ensembles from different aspects and thus they may produce divergent rankings of classifiers and ensembles. Hence, there is a need to resolve these disagreements by utilizing hybrid approaches. The most critical aspect of any MCDM problem is the determination of weights for criteria, that is, the attributes. These attributes are referred to as performance metrics in Machine Learning. In criteria weights' determination, unbiased expert decision makers' judgments are crucial.

This study employs seven machine learning algorithms to evaluate the benefit and cost performance metrics (attributes of the algorithms) using Waikato Environment for Knowledge Analysis (WEKA) and then utilized Fuzzy Analytical Hierarchy Process (FAHP) an MCDM method to determine the weight of each criterion through expert Decision Makers' (DM) judgments. Also, a hybrid MCDM methods such as SAW and TOPSIS were implemented in MATLAB separately to rank the classifiers. The results were compared after which the best classifier was chosen.

The organization of the research is presented as follows: Section II provides literature review of machine learning algorithms and multi-criteria classification problems, section III characterizes related work, Section IV explains the methodology, Section V presents experimental, results and lastly, Section VI gives conclusion and recommendation for further studies.

II. LITERATUTRE REVIEW

A. *Machine Learning Overview*

Supervised machine learning algorithms classification is very popular as a result of the various application areas such as credit risk analysis, development of intelligent data analytic software, clinical decision support development, inventory classification, personnel selection and food choice determination among others. Machine learning algorithm are very sensitive to the characteristics and structure of the datasets being used for any classification. Different algorithm selection have been proposed by different researchers. However, these methods do not consider the uncertainty in input datasets. This uncertainty can be corrected by using Multi Criteria Decision Making (MCDM) techniques in the ranking of the performance metrics and algorithm selection. Therefore, it is highly essential to choose appropriate metrics in algorithm selection. This study is concerned with supervised classification algorithms and not the unsupervised classification algorithm which do not consider the class label information in the dataset.

To study machine learning mathematically, a formal definition of the problem is needed and the learning model should answer the following questions: 1) what is being learned? 2) How is the data being generated? 3) How is the data presented to the learner? 4) What is the goal of learning in this model? [14]. Machine learning has been applied in solving the problems of classification, regression, ranking, clustering and dimensionality or manifold learning. The supervised learning approach of machine learning uses reinforcement learning, dimension reduction learning, regression, time series prediction and classification algorithms for learning [73] and [40]. Algorithm selection is defined as learning a mapping from feature space to algorithm performance space, and acknowledge the importance of selecting the right features to characterize the hardness of problem instances [54].

B. Machine Learning Algorithms for Classification

This study utilized seven classification algorithms for evaluation of performance metrics which belong to the following four categories / classes of classifiers accordingly such as ***Bayes*** (Bayes Network (BNK) and Naive Bayes (NBS)), ***function*** (Logistic Regression (LRN), Sequential Minimal Optimization (SMO) and Multilayer Perceptron (MLP)), ***Tree*** (J48) and ***Lazy*** **(**Instance Based Learner (IBK)). It is to be noted that SMO is a variant of Support Vector Machine (SVM). Also MLP is a variant of Artificial Neural Networks (ANNs). All these algorithms are implemented using WEKA on public-domain credit dataset.

1) Bayes classifiers

Bayes Network (BNK) and Naive Bayes (NBS) are the Bayes classifiers used in this study. Bayes classifiers are probabilistic classifier based on Thomas Bayes' basic law of probability which is referred to as Bayes theorem that is shown in (1)

$$P\,(B/A) = \frac{P(B/A)\ \ x\,P(A)}{P(B)} \tag{1}$$

Equation 1 presents the relationship between the probabilities and the conditional probabilities of A and B. A Naïve Bayes classifier is a simple algorithm with the assumption of independent attributes, which means the algorithm assumes that attributes do not affect each other by means of probability. Bayesian networks are relatively sophisticated algorithms to analyze probabilities under ambiguity and consequently, they allow capturing more complex information from the data analyzed. Accordingly, in specific, circumstances where Naïve Bayes classifiers achieve poorly, a Bayesian Network may be expected to attain better learning outcomes [19]. Bayesian Network encodes probabilistic relationships for a set of interest nodes in uncertain conditions using graphical models. Naive Bayesian (NB) Networks are very simple Bayesian Networks which are composed of directed acyclic graphs with only one parent (representing the unobserved node) and several children (corresponding to observed nodes) with a strong assumption of independence among child nodes in the context of their parent [24]. Bayes classifiers are usually less accurate than other more sophisticated learning algorithms (such as ANNs).

2) Function Classifiers

The function classifiers used in this study are Logistic Regression (LRN), Sequential Minimal Optimization (SMO) and Multilayer Perceptron (MLP).

a) Logistic Regressi on: This is a classification function that uses class for building and a single multinomial logistic regression model with a single estimator. Logistic regression usually states where the boundary between the classes exists. Also, it states that the class probabilities depend on the distance from the boundary, in a specific approach [39]. This moves towards the extremes (0 and 1) more rapidly when dataset is larger. These statements about probabilities which make logistic regression more than just a classifier. It makes stronger, more detailed predictions, and can be fit in a different way; but those strong predictions could be wrong. Logistic regression is an approach to prediction, like Ordinary Least Squares (OLS) regression. However, with logistic regression, prediction results in a dichotomous outcome [42]. Logistic regression is one of the most commonly used tools for applied statistics and discrete data analysis. Logistic regression is linear interpolation.

b) Sequential Minimal Optimization (SMO): This is a variant of Support Vector Machines (SVMs). SVM models are closely related to classical multi-layer perceptron neural networks. SVMs revolve around the notion of a margin in either side of a hyperplane that separates two data classes [3]. Maximizing the margin and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it has been proven to reduce an upper bound on the expected generalisation error [37]. SVM classifier takes the inputs of different classes, and then builds input vectors into a feature space to find the best separating hyperplane. The hyperplane which places at the maximum distance from the nearest points of the dataset is defined as optimal [33], [66].

c) Multi-Layer Perceptron (MLP): MLP is a variant of Artificial Neural Network (ANN). This is a classifier in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem as in standard neural network training [72]. ANN is a learning algorithm which is capable of solving classification problems. An ANN model is composed of a number of parallel, dynamic, and interconnected neuron network systems. A neuron operates a defined mathematical processor using inputs to produce outputs [35]. Other well-known algorithms are based on the notion of perceptron [55]. Perceptron algorithm is used for learning from a batch of training instances by running the algorithm repeatedly through the training set until it finds a prediction vector which is correct on all of the training set. This prediction rule is then used for predicting the labels on the test set [37].

3) Tree Classifier

J48 is an extension of Dichotomiser 3 (ID3). The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, and so on. It is a decision tree algorithm. Decision Tree algorithm is used to find out the way the attributes-vector behaves for a number of instances. Also on the bases of the training instances, the classes for the newly generated instances are being found [36]. This algorithm generates the rules for the prediction of the target variable. With the help of tree classification algorithm, the critical distribution of the data is easily understandable [41].

4) Lazy Classifiers

Instance Based Learner (IBK) belongs to the class of Lazy classifiers. It is a k-Nearest Neighbor (k-NN) algorithm. The procedure follows a simple and easy way to classify a given dataset through a certain number of clusters (assume k clusters) fixed apriori. K-means algorithm is be employed when labeled data is not available [1]. It uses general method of converting rough rules of thumb into highly accurate prediction rule. Given weak learning algorithm that can consistently find classifiers (rules of thumb) at least slightly better than random, say, accuracy of 55%. But with sufficient data, a boosting algorithm can probably construct single classifier with very high accuracy, say, 99% [63].

C. Multi-Criteria Decision Making (MCDM) Overview

Multi-Criteria Decision Making (MCDM) refers to screening, prioritizing, ranking, or selecting a set of alternatives usually under independent, incommensurate or conflicting attributes [4]. MCDM can also be referred to as Multi-Criteria Decision Aiding (MCDA), Multiple Attribute Decision Analysis (MADA) or Multi-Attribute Decision Making (MADM). Decision making problems involve attributes, which can be single or multiple. Where many attributes are involved, this is called Multi-Criteria Decision Making (MCDM). The Multi-Attribute Decision Making (MADM) problem has been found to be of high theoretical value and has been applied in different research areas both in the academia and industry [20]. MCDM techniques can aid decision making in cases where conflicting multiple criteria exists and it can efficiently take care of both the quantitative and qualitative choices. It also has the ability to merge historical data and expert judgments by measuring subjective opinions [48].

Classification, selection and assessment problems have limited number of alternative solutions. Whereas in design problems, an attribute may take any value in a range. Therefore, the potential alternative solutions could be infinite. If this is the case, the problem is referred to as multiple objective optimization problem instead of multiple attribute decision problem. This study is concerned with problems with a finite number of alternatives such as selecting a classifier from a machine learning algorithms classification. It should be noted that, there is not always definite criteria of selection, and decision makers have to take into account a large number of criteria. There is a need for simple, systematic and logical methods or mathematical tools to guide decision makers in considering a number of selection criteria and their interrelations. Depending upon the domain of alternatives, MCDM problems are usually subdivided into continuous and discrete types. MCDM problems have two classifications such as Multiple Objective Decision Making (MODM) and Multiple Attribute Decision Making (MADM). MODM methods have decision variables values that are determined in a continuous or integer domain, with a large number of alternative choices. MADM methods are generally discrete, with limited number of pre-specified alternatives. The MODM and MADM problems are formulated into decision matrix

Each decision matrix has four main parts, namely: (a) alternatives (b) attributes or criteria (c) weight or relative importance of each attribute and (d) measures of performance of alternatives with respect to the attributes.

