Date of publication JUN-30, 2022, date of current version MAY-22, 2022.
www.computingonline.net / computing@computingonline.net

Print ISSN 1727-6209
Online ISSN 2312-5381
DOI 10.47839/ijc.21.2.2593

Using Class Membership based Approach to Improve Predictive Classification in Customer Relationship Management Systems

STEPHANE C. K. TEKOUABOU[1,2], WALID CHERIF[3], HAMZA TOULNI[4,5], ELARBI A. ABDELAOUI[6], HASSAN SILKAN[2]

[1]Center of Urban Systems (CUS), Mohammed VI Polytechnic University (UM6P), Lot 660, Hay Moulay Rachid, 43150, Ben Guerir, Morocco
[2]Laboratory LAROSERI, Department of Computer Science, Faculty of Sciences, B.P. 20, 24000, El Jadida, Morocco
ctekouaboukoumetio@gmail.com, silkan_h@yahoo.fr
[3]National Institute of Statistics and Applied Economics / Department of Computer Science, B.P. 6217, Rabat, Morocco, chrf.walid@gmail.com
[4]LIMSAD Laboratory, Ain Chock Hassan II University of Casablanca, Casablanca, Morocco.
hamza.toulni@outlook.com
[5] EIGSI, 282 Route of the Oasis, Maârif 20140, Casablanca, Morocco.
[6]Department of Science, Ecole Normale Supérieure, Moulay Ismail University, Meknes, Morocco
abdellaoui.e@gmail.com

Corresponding author: Stéphane C. K. Tékouabou (e-mail: ctekouaboukoumetio@gmail.com).

ABSTRACT Recently, the diversity of data collected on both social networks and digital interfaces is extremely increased. This diversity of data raises the problem of heterogeneous variables that are not favourable to classification algorithms. Although machine learning and predictive analysis have significantly improved the efficiency of the classification in customer relationship management (CRM) systems, their performance remains very limited by heterogeneous data processing. In this paper, we propose a new predictive classification approach well adapted for targeting actual CRM systems. Our approach consists of preprocessing each type of feature and constructing a reduced array. From this reduced array, the class membership computations become very faster and perform the predictive targeting of a new instance great accurately. The results of the experiments carried out on four types of data from the CRMs showed that the proposed algorithm is a good tool for strengthening these systems not only to optimize their loyalty actions but also to efficiently acquire new customers.

KEYWORDS CRM; Targeting; Machine learning; Predictive classification; Predictive analysis; Data mining; Classification; CMB.

I. INTRODUCTION

CUSTOMER relationship management (CRM) tools are increasingly enriched with the use of Artificial Intelligence (AI) algorithms for real-time prediction [1, 2]. For example, this use may be intended to help companies: predict scores on their opportunities; predict potential customers and purchases; customer reliability and loyalty, [3] detect credit and insurance fraud [5, 6], Customer Satisfaction [7, 8] predict personalized recommendations for products and services [7], [10] or the risk of churn (loss of customers or subscribers) [9–12]. Faced with these major challenges and tough competition, these algorithms are better than simple tools now at the centre of the digitalization of all companies.

Different machine learning algorithms have been implemented directly in CRM; however, the quality of their prediction closely depends on the dataset present in the CRM database. Nowadays, the CRM data is not only limited to information entered manually by the customers (sometimes not very reliable) but also includes other data from very operational activities such as interactions, appointments, contacts, opportunities, requests and complaints, difficulties encountered by salespeople, and also data collected from external sources (such as social networks) to enrich the CRM [13, 14]. We speak of heterogeneous data. The main characteristic of heterogeneous data is that they are of several types (numerical, Boolean, scaled, nominal…). To achieve the best performance of prediction algorithms, a transformation of these heterogeneous data is needed. Its optimization is one of the main challenges as currently no machine learning algorithm well processes non-digital data. Optimizing the transformation of this mosaic dataset for better prediction performance is the first goal we seek to achieve during preprocessing.

The peculiarity of our approach compared to existing algorithms is that the preprocessing of nominal type features that are difficult to process in the existing approaches so far simply involves imputing the missing values. The nominal features are directly used in the class membership-based (CMB) classification phase process. This and the elimination of non-significant attributes allow the processing time, which constitutes the second challenge of our approach.

Generally, transformed data presents a problem with different variable scales. Standardization is an additional step in dealing with this problem but it has an adverse or favourable effect on the performance of certain algorithms that are said to be unstable to variable scales; knowing that a very large-scale difference is more favourable for overfitting. We have therefore created the proposed classification algorithm, which, from the reduced table of the training database, independently classifies each attribute before predicting the class of the individual. This is the third contribution of our paper.

The binary classification constitutes a major part of the predictive classification problems [15, 16], especially concerning the CRM, our global approach was then simplified in this work in the case of binary classification and then optimized by variable weighting which is the fourth challenge of this work. We finally evaluated the proposed approach on three classics in the fifth contribution and tested the algorithm on three classic predictive classification problems that are: customer churn prediction [19, 14] targeting before a telemarketing banking campaign, the own credit rating [20–22].

The rest of this paper is organized as follows: Sections 2 and 3 cover the research contributions and related works respectively; Section 4 details the proposed approach. In Section 5, the obtained results are analyzed. Finally, Section 6 concludes this work.

II. RESEARCH CONTRIBUTION

We introduce a new approach that processes heterogeneous data by transforming separately each type of feature (numerical, Boolean, scaled and nominal), then a hybrid technique to replace missing values and implicitly selects the most significant features. This helps to optimise the classification in terms of processing time and accuracy. Apart from the replacement of missing values, we do not transform nominal attributes in this step because they are directly treated in the classification; this allows reducing the processing time.

We construct the reduced table of training data. For each class, this table contains the averages of transformed features and the favourable attributes for nominal features.

We propose a simplification of the overall approach for the special case of binary classification. It incorporates a weighting scheme that improves performance.

III. RELATED WORKS

Based on previous research, data mining models are currently very much needed to support or apply the effects of a customer relationship management strategy [1] and this need does not date from today but it is growing with the central place that data has taken in recent years. The choice of a data-mining model is based on the key customer relationship management issues that the article wants to address. For example, clustering models are used more often for problems of automatic recommendation of products and services to customers [2, 23], regression models can predict customer scoring, while classification models are more useful for predicting potential targets, reliability, loyalty and satisfaction [2] or customers churn prediction [9–12].

However, classification algorithms remain the most widely used. A certain part of the leads can be segmented into groups to form the initial classes of the classification model strategy [1, 24]. E.W.T. Ngai's strategy [1] assert for example that concerning loyalty programs, 83.3% used classification models to assist in decision making. And these problems often come down to the binary classification for which the most used methods are: SVM [24]: DT [27, 28], ANN [18] and NB [29], [30]. E. Diaz-Aviles et al. [31] used for example classification techniques such as SVM, DT, and Random Forest to predict real-time customer experience which was a big challenge for traditional methods.

