Optimizing the prediction of telemarketing target calls by a classification technique

Conference Paper · October 2018

DOI: 10.1109/WINCOM.2018.8629675

3 authors:

Cédric Stéphane Koumetio Tekouabou
Mohammed VI Polytechnic University

Walid Cherif
School of Information Sciences

Silkan Hassan

Optimizing the prediction of telemarketing target calls by a classification technique

Cédric Stéphane Tékouabou Koumétio
Laboratory LAROSERI,
Department of Computer Science,
Faculty of Sciences.
B.P. 20, 24000, El Jadida, Morocco
ctekouaboukoumetio@gmail.com

Walid Cherif
Laboratory SI2M, Department of
Computer Science, National Institute of
Statistics and Applied Economics,
B.P. 6217, Rabat, Morocco.
chrf.walid@gmail.com

Silkan Hassan
Laboratory LAROSERI,
Department of Computer Science,
Faculty of Sciences.
B.P. 20, 24000, El Jadida, Morocco
silkan_h@yahoo.fr

***Abstract*—-This paper presents a new classification technique to optimize the prediction of telemarketing target calls for selling bank long-term deposits. A Portuguese retail bank addressed, from 2008 until 2013, data on its clients, products and social-economic attributes including the effects of the financial crisis. An original set of 150 features has been explored and 21 features are retained for the proposed approach. This paper introduces a new technique that implicitly fosters most significant features and predicts the classes of clients according to the types of these features. Moreover, experiments showed that either these features are normalized or not, the proposed technique proved stable and accurate. To evaluate the obtained results, this paper compares them to those of most known machine learning models: Naïve Bayes (NB), Decision Trees (DT), Artificial Neural Network (ANN) and Support Vector Machines (SVM). Thus, the proposed approach yielded the best performance in terms of f-measure, and it allowed reaching 60.12% of the subscribers.**

Keywords—Bank telemarketing; Data Mining; Machine Learning; supervised classificaion; Optimization,similarities computation.

I. INTRODUCTION

Recently, technological progress deeply affects Marketing domain with specific demands for specific targets. Thus, many companies prefer target marketing campaign instead of mass marketing campaign which become very less effective because of the intensive competition when facing challenges from a rapidly changing market situation [1].

Direct marketing methods such as telemarketing is in the main strategy of many banks and insurance companies for interacting with their customers and making new business [2]. Huge information collected on clients became very important and useful for the target marketing strategy such as in telemarketing [2, 3]. Because of its remoteness characteristic, telemarketing is an operationalized direct marketing through a contact center which consists of an interactive technique to solicit prospective customers via phone, mails, social medias… in order to make direct sales of products or services [3, 4, 5].

The success of a telemarketing campaign, is to focus on the quality of the prospect data, attempting to understand costumers behavior and predict the expected customers that could have higher probability to be a client by using machine learning techniques [4, 6]. From one case to another, depending on the features selection attributes, different machine learning techniques can be used.

Machine learning is a subdomain of Data Mining that studies automatic techniques for learning to make accurate predictions based on past observations. Machine Learning uses two types of techniques: supervised learning (classification and regression), which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning (clustering), which finds hidden patterns or intrinsic structures in input data. Classification techniques (Naïve Bayes, k-Nearest Neighbors, Decision Trees, Artificial Neural Network, Support Vector Machines,…) [7, 8, 9] have been used in previous works on Portuguese Bank telemarketing campaign prediction.

Based on the results obtained in those previous works on Portuguese dataset, we noticed that the importance of the features may vary and consequently, taking into account this fact, a preliminary classification of clients should be adopted. In this sense, this paper introduces a new technique that implicitly fosters the most significant features and predicts the class of clients according to their importance. This allowed reaching satisfying results.

The rest of this paper is organized as follows: Section II summarizes main works that applied machine learning techniques in bank telemarketing; section III introduces the proposed technique and details the process to adapt for each type of features. Section IV analyzes the obtained results and compares the proposed approach to other machine learning models. Finally, section V concludes this work.

II. BACKGROUND

In previous researches of the last decade, many authors have focused their works on bank telemarketing success prediction with data mining techniques applied mainly on Portuguese bank database. The original set of 150 features has been explored and 22 most significant features are retained for the proposed approach.

