Data Pre-Processing Algorithm for Neural Network Binary Classification Model in Bank Tele-Marketing

Khairul Nizam Abd Halim, Abdul Syukor Mohamad Jaya, Ahmad Firdaus Ahmad Fadzil

Abstract*: Tele-marketing presents a huge challenge in identifying potential customers with lack of effective marketing strategy may led a company to succumbs to problems such as prolonged marketing campaign. Various attempts to improve the performance of binary classification model for bank tele-marketing data. Previous researches indicate that the neural network is the most common algorithms being employed and able to produce commendable results with higher accuracy percentages compared to other algorithms. Despite several attempts to improve the model through treatment of imbalance dataset and features selection, this research argues that they are incomplete. Therefore, this research proposes a data pre-processing algorithm for bank tele-marketing binary classification neural network. Three datasets have been employed (19, 16, and 20 features) to evaluate the performance of the algorithm towards the classification model. The data pre-processing algorithm is divided into three phases; data cleaning, data imbalance treatment and finally data normalization. In this paper, the result indicated that binary classification model complemented with data cleaning techniques such as Missing common (MC) and Tomek Links (TL) shows a better result compared to Ignore Missing (IM). In terms of data normalization, techniques such as MaxAbsScaler (MAS) and MinMaxScaler (MMS) consistently indicated better performance from other normalization techniques. The classification model employed in this paper utilize data pre-processing algorithm combination of MC-TL-MMS. The algorithm using this approach able to record an area of the receiver operating characteristic curve (AUC) of 0.9129 and 0.9464 by using 16 features and 20 features respectively. This result presents the highest figure in terms of performance accuracy compared to other previous researches.

***Keywords*: Classification, neural network, data pre-processing, tele-marketing.**

I. INTRODUCTION

Tele-marketing is one of the marketing mechanisms that serves as a tool to increase sales of a product [1]. Tele-marketing which also known as target marketing or direct marketing has gained popularity with many companies than the conventional mass marketing due to the latter becomes less effective when there exist many competitions [2]. However, tele-marketing presents a huge challenge in identifying potential customers with lack of effective marketing strategy may led a company to succumbs to problems such as prolonged marketing campaign. Machine learning is one of the tools that can help in terms of reducing both time and cost of a marketing campaign [3]. Research by [4] suggested that machine learning can help processes complicated marketing tasks automatically by analyzing large amount of data and learning patterns of data that can be utilized for predictions and simulations.

Machine learning is a subdivision of artificial intelligence and a subdomain of data mining. The main activity of machine learning is essentially learning or training from past data either by supervised leaning or unsupervised learning. Supervised learning is usually employed for classification while unsupervised learning is catered towards clustering [2]. In order to solve the problem of identifying potential customers in marketing, a machine learning model must be developed. Classification algorithm employed will learn from history of labelled data records or dependent variable (DV) and attribute or independent variable (IV). The development of a classification algorithm involves various learning or training processes and once completed will be able to make prediction based on new data. Various classification algorithm has been employed to yield potential customer prediction which includes k-Nearest Neighbors (KNN), Neural Network (NN), Support Vector Machine (SVM), Logistic regression, Decision Trees (DT), and Naive Bayes (NB) [2], [3], [5].

In this paper, the proposed classification algorithm employs neural network with the emphasis on data pre-processing for data mining. This algorithm is then employed in order to develop the binary classification model for bank tele-marketing. Previous researches employed various machine learning algorithm to accommodate this problem. Results from these researches indicate that neural network and support vector machine are two of the most common algorithms being employed. This is largely due to both algorithms able to produce commendable results with higher accuracy percentage compared to other algorithms [3], [6]–[9]. Subsequently, this research proposes the employment of neural network to develop classification model for bank tele-marketing, which has also been done previously [10].In this paper, this research answers whether the approach of

Revised Manuscript Received on January 05, 2020

Khairul Nizam Abd Halim, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.

Abdul Syukor Mohd Jaya, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.