A. DECISION TREES

The basic algorithm of induction of the decision tree is based on the descending recursive method of constructing a decision tree. The algorithm uses information gain based on the measurement of entropy as heuristic information and selects a sample of classification attributes that can be called. The attribute becomes the test or decision attribute of the node [32, 28]. Very simple and fast, the algorithm becomes very inefficient for large databases but remains very much used in CRM issues such as targeting potential customers [33, 34].

B. SUPPORT VECTOR MACHINE (SVM)

SVM makes it possible to make predictions based on an important parameter which is the function of the kernel [35]. SVM contains several types of the kernel but the most commonly used are linear, polynomial and radial (RBF) kernel [36]. The RBF kernels implemented by default on python have the advantage of being able to model nonlinear relations whereas the linear kernel can not. It has fewer hyperparameters than the polynomial function and it has fewer computation problems (like numerical overfitting) because the kernel values are linked by zero and one, while the value of the polynomial kernel can go to infinity [37]. The RBF nucleus requires the choice of a single hyperparameter, the width of the Gaussian [38]. One of his strengths is the parameters of penalty factor C which specifies the compromise between the hyperplane violations and the margin size. SVM is very used in CRM because it has the advantage of being better suited for binary classification problems and relatively adapts to large volumes of data but becomes inefficient and very slow in dealing with large scales of variables [26, 27, 37].

C. ARTIFICIAL NEURAL NETWORKS (ANN)

The algorithm of artificial neural networks (ANN) is very schematically inspired by the functioning of biological neurons, which is the origin of its name. Neurons receive signals (electrical impulses) through highly branched extensions of their cell bodies (dendrites) and transmit information through long extensions (axons) [33]. The ANN algorithm can learn automatically a model using a network of reaction neurons formed by a retro propagation algorithm. These neurons are mathematical operators that perform weighted sum, followed by nonlinear functions [34]. Some advantages of a neural network's ability in CRM system are: significantly reduces losses by working in real-time, online or batch modes and will reinforce customer trust; improve operational efficiencies; and also give flexibility to easily incorporate data from many sources to the neural models [35].

Faced with the diversification and the volume of collected data, which are increasingly heterogeneous, the classification algorithms currently used are limited by the size of the data (which not only makes them slower but also weakens their performance), the gap of scales between them varies (that makes them unstable), heterogeneity of data and especially non-numerical data that require.

IV. PROPOSED APPROACH

This part presents our proposed approach, which consists of two mean steps: preprocessing and classification (see Fig. 1).

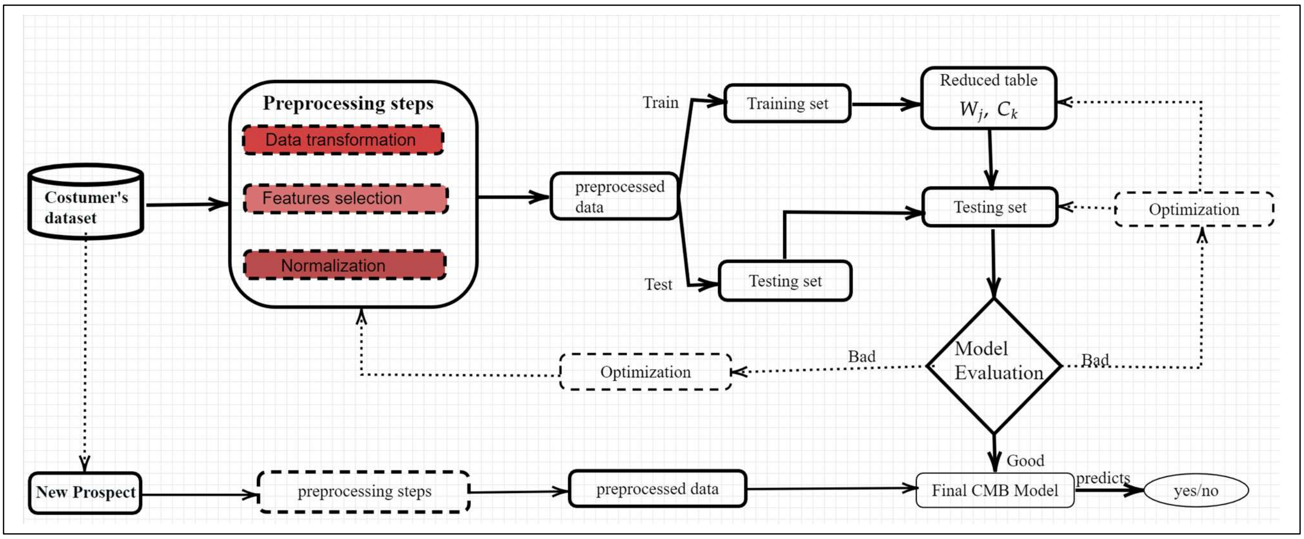

Figure 1. Global CMB approach performance on four CRM datasets.

The preprocessing transforms different types of features into specific formats reducing thus processing time and improving the classification performance while remaining very stable against variables scales by dealing with each attribute separately.

To clearly illustrate these different steps, let us consider an example of instances of the following Table 1 which contain some common features available in major customer datasets.

Table 1. Example of the initial dataset

Inst.	f_1	f_2	f_3	f_4	Y
I_1	59	married	No	regular	**no**
I_2	39	single	Yes	Very regular	**yes**
I_3	59	married	No	irregular	**no**
I_4	41	divorced	No	Unknown	**no**
I_5	44	married	No	irregular	**yes**

A. DATA PREPROCESSING

The preprocessing step turns out to be very important for the prediction process because the performance closely depends on it. The preprocessing combines data cleaning and feature engineering techniques to transform original features into a well-adapted machine learning algorithm. Many sources consider preprocessing steps as approximately 80% of data modelling tasks [34, 39], 40]. With the advent of big data, data are collected from several sources and are of different types: we are treating heterogeneous data [30, 34]. Here we propose a specific preprocessing in several phases: data transformation, replacing missing values, features selection and finally standardization again later.

A.1. DATA TRANSFORMATION

The first step of our proposed technique consists of defining the type of the feature as the technique differs for each type. Table 1 illustrates 4 major types of features: numeric, Boolean, scaled and nominal.

- For Numerical features (f_1): We calculate directly statistical parameters (min, max, mean, variance, standard deviation) of each numerical feature.
- For scaled Features (f_4): We substitute items by their ordinal number. After that substitution, we calculate the statistical parameters.
- For Boolean features such as (f_3): we have only two possibilities yes or no (1/0); success or failure (1/0); telephone/cellular (1/0).
- For Nominal features (f_2): These features are considered independent features and are directly associated with the classification step for our approach. This reduces the processing time of all nominal features by almost half while improving performance. Such features will be processed in the classification step.