Decision Matrix

A MCDM problem may be described using a ***decision matrix***. Suppose there are m alternatives to be assessed based on n attributes, a decision matrix is a $m \times n$ matrix with each element Xij being the j-th attribute value of the i-th alternative. That is, Xij is the performance rating of i-th alternative with respect to j-th attribute,

where $A1, A2,..., Am$ are feasible alternatives, $C1, C2,...,Cn$ are attributes (criteria), and it is also assumed that the decision maker has determined the weights of relative performance of the decision criteria (denoted as Wj, for j = 1,2,3,..., N), where wj is a weight (significance) of j-th attribute. This assertion is best summarized in Figure 1. An MCDM problem can be concisely expressed in the matrix format as shown below:

$$A = \{a_i \mid i = 1, 2, 3, \ldots.n \} \quad (2)$$

$$C = \{c_j \mid j = 1, 2, 3, \ldots.. m \} \quad (3)$$

$$W = \{w_1, w_2, w_3, \ldots.., w_n\} \quad (4)$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

(5)

Given the previous definitions, then the general MCDM problem can be defined as follows [81]:

Definition 1:
Let A = { Ai, for i = 1,2,3,... , N} be a (finite) set of decision alternatives (Alt) and G = {gj, for j = 1,2,3,..., M} a (finite) set of goals (gj are also called decision criteria, or just criteria) according to which the desirability of an action is judged. Determine the optimal alternative A with the highest degree of desirability with respect to all relevant goals gj.*

	Criteria					
	C_1	C_2	C_3	C_4	…..	C_N
Alt	W_1	W_2	W_3	W_4	…..	W_N
A_1	a_{11}	a_{12}	a_{13}	a_{14}	…..	a_{1N}
A_2	a_{21}	a_{22}	a_{23}	a_{24}	…..	a_{2N}
A_3	a_{31}	a_{32}	a_{33}	a_{34}	…..	a_{3N}
.	.	.	.	.	…..	.
.	.	.	.	.	…..	.
A_M	a_{M1}	a_{M2}	a_{M3}	a_{M4}	…..	a_{MN}

Fig. 1 A Typical Decision Matrix.

In a typical MCDM evaluation, attributes can be classified into two main categories such as cost attributes and benefit attributes. In this study, the benefit criteria (attributes) utilized are Accuracy, Kappa Statistic, True Positive Rate (TPR), True Negative Rate (TNR), Precision, *F*-measure, and Area Under Curve (AUC) while the cost criteria (attribute) used are Mean Absolute Error (MAE), Training Time and Test Time. As regards benefit attributes, the higher score is assigned to the alternative in which performance rating is higher, i.e., preferable is a maximum of *j-th* attribute. In contrast to the previous, in relation to cost attributes, higher score is assigned to the alternative which performance rating is lower, i.e., the minimum of *j-th* attribute is preferable.

There are three approaches that can be used to find the value of weight of an attribute (criterion), namely subjective approach, objective approach and integration approach. Integration approach is between subjective approach and objective approach. Each approach has advantages and disadvantages. On the subjective approach, weighting value is determined based on decision makers, so some of the factors in the process of rank alternative can be determined freely. On the objective approach, value of weight is calculated mathematically, so as to ignore subjective from the decision makers. This study adopts the use of integration between objective approach and subjective approach.

D. MCDM Methods

MCDM can be grouped as either Multi Objective Decision Making (MODM), Multi Attribute Decision Making (MADM) or a combination of both. Over the years, different MCDM techniques have been theorized as depicted in the hierarchical structure in Fig. 2.

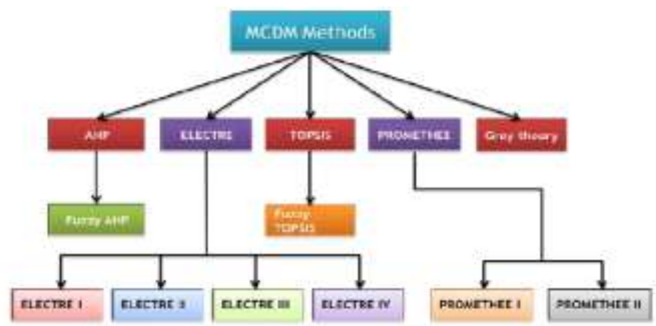

Fig. 2 Various MCDM methods [6]

Other methods include the Multi-Attribute Utility Theory (MAUT), Data Envelope Analysis (DEA), Goal Programming (GP), Analytic Network Process (ANP), Simple Additive Weighting (SAW), Technique or Order of Preference by Similarity to Ideal Solution (TOPSIS), Simple Multi-Attribute Rating Technique (SMART), Elimination and Choice Translating Reality (ELECTRE I, II, III and IV), Fuzzy Set Theory (FST), Preference Ranking Organization Method for Enrichment Organization (PROMETHEE I and II), Case Based Reasoning (CBR) and Vlse-Kriterijusca Optimizacija I Kompromisno Resenje in Serbian (VIKOR). These techniques can also be categorized broadly into three sections as outlined by [75]:

a) *Value measurement models:* These models measure a crisp value for each alternative and a weight w, which denotes the significance of the criterion is assigned to each criterion. Examples include SAW and AHP.

b) *Goal and reference level models:* These models assess how good alternatives reach established goals. TOPSIS is a good example of this.

c) *Outranking models:* These models assess the alternatives pairwise for each criterion by discovering the strength of selecting one over the other. Examples include ELECTRE and PROMETHEE.

E. AHP and FAHP

AHP is a useful mathematical technique that is employed to solve MCDM problems, where a choice has to be made from a number of alternatives based on their relative importance [7], [71]. It enables the development of numerical score or weight to rank each decision alternative based on criteria and the computation of consistency ratio to ascertain the reliability of the comparative judgments represented in the comparison matrix. [30] presented a detailed review of AHP, application areas and developments since its inception in the 70s. The AHP is combined with other methods or techniques, such as mathematical programming, data envelopment analysis, fuzzy theory, and meta-heuristics [26], [27]. AHP has some characteristics which include the ability to handle decision situations involving subjective judgments, multiple decision makers and the ability to provide measures of consistency of preference [74]. It relies on the judgments of experts to derive priority scales through pairwise comparison of decision elements at each level as shown in Figure 3.

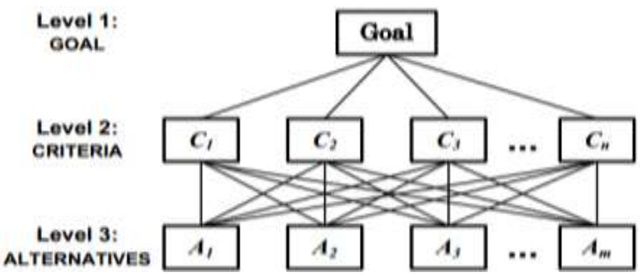

Fig. 3 Hierarchical Decision Structure [8]

The Analytic Hierarchy Process (or AHP) [56], [57], [58], [59], [60] is based on decomposing a complex MCDM problem into a system of hierarchies (more on these hierarchies can be found in [57]). The final step in the AHP deals with the structure of an M×N matrix (where M is the number of alternatives and N is the number of criteria). This matrix is constructed by using the relative importance of the alternatives in terms of each criterion. The vector *(ai1, ai2, ai3, ..., aiN)* for each i is the principal eigenvector of an N×N reciprocal matrix which is determined by pairwise comparisons of the impact of the M alternatives on the i-th criterion (more on this, and some other related techniques, is presented in section IV.

The similarity between the Weighted Sum Method (WSM) and the AHP is evident. The AHP uses relative values instead of actual ones. Thus, it can be used in single or multi-dimensional decision making problems. Some evidence is presented in [57] which supports the technique for eliciting numerical evaluations of qualitative phenomena from experts and decision makers. As a result, the judgments are becoming unreliable and subjective. However, to deal with the impreciseness of experts' judgments, the fuzzy set theory has been selected. Unlike the classic Analytic Hierarchy Process (AHP), Fuzzy AHP (FAHP) uses fuzzy logic, which allows a more accurate evaluation of linguistic criteria. The FAHP model is based on fuzzy sets theory, in which the membership of the given element is determined by the membership function. Fuzzy decision variable values are described by a membership function which is between zero and one. The theory was developed by Zadeh and has become widely used in pairwise comparison [79].

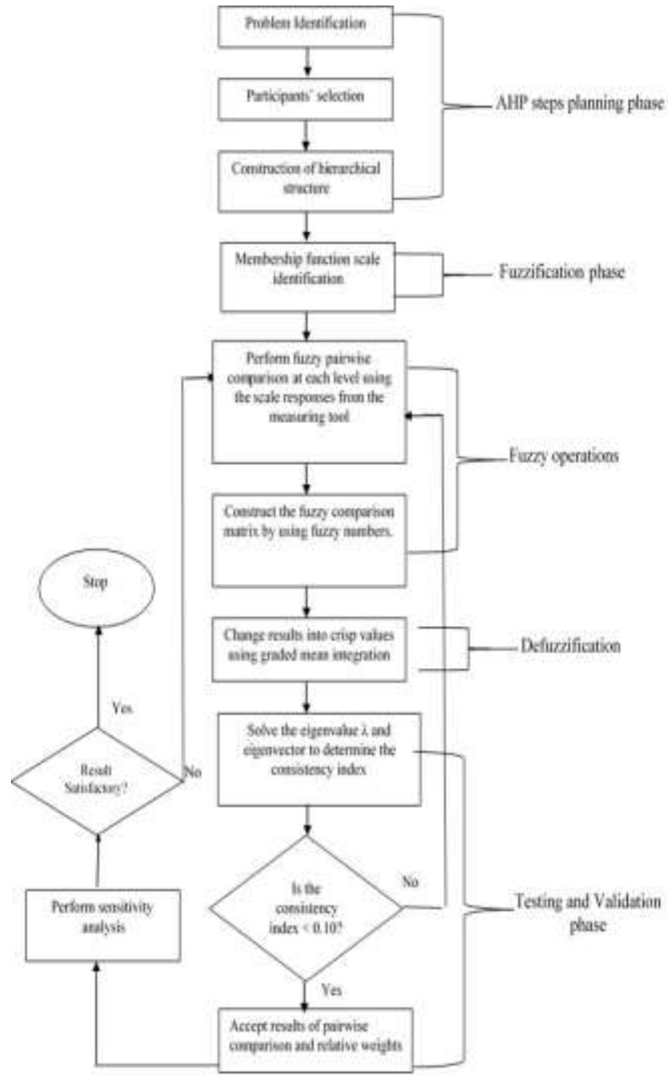

Fig. 4 AHP and Fuzzy Logic Schematic Diagram

The Fuzzy AHP approach is represented by Triangular Fuzzy Numbers (TFNs). The numbers can be identified as triple M = (l, m, u), where its membership function is defined in [12] as:

