Jurnal Natural
Vol. 23, (3) 2023
DOI 10.24815/jn.v23i3.32467
pISSN 1411-8513
eISSN 2541-4062
ORIGINAL RESEARCH

A hybrid intelligent model based on logistic regression and fuzzy multiple-attribute decision-making for credit evaluation

IRVANIZAM IRVANIZAM[1], ZAKIAL VIKKI[1], SUTARMAN[2*], OPIM SALIM SITOMPUL[3]

[1]Department of Computer Science, Universitas Sumatera Utara, Medan, Indonesia
[2]Department of Mathematics, Universitas Sumatera Utara, Medan, Indonesia
[3]Department of Information Technology, Universitas Sumatera Utara, Medan, Indonesia

Abstract. One of the crucial issues in data mining is to select an appropriate classification algorithm. Due to it usually involves many criteria, the duty of algorithm selection can be widely described as multiple-attribute decision-making (MADM) problems, including credit risk evaluation. Many different MADM approaches select classifiers based on different perspectives, and hence they might generate diverse classifiers' rankings. This paper aims to propose a hybrid intelligent model to overcome credit risk assessment problems based on logistic regression and the fuzzy MADM method. Firstly, the Ordinal Priority Approach (OPA) method evaluates attributes involved in credit risk problems by considering professional assessments of a decision-maker and calculates a weight for each criterion. Secondly, all categorical data converted into triangular-fuzzy numbers (TFNs) and numerical data are evaluated using the MADM instrument to obtain an optimal solution dataset and logistic regression to calculate the probabilities of the optimal dataset. In this experimental study, three existing classification techniques and the proposed intelligent model evaluate three banking credit datasets with a different number of criteria under numerical and categorical data types. The prediction accuracy results generated by the proposed model are compared with the three existing classification methods. The results exhibit that there are slight differences between the three datasets. The experimental results demonstrate the proposed intelligent model has superiority in classifying the credit loan recipients especially for categorical datasets.

Keywords: classification, credit loans, fuzzy multiple-attribute decision-making, logistic regression.

INTRODUCTION

The bank aims to be a financial intermediary with the main business of collecting and channeling public funds and providing other services in payment traffic. As a business company, a bank will always try to get the maximum profit from the business. One of the banking businesses is to provide credit loans to clients. During the proposing process and decision-making on whether the clients can be approved or rejected for a credit loan, the bank usually examines a client's creditworthiness. Decades ago, a bank's expert determined the decision based on the expert's individual and conventional assessment. The expert qualitatively evaluated the risk after reviewing the company's financial reports, studying business scenarios, and interviewing the client. During this period, it became evident that such a scheme was inefficient in more complex conditions and growing competition among banks. Therefore, engineers and scientists have begun to create qualitative measures and analytical strategies in credit risk evaluation. These strategies were evolving more and more famous gratitude to information technology products and credit evaluation techniques. There are many descriptions of it, but the authors will observe the one remarking that credit scoring is the procedure supporting the expert, such as a manager, who is responsible for credit loans to decide whether clients can receive loans or not according to a set of pre-requested criteria.

The credit evaluation system has been widely used in retail, corporate, and small enterprise lending. Most credit scoring systems deviate concerning the kind and quantity of the data required for decision-making. Individual and

*Corresponding Author:
sutarman@usu.ac.id

Received: June 2023 | Revised: October 2023 | Accepted: October 2023

trade activities have both been found suitable in a small enterprise credit evaluation system [1][2]. In evaluating business activities, economist attempts to find financial ratios that are important in deciding the repayment possibilities of the company [3]. Many credit evaluation models utilize parameters or financial prerequisites classified into five categories: solvency, utilization, profitability, leverage, and performance.

Many prestigious researchers have studied and developed credit risk evaluation models for solving their different cases. For instance, Li [4] and Jinjuan [5] designed a computerized credit evaluation system based on a machine-learning approach for predicting and evaluating the credit scores of bank clients. Yang *et al.* [6] utilized the neural networks (NNs) model to examine personal credit business. Additionally, they improved the model by proposing a Flexible Neural Tree-NNs model for classifying bank clients. Liu *et al.* [7] also researched credit evaluation and focused on model accuracy improvement. They utilized some machine learning techniques to evaluate the performance in classification. Lin *et al.* [8] developed a credit evaluation index instrument for power market user cases. They compared the experimental results with K-means and Silhouette Coefficient models to demonstrate the feasibility of the designed tool. Dattachaudhuri *et al.* [9] developed an expert system for generating rules of credit financial information using white box neural, classified, and unclassified data models. Shi *et al.* [10] discovered that it is hard for micro, small, and medium enterprises (MSMEs) to eliminate the development of financial dilemmas. To overcome this, they established a mathematical model to evaluate credit risk for MSMEs loans. Egger *et al.* [11] selected Support Vector Machine (SVM) and K-Means models as tools to evaluate credit risk for supply chain finance. An integrated SVM-K-Means model is superior for this case. Hassija *et al.* [12] introduced the ways of blockchain to decentralize credit scoring and evaluate the number of dependence features. Li [13] developed an application based on multi-level deep neural network to assess and predict a bank credit risk.

One way to define and estimate the association between one binary dependent variable (an outcome variable) and one or more independent variables (covariates variables) is to use a logistic regression model [14]. In many different fields of study, phenomena have been described using logistic regression models because they are flexible, have a strong interpretation, and are easy to use [15]. A logistic regression model is frequently used to assess predictors and to account for confounders and/or interactions, much like other regression models. In addition to creating prediction algorithms that can be represented in nomograms or online calculators that convey the probability of an event, these models are used to examine retrospective data, including credit risk studies. Thus, since the outcome variable is binary, logistic regression can be utilized in 2-class classification issues [16].

In recent years, there are some other researchers who have conducted studies for credit evaluation using logistic regression. Yan [14] calculated credit scores using logistic regression and focused on predicting the default rate of credit clients. Dinh and Thanh [15] utilized logistic regression to predict the ability of bank clients to repay credit loans in accordance with a predetermined time. Assef *et al.* [16] compared logistic regression with other machine-learning classifiers in evaluating credit risk for a supervised financial dataset. Manglani *et al.* [17] attempted to minimize the credit risk by selecting appropriate clients to avoid misuse of credit loans. They developed a logistic regression application for this problem. Zhang and Han [18] conducted an artificial model based on logistic regression to measure credit risk of MSMEs in China by comparing financial ratios between enterprises and market. Based on the observations of the studies, the logistic regression summarizes at least three advantages in the classification process. First, it is easy to design a predictive application and quite efficient for training datasets. Second, it does not need to assume distributing classes for the dataset's features. Third, it performs good accuracy in classification for many datasets, especially linearly separable datasets [19].

Unfortunately, the above studies did not consider the case of credit risk as a multi-attribute decision-making (MADM) problem. This evaluation problem involves many parties, such as bank experts. They are instrumental in observing how potential and confident credit is given to customers, evaluating which attributes affect credit risk, and predicting who is entitled to receive credit. These factors will affect the smoothness or failure of the credit return process and increase the competitiveness and business growth of the bank.