This transformation does not deal with missing values. Unlike most works, our approach incorporates a hybrid imputation technique adapted to the different types of variables.

A.2 REPLACEMENT OF MISSING VALUES

The problem of missing values is often very common in heterogeneous datasets especially those collected from CRM and is a challenge for optimal data modelling. One of the advantages of our approach is that it proposes a function allowing a fast and optimal imputation of the missing values according to the type of features to which they belong [41]. Such function consists of replacing all the missing values $V_{ij}$of the feature V_j by the average if they are scaled or numerical variables or the mode if they are Boolean or nominal variables inside the class k:

$$(V_{ij})_{k\,miss} \leftarrow \begin{Bmatrix} Mode(V_j)_k \text{ if boolean or nominal} \\ Mean(V_j)_k \, if\ numeric\ or\ scaled \end{Bmatrix}. \quad (1)$$

For example, if we consider table 2 bellow: V_j="f_4" is scaled variable; I_4 takes the value "$unknown$" for f_4 and k= "yes", $Mean(V_j)_{yes} = \frac{(2+0)}{2} = 1$; The missed value will be replaced by 1.

A.3 FEATURES SELECTION

The analysis of correlations $(CC(f_j, Y))$ between feature (f_j) and class Y allows not only to identify those correlated strongly to the label and therefore that would contribute enough to the classification [42] but also reduces the feature space by setting a threshold. Weakly correlated variables that do not contribute to the decision and consume processing time are then ignored. They are neutral for classification decisions.

$$C_j=(CC(f_j,Y)) = \frac{\sum_{i=0}^{i=n}(f_{ij}-\bar{f_j}).(y_i-\bar{y})}{\sqrt{\sum_{i=0}^{i=n}(f_{ij}-\bar{f_j})}.\sqrt{\sum_{i=0}^{i=n}(y_i-\bar{y})}}. \quad (2)$$

In our case, we set a threshold of 1%, which allowed us to eliminate some variables (having $CC_j < 0.01$) and use only other ones which are significant for classification [40]. This only affects attributes initially numerical or Boolean types that have not been transformed during the transformation phase.

After all, preprocessing steps, the preprocessed table becomes as follows:

Table 2. Data Preprocessed

Inst.	f_1	f_2	f_3	f_4	Y
I_1	59	married	0	1	**0**
I_2	39	single	1	2	**1**
I_3	59	married	0	0	**0**
I_4	41	divorced	0	1	**0**
I_5	44	married	0	0	**1**

A.4 NORMALIZATION

The normalization is used to eliminate the impact of the order of magnitude and is very important to optimize machine learning model performance. It reduces each V_{ij} to the interval [0,1] by the formula:

$$V_{ij} \leftarrow \frac{V_{ij}-min_j(V_{ij})}{max_j(V_{ij})-min_j(V_{ij})}. \quad (3)$$

B. CLASSIFICATION

The classification principle of the CMB approach is that for each feature, it compares the value of this feature for a new instance with the averages of this feature's value in different classes. Then, based on this feature, it tentatively assigns the instance to the class for which the value of this feature is the closest to the corresponding average. Then, it adds weights of all the classes to which the object is tentatively based on different features, and for whatever class the weighted sum is largest, this is the class to which we classify the given object.

Thus, the proposed classification algorithm (Fig. 2) contains two phases. The first is to build the Reduced Table of training data centred on each class, and the second is the decision phase.

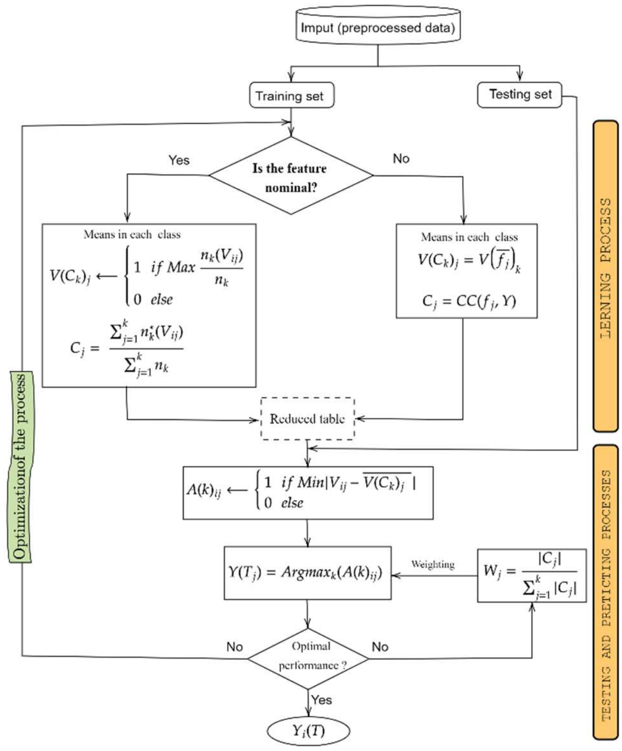

Figure 2. Flowchart of the classification steps of the CMB algorithm.

B.1 TRAINING: REDUCED TABLE CONSTRUCTION

The training step consists of building the reduced table by calculating the centre C_k of each class which contains the average $V\overline{(f_j)}_k$ of previously transformed numeric, scaled and Boolean features.

The particularity of this approach is how it deals with nominal independent features (f_2 in table 1). Each value of the nominal features V_j (for example i) can belong to class C_k if its maximum class frequency is reached for this class in the training set; this frequency is given by the formula:

$$V(C_k)_{ij} \leftarrow \begin{Bmatrix} 1\; if\; Max \frac{\boldsymbol{n_k}(V_{ij})}{\boldsymbol{N_k}} \\ 0\quad else \end{Bmatrix}, \quad (4)$$

where:

-$n_k(V_{ij})$ is the number of V_{ij} variable j in the class C_k

-N_k is the total number of k

The value 1 means that this attribute is more favourable to the class k and will be placed in C_k for this variable in the reduced table, and so on for all attributes of nominal variables. Therefore, nominal attributes take the value 1 in this class k and 0 in other classes.