In equation (6) l, m and u stand for the lower, medium and upper values of M respectively ($l \leq m \leq u$). In special case where all three numbers are equal ($l = m = u$), then we are dealing with no-fuzzy numbers. The operations on TFNs include addition, multiplication and inverse operations according to the extension principles. Suppose M_1 and M_2 are two non-negative TFNs where $M_1 = (a_1, a_2, a_3)$ and $M_2 = (b_1, b_2, b_3)$, and $\alpha \in \mathbf{R}^+$ then the following holds:

i. $M_1 (+) M_2 = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$
ii. $M_1 (-) M_2 = (a_1 – b_3, a_2 - b_2, a_3 – b_1)$
iii. $\alpha M_1 = (\alpha a_1, \alpha a_2, \alpha a_3)$
iv. $M_1 (x) M_2 \approx (a_1 \cdot b_1, a_2 \cdot b_2, a_3 \cdot b_3)$
v. $M_1^{-1} \approx (a_1, a_2, a_3)^{-1} \approx (1/a_3, 1/a_2, 1/a_1)$
vi. $M_1/M_2 \approx (a_1/b_3, a_2/b_2, a_3/b_1)$

$$\mu_M(x) = \begin{cases} \frac{x}{m-l} - \frac{l}{m-l}, x \in [l,m], \\ \frac{x}{m-u} - \frac{u}{m-u}, x \in [m,u], \\ 0, \quad otherwise. \end{cases} \tag{6}$$

The Table II shows Saaty's scale and the corresponding Triangular Fuzzy Numbers (TFNs) and Figure 4 shows AHP and fuzzy logic schematic diagram.

F. Simple Additive Weighting (SAW)

Simple Additive Weighting (SAW) method is also referred to as Weighted Sum Model (WSM) [17] and is probably the best known and most widely used MCDM method [4]. SAW method also known as Weighted Linear Combination (WLC) or Scoring method is one of the best and simplest type of multiple attribute decision making method. The basic logic of the SAW method is to obtain a weighted sum of performance ratings of each alternative over all attributes.

An evaluation score is calculated for each alternative by multiplying the scaled value given to the alternative of that attribute with the weights of relative importance directly assigned by decision maker followed by summing of the products for all criteria. The advantage of this method is that it is a proportional linear transformation of the raw data, which means that the relative order of magnitude of the standardized scores remains equal. In this method, each attribute is given a weight and the sum of all weights must be 1. If there are *M* alternatives and *N* criteria then, the best alternative is the one that satisfies (in the maximization case) the following expression [17]. The overall score of an alternative is given by (7).

$$P_i = \sum_{j=1}^{m} W_j m_{ij} \tag{7}$$

where m_{ij} signifies the normalized value of the attributes, *Wj* is the weight of importance of the *j*-th criterion. The assumption that governs this model is the **additive utility assumption.** That is, the total value of each alternative is equal to the sum of products. The alternative with the highest value of P_i is chosen as the alternative with the highest priority [22]. Detailed procedures for this method is found in [43]. This study implemented SAW in determining the best alternative in order to make a final decision in selecting the best fit classifier.

G. Technique or Order of Preference by Similarity to Ideal Solution (TOPSIS)

TOPSIS was proposed by [29] to assist decision makers to organize, analyze, compare and rank alternatives [65] in a problem under study. It compares each alternative precisely based on information obtained from evaluation matrices and weights [13]. The basic idea behind TOPSIS is that, aside a chosen alternative having the shortest distance from the positive ideal solution, it should also have the farthest distance from the negative ideal solution (as referenced in [28]).

TOPSIS is easy to apply mathematically, yields a relatively reasonable output but the consistency of decision makers opinions cannot be ascertained [77] which is complimented by combining it with AHP technique. That is why this study utilized F-AHP in determining the priority weights for the criteria, then TOPSIS is implemented to rank the alternatives after the decisions must have been made by experts' judgment.

III. RELATED WORK

This section of the study presents the related work on how various techniques have been utilized in the selection of best classification algorithms. Machine learning algorithms have several features. Therefore choosing a particular feature by random selection is not a good strategy. Choosing this particular feature upon which decision will be based has been so challenging and somewhat ambiguous. Machine learning algorithms have been used in diverse areas of applications. The methodology is firmly rooted in machine learning with consideration on various machine learning algorithms classification metrics commonly used in algorithm performance measurement by firstly carrying out data mining on the datasets.

A. Algorithm Performance Measures and MCDM Approaches for Algorithms Comparison and Selection

[2] carried out evaluation amongst eight classifiers with 100 different classification problems by utilizing extended measures of average accuracy (True Positive Rate, True Negative Rate and Accuracy) as well as time complexity (training time and testing time). The drawback of True Positive Rate and True Negative Rate is that they can easily be ruled out by further test *or* investigation which often require good judgment. Another flaw is their failure to tell exactly what we wish to know. At the same time, Accuracy has inability to find the relevant cases within a dataset. Also, using training time and test time in choosing the optimal ML algorithm could give biased estimate of the error rate of the classifier. Hence, the need for MCDM approach. Likewise, for numerous real-world applications, the performance evaluation of various classifiers have been done, for instance, handwritten recognition [67], color prediction of rice paddy plant leaf [68], prediction of diabetes mellitus [32]; [47]. The most generally applied criteria for algorithms evaluation are the Adjusted Ratio of Ratio

(ARR) [9] and performance of algorithm (PAlg) on dataset [70], which used accuracy and time. [52] applied Root Mean Squared Error (RMSE) and Pearson Product-Moment Correlation Coefficient (PMCC) [21] in order to evaluate and in turn recommend the best classification algorithm. The flaw with RMSE is that it gives a relatively high weight to large errors which might be misjudging. Also, the determination of what exactly should be the threshold of lesser (smaller) error in RMSE in order to choose the best classifier. The disadvantages of PMCC is that it assumes that there is always zero linear relationship between the variables which might not be the case at all times.

[18] utilized the Analytical Hierarchy Process (AHP) [57] to study multiple criteria of ABC analysis. Later, [44] and [45] also applied AHP-based multi-criteria inventory classification methods. Since the AHP provided an easy way to combine various kinds and number of attributes, numerous applications have been utilized using many different attributes. On the other hand, as in most real-cases, constraints are not certain, and some criteria are difficult to determine accurately. Therefore, fuzzy logic and its theory were implemented to have many flexibility in decision parameters or attribute values which had such ambiguity [80]; [11]. However, this technique should not be applied in isolation. In the same vein, [49] proposed a fuzzy model and a probabilistic model for ABC analysis with uncertain data. In the study, demand which is also referred to as benefit and cost attributes were considered as fuzzy information. [53] presented a fuzzy model for multi-attribute inventory classification. [10] Proposed a web-based inventory classification system model by considering the fuzzy AHP (FAHP) approach. The study combined fuzzy theory and multi-attribute inventory analysis for a real case study. However, consideration was not given to hybrid approach for effective results comparison.

[15] proposed PROMETHEE II for ranking machine learning classifiers using multi-criteria approach by evaluating five classification algorithms (C4.5, Naive Bayes, SMO, kNN and BayesNet) using UCI database. The study showed that SMO achieved the highest overall ranking. The drawback is that it does not provide a clear method by which to assign weights and it requires the assignment of values but does not provide a clear method by which to assign those values. Therefore there will not be consistency in weights assignment. Hence, its choice require previous use of hierarchical process which the study did not consider.

B. *Statistical Test Approaches for Algorithms Comparison and Selection*

The three basic subtasks in machine learning are Model Evaluation, Model Selection, and Algorithm Selection [51], [50]. Several techniques have been suggested in machine learning Model Comparison (MC) and Algorithm Comparison (AC) as well as selection such as McNemar's test, Cochran's Q test, resampled paired t-test, k-fold cross-validated t-test, the difference in proportions test, Nested cross-validation, Multiple independent test, 5x2 Cross Validated (5x2cv) paired t-test and Combined 5x2 Cross Validated (5x2cv) F test.

1) McNemar's test: this is a non-parametric statistical test for paired comparisons that can be applied to compare the performance of two machine learning classifiers. It is a MC technique. It has low false positive rate and fast, only needs to be executed once. It is applicable to large dataset. The drawbacks are it is not repetitive and cannot handle more than two algorithms.

2) Cochran's Q test: this can be regarded as a generalized version of McNemar's test that can be applied to compare three or more classifiers. It is a MC technique which is efficient for large dataset. It only tells that a group of models differs or not but not how models differ. Hence, there is still a need to perform post hoc tests

3) Resampled paired t-test: Thi is also called k-hold-out paired t-test. It is a popular method for comparing the performance of two models (classifiers or regressors). This technique has many drawbacks such as high false positive rate and computationally very expensive aside the fact that it cannot be used for multiple model comparison. Therefore, it is not recommended to be used in practice.

4) k-fold Cross-validated Paired t-Test: this is similar to the resampled paired t-test. It is an AC technique. It is a statistical testing technique which addresses some of the drawbacks of the resampled paired t-test procedure, however, this method still has problem in that the training sets overlap and is hence also not recommended to be used in practice. Additionally, it gives somewhat elated false positive rate and requires refitting to training sets. It has k times more computations than McNemar's test which could be very expensive.