S. Crone & al. [9] have investigated the influence of different preprocessing techniques of attribute scaling, sampling, coding of categorical as well as coding of continuous attributes on the classifier performance of decision trees, neural networks and support vector machines. Supported by a multifactorial analysis of variance, they have provided empirical evidence that DPP has had significant impact on predictive accuracy. But this work published in 2006 was not on Portuguese dataset as well as those that will be presented in the rest of this section

S. Moro & al. [5], have compared in their first paper four Data mining models (Logistic Regression, k-Nearest Neighbors, Decision Trees, Neural Network, Support Vector Machines) by using two metrics: area of the receiver operating characteristic curve (AUC) and area of the LIFT cumulative curve (ALIFT). Those four models were also tested on an evaluation dataset, using the most recent data (after July 2012) and a rolling window scheme. He has obtained the best results with NN (AUC=0.8 and ALIFT= 0.7), allowing to reach 79% of the subscribers by selecting the half better classified clients. In their next papers [10], he has used Customer Lifetime Value (CLV) to improve predictive performance without even having to ask for more information to the companies they serve. He has used CRISP-DM methodology [11] to show how we can increase campaign efficiency by identifying the main characteristics that affect success and this has been also shown in Z. Chu Thesis [12].

H. A.Elsalamony [13] has used Multilayer perceptron neural network (MLPNN) and Ross Quinlan new decision tree model (C5.0) in his paper on the same bank data base to increase the campaign effectiveness by identifying the main characteristics that affect a success. He has estimated the performances by three statistical measures: classification accuracy, sensitivity, and specificity.

However, D. Grzonka & al. [14] have discussed and compared four classification methods (decision trees, bagging, boosting, and random forests). Authors suggested that using decision tree-based methods to support planning and management of bank marketing campaigns offers the highest effectiveness.

R. Farooqi [15] has attempted to analyze the data mining techniques and its useful application in banking industry like marketing and retail management, CRM, risk management and fraud detection.

In another work, T. Palar & al. [16] have used the same dataset considering two feature selection methods namely information gain and Chi-square method to select the most important features. The methods have been compared using Naive Bayes model and their experimental results have shown that the reduced set of features improves the classification performance. B. Famina & E. Sudheep [17] have proposed efficient CRM-data mining framework and have studied two classification models: Naïve Bayes and Neural Networks to show that the accuracy of Neural Network is comparatively better.

Y. Kawasaki & M. Ueki [18] have examined predictive modeling with several sparse regression methods for bank telemarketing success they have concluded that an effective predictive model may help in reducing costs for marketing in companies.

During last years, many authors have been interested by predicting potential future clients in telemarketing campaigns. Many data mining approaches have been studied and applied on the same dataset giving different performances that remain also optimizable. To improve these performances, this paper introduces a new classification technique which will be evaluated on the same Portuguese dataset. The proposed approach will be presented in the following part.

III. The Proposed Approach

In this paper, the direct marketing campaign of “Portuguese bank dataset” is used. In what follows, we will describe it.

A. Dataset

The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution and provided by The UCI Machine Learning Repository (Lichman, 2013). For evaluation purposes, a time ordered split was initially performed, where the records were divided into training (four years) and test data (one year). Note that 2008 economy crisis made this dataset complicated because time has to be considered as a factor in the analysis. The training data is used for feature and model selection and includes all contacts executed up to June 2012, in a total of 41188 examples. The test data (4119 examples) is used for measuring the prediction capabilities of the selected data-driven model, including the most recent 1293 contacts, from July 2012 to June 2013[5]. The dataset is unbalanced, as only 4638 (11.26%) records are related with successes. The total number of training phone contacts database (the sample size) is 41188 and the marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be yes ($y = 1$) or no($y = 0$) for subscribe [3] [18]. The choice of features is based on a semi-automatic feature selection explored in the modeling phase, performed with the data prior to July 2012 and that allowed to select a reduced set of 21 most significant features for this approach. The dataset includes 21 variables (including binary label variable y) where numeric, binomial and categorical variables are mixed as described in Table 1.