B.2 BELONGING COEFFICIENT OF NOMINAL FEATURES

As the correlation coefficient (Eq. 2) was used to weight the numerical variables in the reduced table, we introduced membership coefficients for the nominal features that are calculated from the belonging coefficients of each unique attribute to class k. It consists of the ratio between the sum of the maximum numbers of each unique attribute in class k and the number of instances (Eq. 5). For this formulation, let us consider $n_k * (V_{ij})$ to denote all the respective $n_k(V_{ij})$ values corresponding to the maximum; we can deduce the belonging coefficient of the feature f_j to the class as follows:

$$C_j = \frac{\sum_{j=1}^{k} n_k*(V_{ij})}{\sum_{j=1}^{k} N_k}. \quad (5)$$

As an example, let us consider table 2:

-The total number of yes is $N_{yes} = 2$

-The total number of yes is $N_{no} = 3$

-The number of "$married$" in the class "no" is $N_{yes}(married) = 2$

-The number of "$married$" in the class "yes" is $N_{yes}(married) = 1$

Therefore, we can build the following Table 3 which summarizes our results.

Table 3. Summary of the number of nominal attributes per class

Unique attribute for f_2	Y	
	0	**1**
married	2	1
divorced	1	0
single	0	1

From Table 3, we can calculate: $\frac{n_{no}(married)}{N_{no}} = \frac{2}{3} = 0.66$; $\frac{n_{yes}(married)}{N_{yes}} = \frac{1}{2} = 0.5$; 0,66>0,5 so, the belonging to the class "no" of all attributes "$married$" will be replaced by 1 and their belonging to the class "yes" will be 0 during the prediction. They will be placed in C_{no} (centre of the class "no" in the reduced table) of the reduced table for the feature f_2.

We can therefore deduce the belonging coefficient of the feature f_2 as follows: $n_k*(married) = 2$, $n_k*(divorced) = 1$, $n_k*(single) = 1$, $C_2 = \frac{2+1+1}{2+3} = 0{,}8$.

Table 4 below represents the reduced table of our example.

Table 4. Reduced table

Inst.	f_1	f_2	f_3	f_4
C_j	0.98	0.8	0,41	0,33
C_{no}	39	married	0	0,5
C_{yes}	59	Single, divorced	0,33	1

B.3 TESTING: DECISION FUNCTION

The decision phase is done by majority voting in two steps. Given a new instance T_i, we start by calculating the membership $A(k)_{ij}$ of each attribute V_{ij} of T_i to the class k centred in C_k. This eliminates the impact of feature scales. The $A(k)_{ij}$ are calculated by the following function except for nominal variables already treated by the previous equation:

$$A(k)_{ij} \leftarrow \begin{Bmatrix} 1 \text{ } if \text{ } Min\left|V_{ij} - \overline{V(C_k)_j}\right| \\ 0 \quad else \end{Bmatrix}, \quad (6)$$

where, T_i is thus implicitly transformed into k instances $A(k)$ making it possible to predict its class $Y(T_i)$ which will be the one having the maximum of the sum of the overall $A(k)_{ij}$ of its k attributes as shown by the following function:

$$Y(T_i) = Argmax_k(A(k)_i) = Argmax_k \sum_{j=1}^{n} w_j \cdot (A(k)_{ij}), \quad (7)$$

where, W_j is the feature weight, calculated as follows:

$$w_j = \frac{|C_j|}{\sum_{j=1}^{k}|C_j|}. \quad (8)$$

This weighting is the most interesting since it allows us to standardize the coefficients so that the class memberships are bounded between 0 and 1. We can therefore set a decision threshold δ according to the number of classes k and the measure of performance used.

In the case of binary classification k=2, the threshold will be:

$$\delta = \frac{1}{k} +/- \epsilon = 0{,}5 \ +/- \epsilon, \quad (9)$$

where, ϵ is an adjustment factor of the optimal threshold δ. We will have:

$$Y(T_i) \leftarrow \begin{Bmatrix} 1 \text{ } if \text{ } A(yes) > \delta \\ 0 \quad else \end{Bmatrix}. \quad (10)$$

For example, let us consider the example in table 1 as our training set. We want to predict the class $Y(I_1)$ of a new instance I_1: {38; Single; yes; irregular}.

The transformation of I_1 by following the preprocessing steps gives us I'_1:{38; Single; 1; 0}.

Considering the corresponding reduced table (table 3), we can deduce the class membership of each feature as follows:

-For the feature f_1 : $\left|V_{11} - \overline{V(C_{no})_1}\right| = |38 - 59| = 21$ and $\left|V_{11} - \overline{V(C_{yes})_1}\right| = |38 - 41{,}33| = 3{,}33 < 21$ and so $A(no)_{11} = 0$; $A(yes)_{11} = 1$ and its relative weight is $w_1 = \frac{0{,}98}{2{,}52} = 0{,}38$.

-For the feature f_3 which is nominal, "$single$" is more favourable to C_{yes}; so, $A(no)_{12} = 0$ and $A(yes)_{12} = 1$ and its relative weight is $w_2 = \frac{0{,}8}{2{,}52} = 0{,}32$.

-For the feature f_3 : $\left|V_{13} - \overline{V(C_{no})_3}\right| = |1 - 0| = 1$ and $\left|V_{13} - \overline{V(C_{yes})_3}\right| = |1 - 0{,}33|$ $= 0{,}66 < 1$ and so $A(no)_{13} = 0$; $A(yes)_{13} = 1$ and its relative weight is $w_3 = \frac{0{,}41}{2{,}52} = 0{,}16$.

-For the feature f_4 : $\left|V_{14} - \overline{V(C_{no})_4}\right| = |0 - 0{,}5| = 0{,}5$ and $\left|V_{14} - \overline{V(C_{yes})_4}\right| = |0 - 1|$ $= 1$; $0{,}5 < 1$ and so $A(no)_{14} = 1$; $A(yes)_{14} = 0$ and its relative weight is

$w_4 = \frac{0{,}33}{2{,}52} = 0{,}13$.

- $A(yes) = 1 * 0{,}38 + 1 * 0{,}32 + 0*0{,}16 + 1*0{,}13 = 0{,}83$

- $A(no) = 0 * 0{,}38 + 0 * 0{,}32 + 1*0{,}16 + 0*0{,}13 = 0{,}16$

$Y(T_i) = Argmax_k(A(yes); A(no)) = A(yes) = 1$ because 0,16<0,83 and that's how I_1 will be predicted for the class "yes".

Referring to the formula in equation 9, since it is binary classification, we simply deduce the class of I_1 as follows: $A(yes)$=0,83 and $\delta = \frac{1}{2} = 0{,}5$ which is less than $A(yes)$, and therefore I_1 will be predicted in the class "yes".

This makes the approach simpler, faster and easy to optimize.

V. RESULTS ANALYSIS

To evaluate the performance of our proposed approach, we use 4 databases related to CRM:

A. DATASETS

The **Clean Credit Scoring Dataset (CCSD):** This dataset is the first one on which the CMB approach will be tested to classify the potential customers who are likely to default on a financial obligation contracted based on their past financial experience, and to decide on this information to accept or refuse a loan "no" [43]. To do this, it is common to use standard binary classification algorithms that we will challenge with our CMB approach. The data is unbalanced with 1250 individuals ranked "bad" out of a total of 4448 and 3198 rated "good".