MADM is a branch of operations research and decision science utilized to specifically assess many competing criteria in decision-making [23]. In real lifetime, conflicting criteria are frequently present. For instance, when

Table 1. Dataset information

No	Datasets	# of Rows	# of Features	Categorical	Numerical	Target	
						Class 1	*Class 0*
1	Germany Credit Data	1000	21	13	8	700	300
2	Australian Credit Approval	1000	11	3	8	206	794
3	Credit Scoring	1225	15	2	13	323	902

managing a portfolio, managers want to maximize returns while minimizing risks; yet, the stocks with the highest return potential generally have the highest risk of losing money. By considering the advantages of MADM, many prestigious scholars have widely utilized this approach to deal with risk management studies. Mendonca *et al.* [24] proposed a MADM model for optimizing a case study financial portfolio in Brazil. The results demonstrated that the model can safe investments for certain period. Syed and Lawryshyn [25] presented a MADM technique for risk-informed in the physical asset management. Pasman *et al.* [26] measured a risk reduction in decision-making complex situations using a MADM method. Sun *et al.* [27] compared and analyzed some MADM models for the flood disaster risk study at a river in China.

Motivated by the above analyses, this paper aims to propose a new hybrid intelligent model based on logistic regression and MADM methods to overcome credit risk problems. It also shows steps to integrate the logistic regression with MADM methodology to design an effective technique for obtaining higher performances. Furthermore, it presents how a bank can classify customers using logistic regression and how the bank can group suitable credit clients using MADM. Important features affected by credit evaluation models for three financial datasets are also provided because particular attributes and their priorities' orders can influence the criteria preference and some intake variables for the strategies used in the paper.

The leftover of this paper neatly unfolds as follows. Section 2 describes the methodology of this paper, such as three dataset information used in this study, data transformation, features' weights determination, the proposed hybrid intelligent model, and performance measurement. Section 3 presents results and discussions consisting of a demonstration of the proposed model through a numerical example, observation of experimental results for the three datasets, and classification results for the three datasets. Section 4 outlines the conclusion and eligible future works.

METHODOLOGY

Datasets

This study uses three datasets with binary classification targets. Two datasets are available at the University of California Irvine (UCI) machine learning repository, and a dataset for 'Credit Scoring' is available in [28]. Table 1 presents the three detailed data set information. The datasets consider various situations for binary classification problems. It is about how to group bank customers who are suitable and not suitable for receiving credit loans.

The dataset (Germany Credit Data) consists of 1000 instances and 21 features. This dataset has 13 categorical with text and eight numerical feature types. The dataset (Australian Credit Approval) contains 1000 instances and 11 feature types. Each row in the dataset consists of three categorical and eight numerical. The dataset (Credit Scoring) consists of 1000 instances and 15 features. Similar to the previous two datasets, this dataset has categorical or numerical features. Each row in this dataset contains two categorical with text features and 13 numerical features.

Classification Models

We utilized three different classifiers to compare with our proposed hybrid model. They include original logistic regression, support vector machine (SVM), and k-nearest Neighbors (k-NN). Logistic regression is a statistical model that is often utilized for classification by estimating the probability of an event occurring [18]. SVM searches the data for the hyperplane that provides the greatest separation between data instances and divides them into two classes [29]. Meanwhile, K-NN is a non-parametric, supervised learning classifier that employs closeness to categorize or anticipate how to group a given data point [30].

To measure performances in this study, we split each dataset into training and testing with 80%:20% composition. We then created a model on the training dataset for each classifier. Later on, we tested each model on the testing dataset and evaluated the performance of each classification model concerning overall

Table 2. Linguistic variables and their TFNs

i	Linguistic Variables	Symbols	Triangular Fuzzy Numbers	Defuzzification
0	Very Inferiority	VI	(0, 0, 0.125)	0.041667
1	Inferiority	I	(0, 0.125, 0.25)	0.125
2	Medium Inferiority	MI	(0.125, 0.25, 0.375)	0.25
3	Fair Priority	FP	(0.25, 0.375, 0.5)	0.375
4	Medium Priority	MP	(0.375, 0.5, 0.625)	0.5
5	Priority	P	(0.5, 0.625, 0.75)	0.625
6	Very Priority	VP	(0.625, 0.75, 0.875)	0.75
7	Very Very Priority	VVP	(0.75, 0.875, 1)	0.875
8	Extremely Priority	EP	(0.875, 1, 1)	0.95833

Table 3. Features' information for dataset 1

Attributes	Criteria	Types	Data Types	Priority	Criteria Weights Determined by OPA
A1	account_status	Benefit	Categorical	5	0.025831
A2	duration	Benefit	numerical	3	0.043049
A3	credit_history	Benefit	Categorical	2	0.064576
A4	purpose	Benefit	Categorical	1	0.129152
A5	credit_amount	Benefit	numerical	1	0.129152
A6	savings_account	Benefit	Categorical	4	0.032288
A7	employment priod	Benefit	Categorical	3	0.043049
A8	installment_rate	Cost	numerical	3	0.043049
A9	personal_status	Benefit	Categorical	6	0.021525
A10	other_debtors	Benefit	Categorical	3	0.043051
A11	residence_period	Benefit	numerical	5	0.025831
A12	Property	Benefit	Categorical	2	0.064576
A13	Age	Cost	numerical	4	0.032288
A14	Installment_plans	Benefit	Categorical	3	0.043049
A15	Housing	Benefit	Categorical	4	0.032288
A16	Number_of_existing_credits	Cost	numerical	1	0.129152
A17	Job	Benefit	Categorical	5	0.025832
A18	Number_of_ responsible_people	Benefit	numerical	4	0.032288
A19	Telephone	Benefit	Categorical	7	0.018449
A20	foreign_worker	Benefit	Categorical	6	0.021525

accuracy, precision, F-measure, mean absolute error, True Positive rate, and True Negative.

Data Transformation

The theory of fuzzy set was first concepted by Zadeh [31] in 1965. This theory can represent uncertain information including categorical dan linguistic data. To represent this set, he defined fuzzy numbers characterized an element with its membership degree. Motivated by the development of this research, many researchers have formulated ways to determine fuzzy numbers. One of them is Jiang *et al.* [32] by introducing formula (1) to determine triangular fuzzy numbers (TFNs) for *n* linguistic variables where $\underline{x}$ is a lower value of $\tilde{x}$, x is a middle value of $\tilde{x}$, $\overline{x}$ is an upper value of $\tilde{x}$, n is a number of linguistic variables, and $i = 0,1,2,...,n-1$.

$$\tilde{x} = (\underline{x}, x, \overline{x}) = \left(\max\left(\frac{i-1}{n-1}, 0\right), \frac{i}{n-1}, \min\left(\frac{i+1}{n-1}, 1\right)\right) \quad (1)$$

In many real MADM applications, TFNs often require a tool to transform them into crisp values, which is usually called defuzzification. One of the defuzzification tools commonly used in the MADM application is the TFN mean based on Beta distribution approach. This statistical tool was first formulated by Rahmani *et al.* [33] as shown as (2).

$$\mu_{\tilde{x}} = \frac{\underline{x} + x + \overline{x}}{3} \quad (2)$$

In this study, an expert identified nine linguistic variables for representing categorical features of the datasets. The detailed linguistic variables' information can be seen in Table 2.