TABLE I. DETAILS OF INPUT VARIABLE OF PORTUGUESE BANK DATASET

N°	Variable Name	Description	Type
1	Age	Age of the client	Numerical
2	Job	Type of client’s job	Categorical
3	Marital	Client’s marital status	Categorical
4	Education	Highest education of the client	Categorical
5	Default	Client credit	Categorical
6	Housing	Housing loan	Categorical
7	Loan	Personal loan	Categorical
8	Contact	Type of contact communication with the client	Categorical
9	Month	Last month of the year contracting to the client	Categorical
10	Day of Week	Last day of the week contracting to the client	Categorical
11	Duration	Duration of client contact	Numerical
12	Campaign	Number of contacts performed during this campaign and for this client	Numerical
13	Pdays	Number of days elapsed after the client's last visit	Numerical
14	Previous	Number of contacts performed before this campaign and for this client	Numerical
15	Poutcome	Outcome of the previous marketing campaign	Categorical
16	Emp.var.rate	Employment variation rate	Numeric
17	Cos.price.idx	Consumer price index	Numeric
18	Cons.conf.idx	Consumer confidence index	Numeric
19	Euribor3m	Euribor 3 month rate	Numeric
20	Nr.employed	Number of employees	Numeric
21	Label	Client subscription	Categorical

B. The proposed approach

The first step of the classification technique consists on defining the type of the feature as the technique differs for either type. The following table presents some of the attributes figuring in the dataset:

TABLE II. INITIAL DATA VALUES

Instance	Age	Job	Housing	Education	Y
I_1	59	Admin.	no	Professional	**no**
I_2	39	Housemade	yes	Basic.9y	**yes**
I_3	59	Retired	no	university	**no**
I_4	41	Management	no	Unknown	**no**
I_5	44	Entrepreneur	no	Basic.4y	**yes**

- For numerical features: (Ref. Table 1)

We calculate directly statistic parameters (min, max, mean, variance, standard deviation) of each numerical feature.

- For categorical features:

We distinguish 3 subtypes of features: scaled features, binomial features and nominal features.

For scaled values (Month, days of week):

We substitute items by their ordinal number. After that substitution, we calculate the statistics parameters as for numerical values.

For Boolean features:

We have only two possibilities yes or no (1/0); success or failure (1/0); telephone/cellular (1/0).

For Nominal features:

Each value of the feature V_j for the example i is replaced by his occurrence frequency which is calculated by:

$$V_{ij} \leftarrow f_{ij} = \frac{m}{N} \quad (1)$$

Where m is the frequency of the class, and N is the total number of instances.

Finally, the class having the highest similarity measure (average of all features) is allocated to the example.

C. Missing Attribute Values:

There are several missing values in some categorical attributes, all coded with the "unknown" label. These missing values have been treated by using completion imputation techniques which consist to replace all the missing values of the variable by the average:

$$(y_{ij}) = y_{i*j*} = \overline{Y}_{j*}$$

TABLE III. TRANSFORMED DATA VALUES

Instance	Age	Job	Housing	Education	Y
I_1	59	0.253	0	13	**no**
I_2	39	0.026	1	9	**yes**
I_3	59	0.042	0	16	**no**
I_4	41	0.071	0	11	**no**
I_5	44	0.035	0	4	**yes**

A common step of the classification process is normalization which permits to eliminate the impact of the order of magnitude. Considering the two following examples: C_0 which is the center of the first class no, and C_1 which is the center of the second class yes:

TABLE IV. NON NORMALIZED DISTANCES TO CENTER

	x_1	x_2	x_3	x_4	x_5	y
C_0	39.91	0.163	220.84	984.11	5176.17	no
C_1	40.91	0.152	553.19	792.04	5093.12	yes
E_{21}	27	0.1	698	999	5228.1	

We can easily estimate the similitude of the instance E_{21} to every one of the classes by calculating the Euclidian distances for each center of the two classes which are: $d(E_{21}; C_0) = 480.38$ and $d(E_{21}; C_1) = 286.73$.

E_{21} is then closer to C_1, we can conclude that E_{21} belong to the class yes regardless of the order of magnitude of the features x_j [19], [20] .