The **churn modelling Dataset (CMD):** The predictive modelling of customer churn consists in estimating the probability that a customer will be defected using historical, behavioural and socio-economic information. This prediction is very important because it can boost customer satisfaction and is likely to churn. It is generally approached using classification algorithms to learn the different models of churn and non-churn [43]. Nevertheless, current overview classification algorithms are not well aligned with business objectives [44]. We will test the CMB approach on churn telecom (CTD) and churn banking data (CBD).

The **Direct Marketing Dataset (DMD):** It is a prediction modelling in direct marketing. The goal is to rank the most likely customers to respond favourably to a marketing campaign. We used a direct marketing dataset that contains 45,000 customers from a Portuguese bank who were contacted by telephone between 2008 and 2012 and who received an offer to open a long-term deposit account with rates of attractive interest. Classes are unbalanced with less than 12% ranked "yes" and the rest classified as "no" [29, 45]. The dataset contains characteristics such as age, employment, marital status, educational level, average annual balance, and current loan status, as well as the class label, indicating whether the client accepted or not.

B. PERFORMANCE METRICS

The predictive accuracy rate (Eq. 8) is the most commonly used measure but it is not an effective tool for evaluating models on unbalanced datasets, because it does not indicate how the model correctly classified the minority class instances that are often the targets. Concerning our databases which are unbalanced [46], we will evaluate our approach in terms of other performance measures:

The *Area Under the Curve (AUC)* is a performance metric generated from the Receiver Operating Characteristic (ROC) curve. The ROC curve is created by plotting the True Positive Rate (TPR) on the y-axis against the True Negative Rate (TNR) on the x-axis. It shows the portion of misclassified instances and then is an ideal performance measurement for imbalance class datasets [21].

$$Accuracy = \frac{a+b}{(a+b+c+d)}, \quad (11)$$

$$FM = \frac{2a}{(2a+c+d)}, \quad (12)$$

where, a: refers to the number of true positives, b: is the number of true negatives and c: is the number of false-positive and d: is the number of false negatives.

C. EXPERIMENTAL PARAMETERS

Two parameters are important to achieve the best performance: the decision threshold δ and the proportion of training and test data. We evaluated the overall performances of a different combination of these two parameters. The optimal δ of 0.44 and the data is divided the data into 4 parts including $\frac{3}{4}$ for the training and $\frac{1}{4}$ for the test. Since the databases are unbalanced, Accuracy, AUC ROC and f1-measure are most suitable for better interpretation of results.

D. RESULTS ANALYSIS

The main objective is to show the influence of the CMB approach in optimizing the performance of predictive analytics integrated into CRM, which aims to conquer, acquire and retain target customers to do more profits.

D.1 RESULTS OF BASIC CMB MODEL ANALYSIS

First, we evaluated the basic model CMB approach without weighting or normalization. The results are illustrated in table 5. Without any optimization, the basic results were already quite interesting compared to several classification algorithms. They exceeded 50% on all databases except in terms of the f1-measure for the churn data.

After normalizing, all the made measurements compared to the basic model gave almost the same results with a deviation not exceeding 2%. This stable performance shows the stability of the model that treats each variable independently both in the pre-treatment phase and during the classification. However, normalization is quite interesting as it reduced the processing time.

Table 5. Combining results of CMB on four CRM datasets

	Model	AUC	Accuracy	f1-measure
DMD	basic	78,1%	90,0%	57,9%
	with FN	76,9%	89,9%	56,8%
	with FNW	**95,9%**	**97,3%**	**93,2%**
BCD	basic	64,8%	73,6%	44,0%
	with FN	64,4%	73,6%	43,9%
	with FNW	**72,8%**	**74,4%**	**67,8%**
CCSD	basic	61,3%	51,4%	55,7%
	with FN	61,3%	51,4%	55,7%
	with FNW	**61,1%**	**65,0%**	**59,2%**
CTD	basic	59,3%	76,9%	30,3%
	with FN	59,3%	76,9%	30,3%
	with FNW	70,0%	79,7%	**65,9%**

Finally, we integrated the selection of significant variables and variables weighting into our global model. This later yielded the best performances as shown in Table 4 and Fig. 2. Compared to the basic model, there is a clear performance improvement. On DMD, we have an improvement of more than 17.0%, 12.5% and 35.5% in terms of AUC, Accuracy and f1-measure respectively. On BCD and CTD, this improvement is almost similar reaching more than 8.7%, 5.3% and 12.6% in terms of AUC, Accuracy and f1-measure respectively. And finally, on CCSD, there is no big improvement in terms of AUC (0.2%) but rather 8.9% and 13.56% for the accuracy and f1 measure respectively.

D.2 OVERALL PERFORMANCE RESULTS

The Overall results in Fig. 3 show that for all three measures, the CMB approach has given a better performance. DMD dataset yielded the highest performance compared to the other databases reaching 95.9% AUC, 97.3% for the accuracy and 93.2% for the f-measure. In terms of accuracy, the CTD dataset obtained the highest performance (79.7%) followed by BCD (74.4%) and finally CCSD with 65.0%. The AUC yielded by the three datasets BCD, CTD and CCSD are relatively close to varying between 72.8% and 61.1%.

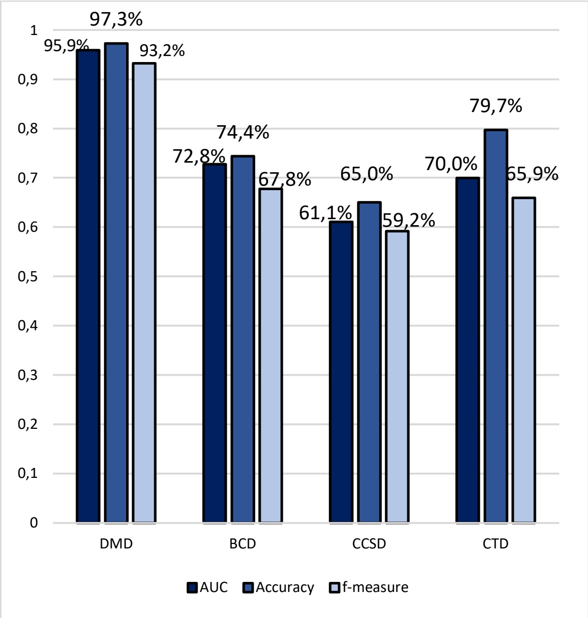

Figure 3. Global CMB approach performance on four CRM datasets.

CTD returned the highest accuracy (79.7%) followed by BCD with 74.4%, and the weakest accuracy 65.0% concerns the CCSD dataset. The weakest f-measure concerns the CCSD dataset with a value not exceeding 59.2%, while the two other datasets are slightly higher with 67.8% for the BCD dataset and 65.9% for CCSD.