Criteria Weights Determination

The classification problem in this study is a MADM issue involving the roles and assessments of the decision-maker or expert. They include determining which criteria (features) are beneficial and cost and providing opinions about the priority of features in this

Table 4. Features' information for dataset 2

Attributes	Criteria	Types	Data Types	Priority	Criteria Weights Determined by OPA
A1	person_age	Cost	numerical	5	0.072202
A2	person_income	Benefit	numerical	3	0.108303
A3	person_home_ownership	Benefit	Categorical	2	0.054152
A4	person_emp_length	Cost	numerical	1	0.216607
A5	loan_intent	Benefit	Categorical	1	0.216606
A6	loan_amnt	Benefit	numerical	3	0.108303
A7	loan_int_rate	Benefit	numerical	3	0.072202
A8	loan_percent_income	Benefit	numerical	3	0.072202
A9	cb_person_default_on_file	Benefit	Categorical	5	0.043322
A10	cb_preson_cred_hist_length	Benefit	numerical	2	0.036101
A11	loan_status	TARGET	Class{1,0}		

Table 5. Features' information for dataset 3

Attributes	Criteria	Types	Data Types	Priority	Criteria Weights Determined by OPA
A1	Year_of_birth (YOB)	Benefit	numerical	4	0.057788
A2	Number_of_children (NKID)	Benefit	numerical	6	0.038525
A3	Number_of_other dependents (DEP)	Benefit	numerical	6	0.038525
A4	Having_a_home phone (PHON)	Benefit	numerical	7	0.033022
A5	Spouses_income (SINC)	Benefit	numerical	5	0.046231
A6	Applicants_employment_status (AES)	Benefit	Categorical	1	0.231149
A7	Applicants_income (DAINC)	Benefit	numerical	3	0.077049
A8	Residential_status (RES)	Cost	Categorical	2	0.115575
A9	Value_of_Home (DHVAL)	Benefit	numerical	3	0.077049
A10	Mortgage_balance_outstanding (DMO)	Benefit	numerical	3	0.077049
A11	Outgoings_on_mortgage_orrent (DOU)	Benefit	numerical	4	0.057788
A12	Outgoings_on_Loans (DOUTL)	Benefit	numerical	4	0.057788
A13	Outgoings_on_Hire_Purchase (DOTHP)	Cost	numerical	5	0.046231
A14	Outgoings_on_credit cards (DOUTCC)	Benefit	numerical	5	0.046231
A15	BAD	TARGET	Class{1,0}		

case study opinions about the priority of features in this case study.

Moreover, based on the given priority values from the expert, we utilize the Ordinal Priority Approach (OPA) [34] method to determine features' weights automatically. Since this study evaluates three datasets, we require three different opinion values and generate their feature weight sets. Table 3, , and Table 5 are opinion decision-maker values for datasets 1, 2, and 3, respectively. The web-based OPA solver [34] then calculates the features' weights for each dataset.

The Proposed Hybrid Intelligent Model

This section describes the procedures of the proposed hybrid intelligent model. It is a model integrating fuzzy MADM with logistic regression. The proposed model consists of three stages. Firstly, the model utilizes the OPA method to measure numeric weights of banking credit prerequisite attributes obtained from experts' professional assessment. Secondly, it calculates a cumulative value for each alternative or bank customer using the MADM and the suitable attributes weights. Finally, the proposed model classifies the approved and not approved alternatives using the logistic regression model. Figure 1 shows the flowchart of the proposed hybrid model.

Furthermore, due to the proposed model involving many attributes and presenting as a MADM problem, this proposed hybrid model will differ into two feature types. One is benefit type, and another is cost type. The benefits feature type is attributes that their higher values' properties are preferable, while the cost feature type is attributes that their smaller values' properties are always desirable [35].

To solve this case, a dataset can be firstly viewed as a matrix $B = \left[b_{ij}\right]_{m\,x\,n}$, where m is the number of credit recipients and n is the number of attributes. The matrix B can be illustrated as (3). Later on, all the MADM procedures should be executed before we perform a classification process.

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} \quad (3)$$

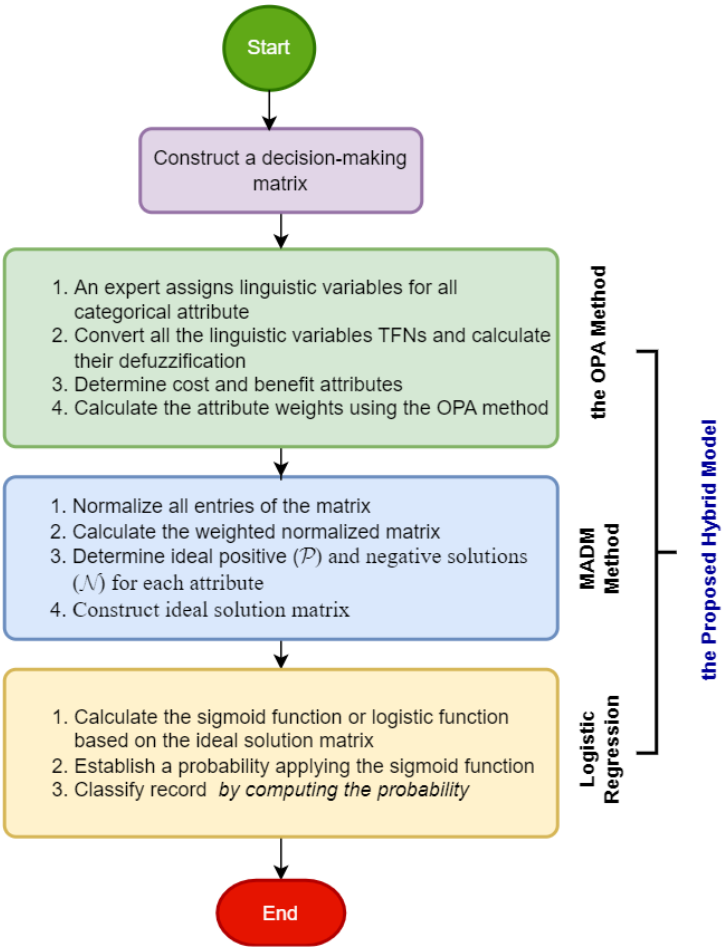

Figure 1. The flowchart of the proposed hybrid model

The detailed computational procedures of the proposed model are presented as the following:

Step 1. An expert assigns linguistic variables for all categorical attribute records of the dataset according to his/her professional judgement.

Step 2. Convert all linguistic variables of the categorical attributes into TFNs and calculate their defuzzification.

Step 3. Determine which attributes are belonging to benefit and cost types.

Step 4. Assign priority order for all the attributes and compute the attribute weights $W = \left[w_j \right]_{1xn}$ using the OPA method.