5) The difference in proportions test: this is a test technique in which the sampling method for each population is simple random sampling. The samples are independent. Each sample includes at least 10 successes and 10 failures. It is cheap to compute. The drawback is that it has high false positive rate which usually incorrectly detect a difference when there is none.

6) Nested cross-validation: this has emerged as one of the popular or somewhat recommended methods for comparing machine learning algorithms. It is an AC technique with reduced bias compared to regular k-fold cross-validation. The drawback is that is it efficient when comparing small dataset which gives biased estimate of the true generalization error when dealing with large dataset.

7) Multiple independent test: this is an AC technique which is applicable on large dataset but it cannot be used to evaluate models, therefore, the need to resort to cross-validation procedures such as k-fold cross-validation, the 5x2cv, or nested cross-validation.

8) The 5x2cv paired t-test: this is a procedure for comparing the performance of two models (classifiers or regressors) that was proposed to address shortcomings in other methods such as the resampled paired t-test and the k-fold cross-validated paired t-test. It is an AC technique which is slightly more powerful than McNemar's test. The drawbacks are low false positive rate (similar to McNemar's test) and more runtime. It is only recommended if computational efficiency is not an issue. That is, it has ten times more computations than McNemar's test.

9) The 5x2cv combined F-test: this is a procedure for comparing the performance of models (classifiers or regressors). It is an AC technique that was proposed as a more robust alternative to 5x2cv paired t-test technique. The combined 5x2cv F-test is available from MLxtend (Raschka, 2018). The drawback is that is it efficient for small dataset.

As a result of these drawbacks, flaws and deficiencies arising from the aforementioned methods and techniques, there is a need for efficient techniques that overcomes these limitations, hence the choice of MCDM for ML algorithm selection.

IV. METHOGOLOGY

A. Implementation Phases

This study implemented MCDM techniques in determining the best machine learning algorithms classification using FAHP, SAW and TOPSIS techniques. The No Free Lunch (NFL) theorem states that no algorithm can outperform all other algorithms when performance is amortized over all measures. Therefore, ML algorithm selection must be critically examined and evaluated. Many studies indicate that classifiers' performances vary under different datasets and circumstances. It is in this light that, efficient algorithm selection techniques must be institutionalized. This research provides a way of choosing the desired algorithm in any problem domain. It makes use of FAHP to assign priority weights to the ten performance metrics which comprise of seven *benefit criteria* such as *Accuracy, Kappa Statistic, True Positive (TP) Rate, True Negative (TN) Rate, Precision, F-measure, and Area Under the Curve (AUC*) which are discussed in (8) to (14). The other three *cost criteria* performance measures are *Mean Absolute Error (MAE), Training time and Test time*. The weights are determined through the decision makers' expert judgment by carrying out pairwise comparisons of the ten criteria. The methodology is broadly divided into four major phases below:

a) Data mining phase: this is concerned with data cleaning, data integration and data transformation. Training and testing of the selected classification algorithms on randomly sampled partitions (i.e., 10-fold cross-validation) were carried out using WEKA for algorithms performance evaluation. Equations 8 to 14 were used in WEKA to arrive at the performance evaluation results of the algorithms shown in Table III.

b) Assigning Criteria weights: An Online Analytic Hierarchy Process (AHP) software was used to formulate the pairwise comparison matrix. Then, five decision makers were consulted that implemented the pairwise comparisons using the ten metrics (both benefits and cost criteria) to determine the respective weights of the criteria which were aggregated and normalized. A total of 45 decisions were made by each expert decision maker. Thereafter, FAHP was employed to compute the priority weights for the criteria. Fuzzification was done using Triangular Fuzzy Numbers (TFN) as shown in Table II and defuzzification was done using Graded Mean Integration (GMI) approach. Equations 15 to 21 were applied to implement the FAHP after the pairwise comparisons from expert decision makers have been conducted. This stage is essential because the weights of criteria are very crucial for raking of the algorithms using SAW and TOPSIS techniques.

c) Ranking of Algorithms: SAW and TOPSIS MCDM methods were implemented on the results obtained from algorithms performance evaluation conducted in (a) above in relation to normalized relative weights obtained in (b) above using MATLAB. Equations 22 to 29 were utilized in MATLAB in order to compute the SAW results shown in Table VII. Likewise, equations 30 to 36 were utilized in MATLAB in order to compute the TOPSIS results shown in Table VIII

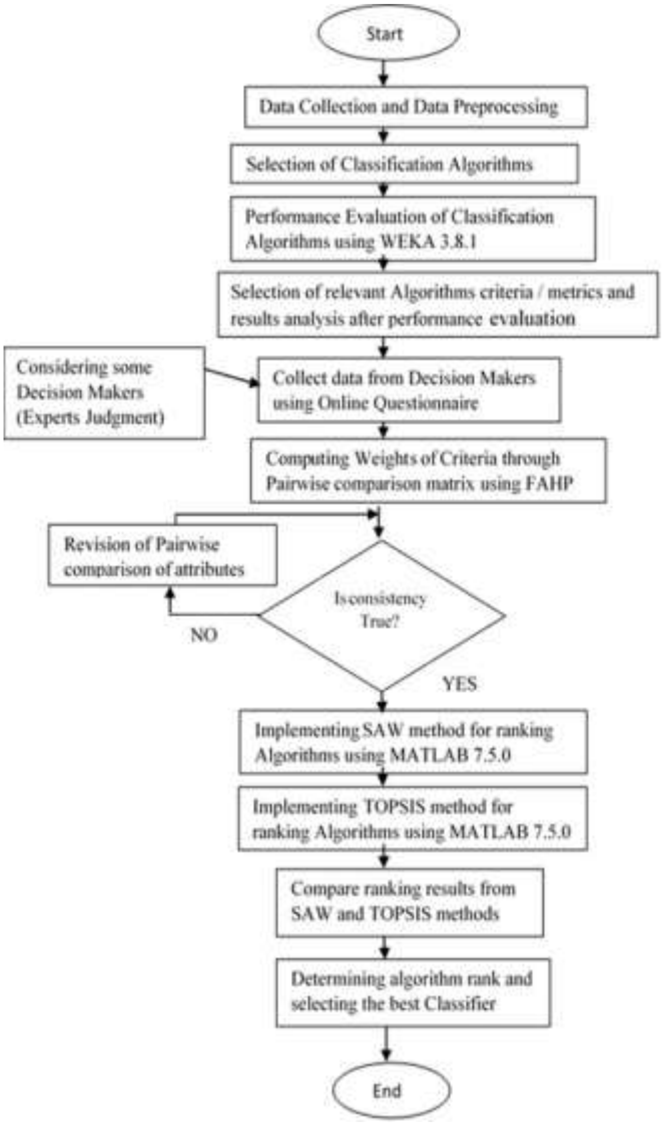

Fig. 5 Research Framework for selecting the best classifier using MCDM techniques

d) Results Comparison and Algorithm Selection: The ranking results obtained in (c) above for both SAW and TOPSIS were compared. Secondary ranking was performed as a result of tie as shown in Table VIII which gives the results in Table IX, after which the final decision on the choice of best fit classifier was made.

The decision process of FAHP is depicted in figure 4 which represents stage (b) above. Figure 4 shows the general process used to give priority weight to the identified machine learning algorithms metrics /attributes for this research. The difference between FAHP and AHP is analyzed in Table II. AHP makes use of Saaty's scale while FAHP makes use of Triangular Fuzzy Numbers (TFNs). Also, instead of calculating simple average for AHP, geometric mean was calculated for FAHP. Figure 5 shows the framework for selecting the best classifier.

B. Dataset

The study utilized public-domain credit dataset which is an Australian credit dataset with features as shown in Table I. This dataset is publicly available at the UCI machine learning repository at: http://archive.ics.uci.edu/ml/datasets/statlog+(australian+credit+approval) The file concerns credit card applications information. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data. It has 690 instances with 44.5% examples of credit worthy customers and 55.5% examples for credit unworthy customers. It contains 14 attributes, where eight are categorical attributes and six are continuous attributes.

TABLE I THE AUSTRALIAN DATASET FOR CREDIT CARD APPLICATIONS

Data name	Total cases	Good cases	Bad cases	Number of attributes
Australian data	690	307	383	14

C. Classification Algorithm Performance Measures

This study employs the following ten commonly used performance measures [46], [38]. These widely used metrics are *Accuracy, Kappa Statistic, TP Rate, TN Rate, Precision, F-measure, AUC, MAE, Training time, and Test time*. These metrics were evaluated using seven machine learning algorithms such as Bayes Network (BNK), Naive Bayes (NBS), Logistic Regression (LRN), J48, IBK, SMO and Multilayer Perceptron (MLP) in WEKA. The results are shown in Table III. The definition and formulae for the metrics are given below.

1) Overall Accuracy (ACC): accuracy is the percentage of correctly classified instances. It is one of the most widely used classification performance metrics.

$$Overall\ Accuracy = \frac{TN + TP}{TP + FP + FN + TN} \tag{8}$$

Where TN, TP, FN and FP represent True Negative, True Positive, False Negative and False Positive, respectively.

2) True Positive Rate (TPR): TPR is the number of correctly classified positive instances or abnormal instances. TPR is also called *sensitivity* measure.

$$True\ Positive = \frac{TP}{TP + FN} \tag{9}$$

*3) **True Negative Rate (TNR):** TNR is the number of correctly classified negative instances or normal instances. TNR is also called specificity measure:*

$$True\ Negative = \frac{TN}{TN + FP} \tag{10}$$

4) Precision: this is the number of classified fault-prone modules that actually are fault-prone modules:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

5) The Area Under receiver operating characteristic (AUC): receiver operating characteristic stands for receiver operating characteristic, which shows the tradeoff between TP rate and FP rate. AUC represents the accuracy of a classifier. The larger the area, the better the classifier.