E. DISCUSSION OF THE RESULTS

In the previous section, we analyzed our performance to build the optimal CMB model. In the state of the art, we have not only talked about previous work but also three algorithms (DT, SVM and ANN) most used in CRM according to which we will compare the performance of our approach. This comparison will be based on two measures most used in the state of the art namely AUC and f1-measure as illustrated in Table 6. Cross-checking Fig. 2 and Table 5 clearly shows that our approach has improved the performance of previous work of Moro et al. [36], C. Bahnsen et al. [34] and C. S. T. Koumetio et al. [26], N. Ghatasheh et al. [47], C. Yan et al [48], on DMD dataset. Compared to [36], in terms of AUC, the CMB approach outperformed DT, SVM and ANN by 20.2%, 18.8% and 16.1% respectively. There is also a clear improvement (by more than 33.08% in terms of f-measurement) in the performance of [26] which used ANN and more than 40% for [34] which used DT. Only [49] shows performances slightly better than those of CMBs' one for DMB dataset but ANN implement is very low, cost-expensive and there is a doubt on results analysis of this work. For the prediction of own credit, CMB outperformed ANN (RBF kernel) [18] by more than 2% in terms of AUC and CS-DT [34] by more than 7% in terms of f-measure. Similarly, for the customer's churn prediction, our approach improves the performance of the state of the art [34] but remains slightly lower than the performance of [12, 50, 51] for churn prediction because they have upsampled instances' number.

Table 6. Performance Comparison with previous works

	Year Publication	Dataset	Accuracy	AUC	FM	Model
Moro et al. [36]	2014	DMD		0,757	-	DT
				0,767	-	SVM
				0,794	-	NN
C. Bahnsen et al.[43]	2015	DMD		-	0,4973	CS-DT
		Churn		-	0,1652	CS-DT
		CCSD		-	0,5154	CS-DT
C. S. T. Koumetio et al. [33]	2019	DMD		-	0,6012	ANN
T. Verbraken et al. [44]	2013	Churn		0,644	-	ANN
S. Höppner et al.[52]	2018	Churn				
N. Ghatasheh et al. [47]	2020	DMD	0.8418			CostSensitive-MLP
C. Yan et al [48]	2020	DMD	0.80	-	-	S_Kohonen network
S. Mokrane [49]	2020	DMD	0.9893		0.95	ANN
S.K. Trivedi [53]	2020	CCSD	0.762		0.7618	RF
P. Pławiak et al. [54]	2020	CCSD	0.949	-	-	Deep Genetic Hierarchical Network of Learners (DGHNL)
K. Shankar et al. [50]	2021	Churn	0.909	-	0.908	SMOTE-OWELM
T. Xu et al. [51]	2021	Churn	0.9612	-	-	XGBoost
X. Huang [21]	2018	CCSD		0,59	-	ANN (RBF)
T. Vafeiadis et al. [12]	2015	Churn		94,15	77,04	DT-C5.0
				94,37	76,18	SVM-RBF

VI. CONCLUSION AND PERSPECTIVES

In this paper, it was discussed to introduce a Machine Learning approach to optimize the performance of customer relationship management systems by automatically targeting customers and loyalty in real-time. The first part was to study the main algorithms applied to CRM problems. After, we introduced the class membership based classification approach to overcome their limitations. Our approach has been very stable against varying scales and effective as a tool for strengthened customer relationship management systems through the actions of automatic optimization of loyalty, but also to acquire new customers. The tests carried out show that the CRM integrating this algorithm can, for example, predict with more than 93.2% chances the probability of acquiring a new customer, with more than 65.9% the risk of churn and more than 59.2% the reliability scores of a customer for a bank loan. Our current challenge is to optimize our classification approach with techniques such as boosting and gradient descent to improve its performance and adapt it to the problems of big data prediction.

References

[1] E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert Syst. Appl.*, vol. 36, no. 2 PART 2, pp. 2592–2602, 2009. https://doi.org/10.1016/j.eswa.2008.02.021.

[2] B. T. Femina and E. M. Sudheep, "An efficient CRM-data mining framework for the prediction of customer behaviour," *Procedia Comput. Sci.*, vol. 46, no. Icict 2014, pp. 725–731, 2015. https://doi.org/10.1016/j.procs.2015.02.136.

[3] J. Ranjan, V. Bhatnagar, "Critical success factors for implementing CRM using data mining," *J. Knowl. Manag. Pract.*, vol. 3, no. 9, pp. 18–25, 2008.

[4] A. S. Dick, "Customer Loyalty: Toward an Integrated Conceptual Framework," *J. Acad. Mark. Sci.*, vol. 22, no. 2, pp. 99–113, 1978. https://doi.org/10.1177/0092070394222001.

[5] M. S. Garver, "Using data mining for customer satisfaction research," *International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 4, pp. 63–66, 2016. https://doi.org/10.18201/ijisae.266801.

[6] B. Oralhan, K. Uyar, and Z. Oralhan, "Customer satisfaction using data mining approach," *Int. J. Intell. Syst. Appl. Eng.*, vol. 4, no. Special Issue, pp. 63–66, 2016. https://doi.org/10.18201/ijisae.266801.

[7] Z. Zhang, H. Lin, K. Liu, D. Wu, G. Zhang, and J. Lu, "A hybrid fuzzy-based personalized recommender system for telecom products/services," *Inf. Sci. (Ny).*, vol. 235, pp. 117–129, 2013. https://doi.org/10.1016/j.ins.2013.01.025.

[8] J. Díez, D. Martínez-Rego, A. Alonso-Betanzos, O. Luaces, and A. Bahamonde, "Optimizing novelty and diversity in recommendations," *Prog. Artif. Intell.*, vol. 8, no. 1, pp. 101–109, 2019. https://doi.org/10.1007/s13748-018-0158-4.

[9] W.-H. Au, K. C. C. Chan, and X. Yao, "A novel evolutionary data mining algorithm with applications to churn prediction," *IEEE Trans. Evol. Comput.*, vol. 7, no. 6, pp. 532–545, 2003. https://doi.org/10.1109/TEVC.2003.819264.

[10] C.-P. Wei, I.-T. Chiu, "Turning telecommunications call details to churn prediction: a data mining approach," *Expert Syst. Appl.*, vol. 23, no. 2, pp. 103–112, 2002. https://doi.org/10.1016/S0957-4174(02)00030-1.

[11] W. Verbeke, D. Martens, B. Baesens, "Social network analysis for customer churn prediction," *Appl. Soft Comput.*, vol. 14, pp. 431–446, 2014. https://doi.org/10.1016/j.asoc.2013.09.017.