Step 5. Normalize all entries of the Matrix B into a matrix $N = \left[n_{ij} \right]_{mxn}$, in which entry n_{ij} can be obtained using (4).

$$n_{ij} = \begin{cases} \dfrac{b_{ij} - \min\limits_{1 \le j \le n} \{b_{ij}\}}{\max\limits_{1 \le j \le n} \{b_{ij}\} - \min\limits_{1 \le j \le n} \{b_{ij}\}}, & C_j \in C_{benefit} \\ 1 - \dfrac{b_{ij} - \min\limits_{1 \le j \le n} \{b_{ij}\}}{\max\limits_{1 \le j \le n} \{b_{ij}\} - \min\limits_{1 \le j \le n} \{b_{ij}\}}, & C_j \in C_{Cost} \end{cases} \quad (4)$$

Step 6. Calculate the weighted normalized matrix $N' = \left[n'_{ij} \right]_{mxn}$, in which entry n'_{ij} can be calculated using (5).

$$n'_{ij} = n_{ij} w_j, \qquad \forall j = 1, 2, ..., n \quad (5)$$

Step 7. Determine ideal positive (Π) and negative solutions (N) for each attribute j using (6) and (7), respectively.

$$P_j = \begin{cases} \max\limits_{1 \le i \le m} \{n'_{ij}\}, & C_j \in C_{Benefit} \\ \min\limits_{1 \le i \le m} \{n'_{ij}\}, & C_j \in C_{Cost} \end{cases} \quad (6)$$

$$N_j = \begin{cases} \min\limits_{1 \le i \le m} \{n'_{ij}\}, & C_j \in C_{Benefit} \\ \max\limits_{1 \le i \le m} \{n'_{ij}\}, & C_j \in C_{Cost} \end{cases} \quad (7)$$

Step 8. Construct ideal solution matrix $X = \left[x_{ij} \right]_{m \times n}$ from the negative and positive solutions. All entries or instances of the ideal solution matrix can be calculated using (8).

$$x_{ij} = \frac{\left| n'_{ij} - N_j \right|}{\left| n'_{ij} - N_j \right| + \left| n'_{ij} - P_j \right|}, \qquad \forall j = 1, 2, ..., n \tag{8}$$

Step 9. Calculate the sigmoid function or logistic function based on the ideal solution matrix. The sigmoid function is represented as (9) where $\boldsymbol{w}$ is the weighted features vector and b is a bias term or intercept.

$$z = \boldsymbol{w} \cdot \boldsymbol{x} + b \tag{9}$$

Step 10. Establish a probability that we apply the sigmoid function to the sum of the weighted features. The probability can be obtained using (10) .

$$\begin{cases} P(y=1) = \dfrac{1}{1+\exp(-z)} \\ P(y=0) = \dfrac{\exp(-z)}{1+\exp(-z)} \end{cases} \tag{10}$$

Step 11. Classify record r_i by computing the probability using (11) to ensure that whether it is belonging to class 1 or 0.

$$decision(r_i) = \begin{cases} 1, & P(y=1|r_i) > 0.5 \\ 0, & Otherwise \end{cases} \tag{11}$$

Performance Measurement

Usually, the instruments used to measure performances in classification cases are overall accuracy, precision, F-measure, mean absolute error (MAE), True Positive (TP) rate, and True Negative (TN) rate. The following are definitions of these instruments.

(a). *Overall accuracy*: It is the percentage of accurately classified records. It can be calculated using (12) in which TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively.

$$Accuracy = \frac{TN + TP}{TP + FP + FN + TN} \tag{12}$$

(b). *True Positive (TP) rate*: It is also known as sensitivity, which is the number of accurately classified positive records. The TP rate can be obtained using (13).

$$TP_{Rate} = \frac{TP}{TP + FN} \tag{13}$$

(c). *True Negative (TN) rate*: It is also known as specificity, which is the number of accurately classified negative records. The TP rate can be obtained using (14).

$$TN_{Rate} = \frac{TN}{TN + FP} \tag{14}$$

(d). Mean absolute error (MAE): It estimates how much the predictions differed from the true probability. Eq (15) is a formula to compute MAE in which x_i is the actual value of instance i, $\tilde{x}_i$ is the predicted value of instance i, and N is the number of instances.

$$MAE = \frac{\sum_{i=1}^{N} \left(x_i - \tilde{x}_i \right)^2}{N} \tag{15}$$

(e). *Precision*: It is an ability of a classification method to specify only the appropriate data points. The precision can be determined by using (16).

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

(f). *Recall*: It is a metric that gauges the number of true positive predictions created by all correct predictions that could have been made. Eq. (17) is a formula to calculate the recall.

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

(g). *F-Measure*: It is defined as the harmonic mean of precision and recall [25]. Eq. (18) is a mathematical equation to calculate F-measure.

$$F_1\_Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{18}$$

RESULTS AND DISCUSSION

A Numerical Example

The MADM problem assumed in this article is the difficulty of credit risk examination of a bank in Germany. Supposes that an expert from the bank, who has professional experience in decision-making, has provided the individual assessment for features in a dataset. To be clear, we will illustrate this problem by demonstrating a numerical example.

Table 6. dataset matrix

A4	A5	A6	A7	A8	A9
A42	7882	A61	A74	2	A93
A40	4870	A61	A73	3	A93
A46	9055	A65	A73	2	A93
A42	2835	A63	A75	3	A93
A41	6948	A61	A73	2	A93
A43	3059	A64	A74	2	A91
A40	5234	A61	A71	4	A94

Table 7. The linguistic variable matrix

A4	A5	A6	A7	A8	A9
I	7882	P	VVP	2	EP
VVP	4870	P	VP	3	EP
VP	9055	I	VP	2	EP
I	2835	VVP	EP	3	EP
P	6948	P	VP	2	EP
FP	3059	EP	VVP	2	VP
VVP	5234	P	MI	4	VP

Table 8. The defuzzification matrix

A4	A5	A6	A7	A8	A9
0.125	7882	0.625000	0.875000	2	0.958333
0.875	4870	0.625000	0.750000	3	0.958333
0.750	9055	0.125000	0.750000	2	0.958333
0.125	2835	0.875000	0.958333	3	0.958333
0.625	6948	0.625000	0.750000	2	0.958333
0.375	3059	0.958333	0.875000	2	0.750000
0.875	5234	0.625000	0.250000	4	0.750000

Table 9. Priority order of attributes based on the expert's professional assessment

	1st Priority	3rd Priority	4th Priority	6th Priority
Expert	A4 and A5	A7 and A8	A6	A9

Table 10. Attributes' weights calculated by the OPA method

Weights	A4	A5	A6	A7	A8	A9
w_j	0.324324	0.324324	0.081082	0.108108	0.108108	0.054054

Additionally, this section demonstrates how the proposed model performs a classification process using the numerical example. Suppose we took dataset 1 (Table 3) and adjusted them to be six features A4, A5, A6, A7, A8, and A9, and seven records (r_i). The adjusted dataset has four categorical features and two numerical features. Based on the professional assessments and experiences of the expert, he has decided that features A4, A5, A6, A7, and A9 are beneficial attributes, and A8 is a cost attribute. The computational steps of the proposed model are the following.

Step 1. The numerical example of a dataset consisting of linguistic variables can be viewed as a matrix *B*, as presented in Table 6.

Step 2. All categorical features' entries of the matrix *B* were converted into TFNs (Table 7) and then calculated their defuzzification. The results can be seen in Table 8.

Step 3. The expert has decided that attributes A4, A5, A6, A7, and A9 are beneficial attributes and A8 is a cost attribute.