Ahmad Firdaus Ahmad Fadzil, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka, 77300 Merlimau, Melaka, Malaysia.

emphasizing data pre-processing in neural network would allow increased performance in terms of accuracy in neural network model. This research also highlights which data pre-processing techniques that would allow such performance improvement. Various attempts to improve the performance of the binary classification model for bank tele-marketing data. Examples of such attempts, comparing the performance between machine learning algorithms [3], [7]–[9], [11]–[15]. There are also various attempts to tweak and hybridize different machine learning algorithms to obtain better performance improvement [4], [5], [16], [17]. In terms of attempts to improve the model via data mining pre-processing, some researches went for the treatment of imbalance dataset [18]-[20]. There are also researches that emphasize on features selection [3], [15], [21], [22]. While there exists attempts to improve the model via data pre-processing, this research argues that the attempt is incomplete, especially when considering there are no approaches in terms of data cleaning and normalization.

This research hypothesizes that the emphasis on data pre-processing in neural network algorithm such as data cleaning and normalization would allow performance improvement towards the accuracy of the classification model. This hypothesis is tested with results from experiments using each data pre-processing technique are recorded, measured and compared. Results from the comparison indicate that this approach is important to improve the performance of the binary classification model.

Researchers have attempted to improve the performance of the model from the different aspect as mentioned earlier. This research however attempted to improve the performance by emphasizing on the data pre-processing approach such as data cleaning and normalization. To accomplish this, this research first demonstrates how the data cleaning process is performed by using Missing Common (MC) and Ignore Missing (IM) while solving the data imbalance problem via undersampling, oversampling, and combination of both. Next, the process of data normalization is explicated using scaler, normalization, and standardize. Finally, the performance of the neural network classification model with emphasis on data pre-processing are assessed. The results from this research indicated that binary classification model neural network complemented with data cleaning via MC and sampling Tomek Links (TL) enables better performance compared to IM technique. Meanwhile, MaxAbsScaler (MAS) and MinMaxScaler (MMS) is shown to be the better method to be employed for data normalization compared to other normalization techniques.

The paper is organized as follows. Section 2 includes background information on NN, cleaning data, data imbalance treatment and data normalization techniques. Section 3 describes the material and methods of the experiment. Experimental results are presented in Section 4, and the paper is concluded in Section 5.

II. NEURAL NETWORK, CLEANING DATA, DATA IMBALANCE TREATMENT AND DATA NORMALIZATION TECHNIQUES

A. Neural Network

Neural network or Artificial Neural Networks (ANN) is a computational model inspired by the human brain's neurons which executed on the computer to perform certain specific tasks like classification, regression, clustering, and pattern recognition. The concept of neural network was introduced by McCulloch and Pitts in 1943 [23], nevertheless without the mechanism of learning or training. In 1957, Rosenblatt introduces the Perceptron based on the concept of neural network by McCulloch and Pitts [24]. Perceptron is the earliest neural network that able to identify and classify an object. This is proceeded by the introduction of Multilayer Perceptron (MLP) that able to accommodate more than one Perceptron layers. MLP able to solve non-linear problems while maintaining the original structure of perceptron of feedforward layered [25]. Neural network has been demonstrated to successfully employed in various field of studies such as medical [26], [27], agriculture [28], [29], industrial [30], [31] and finance [32], [33].

B. Data Cleaning Technique and Imbalance Data Treatment

Data cleaning is one of the activities in data pre-processing. The main activity of data cleaning is the remediation of data loss and data noise. Two of the common data cleaning techniques are Ignore Missing (IM) and Missing Common (MC) [34]. IM technique is employed to remove records that contains at least one attribute loss. Meanwhile, MC (i.e. imputation) is contrary with IM, whereas this technique replaces the lost data with the most frequent data in a specific attribute [34]. Imbalance data treatment are then continued after the process of data cleaning. Examples of imbalance data treatment includes Random Over Sampler (ROS) [34], Synthetic Minority Over-sampling Technique (SMOTE) [35], Random Under Sampler (RUS) [34], Tomek links (TL) [36], and Synthetic Minority Over-sampling Technique and Edited Nearest Neighbors (SMOTENN) [37].

C. Data Normalization Technique

Data normalization is an activity in data pre-processing that changes the attribute value according to a common scale or range in order to improve the performance of a machine learning algorithm. There are various examples of normalization technique such as min max, z-score dan decimal scaling [34]. In python environment, there exists machine learning framework such as sklearn [38]. The framework contains numerous useful normalization techniques such as MinMaxScaler (MMS), MaxAbsScaler (MAS), StandardScaler (SS), RobustScaler (RS) and Normalizer (NM).