[12] T. Vafeiadis, K. Diamantaras, T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. Ch Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," *Simul. Model. Pract. Theory*, vol. 55, pp. 1–9, 2015. https://doi.org/10.1016/j.simpat.2015.03.003.

[13] M. Cioca, G. Andrada-Iulia, C. Lucian-Ionel, G. Daniela, and L. Blaga, "Machine learning and creative methods used to classify customers in a CRM systems," *Appl. Mech. Mater. Trans Tech Publ.*, vol. 371, pp. 769–773, 2013. https://doi.org/10.4028/www.scientific.net/AMM.371.769.

[14] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decis. Support Syst.*, vol. 50, no. 3, pp. 602–613, 2011. https://doi.org/10.1016/j.dss.2010.08.008.

[15] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4. Springer Verlag, pp. 221–232, 2016. https://doi.org/10.1007/s13748-016-0094-0.

[16] S. Fletcher and Z. Islam, "An anonymization technique using intersected decision trees," *J. King Saud Univ. – Comput. Inf. Sci.*, vol. 27, no. 3, pp. 297–304, 2015. https://doi.org/10.1016/j.jksuci.2014.06.015.

[17] H. Elmandili, H. Toulni, and B. Nsiri, "Optimizing road traffic of emergency vehicles," *Proceedings of the 2013 Int. Conf. Adv. Logist. Transp. ICALT 2013*, 2013, pp. 59–62. https://doi.org/10.1109/ICAdLT.2013.6568435.

[18] M. Boudhane, B. Nsiri, and H. Toulni, "Optical fish classification using statistics of parts," *Int. J. Math. Comput. Simul.*, vol. 10, no. 1, pp. 18–22, 2016.

[19] S.-Y. Hung, D. C. Yen, and H.-Y. Wang, "Applying data mining to telecom churn management," *Expert Syst. Appl.*, vol. 31, pp. 515–524, 2006. https://doi.org/10.1016/j.eswa.2005.09.080.

[20] C.-L. Huang, M.-C. Chen, and C.-J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert Syst. Appl.*, vol. 33, pp. 847–856, 2007. https://doi.org/10.1016/j.eswa.2006.07.007.

[21] X. Huang, X. Liu, and Y. Ren, "Enterprise credit risk evaluation based on neural network algorithm," *Cogn. Syst. Res.*, vol. 52, pp. 317–324, 2018. https://doi.org/10.1016/j.cogsys.2018.07.023.

[22] S. F. Crone, S. Lessmann, and R. Stahlbock, "The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing," *Eur. J. Oper. Res.*, vol. 173, no. 3, pp. 781–800, 2006. https://doi.org/10.1016/j.ejor.2005.07.023.

[23] A. Alphy and S. Prabakaran, "A dynamic recommender system for improved web usage mining and CRM using swarm intelligence," *Sci. World J.*, vol. 2015, no. Article ID 193631, pp. 1–16, 2015. https://doi.org/10.1155/2015/193631.

[24] Y. L. Chen, C. L. Hsu, and S. C. Chou, "Constructing a multi-valued and multi-labeled decision tree," *Expert Syst. Appl.*, vol. 25, no. 2, pp. 199–209, 2003. https://doi.org/10.1016/S0957-4174(03)00047-2.

[25] K. K. Lai, L. Yu, S. Wang, and W. Huang, "An intelligent CRM system for identifying high-risk customers: An ensemble data mining approach," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2007, vol. 4488 LNCS, no. PART 2, pp. 486–489. https://doi.org/10.1007/978-3-540-72586-2_70.

[26] M. A. H. Farquad, V. Ravi, and S. B. Raju, "Analytical CRM in banking and finance using SVM: A modified active learning-based rule extraction approach," *Int. J. Electron. Cust. Relatsh. Manag.*, vol. 6, no. 1, pp. 48–73, 2012. https://doi.org/10.1504/IJECRM.2012.046470.

[27] E. Arbi *et al.*, "Improvement in automated diagnosis of soft tissues tumors using machine learning," *Big Data Min. Anal.*, vol. 4, no. 1, pp. 33–46, 2021. https://doi.org/10.26599/BDMA.2020.9020023.

[28] E. Arbi, A. Alaoui, S. Cedric, and K. Tekouabou, "Intelligent management of bike sharing in smart cities using machine learning and Internet of Things," *Sustain. Cities Soc.*, vol. 67, no. April 2020, p. 102702, 2021. https://doi.org/10.1016/j.scs.2020.102702.

[29] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decis. Support Syst.*, vol. 62, pp. 22–31, 2014. https://doi.org/10.1016/j.dss.2014.03.001.

[30] C. S. T. Koumetio, W. Cherif, and S. Hassan, "Optimizing the prediction of telemarketing target calls by a classification technique," *Proceedings of the 2018 International Conference on Wireless Networks and Mobile Communications, WINCOM 2018*, 2019. https://doi.org/10.1109/WINCOM.2018.8629675.

[31] E. Diaz-Aviles *et al.*, "Towards real-time customer experience prediction for telecommunication operators," *Proceedings of the IEEE International Conference on Big Data*, 2015, pp. 1063–1072. https://doi.org/10.1109/BigData.2015.7363860.

[32] H. Xu, L. Wang, and W. Gan, "Application of improved decision tree method based on rough set in building smart medical analysis CRM system," *Int. J. Smart Home*, vol. 10, no. 1, pp. 251–266, 2016. https://doi.org/10.14257/ijsh.2016.10.1.23.

[33] C. S. T. Koumetio, W. Cherif, and S. Hassan, "Optimizing the prediction of telemarketing target calls by a classification technique," *Proceedings of the 2018 International Conference on Wireless Networks and Mobile Communications, WINCOM 2018*, 2019, pp. 1–6. https://doi.org/10.1109/WINCOM.2018.8629675.

[34] S. C. Koumetio, "A data modeling approach for classification problems : application to bank telemarketing prediction," *Proceedings of the NISS19*, March 27–29, 2019, Rabat, Morocco, pp. 1–7.

[35] B. Martin-Barragan, R. Lillo, and J. Romo, "Interpretable support vector machines for functional data," *Eur. J. Oper. Res.*, vol. 232, pp. 146–155, 2014. https://doi.org/10.1016/j.ejor.2012.08.017.

[36] D. V. den P. Michel Ballings, "Evaluating multiple classifiers for stock price direction prediction," *Eur. J. Oper. Res.*, vol. 244, no. 1, pp. 248–260, 2015.

[37] M. Ballings, D. Van den Poel, "CRM in social media: Predicting increases in Facebook usage frequency," *Eur. J. Oper. Res.*, vol. 244, no. 1, pp. 248–260, 2015. https://doi.org/10.1016/j.ejor.2015.01.001.