Step 4. The expert has assigned the priority orders of the attributes (see Table 9). The OPA-solver [34] then calculated the attributes' weights computationally based on the result from Table 9. The obtained attributes' weights can be seen in Table 10.

Step 5. The proposed model constructed the normalized matrix by using (4), as presented in Table 11.

Step 6. The proposed model utilized (5) and the result from Table 10 to compute all entries of the weighted normalized matrix.

Step 7. The proposed model calculates the positive (Π) and negative ideal (N) solution for each attribute. The negative

Table 11. The normalized matrix

A4	A5	A6	A7	A8	A9
0.000000	0.811415	0.6	0.882353	1.0	1.0
1.000000	0.327170	0.6	0.705883	0.5	1.0
0.833333	1.000000	0.0	0.705883	1.0	1.0
0.000000	0.000000	0.9	1.000000	0.5	1.0
0.666667	0.661254	0.6	0.705883	1.0	1.0
0.333333	0.036013	1.0	0.882353	1.0	0.0
1.000000	0.385691	0.6	0.000000	0.0	0.0

Table 12. The weighted normalized matrix

A4	A5	A6	A7	A8	A9
0.000000	0.263161	0.048649	0.095389	0.108108	0.054054
0.324324	0.106109	0.048649	0.076312	0.054054	0.054054
0.270270	0.324324	0.000000	0.076312	0.108108	0.054054
0.000000	0.000000	0.072974	0.108108	0.054054	0.054054
0.216216	0.214461	0.048649	0.076312	0.108108	0.054054
0.108108	0.011680	0.081082	0.095389	0.108108	0.000000
0.324324	0.125089	0.048649	0.000000	0.000000	0.000000

Table 13. The negative and positive ideal solutions for each attribute

	A4	A5	A6	A7	A8	A9
Π_j	0.324324	0.324324	0.081082	0.108108	0	0.054054
N_j	0	0	0	0	0.108108	0

Table 14. The ideal solution matrix for each record

A4 (x_1)	A5 (x_2)	A6 (x_3)	A7 (x_4)	A8 (x_5)	A9 (x_6)
0.000000	0.811415	0.6	0.882353	0.0	1.0
1.000000	0.327170	0.6	0.705883	0.5	1.0
0.833333	1.000000	0.0	0.705883	0.0	1.0
0.000000	0.000000	0.9	1.000000	0.5	1.0
0.666667	0.661254	0.6	0.705883	0.0	1.0
0.333333	0.036013	1.0	0.882353	0.0	0.0
1.000000	0.385691	0.6	0.000000	1.0	0.0

$$z = 0.6417x_1 - 0.1545x_2 - 0.01033x_3 - 0.44097x_4 + 0.6545x_5 - 0.2034x_6 - 0.9856 \quad (19)$$

Table 15. The possibility of each record

Records	P(y=1)	Class
r_1	0.84679	1
r_2	0.64290	1
r_3	0.75398	1
r_4	0.78785	1
r_5	0.76508	1
r_6	0.76434	1
r_7	0.43908	0

and positive ideal solution can be seen in Table 13.

Step 8. All entries of the ideal solution matrix were calculated by using (8). Thus, the ideal solution matrix can be presented as Table 14.

Step 9. The proposed model used the result from Table 14 to determine the sigmoid function. After running the logistic regression model using a Python code, the obtained intercept value (bias term) was -0.9856 and a weight vector was [0.6417, -0.1545, -0.01033, -0.44097, 0.6545, -0.2034]. Eq. (19) is the sigmoid function received from this step.

Step 10. Eq. (10) calculated the possibilities for all the record. The result can be presented in Table 15.

Step 11. Based on the result from Table 15 and Eq. (11), we can classify all the record as follows. Records r_1, r_2, r_3, r_4, r_5, and r_6 are belonging to class 1, while record r_7 is belong to class 0.

Experimental Results for Datasets

We converted categorical features into TFNs based on the expert's professional judgments and experiences for all datasets. We then executed three classification methods for each dataset to examine the prediction performance measures separately. Next, we utilized F-measure to measure the classification accuracy for the dataset's performance in predicting a biner classification.

Table 16 exhibits the accuracy value for logistic regression, Support Vector Machine (SVM), k-NN, and the proposed hybrid models. As can be seen in Table 16, the proposed hybrid model delivers better accuracy than other methods in two datasets, except for dataset 2. However, in dataset 2, the proposed model shows slightly lower performance than the SVM model. Figure

2 shows the accuracy performances of logistic regression, SVM, k-NN, and the proposed hybrid models for datasets 1, 2, and 3.

Moreover, this section estimates credit risk and extracts crucial features in the prediction for each dataset. Since the proposed model is based on logistic regression, we present only comparison results between the original logistic regression and the proposed hybrid models. We utilized logistic regression computation functions provided in Python with all standard procedures to show the results. We also embedded those standard procedures to develop our hybrid model. We will discuss the prediction performances for all three datasets.

***Dataset 1*:**

Dataset 1 has 20 independent features with 13 categorical features and seven numerical features to predict credit risk. We know that not all the features can contribute toward the prediction. Therefore, we will find the features that can impact credit risk prediction. To do this, we first split the dataset into training and testing with 80%:20% portions. We then created the model on the training dataset and tested it on the testing dataset. Table 16 shows the result of prediction performance for all the features with their significance levels for dataset 1 using the original logistic regression and the proposed hybrid models, respectively.

As can be seen from Table 16, the most crucial features exhibit p-values less than 0.05 [27]. It indicates that these independent features can significantly impact dependent variable prediction [28]. Thus, for the logistic regression model, the most important ones affecting the credit risk prediction are *credit_history*, *installment_rate*, *personal_status*, *duration*, *telephone*, *savings_account*, *employment_period*, *installment_plans*, *property*, *purpose*, *credit_amount*, and *account_status*. Meanwhile, the features *credit_history*, *employment period*, *installment_plans*, *personal_status*, *property*, *duration*, and *age* are the most impact features on the credit risk prediction suggested by the proposed hybrid model.

***Dataset 2*:**

Dataset 2 consists of three independent categorical features and seven independent numerical features. Similar to dataset 1, we performed a prediction test after splitting the dataset into training and testing datasets under the 80%:20% portion. We obtained that features that have p-values less than 0.05 for logistic regression are features *loan_int_rate*, *loan_percent_income*, *cb_person_default_on_file*, *person_income*, and *loan_intent*. Meanwhile, the most important features generated by the proposed model are features *person_income, loan_int_rate, cb_person_default_on_file, loan_amnt*, *person_emp_length*, and *loan_intent*. Table 17 shows the prediction performances of dataset 2 for both models.