III. MATERIAL AND METHODS

A. Dataset

Experiments performed in this research utilizes a bank marketing dataset (file 'bank-additional-full.csv') by [3]. This dataset is published by California University under the machine learning repository. This dataset is very commonly utilized among machine learning and data mining researches for performance assessment [3], [7]–[9], [11]–[15] and has also being employed for feature selection [3], [15], [21], [22]. This dataset contains 41,188 tele-marketing record of phone calls made by bank agents to sell long-term deposit products. This dataset came from a record ranging from year 2008 until 2010. Each record has 20 features individually. Table 1 illustrates the list of features which relates to a customer's private information, phone call transactions, call history, and social and economic information [9]. The target or features labels are in the form of binary response of "YES" or "NO". The response frequency differs with "YES" shows a much higher value of 36548 while "NO" response only 4640. "YES" response indicates that the customer accepts the product while "NO" otherwise.

Table I: Dataset features

Features	Data type	Description
1. Age	Numeric	Client age
2. Job	Categorical	Job type
3. Marital	Categorical	Marital of status
4. Education	Categorical	Education level
5. Default	Categorical	Credit record
6. Housing	Categorical	Housing loan record
7. Loan	Categorical	Personal loan record
8. Contact	Categorical	Communication type
9. Month	Categorical	Last contact (month)
10. Day of Week	Categorical	Last contact (day of week)
11. Duration	Numeric	Call duration (in seconds)
12. Campaign	Numeric	Number of contacts (for campaign)
13. pdays	Numeric	Number of days that passed by after the client was last contacted from a previous campaign
14. Previous	Numeric	Number of contacts performed before this campaign and for this client
15. Poutcome	Categorical	Previous marketing campaign outcome
16. Emp_Var_Rate	Numeric	Employment variation rate (quarterly indicator)
17. cons_price_idx	Numeric	Consumer price index (monthly)
18. cons_conf_idx	Numeric	Consumer confidence index (monthly)
19. euribor3m	Numeric	Euribor rate (3 month)
20. nr_employed	Numeric	Employees (quarterly) number
y (desired target)	Binary	Client subscribed (Yes/No)

B. Data Pre-Processing Algorithm

This research proposes data pre-processing algorithm to produce binary classification model of bank tele-marketing. Once the data pre-processing is completed, the dataset can be splitted into training and testing dataset which can be later utilized to evaluate the performance of neural network for bank tele-marketing classification model. The algorithm is as follows (Fig. 1.):

Fig. 1. Data pre-processing algorithm

```
Start
Exist missing data
//Data cleaning
If(IM)
   Execute IM technique
   Split data e.g. 80% training, 20% testing
   //Data imbalance treatment
   If(No imbalance treatment)
      Do nothing for data imbalance treatment
      //Data normalization
      If (MinMaxScaler)
         Execute MinMaxScaler and fit for training data
         Execute MinMaxScaler and transform for testing data
      Else If(MaxAbsScaler)
      .
      .
      .
      Else(RobustScaler)
         Execute RobustScaler and fit for training data
         Execute RobustScaler and transform for testing data
   Else If(ROS)
      //Excute ROS on data training
      //Data normalization
   Else If(SMOTE)
      //Execute SMOTE on data training
      //Data normalization
   .
   .
   .
   Else (SMOTENN)
      //Execute SMOTENN on data training
      //Data normalization

Else (MC)
   Execute MC technique
   Split data e.g. 80% training, 20% testing
   //Data imbalance treatment
      //Data normalization
End
```

The algorithm starts with the selection of data cleaning technique of either IM or MC. Both techniques utilize the python environment whereas IM dropna(inplace=True) while MC SimpleImputer sklearn [38]. Treatment of imbalance data follows the data cleaning process. Examples of data treatment techniques employed includes No imbalance treatment (NIT), Random Over Sampler (ROS), Synthetic Minority Over-sampling Technique (SMOTE), Random Under Sampler (RUS), Tomek links (TL), and Synthetic Minority Over-sampling Technique and Edited Nearest Neighbors (SMOTENN). These techniques utilize the python environment via imbalanced-learn library that is compatible with sklearn [39]. The algorithm is proceeded with normalization techniques such as MinMaxScaler, MaxAbsScaler, Normalizer, StandardScaler, and RobustScaler. These techniques also utilize the python environment and sklearn [38]. Table 2 shows the total of 60 set of pre-processed data that has been produced by executing the algorithm in Fig. 1. This dataset will allow thorough searches towards finding the best results that can be utilized for neural network binary classification model.