6) F-measure: it is the harmonic mean of precision and recall. *F*-measure has been widely used in information retrieval:

$$F - measure = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \quad (12)$$

7) Mean Absolute Error (MAE): this measures how much the predictions deviate from the true probability. (*i*, *j*) is the estimated probability of *i* module to be of class *j* taking values in [0, 1]:

$$\text{MAE} = \frac{\sum_{j=1}^{c} \sum_{i=1}^{m} |f(i,j) - P(i,j)|}{m \cdot c} \quad (13)$$

8) *Kappa Statistic (KapS):* this is a classifier performance measure that estimates the similarity between the members of an ensemble in multi classifiers systems:

$$Kappa\ Statistic\ (Kaps) = \frac{P(A) - P(E)}{1 - P(E)} \quad (14)$$

where (A) is the accuracy of the classifier and $P(E)$ is the probability that agreement among classifiers is due to chance.

9) Training time: is the time needed to train a classification algorithm or ensemble method.

10) Test time: is the time needed to test a classification algorithm or ensemble method.

D. *Implementation of FAHP*

Ten criteria were used in the study which are A*ccuracy, Kappa Statistic, TP Rate, TN Rate, Precision, F-measure, AUC, MAE, Training Time, and Test Time.* Also, seven alternatives were considered in this study which represents the seven classifiers in this research such as Bayes Network (BNK), Naive Bayes (NBS), Logistic Regression (LRN), J48, IBK, SMO and Multilayer Perceptron (MLP).

FAHP is implemented to determine the weights of performance criteria using judgments made by decision makers. The Analytic Hierarchy Process Model is depicted in Figure 6. Level 1, level 2 and level 3 represents the goal, criteria and alternatives respectively.

If the number of criteria is given as n, then, the number of comparison is given as n* (n-1) / 2 [25] e.g. if n = 10, since we used 10 performance metrics, then number of pairwise comparisons was 45. This was used to construct the pairwise comparison matrix in order to determine the priority weights of the criteria. Six decision makers were used in this study.

Decision makers were not utilized to obtain the weights of the alternatives in relation to the criteria. This is because the values of alternatives in relation to the criteria have been achieved with the implementation of performance measures obtained from WEKA, and the results are shown in Table III. These values were used in the computation of SAW and TOPSIS.

The model was implemented using AHP- Online Software (OS) which offers tools to generate and handle AHP models and the judgments made by decision makers' were inserted into it [23]. AHP- Online Software (OS) gives reliable results of analysis, and sensitivity analysis can be performed on the results given. Pairwise comparison matrix was formulated based on decision makers' judgments that were aggregated and normalized. Fuzzification was done using Triangular Fuzzy Numbers (TFNs) and defuzzification was done using Graded Mean Integration (GMI) approach. Consistency of the decision makers judgments were obtained using Saaty's Eigen value and Eigen vector approach.

The triangular fuzzy scale implemented in this study and the linguistic variable for criteria are shown in Table II.

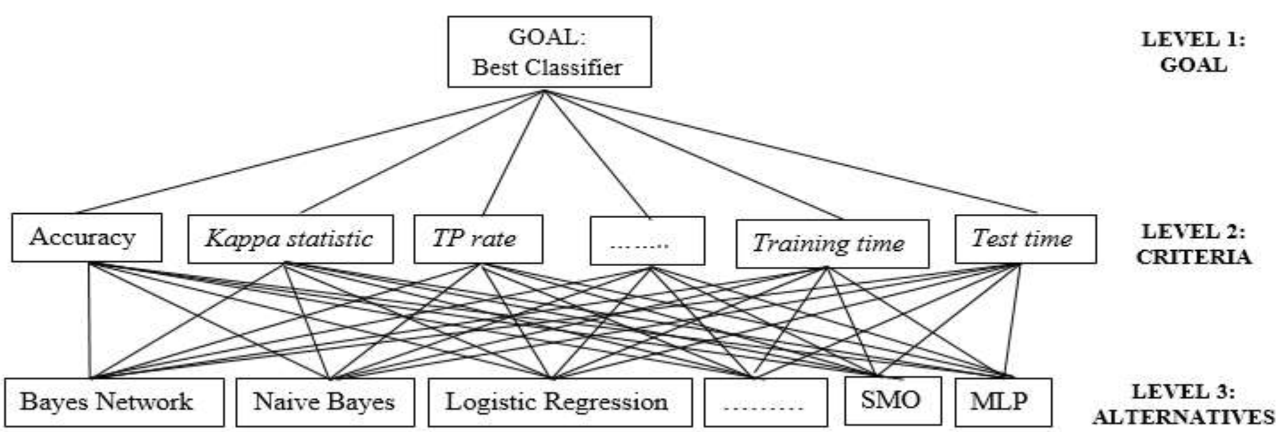

Fig. 6 Analytical Hierarchy Model for Classification Algorithm Selection

TABLE II LINGUISTIC SCALE AND THE CORRESPONDING TRIANGULAR FUZZY NUMBERS (TFNS) OF FUZZY AHP PAIR-WISE COMPARISONS

Saaty's scale	Linguistic Variable for criteria	Linguistic Variable for alternatives	Corresponding TFN membership values (l, m, u)	Reciprocal of TFN
1	Same level of significance	equally favored	(1,1,1)	(1,1,1)
2	between equal and moderately significant	between equal and moderately favored	(1,2,3)	(1/3,1/2,1)
3	moderately significant	moderately favored	(2,3,4)	(1/4,1/3,2)
4	between moderately and more significant	between moderately and more favored	(3,4,5)	(1/5,1/4,1/3)
5	more important	more favored	(4,5,6)	(1/6,1/5,1/4)
6	between more and strongly significant	between more and strongly favored	(5,6,7)	(1/7,1/6,1/5)
7	strongly significant	strongly favored	(6,7,8)	(1/8,1/7,1/6)
8	between strongly and extremely significant	between strongly and extremely favored	(7,8,9)	(1/9,1/8,1/7)
9	extremely significant	extremely favored	(9,9,9)	(1/9,1/9,1/9)

For example, let us assume that the criterion *i* has been ranked by an expert as Criterion 1 (C1) between strongly and extremely important than Criterion 2 (C2) in comparison with criterion *j*. In latter case, based on values given in table II, the criterion *i* will be evaluated with fuzzy number M = (7, 8, 9). Alternatively, in the case where criterion *j* appears less important than criterion *i*, the pairwise comparison between criteria *j* and *i* of C2 to C1 could be represented by the reciprocal fuzzy number M = (1/9, 1/8, 1/7).

The steps for implementing FAHP are outline below:

Step 1: Decision Maker compares the criteria (in case of alternatives) via linguistic terms shown in Table I.

The pairwise comparisons matrix is shown in (15), where $\widetilde{d_{ij}^{k}}$ indicates the *kth* decision maker's preference of *ith* criterion over *jth* criterion, via fuzzy triangular numbers. Here, "tilde" represents the triangular number demonstration and for the example case, $\widetilde{d_{12}^{1}}$ represents the first decision maker's preference of first criterion over second criterion, and equals to $\widetilde{d_{12}^{1}}$= (7, 8, 9).

$$\tilde{A}^k = \begin{bmatrix} \tilde{d}_{11}^k & \tilde{d}_{12}^k & \cdots & \tilde{d}_{1n}^k \\ \tilde{d}_{21}^k & \cdots & \cdots & \tilde{d}_{2n}^k \\ \cdots & \cdots & \cdots & \cdots \\ \tilde{d}_{n1}^k & \tilde{d}_{n2}^k & \cdots & \tilde{d}_{nn}^k \end{bmatrix} \tag{15}$$

Step 2: If there is more than one decision maker, preferences of each decision maker ($\widetilde{d_{ij}^{k}}$) are averaged and ($\widetilde{d_{ij}}$) is calculated as in (16).

$$\widetilde{d_{ij}} = \frac{\sum_{k=1}^{K} \widetilde{d_{ij}^{k}}}{K} \tag{16}$$

Step 3: According to averaged preferences, pair wise contribution matrix is updated as shown in (17).

$$\tilde{A} = \begin{bmatrix} \widetilde{d_{11}} & \cdots & \widetilde{d_{1n}} \\ \vdots & \ddots & \vdots \\ \widetilde{d_{n1}} & \cdots & \widetilde{d_{nn}} \end{bmatrix} \tag{17}$$

Step 4: The geometric mean of fuzzy comparison values of each criterion is calculated as shown in (18). Here, $\tilde{r}_i$ still represents triangular values.

$$\tilde{r}_i = \left(\prod_{j=1}^{n} \tilde{d}_{ij} \right)^{1/n}, \quad i=1,2,\ldots,n \tag{18}$$

Step 5: The fuzzy weights of each criterion can be found with (19), by incorporating next 3 sub-steps.

Step 5a: Find the vector summation of each $\tilde{r}_i$

Step 5b: Find the (-1) power of summation vector. Replace the fuzzy triangular number, to make it in an increasing order.

Step 5c: To find the fuzzy weight of criterion *i* ($\widetilde{w_i}$), multiply each $\tilde{r}_i$ with this reverse vector.

$$\begin{aligned} \tilde{w}_i &= \tilde{r}_i \otimes (\tilde{r}_1 \oplus \tilde{r}_2 \oplus \cdots \oplus \tilde{r}_n)^{-1} \\ &= (lw_i, mw_i, uw_i) \end{aligned} \tag{19}$$

Step 6: Since $\widetilde{w_i}$ are still fuzzy triangular numbers, they need to de-fuzzified by using Graded Mean Integration method proposed by [13], via applying the (20).

$$M_i = \frac{lw_i + 4mw_i + uw_i}{6} \qquad (20)$$

Step 7: *Mi* is a non-fuzzy number. But it needs to be normalized by following (21).

$$N_i = \frac{M_i}{\sum_{i=1}^{n} M_i} \qquad (21)$$

NOTE: the weights of alternatives were not calculated with the FAHP but only that of criteria. Since SAW and TOPSIS were implemented to determine the final ranking of the weights of alternatives with respect to each criterion. This is because only the relative weights of the criteria (performance metric) are needed to be determined using FAHP.