Table 16. Features in logistic regression and the proposed hybrid models with their significances for dataset 1

No.	The Logistic Regression		The Proposed Hybrid Model	
	Features	***p*-value**	**Features**	***p*-value**
1	credit_history	0	credit_history	0
2	installment_rate	0.0003	Employment_period	0.0004
3	personal_status	0.0021	installment_plans	0.0011
4	duration	0.0031	installment_rate	0.0038
5	telephone	0.0042	personal_status	0.0046
6	savings_account	0.0061	property	0.0054
7	Employment_period	0.0092	duration	0.0058
8	Installment_plans	0.0198	age	0.0149
9	property	0.0218	telephone	0.0222
10	purpose	0.0268	savings_account	0.0994
11	credit_amount	0.0366	purpose	0.1275
12	account_status	0.0448	credit_amount	0.1358
13	foreign_worker	0.0508	foreign_worker	0.1358
14	number_of_existing_credits	0.0683	account_status	0.1454
15	age	0.0855	number_of_existing_credits	0.1913
16	number_of_ responsible_people	0.1913	number_of_ responsible_people	0.2041
17	housing	0.5371	job	0.3807
18	other_debtors	0.6424	other_debtors	0.6025
19	residence_period	0.6651	residence_period	0.6301
20	job	0.8927	housing	0.7851

Table 17. Features in logistic regression and the proposed hybrid models with their significances for dataset 2

No.	The Logistic Regression		The Proposed Hybrid Model	
	Features	***p*-value**	**Features**	***p*-value**
1	loan_int_rate	0	person_income	0
2	loan_percent_income	0.0006	loan_int_rate	0
3	cb_person_default_on_file	0.0193	cb_person_default_on_file	0.0058
4	person_income	0.0244	loan_amnt	0.0059
5	loan_intent	0.0378	person_emp_length	0.0161
6	person_emp_length	0.0647	loan_intent	0.0218
7	person_age	0.2461	person_home_ownership	0.3179
8	loan_amnt	0.5854	cb_preson_cred_hist_length	0.3354
9	cb_preson_cred_hist_length	0.7545	loan_percent_income	0.7705
10	person_home_ownership	0.8919	person_age	0.7928

Table 18. Features in logistic regression and the proposed hybrid models with their significances for dataset 3

No.	The Logistic Regression		The Proposed Hybrid Model	
	Features	***p*-value**	**Features**	***p*-value**
1	Applicants_income	0.0001	Applicants_income	0.0001
2	Outgoings_on_credit_cards	0.0323	Outgoings_on_credit_cards	0.0322
3	Year_of_birth	0.0455	Year_of_birth	0.0849
4	Residential_status	0.0791	Spouses_income	0.0888
5	Spouses_income	0.0807	Residential_status	0.2142
6	Applicants_employment_status	0.2065	Outgoings_on_Hire_Purchase	0.2263
7	Outgoings_on_Hire_Purchase	0.2297	Applicants_employment_status	0.2557
8	Outgoings_on_Loans	0.4625	Outgoings_on_Loans	0.4549
9	Number_of_other_dependents	0.5036	Number_of_other_dependents	0.4868
10	Outgoings_on_mortgage_or_rent	0.5696	Outgoings_on_mortgage_or_rent	0.5741
11	Value_of_Home	0.5965	Number_of_children	0.6725
12	Number_of_children	0.6428	Value_of_Home	0.7067
13	Having_a_home_phone	0.7082	Mortgage_balance_outstanding	0.7267
14	Mortgage_balance_outstanding	0.7125	Having_a_home_phone	0.9764

***Dataset 3*:**

Dataset 3 contains two independent categorical features and 12 independent numerical features. Similar to datasets 1 and 2, after splitting the dataset into training and testing datasets, we modeled this dataset to predict the most crucial features in the credit risk prediction suggested by logistic regression and our proposed models. The logistic regression model suggested features *Applicants_income*, *Outgoings_on-credit_cards*, and *Year_of_birth* as the most crucial features, while the proposed model generated *Applicants_income* and *Outgoings_on_credit_cards* as the most important ones. Table 17 demonstrates the

Table 19. Classification results of dataset 1

Methods	Accuracy	MAE	TP	TN	Precision	Recall	F Measure
Logistic Regression	0.73	0.27	0.9014	0.3103	0.6622	0.6059	0.6129
SVM	0.695	0.305	**0.9424**	0.1311	0.606	0.5368	0.5095
k-NN with k=7	0.725	0.275	0.8836	0.2963	0.6287	0.5899	0.596
the Proposed Hybrid	**0.76**	**0.24**	0.9225	**0.3621**	**0.718**	**0.6423**	**0.6559**

Table 20. Classification results of dataset 2

Methods	Accuracy	MAE	TP	TN	Precision	Recall	F Measure
Logistic Regression	0.83	**0.165**	0.2683	0.9811	0.8122	0.6247	0.6522
SVM	0.8	0.2	**0.2941**	0.9732	0.7953	**0.6336**	**0.6537**
k-NN with k=7	0.795	0.205	0.2766	0.9542	0.7306	0.6154	0.6325
the Proposed Hybrid	**0.85**	0.195	0.0488	**1**	**0.9015**	0.5244	0.4919

Table 21. Classification results of dataset 3

Methods	Accuracy	MAE	TP	TN	Precision	Recall	F Measure
Logistic Regression	0.7429	0.2571	0	**1**	0.3714	0.5	0.4262
SVM	0.7633	**0.2367**	0.1167	0.973	0.6779	0.5448	0.5278
k-NN with *k*=7	0.7143	0.2857	**0.2881**	0.8495	0.5839	**0.5688**	**0.5728**
the Proposed Hybrid	**0.7695**	0.2408	0.0952	0.989	**0.7547**	0.5421	0.5141

feature performances of dataset 3 for both models.

Classification Results for Datasets

Dataset 1: Table 19 demonstrates the testing results of the German Credit dataset utilizing logistic regression, SVM, k-NN, and our hybrid methods. Each method is measured by the seven performance instruments provided in Section Performance Measurement. The values with boldface in each performance instrument's column are the best result. The proposed hybrid method performs well on this dataset. This method achieves almost all the performance instruments except for instrument TP.

Dataset 2: Table 20 shows different superiority results for each measurement tool. For the measurement tool of MAE, the logistic regression model is slightly superior to our proposed hybrid model, which is different from 0.03 only. While for measurement tools TP, recall, and F-Measure, SVM has relatively higher than the proposed model. However, for accuracy, TN, and precision, the proposed hybrid model demonstrates relatively far better than the others. It confirms that the proposed model achieves better results for almost four measurement tools.

Dataset 3: Table 21 presents the testing results for dataset Credit Scoring using logistic regression, SVM, k-NN, and our hybrid methods. The proposed hybrid model shows better results for accuracy and precision. SVM achieves a better value for MAE. Meanwhile, k-NN demonstrates its best performance for the others. Nevertheless, in general, it can be seen that, for this dataset, the proposed model is a bit superior because it can achieve the best overall accuracy result compared with the others. Additionally, the different MAE value obtained by the proposed model and SVM is not far, that is only 0.0041. Thus, based on the analyses of these three datasets, it is clear that the proposed model has shown better performance for the three datasets, especially for a categorical dataset.

CONCLUSION

At the beginning of the paper, we stated that this study used two strategies to deal with credit risk evaluation. The first strategy is logistic regression, and the second is viewing the problem as a MADM problem. The objective of using these two strategies is as follows. Every strategy has different results, and no strategies provide the optimal output. Sometimes, in practice cases, it is challenging for the expert in a bank to deliver a decision according to only one strategy, due to in practice, the experts are considering many qualitative impacts. In this situation, the expert wants to listen to arguments and suggestions for deciding a policy.