Table II: 60 sets of pre-processed data

Set	Data cleaning	Data imbalance treatment	Data normalization
1.	IM	NIT	MinMaxScaler
2.	IM	NIT	MaxAbsScaler
3.	IM	NIT	Normalizer
4.	IM	NIT	StandardScaler
5.	IM	NIT	RobustScaler
6.	IM	ROS	MinMaxScaler
7.	IM	ROS	MaxAbsScaler
8.	IM	ROS	Normalizer
9.	IM	ROS	StandardScaler

10.	IM	ROS	RobustScaler
.	.	.	.
.	.	.	.
.	.	.	.
56.	MC	SMOTENN	MinMaxScaler
57.	MC	SMOTENN	MaxAbsScaler
58.	MC	SMOTENN	Normalizer
59.	MC	SMOTENN	StandardScaler
60.	MC	SMOTENN	RobustScaler

C. Data Training and Testing

In data pre-processing, data splitting will be performed once data cleaning has been completed. In this research, the 41,188 records are divided into two parts. A ratio of 80-20 (80% training and 20% testing) is levied hence dividing the number of datasets into 32,950 and 8,238 of training and testing data respectively.

D. Modeling

Cross Validation

This research utilizes Kfold (via StratifiedKFold) technique for cross validation (CV) purpose. Kfold is divided into 5 different parts for data training. Each Kflod is developed with bank tele-marketing binary classification model and evaluated via receiver operating characteristics – area under the curve (ROC-AUC). The best hyperparameter neural network (e.g. number of hidden layers, solver, alpha, activation etc.) is obtained from the results of the highest ROC-AUC. The hyperparameter will be recorded and employed to develop bank tele-marketing binary classification model using the original training data of 32,950 records.

Multilayer Perceptron Algorithm

Multi-Layer Perceptron (MLP) employed in this research has 3 layers structure (input layer, hidden layer, and output layer). Input layer has several neurons that multiply the data input with specific weightage which results will be carried over into the hidden layer. Hidden layer will also multiply the previous results analogous to the input layer. A configuration of -1 until 1 using activation function hyperbolic tangent will then follows the previous results. This research employs 5-12 number of hidden layers. The output layer is configured with only one neuron where the neuron also performs results calculation like other layers. In this research, the output layer utilizes logistic function with range from 0 to 1. Values larger or equal to 0.5 are rounded to 1, otherwise to 0. In this research, MLP is trained by using Backpropagation with Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS or LM-BFGS) as an optimizer which it minimizes the Cross-Entropy loss function. For the hyperparameter, all parameters had been set as default under python environment and sklearn except alpha (i.e. 0.1, 0.01, 0.001, 0.0001). Sklearn is chosen due to its superior performance compare to other machine learning frameworks such as Weka dan Apache Spark [40].

E. Evaluation

Several researches have employed ROC-AUC instrument to measure the performance of bank tele-marketing classification model. ROC-AUC is a popular instrument [3] with probability curve represents degree or measure of separability. ROC curve uses the true positive rate (sensitivity) as the vertical axis, the false positive rate (1-specificity) as the abscissa of the curve. From the ROC curve, AUC can be calculated, the size of AUC is between 1.0 and 0.5. In the case of AUC>0.5, the closer to 1 the AUC, the better the diagnosis [5].

IV. RESULTS AND DISCUSSION

This section verifies the hypothesis of this research; neural network with emphasis on data pre-processing using data cleaning and normalization can improve the accuracy percentage of binary classification model of bank tele-marketing. This section also deliberates which data pre-processing technique able to help improve accuracy of neural network model.

The experiment in this research is divided into three different setups. The experimental setup of the research is explicated in Table 3 with each setup is crucial to enable comparisons between results from previous researches.