[38] A. Ben-Hur and J. Weston, "A user's guide to support vector machines," *Methods Mol. Biol.*, vol. 609, pp. 223–239, 2010. https://doi.org/10.1007/978-1-60327-241-4_13.

[39] J. Yang and S. Olafsson, "Optimization-based feature selection with adaptive instance sampling," *Comput. Oper. Res.*, vol. 33, no. 11, pp. 3088–3106, 2006. https://doi.org/10.1016/j.cor.2005.01.021.

[40] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," *Proceedings of the 20th International*

Conference on Machine Learning (ICML-03), 2003, pp. 856–863.

[41] K. Lakshminarayan, S. A. Harp, and T. Samad, "Imputation of missing data in industrial databases," *Applied Intelligence*, vol. 11, pp. 259–260, 1999. https://doi.org/10.1023/A:1008334909089.

[42] D. Tripathi, D. R. Edla, V. Kuppili, A. Bablani, R. Dharavath, "Credit scoring model based on weighted voting and cluster based feature selection," *Procedia Comput. Sci.*, vol. 132, pp. 22–31, 2018. https://doi.org/10.1016/j.procs.2018.05.055.

[43] A. C. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive decision trees," *Expert Syst. Appl.*, vol. 42, no. 19, pp. 6609–6619, 2015. https://doi.org/10.1016/j.eswa.2015.04.042.

[44] T. Verbraken, W. Verbeke, and B. Baesens, "A novel profit maximizing metric for measuring classification performance of customer churn prediction models," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 961–973, 2013. https://doi.org/10.1109/TKDE.2012.50.

[45] S. Moro, P. Cortez, and P. Rita, "Using customer lifetime value and neural networks to improve the prediction of bank deposit subscription in telemarketing campaigns," *Neural Comput. Appl.*, vol. 26, no. 1, pp. 131–139, 2014. https://doi.org/10.1007/s00521-014-1703-0.

[46] G. Marinakos and S. Daskalaki, "Imbalanced customer classification for bank direct marketing," *J. Mark. Anal.*, vol. 5, no. 1, pp. 14–30, 2017. https://doi.org/10.1057/s41270-017-0013-7.

[47] N. Ghatasheh, H. Faris, I. Altaharwa, Y. Harb, A. Harb, "Business analytics in telemarketing: Cost-sensitive analysis of bank campaigns using artificial neural networks," *Appl. Sci.*, vol. 10, pp. 8–13, 2020. https://doi.org/10.3390/app10072581.

[48] C. Yan, M. Li, and W. Liu, "Prediction of bank telephone marketing results based on improved whale algorithms optimizing S_Kohonen network," *Appl. Soft Comput. J.*, vol. 92, p. 106259, 2020. https://doi.org/10.1016/j.asoc.2020.106259.

[49] M. Selma, "Predicting the success of bank telemarketing using artificial neural network," *Int. J. Econ. Manag. Eng.*, vol. 14, no. 1, pp. 1–4, 2020.

[50] I. V Pustokhina, D. A. Pustokhin, P. Thanh, N. Mohamed, and K. Shankar, "Multi - objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector," *Complex & Intell. Syst.*, 2021. https://doi.org/10.1007/s40747-021-00353-6.

[51] T. Xu, Y. Ma, and K. Kim, "Telecom churn prediction system based on ensemble learning using feature grouping," *Appl. Sci.*, vol. 11, no. May, p. 4742, 2021. ttps://doi.org/10.3390/app11114742.

[52] E. Stripling and B. Baesens, "PT US CR," *Eur. J. Oper. Res.*, vol. 284, no. 3, pp. 920–933, 2018.

[53] S. K. Trivedi, "A study on credit scoring modeling with different feature selection and machine learning approaches," *Technol. Soc.*, vol. 63, no. September, p. 101413, 2020. https://doi.org/10.1016/j.techsoc.2020.101413.

[54] P. Pławiak, M. Abdar, J. Pławiak, and V. Makarenkov, "DGHNL : A new deep genetic hierarchical network of learners for prediction of credit scoring," *Inf. Sci. (Ny).*, vol. 516, pp. 401–418, 2020. https://doi.org/10.1016/j.ins.2019.12.045.

***STEPHANE C. K. TEKOUABOU** is a PostDoc Researcher at The Center of Urban Systems (CUS) of Mohammed VI University of Benguérir, Morocco. He received his PhD degree in 2020 from the Faculty of Sciences, Chouaib Doukkali University, Morocco. Prior, he obtained his State Engineering degree in 2016 from the National School of Applied Sciences El Jadida (ENSAJ) of the same university. He has also participated in the Scientific Committee of many international conferences. His research interests include: computer vision (recognition, detection and classification problems); artificial intelligence; machine learning, deep learning and optimization;*

***WALID CHERIF** has completed his PhD in Data Science from Chouaib-Doukkali University (Morocco). He is an IT engineer from the National Institute of Statistics and Applied Economics (Rabat, Morocco) in which he is now a member of the SI2M laboratory. He is the author of many papers in reputed journals and international conferences, he has been serving as a reviewer at international journals and as a session chair at international conferences on Data Mining. He has also participated in the Scientific Committee of international conferences. His research interests include data science, artificial intelligence, machine learning, deep learning, text and data mining.*

***HAMZA TOULNI** received his BSc in Mathematics and Computer Science in 2010 from the Faculty of Sciences Ain Chock, Hassan II University of Casablanca, Morocco. And in 2012, he got his master's in Computer and Internet Engineering from the same university. While in 2018, he received his PhD degree from the Faculty of Sciences Ain Chock, Hassan II University of Casablanca, Morocco. Currently, he is a Professor in EIGSICA, Casablanca, Morocco, and a member associate in LIMSAD Laboratory at Faculty of Sciences Ain Chock, Hassan II University of Casablanca, Morocco. His research interests include, but are not restricted to, VANET communication and its application, smart cities, IoT, and artificial intelligence.*

***ELARBI A. ABDELLAOUI** received a PhD degree in Computer Science in 2017 from the Faculty of Sciences and Technology – Errachidia, University of Moulay Ismaïl, Meknès, Morocco. Before this, he received the Master's degree in Telecommunication in 2013 from the National School of Applied Sciences, University of Sidi Mohamed Ben Abdallah, Fès, Morocco. He is currently a Research Professor at ENS of Meknes; Mouley Ismail University, Morocco. His research publications include mainly wireless networking, ad hoc networking, DTN networks, game theory, Internet of Things (IoT), smart cities and optimization.*

***HASSAN SILKAN** is a professor in the Department of Computer Science, Faculty of Science, Chouaib Doukkali University. Its research areas are Shape Representation and Description, similarity search, content-based image retrieval, database indexing, 2D/3D Shapes Indexing and Retrieval, Multimedia Databases.*