These 7 steps can be performed to find the normalized weights of both criteria and the alternatives. Then by multiplying each alternative weight with related criteria, the scores for each alternative is calculated. According to these results, the alternative (performance algorithm) with the highest score is suggested to the decision maker as the best classifier.

E. Implementation of SAW

The SAW MCDM technique was implemented in MATLAB. The basic logic of the SAW method is to obtain a weighted sum of performance ratings of each alternative over all attributes. The SAW method was used to determine the ranks of the alternatives using the relative weights of criteria derived from FAHP. Equations (25) to (29) were implemented in MATLAB to obtain the results in Table VIII. The step wise processes used in SAW are elucidated below:

$$A = (a_1, a_2, a_3, \ldots\ldots a_n) \qquad (22)$$

Let A = $(a_1, a_2, a_3, \ldots\ldots a_n)$ be a set on alternatives

$$C = (c_1, c_2, c_3, \ldots\ldots c_n) \qquad (23)$$

Let C = $(c_1, c_2, c_3, \ldots\ldots c_n)$ be a set of criteria

Step 1: Construct the decision matrix:

$$\begin{matrix} d_{11} & d_{12} & \ldots & d_{1n} \\ d_{21} & d_{22} & \ldots & d_{2n1} \\ \ldots & \ldots & \ldots & \ldots \\ d_{n1} & d_{n2} & \ldots & d_{nn} \end{matrix} \qquad (24)$$

Where d_{ij} is the rating of alternative *Ai* with respect to criterion *Ci.*

For benefit criteria (criteria of benefit):

$$rij = \frac{d_{ij}}{d_{ij}{}^{max}} \qquad (25)$$

For non-benefit criteria (criteria of cost):

$$rij = \frac{d_{ij}{}^{min}}{d_{ij}} \qquad (26)$$

Step 3: Construct weighted normalized decision matrix

$$v_{ij} = w_{ij} * r_{ij}, \quad \sum_{i=1}^{n} w_i = 1 \qquad (27)$$

Step 4: Calculate the score of each alternative

$$s_i = \sum_{j=1}^{m} v_{ij}, \quad i = 1, 2, 3, \ldots, n \qquad (28)$$

Step 5: Select the best alternative.

$$BA_{saw} = max_{i=1}^{n} S_i \qquad (29)$$

Where BA_{saw} is Best Alternative in Simple Additive Weighting (SAW) method and S_i is matrix score.

F. Implementation of TOPSIS

The TOPSIS MCDM technique was implemented in MATLAB. It is an MCDM method to rank alternatives over multiple criteria. It finds the best alternatives by minimizing the distance to the ideal solution and maximizing the distance to the nadir or negative ideal solution. The TOPSIS method was used to calculate the best alternative based on the relative weights of criteria derived from FAHP. The TOPSIS method was used to determine the ranks of the alternatives using the relative weights of criteria derived from FAHP. Equations (30) to (36) were implemented in MATLAB to obtain the results in Table VIII. The TOPSIS procedures used in this study are listed thus:

Step 1: Calculate the normalized decision matrix. The normalized value $\bar{X}_{ij}$ is calculated as

$$\bar{X}_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}} \qquad (30)$$

where j and n denote the number of alternatives and the number of criteria, respectively. For alternative A_j, the performance measure of the *ith* criterion *Ci* is represented by X_{ij} .

Step 2: Develop a set of weights w_i for each criterion and calculate the weighted normalized decision matrix. That is, calculate weighted normalized matrix. The weighted normalized value v_{ij} is calculated as:

$$V_{ij} = \bar{X}_{ij} \; x \; W_j, \quad j = 1, 2, \ldots, j \; ; \; i = 1, 2, \ldots, n \qquad (31)$$

where w_i is the weight of the *ith* criterion, and $\sum_{i=1}^{n} w_i = 1$

Step 3: Calculate the ideal best and ideal worst value. Find the ideal alternative solution S^+, which is calculated as

$$S^+ = \{v_1^+, \ldots, v_n^+\} = \left\{ \left(\max_j v_{ij} | i \in I' \right), \left(\min_j v_{ij} | i \in I'' \right) \right\}, \qquad (32)$$

where I' is associated with benefit criteria and I'' is associated with cost criteria.

Step 4: Find the negative-ideal alternative solution S^-, which is calculated as

$$S^- = \{v_1^-, \ldots, v_n^-\} = \left\{\left(\min_j v_{ij} | i \in I'\right), \left(\max_j v_{ij} | i \in I''\right)\right\}. \quad (33)$$

Step 5: Calculate the Euclidean distance from the ideal best. The separation of each alternative from the ideal solution is calculated as

$$S_i^+ = \sqrt{\sum_{i=1}^{n} (v_{ij} - v_i^+)^2}\ , \quad j = 1, 2, \ldots\ldots, J \qquad (34)$$

The separation of each alternative from the negative-ideal solution is calculated as

$$S_i^- = \sqrt{\sum_{i=1}^{n} (v_{ij} - v_i^-)^2}\ , \quad j = 1, 2, \ldots\ldots, J \qquad (35)$$

Step 6: Calculate Performance Score. Calculate a ratio P_i that measures the relative closeness to the ideal solution and is calculated as

$$P_i = \frac{S_i^-}{S_i^+ + S_i^-} \qquad (36)$$

Step 7: Rank alternatives by maximizing the ratio P_i. That is, arrange the rank in order.

V. EXPERIMENTAL RESULT AND DISCUSSION

A. Machine Learning Classification Results

Table III shows the performance measure results of the Australian credit dataset using the selected classifiers as implemented in WEKA. Each classifier was evaluated using ten performance metrics. The best result of each performance measure is highlighted in boldface. The benefit criteria (metrics) are supposed to be as high as possible while cost criteria (metrics) are supposed to be as low as possible.

SMO algorithm has the best performance from this dataset based on ACC, TPR, F-Measure and MAE likewise NBS has the best performance based on TNR and BNK based on Precision as shown in Table III.

Thus, there is no classification algorithm that achieves the best results across all measures of the metrics.

Therefore, the No Free Lunch (NFL) theorem is established. Hence, the need for a MCDM techniques is essential to utilize the various results from the performance measure in the final ranking of the algorithms

B. Weights of Criteria Determination Using FAHP

Six expert decision makers in the field of machine learning participated in this study to rate the algorithm performance metrics using 45 pairwise comparisons of the performance metrics. The results were aggregated and normalized. Table IV shows weighted criteria which were arrived at using steps 1 to step 7 of (15) to (21).

Consistency Ratio (CR) obtained from FAHP had a value of 1.7% that is 0.017, which is less than 10%. Hence, criteria weights results are reliable. Table IV shows FAHP results for relative criteria weights for each algorithm performance metric where Kappa Statistics has the highest relative weight of 0.213 followed by Accuracy with 0.193 according to experts' decision makers' judgments. Total criteria weights must be equal to 1, which is 0.995 ≈ 1 as shown in Table IV.

Table VI shows the Eigenvectors and Eigenvalues obtained from six decision makers' judgments with their corresponding criteria where 1 means Accuracy, 2 means Kappa Statistic and so on, as shown in Table V

TABLE III PERFORMANCE EVALUATION RESULTS OF AUSTRALIAN CREDIT DATASET

Algorithms	ACC	TPR	TNR	Precision	F-measure	AUC	KapS	MAE	Training Time	Test Time
BNK	0.852	0.798	0.896	**0.860**	0.828	0.913	0.6986	0.1702	0.0125	0.0009
NBS	0.772	0.586	**0.922**	0.857	0.696	0.896	0.5244	0.2253	0.0055	0.0014
LRN	0.862	0.866	0.859	0.831	0.848	**0.932**	**0.7224**	0.1906	0.0508	0.0005
J48	0.835	0.795	0.867	0.827	0.811	0.834	0.6642	0.1956	0.0398	**0.0002**
IBK	0.794	0.775	0.809	0.765	0.770	0.792	0.5839	0.2067	**0.0003**	0.0164
SMO	**0.885**	**0.925**	0.799	0.787	**0.850**	0.862	0.7116	**0.1449**	0.3744	0.0008
MLP	0.825	0.818	0.830	0.794	0.806	0.899	0.6460	0.1807	5.6102	0.0014

TABLE IV WEIGHTED CRITERIA FOR PERFORMANCE METRICS

Performance Metrics	**Accuracy**	**TPR**	**TNR**	**Precision**	**F-Measure**	**AUC**	**KapS**	**MAE**	**Training Time**	**Testing Time**	**TOTAL**

Criteria Weights	0.193	0.099	0.092	0.101	0.137	0.054	0.213	0.055	0.028	0.023	0.995

TABLE V CATEGORY ANALYSIS

Category	1	2	3	4	5	6	7	8	9	10
Metric / Criteria	Accuracy	KapS	TPR	TNR	Precision	F-Measure	AUC	MAE	Training Time	Testing Time

TABLE VI AGGREGATION OF JUDGMENTS BY SIX DECISION MAKERS FROM CATEGORY ANALYSIS

	1	2	3	4	5	6	7	8	9	10
1	1	1.18	2.47	2.82	1.86	1.2	2.93	3.05	6.06	6.56
2	0.85	1	2.88	3.14	3.13	1.56	3.29	3.31	6.24	6.17
3	0.41	0.35	1	1.26	0.92	0.68	2.85	2.32	2.82	3.55
4	0.35	0.32	0.79	1	0.86	0.96	1.94	2.26	3.23	3.62
5	0.54	0.32	1.08	1.16	1	0.68	1.91	1.26	5.73	4.6
6	0.83	0.64	1.48	1.04	1.47	1	3.62	2.4	4.77	4.43
7	0.34	0.3	0.35	0.51	0.52	0.28	1	1.51	1.94	2.24
8	0.33	0.3	0.43	0.44	0.79	0.42	0.66	1	2.46	2.04
9	0.17	0.16	0.35	0.31	0.17	0.21	0.52	0.41	1	1.35
10	0.15	0.16	0.28	0.28	0.22	0.23	0.45	0.49	0.74	1

C. *Ranking of Classification Algorithms Using SAW*

The result of SAW technique implemented in MATLAB shows that SMO has the highest ranking followed by LRN and lastly NBS as shown in Table VII.