Table III: List of experiments setup

Set	Dataset sources	Number of attributes (IV)
1.	bank-additional-full.csv [3](drop duration attribute)	19
2.	bank-full.csv [3]	16
3.	bank-additional-full.csv [3]	20

The first experiment utilizes dataset containing 20 IV as depicted in Table 1. The data contributor suggested that the attribute “duration” should be negated in order to produce a practical predictive model. This is because the attribute can be classified as DV which can affect the model and is impractical for real-world applications. This suggestion is taken into consideration hence the number of attributes of 19. The results from this experiment indicates that binary classification model via emphasizing data pre-processing shows better performance when employed with MC and TL compared to IM. Meanwhile, in terms of data normalization technique MAS and MMS shows better performance compared to other normalization techniques. Table 4 illustrate the performance evaluation via ROC-AUC with bolded texts highlight the technique with the best performance using the pre-processed dataset in Table 2. This is followed by figure 2 which depicts the ROC-AUC results for the algorithm using MAS and MC for each data imbalance treatment techniques.

Table IV: Performance of data pre-processing algorithm

ALGORITHM	MAS	MMS	NM	SS	RS
NIT - IM	**0.7933**	0.7912	0.7623	0.7921	0.7927
ROS - IM	**0.7862**	0.7789	0.7380	0.7727	0.7682
SMOTE - IM	**0.7815**	0.7770	0.7328	0.7660	0.7681
RUS - IM	0.7811	**0.7847**	0.7384	0.7782	0.7767
TL - IM	**0.7944**	0.7890	0.7636	0.7876	0.7819
SMOTEENN - IM	0.7759	**0.7780**	0.7545	0.7729	0.7730
NIT - MC	**0.7946**	0.7937	0.7636	0.7851	0.7908
ROS - MC	**0.7817**	0.7653	0.7473	0.7765	0.7802
SMOTE - MC	0.7676	0.7693	0.7305	0.7753	**0.7755**
RUS - MC	0.7711	0.7781	0.7582	**0.7831**	0.7702

TL - MC	0.7946	**0.7962**	0.7650	0.7921	0.7939
SMOTEENN - MC	0.7814	0.7835	0.7486	**0.7888**	0.7788
Winner	**6**	**3**	0	2	1
Highest	0.7946	**0.7962**	0.7650	0.7921	0.7939
Average	**0.7836**	0.7821	0.7502	0.7809	0.7792

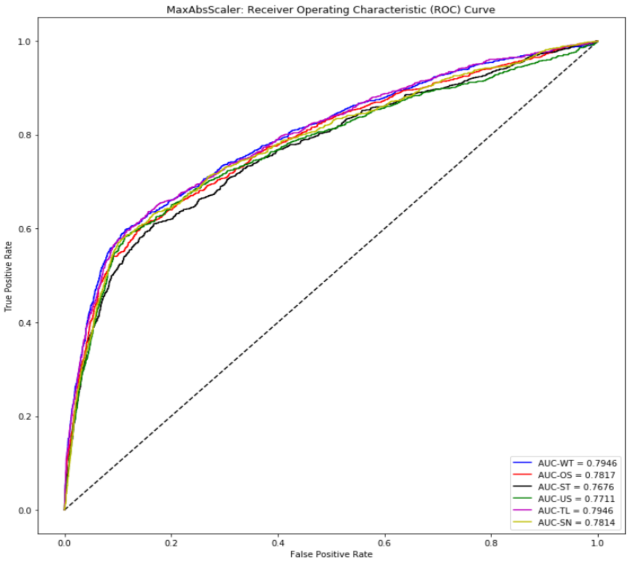

Fig. 2. ROC-AUC results for the algorithm using MAS and MC for each data imbalance treatment techniques

The second experimental setup utilizes 16 IV as shown in Table 3. This dataset contains older dataset compared to the later dataset of 20 IV. The results from previous experiment is also carried over this experiment where superior data pre-processing technique such as MAS, MMS, TL, and MC are employed for the experiment.

Table 5 illustrates the result from the second experiment. The hypothesis of this research is verified with ROC-AUC results of 0.9158 via neural network using MC-TL-MAS which is higher compared to previous research without emphasis on data pre-processing. However, Random Forest (RF) algorithm performs better with ROC-AUC results of 0.9270. The difference however, are insignificant and this research nevertheless confirms that employing data pre-processing for neural network algorithm enable a significant performance increases towards the conventional algorithm making it comparable with some of the best neural network implementations.