Therefore, to help the expert, this study aims to propose a hybrid intelligent model that blends the logistic regression algorithms and MADM strategy, which have hard work to classify a binary classification effectively in large data sets. The results promised and demonstrated that this hybrid model performs more effectively and has higher accuracy than logistic regression, SVM, and *k*-NN models for the dataset containing more categorical features. It confirms that the proposed hybrid model is a model that develops the capability of prediction of these two approaches, especially for predicting a dataset containing many

categorical features and viewing it as a MADM problem. However, we realize that the testing process should be expanded with balanced datasets between the two classes, and it involves many experts in contributing their opinions on the priority of features and evaluating the features. Additionally, this study can conduct a valuable comparison between the proposed hybrid model with other existing classifiers.

It is already known that the logistic regression model has a better classification performance for the training data sets containing a balance between the number of acceptable and unacceptable credit loan recipients. It is possible to increase the accuracy of the proposed hybrid model by concerning sampling models. If each independent feature of the balanced training datasets has approximately 50 % acceptable and 50 % unacceptable observations, the classification performance of the logistic regression model could be improved from the current sampling. Hence, this increase in performance will give better predictive probabilities, and this could also efficiently influence the overall model's conquest.

For future studies, to improve the accuracy of the proposed hybrid model, some primary studies can be conducted effectively. Before applying logistic regression to the dataset, some additional stages in MADM methodology have to be considered, such as selecting spherical fuzzy numbers [38], utilizing neutrosophic numbers [39] that can handle not only uncertain information but also non-membership, indeterminacy, and inconsistency information, selecting optimal features' weights automatically by using genetic algorithm to increase the accuracy of the proposed hybrid method, and improving MADM steps based on distance from average solution [31], and other methods [41][42][43] to be more reasonable in decision-making.

ACKNOWLEDGMENT

The authors would like to show thankfulness to the Doctoral Study Program in Computer Science and the Department of Computer Science, the Faculty of Computer Science and Information Technology of Universitas Sumatera Utara for supporting this experimental study, and the authors also thank anonymous reviewers for their recommendations and commentaries on the initial paper.

REFERENCE

[1] Chortareas, G.; Magkonis, G.; Zekente, K.-M. 2020. Credit risk and the business cycle: What do we know?. *Int. Rev. Financ. Anal.*, **67** 101421, DOI: 10.1016/j.irfa.2019.101421.

[2] Bannier, C. E.; Bofinger, Y.; Rock, B. 2022. Corporate social responsibility and credit risk. *Financ. Res. Lett.*, **44** 102052, DOI: 10.1016/j.frl.2021.102052.

[3] Naili, M; Lahrichi, Y. 2022. Banks' credit risk, systematic determinants and specific factors: recent evidence from emerging markets. *Heliyon*, **8**, no. 2, e08960, DOI: 10.1016/j.heliyon.2022.e08960.

[4] Li, Y. 2019. Credit Risk Prediction Based on Machine Learning Methods. in *2019 14th International Conference on Computer Science & Education (ICCSE)*, pp. 1011–1013. DOI: 10.1109/ICCSE.2019.8845444.

[5] Jinjuan, L. 2017. Research on Enterprise Credit Risk Assessment Method Based on Improved Genetic Algorithm. in *2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pp. 215–218. DOI: 10.1109/ICMTMA.2017.0058.

[6] Yang, P.; Wang, W.; Chen, Y. 2020. Application of Neural Network Based on Flexible Neural Tree in Personal Credit Evaluation. in *2020 12th International Conference on Advanced Computational Intelligence (ICACI)*, pp. 218–223. DOI: 10.1109/ICACI49185.2020.9177759.

[7] Liu, S.; Wang, R.; Han, Y. 2021. Research on Personal Credit Evaluation Based on Machine Learning Algorithm. in *2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT)*, pp. 48–52. DOI: 10.1109/ISCIPT53667.2021.00016.

[8] Lin Z; Xingzhong, B; Yajun, C; Ting, W; Fei-hu, H; Mingzhu, L; Jian, P. 2020. Research on Power Market User Credit Evaluation Based on K-Means Clustering and Contour Coefficient. in *2020 3rd International Conference on Robotics, Control and Automation Engineering (RCAE)*, pp. 64–68. DOI: 10.1109/RCAE51546.2020.9294725.

[9] Dattachaudhuri, A.; Biswas, S.; Sarkar, S.; Boruah, A. N. 2020. Transparent Decision Support System for Credit Risk Evaluation: An automated credit approval system. in *2020 IEEE-HYDCON*, pp. 1–5. DOI: 10.1109/HYDCON48903.2020.9242905.

[10] Shi, H.; Wang, Z.; Wang, X. 2020. Credit

Risk Intelligent Assessment Model Based on Machine Learning. in *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*, pp. 735–738. DOI: 10.1109/ICETCI55101.2022.9832083.

[11] Egger, D. J.; García Gutiérrez, R.; Mestre, J. C.; Woerner, S. 2021. Credit Risk Analysis Using Quantum Computers. *IEEE Trans. Comput.*, **70** (12) 2136–2145, DOI: 10.1109/TC.2020.3038063.

[12] Hassija, V.; Bansal, G.; Chamola, V.; Kumar, N.; Guizani, M. 2020. Secure Lending: Blockchain and Prospect Theory-Based Decentralized Credit Scoring Model. *IEEE Trans. Netw. Sci. Eng.*, **7** (4) 2566–2575, 2020, DOI: 10.1109/TNSE.2020.2982488.

[13] Li, Q. 2023. Research on Bank Credit Risk Assessment Based on BP Neural Network. in *2023 2nd International Conference on 3D Immersion, Interaction and Multi-sensory Experiences (ICDIIME)*, pp. 322–326. DOI: 10.1109/ICDIIME59043.2023.00068.

[14] Dumitrescu, E.; Hué, S.; Hurlin, C. Tokpavi, S. 2022. Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *Eur. J. Oper. Res.*, **297** (3) 1178–1192, DOI: 10.1016/j.ejor.2021.06.053.

[15] Zabor, E. C.; Reddy, C. A.; Tendulkar, R. D.; Patil, S. 2022. Logistic Regression in Clinical Studies. *Int. J. Radiat. Oncol.*, **112** (2) 271–277, DOI: 10.1016/j.ijrobp.2021.08.007.

[16] Yoshida, R.; Hara, H.; Saluke, P. M. 2019. Sequential Importance Sampling for Logistic Regression Model. in *Computational Models for Biomedical Reasoning and Problem Solving*, C.-H. Chen and S.-C. S. Cheung, Eds., Hershey, PA, USA: IGI Global, pp. 231–255. DOI: 10.4018/978-1-5225-7467-5.ch009.

[17] Yan, M. 2019. Personal Credit Rating System Based on the Logistic Regression Method. in *2019 International Conference on Economic Management and Model Engineering (ICEMME)*, pp. 156–163. DOI: 10.1109/ICEMME49371.2019.00040.

[18] Dinh, T. N.; Thanh, B. P. 2022. Loan Repayment Prediction Using Logistic Regression Ensemble Learning With Machine Learning Algorithms. in *2022 9th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, pp. 79–85. DOI: 10.1109/ISCMI56532.2022.10068483.