TABLE VII SAW RANKING

Algorithm	**Value**	**Rank**
BNK	0.90655985	3
NBS	0.78262099	7
LRN	0.91883314	2
J48	0.8921599	4
IBK	0.83566215	6
SMO	0.94181068	1
MLP	0.86339249	5

D. *Ranking of Classification Algorithms Using TOPSIS*

The result of TOPSIS technique implemented in MATLAB shows that SMO and LRN has the highest ranking (there was a tie) followed by BNK and lastly IBK as shown in Table VIII.

From Tables VII and VIII, LRN and SMO are the algorithms that could be selected for classification. However, there is a tie between SMO and LRN, which necessitated the second ranking. Table IX shows the result of the secondary ranking where LRN was the best classifier, followed by SMO and then BNK.

TABLE VIII TOPSIS RANKING

Algorithm	**Value**	**Rank**
BNK	0.855	3
NBS	0.523	5
LRN	0.881	1
J48	0.781	4
IBK	0.510	7
SMO	0.881	1
MLP	0.515	6

TABLE IX SECONDARY RANKING USING TOPSIS

Algorithm	Value	Rank
BNK	0.766663	3
NBS	0.269419	5
LRN	0.767579	1
J48	0.670448	4
IBK	0.189955	6
SMO	0.766744	2
MLP	0.148782	7

VI. CONCLUSION AND RECOMMENDATIONS

The three basic subtasks in machine learning involves model evaluation, model selection and algorithm selection. Algorithm selection tasks in ML have been a major concern due to subjective nature of human judgment. Several techniques have been suggested for dealing with these problems such as McNemar's test, Cochran's Q test, resampled paired t-test, k-fold cross-validated t-test, the difference in proportions test, Nested cross-validation, Multiple independent test, 5x2 Cross Validated (5x2cv) paired t-test and Combined 5x2 Cross Validated (5x2cv) F test. Nonetheless, these techniques exhibit one limitation or the other. Hence the need for techniques that deals with the subjective nature of human in decision making such as MCDM is imperative.

MCDM methods are practicable tools for choosing machine learning classification algorithms. Meanwhile diverse MCDM methods evaluate classifiers from different perspectives, which in turn produce divergent rankings. Therefore, application of several coherence MCDM techniques is desirable for algorithm selection. ML algorithms for evaluation of performance metrics which belong to the following four categories / classes of classifiers such as Bayes (Bayes Network (BNK) and Naive Bayes (NBS)), function (Logistic Regression (LRN), SMO and Multilayer Perceptron (MLP)), Tree (J48) and Lazy (IBK) are standard algorithms for classification and regression. Performance evaluation of benefit metrics as well as cost metrics can be utilized in choosing the best classifier. Performance evaluation metrics can be considered as criteria and likewise algorithms can be considered as set of alternatives in MCDM problems.

The proposed Fuzzy Analytical Hierarchical Process (FAHP) and Technique or Order of Preference by Similarity to Ideal Solution (TOPSIS) model can reduce the degrees of disagreements for decision optimization, especially when different evaluation algorithms generate conflicting results. Based on proposed approach and numerical results obtained from the study, ML algorithms performance evaluation can be modelled based on AHP and fuzzy logic principle.

The study shows that logistics regression (LRN) is the best classifier based on MCDM methodology. Unlike the usual practice, in addition to Accuracy as the benchmark for selecting an algorithm, the Kappa Statistic measure can also be considered. It has been observed in this study that Logistic Regression (LRN) from Waikato Environment for Knowledge Analysis (WEKA) has the highest Kappa Statistic. Also, FAHP result for criteria weights determination shows that Kappa Statistic has the highest priority weight followed by Accuracy based on experts' decision makers' judgments. Therefore, the implementation of hybrid MCDM techniques that are closely applicable in the type of problem such as performance evaluation of supervised machine learning algorithms ensures that the disadvantage of one technique is compensated for in the other. This study therefore recommends that further study should involve more classification algorithms and ensembles. Additionally other MCDM techniques not utilized in this study can be implemented for critical analysis and comparison.

REFERENCES

[1] Alex S. & Vishwanathan, S.V.N. (2008). *Introduction to Machine Learning*. Published by the press syndicate of the University of Cambridge, Cambridge, United Kingdom. Copyright © Cambridge University Press 2008. ISBN: 0-521-82583-0. Available at KTH website: https://www.kth.se/social/upload/53a14887f276540ebc81aec3/online.pdf Retrieved from website: http://alex.smola.org/drafts/thebook.pdf

[2] Ali, S. & Smith, K. A. (2006). On learning algorithm selection for classification. *Applied Soft Computing, 6*, 119–138.

[3] Ali, S. & Smith-Miles, K. A. (2006). A meta-learning approach to automatic kernel selection for support vector machines. *Neurocomputing, 70*, 173–186.

[4] Anupama, K. S. S., Gowri, S. S., Rao, B. P. & Rajesh, P. (2015). "Application of MADM algorithms to network selection", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, Vol. 3, Issue 6, 64-67

[5] Anwaar, A., Junaid, Q., Raihan, R., Arjuna, S., Andrej, Z. & Jon C. (2016). Big data for development: applications and techniques. *Big Data Analytics, 1 / 2,* 1- 24. ISSN: 2058-6345. doi: 10.1186/s41044-016-0002-4

[6] Aruldoss, M., Lakshmi, T. M., & Venkatesan, V. P. (2013). A survey on multi criteria decision making methods and its applications. *American Journal of Information Systems, 1*(1), 31-43.

[7] Asuquo, D. E., & Onuodu, F. E. (2016). A fuzzy AHP Model for selection of university academic staff. *International Journal of Computer Applications, 141*(1), 19-26. doi:10.5120/ijca2016908969

[8] Bacudio, L., Esmeria, G. J., & Promentilla, M. A (2016). *A fuzzy analytic process approach for optimal selection of manufacturing layout*. Proceedings of the De La Salle University Research Congress (vol. 4). Manilla, Philippines.

[9] Brazdil, P. B., Soares, C. & da Costa, J. P. (2003). Ranking learning algorithms: Using IBL and meta-learning on accuracy and time results. *Machine Learning, 50*, 251–277.

[10] Cakir, O., & Canbolat, M. (2008). A web-based decision support system for multi-criteria inventory classification using fuzzy AHP methodology. Expert Systems with Applications, 35(3), 1367–1378.

[11] Cebi, F., & Kahraman, C. (2012). Single and multiple attribute fuzzy Pareto models. Journal of Multi-Valued Logic & Soft Computing, 19, 565–590.

[12] Chang, D.Y. (1996). Applications of the extent analysis method on fuzzy AHP. *European Journal of Operation Research*, 95, 649–655

[13] Cheng, S., Chan, C. W., & Huang, G. H. (2002). *Using multiple decision analysis for supporting decision of solid waste management. Journal of Environmental Science and Health, Part A, 37*(6), 975-990. doi: 10.1081/ESE-120004517

[14] De Houwer, J., Barnes-Holmes, D. & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychon Bull Rev* 20: 631. https://doi.org/10.3758/s13423-013-0386-3

[15] de Moura Rezende dos Santos, F., Guedes de Oliveira Almeida, F., Pereira Rocha Martins, A., Bittencourt Reis, A. and Holanda, M. (2018). Ranking Machine Learning Classifiers Using Multi-criteria Approach. *2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC)*.

[16] Figueira, J., Greco, S. & Ehrgott, M. (2005). Multiple criteria decision analysis: state of the art surveys, *Vol. 78)*. Springer Science & Business Media.

[17] Fishburn, P.C., (1967). Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments, *Operations Research Society of America (ORSA) Publication*, Baltimore, MD,

[18] Flores, B. E., Olson, D. L., & Doria, V. K. (1992). Management of multi-criteria inventory classification. Mathematical and Computer Modeling, 16, 71–82.

[19] Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers.

[20] Fu, S., & Fan, G. (2016). *A multiple attribute decision-making method based on exponential fuzzy numbers*. *Mathematical and Computational Applications, 21*(19), 1-9.

[21] Gayen, A. K. (1951). The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika, 38*, 219–247.

[22] Geetha, N. and Sekar, P. (2015). Graph Theory Matrix Approach – A Qualitative Decision Making Tool. *Materials Today: Proceedings*, 4(8), 7741-7749.

[23] Goepel, K. D. (2017). BPMSG's AHP Online System. Rational Decision Making Made Easy. Business Performance Management Singapore (BPMSG)

[24] Good, I.J. (1951). Probability and the Weighing of Evidence, Philosophy Volume 26, Issue 97. Published by Charles Griffin and Company, London 1950.Copyright © The Royal Institute of Philosophy, 163-164.doi: https://doi.org/10.1017/S0031819100026863. Available at Royal Institute of Philosophy website: https://www.cambridge.org/core/journals/philosophy/article/probability-and-the-weighing-of-evidence-by-goodi-j-london-charles-griffin-and-company-1950-pp-viii-119-price-16s/7D911224F3713FDCFD1451BBB2982442

[25] Hahn. (n.d.). *The analytic hierarchy process* - Computer science. Retrieved from muslimin-mcs.web.id>uploads>2017/04

[26] Ho, W. (2008). Integrated analytic hierarchy process and its applications - A literature review. European Journal of Operation Research, 186, 211–228.

[27] Hsieh, C. H., & Chen, S. H. (1999). Similarity of generalized fuzzy numbers with graded mean integration representation. *Proceedings of 8th International Fuzzy Systems Association World Conference*, Taipei, Taiwan, 551-555.