In order to overcome such problems, a normalization is adopted for each value v_{ij} of the attribute x_j for the example E_i. It reduces each V_{ij} to the interval [0,1] by the formula:

$$V_{ij} \leftarrow \frac{V_{ij} - min_j(V_{ij})}{max_j(V_{ij}) - min_j(V_{ij})} \quad (2)$$

TABLE V. NORMALIZED DISTANCES TO CENTER

	x_1	x_2	x_3	x_4	x_5	y
C_0	0.28	0.61	0.04	0.99	0.80	no
C_1	0.30	0.56	0.11	0.79	0.49	yes
E_{21}	0.12	0.32	0.14	1	1	

With this normalization $d(E_{21}; C_0) = 0.96$ and $d(E_{21}; C_1) = 1.05$. The example E_{21} belongs, after normalization, to the class no as it is closer to C_0 than C_1. Thus, the normalized data values can be presented as follows:

TABLE VI. NORMALIZED DATA VALUES

Instance	Age	Job	Housing	Education	Y
I_1	0.519	1	0	0.875	**no**
I_2	0.272	0.019	1	0.562	**yes**
I_3	0.519	0.089	0	1	**no**
I_4	0.296	0.215	0	0.688	**no**
I_5	0.333	0.061	0	0.250	**yes**

The performance of classification algorithms depends in particular on the normalization and the similarity which is based on the choice of the distance.
Two most common distances have been compared, namely the Manhattan distance and the Euclidean distance, and this latter has been retained as it offered better performance.

Let's consider the normalized training dataset which contains 41 188 examples.
Each example from the normalized test dataset will be predicted by computing its overall similarities (average of numerical, scaled, binomial and nominal features) as detailed above.

IV. RESULTS AND ANALYSIS

A. Performance Measure

To evaluate the performance of our approach, the first used metric is the accuracy, it expresses the rate of correct predictions; and the second metric is the F1-measure, which is calculated from precision and recall:

$$precision = \frac{a}{(a+b)} \quad (3)$$

$$recall = \frac{a}{(a+c)} \quad (4)$$

$$FM = \frac{(2 \times recall \times precision)}{(recall + precision)} \quad (5)$$

a : refers to the set of clients that are correctly predicted, b: is the number of false positives, and c : is the number of false negatives.

B. Results

The proposed approach is compared to the four algorithms: DT, NB, ANN and SVM in terms of precision, recall, f-measure and accuracy.

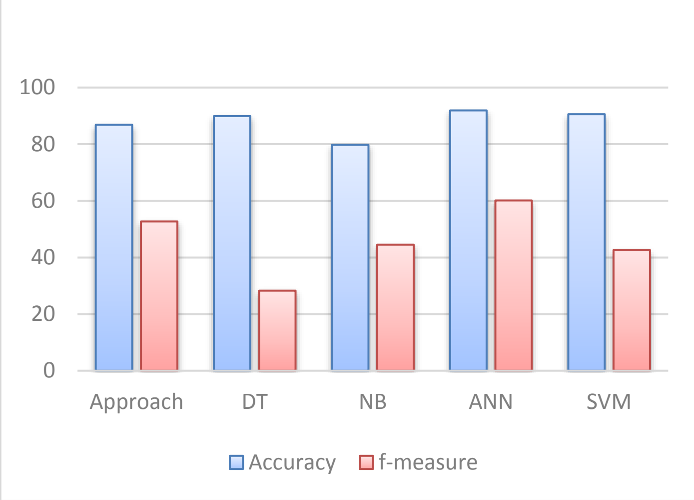

Fig. 1. Result without normalization (accuracy – f-measure)

The proposed algorithm performed the highest f-measure 55.35% followed by Naïve Bayes with 44.52% then SVM with 42.60% and finally Decision Tree (C4.5 algorithm) with the lowest f-measure 28.28%. The only model that yielded better results is ANN with 60.12%, but it consumed the most execution time. Finally, on normalized data, the proposed technique offered satisfying classification performance.