[19] Assef, F.; Steiner, M. T.; Steiner Neto, P. J.; Franco, D. G. B. 2019. Classification Algorithms in Financial Application: Credit Risk Analysis on Legal Entities. *IEEE Lat. Am. Trans.*, **17** (10) 1733–1740, DOI: 10.1109/TLA.2019.8986452.

[20] Manglani R.; Bokhare, A. 2021. Logistic Regression Model for Loan Prediction: A Machine Learning Approach. in *2021 Emerging Trends in Industry 4.0 (ETI 4.0)*, pp. 1–6. DOI: 10.1109/ETI4.051663.2021.9619201.

[21] Zhang Z.; Han, Y. 2020. Detection of Ovarian Tumors in Obstetric Ultrasound Imaging Using Logistic Regression Classifier With an Advanced Machine Learning Approach. *IEEE Access*, **8** 44999–45008, DOI: 10.1109/ACCESS.2020.2977962.

[22] Chang, C.-C.; Chen, C.-H.; Hsieh, J.-G.; Jeng, J.-H. 2021. Survival Prediction in Patients with Diffuse Large B-cell Lymphoma Using Logistic and Neural Network Models, in *2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, pp. 67–68. DOI: 10.1109/ECBIOS51820.2021.9510518.

[23] Schryvers, S.; De Bock, T.; Uyttendaele, M.; Jacxsens, L. 2023. Multi-criteria decision-making framework on process water treatment of minimally processed leafy greens. *Food Control*, **148** p. 109661, DOI: 10.1016/j.foodcont.2023.109661.

[24] Mendonça, G. H. M.; Ferreira, F. G. D. C.; Cardoso, R. T. N.; Martins, F. V. C. 2020. Multi-attribute decision making applied to financial portfolio optimization problem. *Expert Syst. Appl.*, **158** p. 113527, DOI: 10.1016/j.eswa.2020.113527.

[25] Syed Z.; Lawryshyn, Y. 2020. Multi-criteria decision-making considering risk and uncertainty in physical asset management. *J. Loss Prev. Process Ind.*, **65** p. 104064, DOI: 10.1016/j.jlp.2020.104064.

[26] Pasman, H. J.; Rogers, W. J.; Behie, S. W. 2022. Selecting a method/tool for risk-based decision making in complex situations. *J. Loss Prev. Process Ind.*, **74** p. 104669, DOI: 10.1016/j.jlp.2021.104669.

[27] Sun, R.; Gong, Z.; Gao, G.; Shah, A. A. 2020. Comparative analysis of Multi-Criteria Decision-Making methods for flood disaster risk in the Yangtze River Delta. *Int. J. Disaster Risk Reduct.*, **51**, p. 101768, DOI: 10.1016/j.ijdrr.2020.101768.

[28] Thomas, L.; Crook, J.; Edelman, D. 2017

Credit Scoring and Its Applications, 2nd ed. Philadelphia, PA, USA: SIAM-Society for Industrial and Applied Mathematics.

[29] Pisner, D. A.; Schnyer, D. M. 2020. Chapter 6 - Support vector machine, in *Machine Learning*, A. Mechelli and S. Vieira, Eds., Academic Press, pp. 101–121. DOI: 10.1016/B978-0-12-815739-8.00006-7.

[30] Wang, Y.; Pan, Z.; Dong, J. 2022. A new two-layer nearest neighbor selection method for kNN classifier. *Knowledge-Based Syst.*, **235** p. 107604, DOI: 10.1016/j.knosys.2021.107604.

[31] Zadeh, L. A. 1975. The concept of a linguistic variable and its application to approximate reasoning—I, *Inf. Sci. (Ny).*, **8** (3) 199–249, DOI: 10.1016/0020-0255(75)90036-5.

[32] Jiang, Y.-P.; Fan, Z.-P.; Ma, J. 2008. A method for group decision making with multi-granularity linguistic assessment information. *Inf. Sci. (Ny).*, **178** (4) 1098–1109, DOI: 10.1016/j.ins.2007.09.007.

[33] Rahmani, A.; Hosseinzadeh Lotfi, F.; Rostamy-Malkhalifeh, M.; Allahviranloo, T. 2016. A New Method for Defuzzification and Ranking of Fuzzy Numbers Based on the Statistical Beta Distribution. *Adv. Fuzzy Syst.*, **2016** p. 6945184, DOI: 10.1155/2016/6945184.

[34] Mahmoudi, A. 2021. OPA Solver: A Solver for Multiple-Attribute Decision-Making Problems. DOI: 10.5281/ZENODO.4453887.

[35] Jahan, A.; Edwards, K. L.; Bahraminasab, M. 2016. 5 - Multi-attribute decision-making for ranking of candidate materials. in *Multi-criteria Decision Analysis for Supporting the Selection of Engineering Materials in Product Design (Second Edition)*, A. Jahan, K. L. Edwards, and M. Bahraminasab, Eds., Second Edition.Butterworth-Heinemann, pp. 81–126. DOI: 10.1016/B978-0-08-100536-1.00005-9.

[36] Buckland, M.; Gey, F. 1994. The relationship between Recall and Precision. *J. Am. Soc. Inf. Sci.*, **45** (1) 12–19, DOI: 10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L.

[37] Jerić, S.; vSarlija, N.; vSori'c, K.; Rosenzweig, V. 2009. Logistic Regression and Multicriteria Decision Making in Credit Scoring,

[38] Kutlu Gündoğdu, F.; Kahraman, C. 2019. Spherical fuzzy sets and spherical fuzzy TOPSIS method. *J. Intell. Fuzzy Syst.*, **36** 337–352, DOI: 10.3233/JIFS-181401.

[39] Smarandache, F. 1998 *Neutrosophy: Neutrosophic Probability, Set, and Logic: Analytic Synthesis & Synthetic Analysis*. American Research Press.

[40] Jana C.; Pal, M. 2021. Extended bipolar fuzzy EDAS approach for multi-criteria group decision-making process. *Comput. Appl. Math.*, **40** (1) p. 9, DOI: 10.1007/s40314-020-01403-4.

[41] Irvanizam, I.; Syahrini, I.; Afidh, R. P. F.; Andika, M. R.; Sofyan, H. 2019. Applying Fuzzy Multiple-Attribute Decision Making Based on Set-pair Analysis with Triangular Fuzzy Number for Decent Homes Distribution Problem. in *2018 6th International Conference on Cyber and IT Service Management, CITSM 2018*, DOI: 10.1109/CITSM.2018.8674290.

[42] Irvanizam, I.; Nazaruddin, N.; Syahrini, I. 2018. Solving Decent Home Distribution Problem Using ELECTRE Method with Triangular Fuzzy Number. in *Proceedings of ICAITI 2018 - 1st International Conference on Applied Information Technology and Innovation: Toward A New Paradigm for the Design of Assistive Technology in Smart Home Care*, DOI: 10.1109/ICAITI.2018.8686768.

[43] Irvanizam, I.; Marzuki, M.; Patria, I.; Abubakar, R. 2018. An Application for Smartphone Preference Using TODIM Decision Making Method. in *Proceedings - 2nd 2018 International Conference on Electrical Engineering and Informatics, ICELTICs 2018*, DOI: 10.1109/ICELTICS.2018.8548820.