[28] Hung, C. C & Chen, L. H. (2009). A Fuzzy TOPSIS Decision Making Model with Entropy Weight under Intuitionistic Fuzzy Environment. *Proceedings of the International Multi-Conference of Engineers and Computer Scientists (*IMECS)*, Vol 1,* Hong Kong

[29] Hwang, C. L., & Yoon, K. (1981). *Multiple attribute decision making: Methods and applications, a state-of-the-art-survey*. Berlin, Heidelberg: Springer-Verlag

[30] Ishizaka A., & Labib, A. (2011). *Review of the main developments in the analytic hierarchy process. Expert Systems with Applications, 38*(11), 14336-14345

[31] Jaya, B. J. Y. and Tamilselvi, J. J. (2014). Simplified MCDM Analytical Weighted Model for Ranking Classifiers in Financial Risk Datasets. *International Conference on Intelligent Computing Applications*. 158-161. Published by IEEE. DOI 10.1109/ICICA.2014.42 [online] Semanticscholar.org. Available at: https://www.semanticscholar.org/paper/Simplified-MCDM-Analytical-Weighted-Model-for-in-Jaya-Tamilselvi/609a4aa69fe78be99ec483401be7e1c2933ff5a0 [Accessed 2 Mar. 2019].

[32] Kandhasamy, J. P. & Balamurali, S. (2015). Performance analysis of classifier models to predict diabetes mellitus. *Procedia Computer Science, 47*, 45–51.

[33] Kecman, V. (2005). *Support vector machines: An introduction in support vector machines: Theory and applications*. In L. Wang (Ed.), 1–47. Berlin: Springer-Verlag. Chapter 1.

[34] Khanmohammadi, S. & Rezaeiahari, M. (2014). AHP based classification algorithm selection for clinical decision support system development. *Procedia Computer Science, 36*, 328–334.

[35] Ko, M., Twari, A., & Mehmen, J. (2010). A review of soft computing applications in supply chain management. *Applied Soft Computing*, 10, 661–664.

[36] Korting, T. S. (n.d)."C4. 5 algorithm and Multivariate Decision Trees." Image Processing Division, National Institute for Space Research--INPE.

[37] Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica* 31 (2007). 249 – 268. Retrieved from IJS website: http://wen.ijs.si/ojs-2.4.3/index.php/informatica/article/download/148/140.

[38] Kou, G., Lu, Y., Peng, Y., & Shi, Y. (2012). Evaluation of classification algorithms using MCDM and rank correlation. *International Journal of Information Technology & Decision Making, 11*, 197–225.

[39] Logistic Regression (LR), (n.d.) 223 – 237. Available at: https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12.pdf

[40] Luckyson, K., Snehanshu, S., Sudeepa, R. D. (2016) Predicting the direction of stock market prices using random forest. Applied Mathematical Finance, 1 (1), 1-20

[41] Nadali, A; Kakhky, E. N.; Nosratabadi, H. E., (2011). "Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system*," 3rd International Conference on Electronics Computer Technology (ICECT), vol.6*, 161,165

[42] Newsom, I. (2015). Data Analysis II: Logistic Regression. Available at: http://web.pdx.edu/~newsomj/da2/ho_logistic.pdf

[43] Olsen, E. (1996). Effect of sampling on measurement errors. *The Analyst*, 121(9), 1155.

[44] Partovi, F. Y., & Burton, J. (1993). Using the analytic hierarchy process for ABC analysis. International Journal of Operations & Production, 13(9), 29–44.

[45] Partovi, F. Y., & Hopton, W. E. (1994). The analytic hierarchy process as applied to two types of inventory problems. Production and Inventory Management Journal, 35(1), 13–19.

[46] Peng, Y., Wang, G., Kou, G., & Shi, Y. (2011). "An empirical financial study of classification algorithm evaluation for financial risk prediction. *In Journal on Applied Soft Computing-AS*, Vol. 11, No. 2, 2906 - 2915.

[47] Perveen, S., Shahbaz, M., Guergachi, A. & Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. *Procedia Computer Science, 82*, 115–121.

[48] Pohekar, S. D., & Ramachandran, M. (2004). *Application of multi-criteria decision making to sustainable energy planning: A Review*. *Renewable & Sustainable Energy Reviews, 8*(4), 365–381

[49] Puente, J., De Lafuente, D., Priore, P., & Pino, R. (2002). ABC classification with uncertain data: A fuzzy

[50] Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to python's scientific computing stack. *The Journal of Open Source Software*, 3(24).
[51] Raschka, S. (2018). Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning.
[52] Reif, M., Shafait, F., Goldstein, M., Breuel, T. & Dengel, A. (2014). Automatic classifier selection for non-experts. *Pattern Analysis and Applications, 17*, 83–96.
[53] Rezaei, J. (2007). A fuzzy model for multi-criteria inventory classification. *In 6th International conference on analysis of manufacturing systems, AMS*, Lunteren, The Netherlands, 11–16.
[54] Rice, J. R., (1976). "The Algorithm Selection Problem" (1975). Computer Science Technical Reports. Paper 99. http://docs.lib.purdue.edu/cstech/99
[55] Rosenblatt, F. (1962), Principles of Neurodynamics. Spartan, New York.
[56] Saaty, T. L. (1977). "A scaling method for priorities in hierarchical structures," *Journal of Mathematical Psychology, 15*, 57-68
[57] Saaty, T. L. (1980). "Axiomatic foundations of the Analytic Hierarchy process," *Management Science*, 32(2), 841- 855
[58] Saaty, T. L. (1983). The Analytic Hierarchy Process, McGraw-Hill International, New York, NY, 30.
[59] Saaty, T. L. (1990). "An exposition of the AHP in reply to the paper 'Remarks on the Analytic Hierarchy Process'," Management Science, 36(3), 259-268
[60] Saaty, T. L. (1994). Fundamentals of Decision Making and Priority Theory with the AHP, RWS Publications, Pittsburgh, PA, U.S.A.
[61] Saaty, T. L. (2003). Decision-making with the AHP: Why is the principal eigenvector necessary. *European Journal of Operational Research 145* 85–91
[62] Schaffer, C. (1994). A conservation law for generalization performance. *In Proceedings of the Eleventh International Conference on Machine Learning*, New Brunswick, USA, 153-178
[63] Schapire, R. (n.d). Machine Learning Algorithms for Classification.
[64] Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding machine learning*. Cambridge: Cambridge University Press.
[65] Shih, H., Shyur, H., & Lee, E. S. (2007). *An extension of TOPSIS for group decision making*. *Mathematical and Computer Modelling, 45*, 7-8. Elsevier
[66] Shiue, Y. (2009). Data-mining-based dynamic dispatching rule selection mechanism for shop floor control systems using a support vector machine approach. *International Journal of Production Research*, 47(13), 3669–3690.
[67] Singh, A. & Singh, M. L. (2016). Performance evaluation of various classifiers for color prediction of rice paddy plant leaf. *Journal of Electronic Imaging, 25*, 061403.
[68] Singh, P., Verma, A. & Chaudhari, N. S. (2016). Performance evaluation of classifier combination techniques for the handwritten devanagari character recognition. In *Information systems design and intelligent applications*, 651–662. India: Springer.
[69] Skinner, D.C. (3rd Ed). (2009). *Introduction to decision analysis*. Gainesville: Probabilistic Publishing
[70] Song, Q., Wang, G. & Wang, C. (2012). Automatic recommendation of classification algorithms based on data set characteristics. *Pattern recognition, 45*, 2672–2689.
[71] Srichetta, P., & Thurachon, W. (2012). *Applying fuzzy analytic hierarchy process to evaluate and select product of notebook computers*. *International Journal of Modeling and Optimization, 2*(2), 168-173. doi: 10.7763/IJMO.2012.V2.105
[72] Taiwo, O. A. (2010). Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), ISBN: 978-953-307-034-6, *InTech,* University of Portsmouth United Kingdom, 3 – 31. Available at InTech open website: http://www.intechopen.com/books/new-advances-in-machine-learning/types-of-machine-learning-algorithms
[73] Tang, J., Liu, R., Zhang, Y. L., Liu, M. Z., Hu, Y. F., Shao, M. J., Zhu, L. J., Xin, H. W., Feng, G. W., Shang, W. J., Meng, X. G., Zhang, L. R., Ming, Y. Z., … Zhang, W. (2017). Application of Machine-Learning Models to Predict Tacrolimus Stable Dose in Renal Transplant Recipients. *Scientific reports*, *7*, 42192. doi:10.1038/srep42192
[74] Triantaphyllou, E. (2000) Multi-Criteria Decision Making Methods: A Comparative Study. Kluwer Academic Publishers, Dordrecht. DOI: http://dx.doi.org/10.1007/978-1-4757-3157-6
[75] Tscheikner-Gratl, F., Egger, P., Rauch, W., & Kleidorfer, M. (2017). *Comparison of multi-criteria decision support methods for integrated rehabilitation prioritization*. *Water, 9*(68), 1-28. doi: 10.3390/w9020068
[76] Tzeng, G. H. & Huang, J.-J. (2011, June 22). *Multiple attribute decision making: methods and applications*. CRC press.
[77] Velasquez, M., & Hester, P. T. (2013). *An analysis of multi-criteria decision making methods*. *International Journal of Operations Research, 10*(2), 56-66
[78] Xu, L. & Yang, J. (2001). Introduction to Multi-Criteria Decision Making and the Evidential Reasoning Approach. *Manchester School of Management, University of Manchester Institute of Science and Technology*. ISBN: 186115 111 X, 1- 21
[79] Zadeh, L. (1965). Fuzzy sets. Inform. Control 8, 109–141
[80] Zadeh, L. A. (2005). Toward a generalized theory of uncertainty (GTU) – An outline. Information Sciences, 172(1–2), 1–40.
[81] Zimmermann, H. J. (1991). *Fuzzy Set Theory and Its Applications*, Kluwer Academic Publishers, Second Edition, Boston, MA.