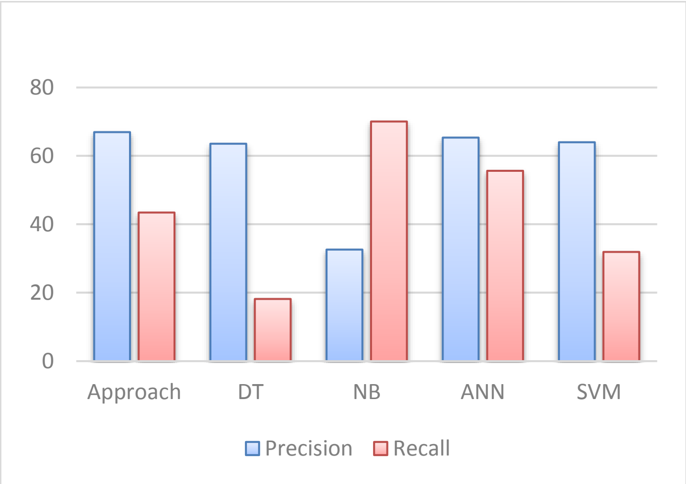

Fig. 2. Result without normalization (recall – precision)

The results show that the proposed algorithm is more efficient than NB, ANN, DT, SVM on points of precision. Its recall is below NB which has a very low precision and ANN which has a very high computational time and relative low precision. Thus, globally the algorithm is more efficient combining precision and recall.

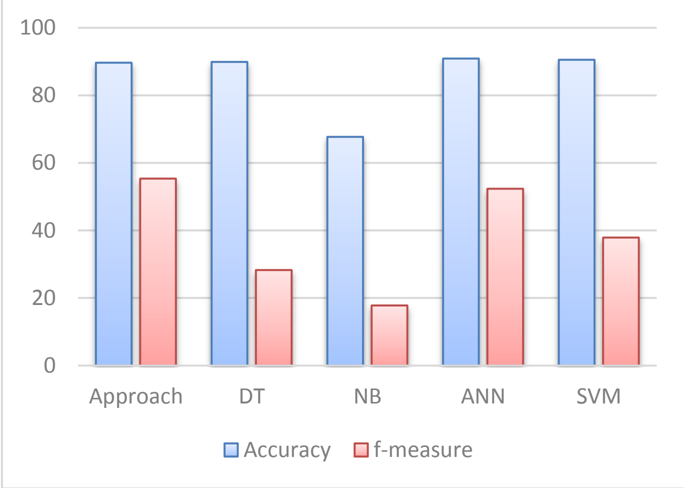

Fig. 3. Results with normalization (accuracy – f-measure)

The normalization improves the performance of our approach in terms of f-measure (from 52.71% to 55.35%) while the performances of other algorithms that are sensible to variable amplitudes decrease. For example, the variable duration, which is not known before the call, contributes a lot to the decision, which is unrealistic. Thus, the advantage of our approach, in addition to its high performance, is its stability relative to the order of magnitude of the variables unlike other algorithms that depend directly on the amplitude of the variables.

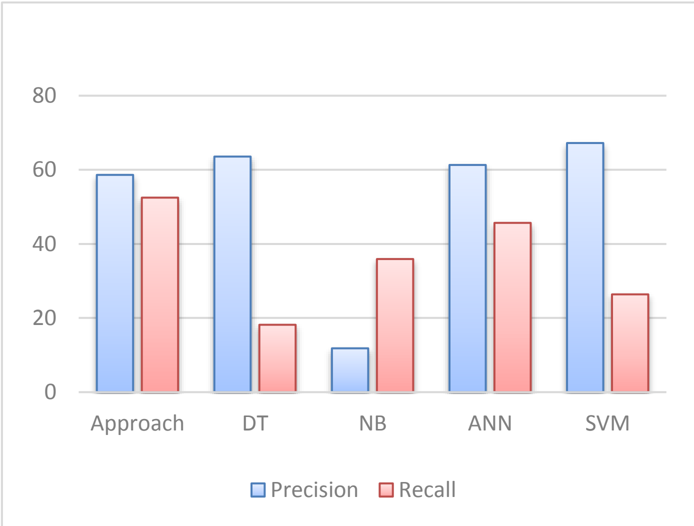

Fig. 4. Results with normalization (recall – precision)

The normalization improves the recall of our approach (from 43.45% to 50.00%) while the recall of all other algorithms decrease hugely. It has slightly increased the precision of SVM (from 64.00% to 67.23%) compared to other approaches. Globally the normalization has improved our approach performance when combining precision and recall due to its stability relative to the order of magnitude of the variables unlike other algorithms.

V. CONCLUSION AND PERSPECTIVES

In this paper we have proposed a new classification technique to improve the prediction of bank direct marketing campaign which is in the main strategy of many banks and insurance companies for interacting with their customers and making new business.

After comparing main works that applied machine learning techniques in bank telemarketing, we have introduced a new technique which computes a specific similarity for each type of features. The proposed technique fostered most significant features and predicts the class of clients more accurately. Moreover, it proved insensitive to the order of magnitude of the features relatively to other algorithms.

This work has been done considering the full set of data where one the most significant attribute was the duration of a call. Unfortunately, in reality this parameter is known only after performing a direct marketing operation. So, these results are just limited for benchmarking but not realistic for exact prediction. Therefore, in contrast to S. Moro et al. research presented in [6], we should omit such parameters in our next study in the construction of very realistic predictive discriminant models with attributes which are known before a direct marketing operation. And the consumption time for each considered algorithm will be compared, as it is a key factor to evaluate models for live telemarketing.

REFERENCES

[1] C. T. Su, Y. H. Chen, and D. Y. Sha, "Linking innovative product development with customer knowledge: a data-mining approach," *Technovation*, vol. 26, no. 7, pp. 784–795, 2006.

[2] H. A. Elsalamony and A. M. Elsayad, "Bank Direct Marketing Based on Neural Network," *Int. J. Eng. Adv. Technol.*, vol. 2, no. 6, pp. 392–400, 2013.

[3] C. Vajiramedhin and A. Suebsing, "Feature selection with data balancing for prediction of bank telemarketing," *Appl. Math. Sci.*, vol. 8, no. 114, pp. 5667–5672, 2014.

[4] R. Vaidehi, "Predictive Modeling to Improve Success Rate of Bank Direct Marketing Campaign," *IJMBS*, vol. 9519, pp. 2230–2232, 2016.

[5] 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.

[6] 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.

[7] B. B. Bezabeh, "The Application of Data Mining Techniques To Support Customer Relationship Management: the Case of Ethiopian Revenue and Customs Authority."

[8] A. K. Muhammed, "Application of Data Mining In Marketing Application of Data Mining In Marketing," *IJCSN Int. J. Comput. Sci. Netw.*, vol. 2, no. 5, pp. 2277–5420, 2013.

[9] 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.

[10] 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.

[11] S. Moro and R. M. S. Laureano, "Using Data Mining for Bank Direct Marketing: An application of the CRISP-DM methodology," *Eur. Simul. Model. Conf.*, no. Figure 1, pp. 117–121, 2011.

[12] Z. Chu, "Bank Marketing with Machine Learning," 2015.

[13] H. A.Elsalamony, "Bank Direct Marketing Analysis of Data Mining Techniques," *Int. J. Comput. Appl.*, vol. 85, no. 7, pp. 12–22, 2014.

[14] B. B. D. GRZONKA, G. SUCHACKA, "Application of selected supervised classification methods to bank marketing campaign d," *Inf. Syst. Manag.*, vol. 5, pp. 36–48, 2016.

[15] R. Farooqi, "Effectiveness of Data mining in Banking Industry: An empirical study," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 5, pp. 827–830, 2017.

[16] T. Parlar and S. K. Acaravci, "Using Data Mining Techniques for Detecting the Important Features of the Bank Direct Marketing Data," *Int. J. Econ. Financ. Issues*, vol. 7, no. 2, pp. 692–696, 2017.

[17] 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.

[18] Y. Kawasaki and M. Ueki, "Sparse Predictive Modeling for Bank Telemarketing Success Using Smooth-Threshold Estimating Equations," *J. Jpn. Soc. Comp. Stat.*, vol. 28, pp. 53–66, 2015.

[19] W. Cherif, "Hybrid Reliability-Similarity-Based Approach for Supervised Machine Learning," vol. 12, no. 3, pp. 170–175, 2018.

[20] W. Cherif, "Optimization of K-NN algorithm by clustering and reliability coefficients: Application to breast-cancer diagnosis," *Procedia Comput. Sci.*, vol. 127, pp. 293–299, 2018.