Table V: Result for second experiment setup

Num.	ML algorithm	ROC-AUC	Number of IV
1.	NN MC-TL-MAS (0.9158), NN MC-TL-MMS (0.9129)	0.9158	16
2.	R-GRNN O (0.8929), NN (0.8704), 2019 [4]	0.8929	16
3.	RF (0.9270), NN (0.9000), 2017 [13]	0.9270	16
4.	KNN (0.9000), NN (0.8200+), 2017 [18]	0.9000	16
5.	FMLP-SVM, 2017 [5]	0.9048	16
6.	NB (0.858), NN (0.847), 2015 [14]	0.8580	16
7.	SVM Ensemble (0.8), NN (0.71), 2015 [8]	0.8000	16

For the third experiment, dataset containing 20 IV is employed as depicted in Table 1. Like the second experiment, data pre-processing techniques such as MAS, MMS, TL and MC are employed.

Table VI: Result for third experiment setup

Num.	ML algorithm	ROC-AUC	Number of IV
1.	NN MC-TL-MAS (0.9437), NN MC-TL-MMS (0.9464)	0.9464	20
2.	NN, 2019 [9]	0.8910	20
3.	NAC DE, 2018 [11]	0.7625	20
4.	NB, 2017 [12]	0.7700	20
5.	NN-PSO, 2017 [16]	0.6000	20
6.	AWS- Logistic regression, 2017 [41]	0.9360	20
7.	C4.5 algorithm, 2014 [21]	0.8840	20

Table 6 illustrates the results of the third experiment. The result shows that employing data pre-processing for neural network allows significantly higher accuracy towards the binary classification model with 0.9464. This result is significantly higher than simply employing neural network at 0.8910 subsequently confirming that data pre-processing can help in terms of improving the accuracy of neural network classification model.

V. CONCLUSION

This research has proven that employing data pre-processing approach improves the performance of bank tele-marketing binary classification model. An algorithm for data pre-processing has been proposed using techniques such as MC, TL, MAS, and MMS in order to improve the ROC-AUC of a model. The algorithm has been demonstrated to able to significantly improve the ROC-AUC performance of a neural network classification model for all three experiments in this research. This research therefore concludes that emphasizing a complete data mining pre-processes algorithm; data cleaning, imbalance data treatment, and data normalization would allow significant performance improvement for neural network binary classification model. In the future, the research can be continued by testing the approach in other area of studies such as finance, industrial, agriculture, medical and others that currently facing the same situation with the problem stated in this research.

ACKNOWLEDGMENT

This research is supported financially by Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia.

REFERENCES

1. S. M. F. D. S. Mustapha and A. Alsufyani, “Application of Artificial Neural Network and information gain in building case-based reasoning for telemarketing prediction,” Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 3, pp. 300–306, 2019.
2. C. S. T. Koumetio, W. Cherif, and S. Hassan, “Optimizing the prediction of telemarketing target calls by a classification technique,” Proc. - 2018 Int. Conf. Wirel. Networks Mob. Commun. WINCOM 2018, 2019.
3. 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.
4. D. Bani-Hani and M. Khasawneh, "A Recursive General Regression Neural Network (R-GRNN) Oracle for classification problems," Expert Syst. Appl., vol. 135, pp. 273–286, 2019.
5. X. Zhang and F. Zhang, "A fuzzy support vector machine algorithm and its application in telemarketing," in Advances in Intelligent Systems and Computing, 2017, vol. 510, pp. 671–679.
6. S. Moro, R. M. S. Laureano, and P. Cortez, "Using Data Mining for Bank Direct Marketing: An application of the CRISP-DM methodology," 25th Eur. Simul. Model. Conf. ESM'2011, no. Figure 1, pp. 117–121, 2011.
7. A. Zakaryazad and E. Duman, "A profit-driven Artificial Neural Network (ANN) with applications to fraud detection and direct marketing," Neurocomputing, vol. 175, no. PartA, pp. 121–131, 2014.
8. M. Amini, J. Rezaeenour, and E. Hadavandi, "A Cluster-Based Data Balancing Ensemble Classifier for Response Modeling in Bank Direct Marketing," Int. J. Comput. Intell. Appl., vol. 14, no. 04, p. 1550022, 2015.
9. M. Rashid Farooqi and N. Iqbal, "Performance evaluation for competency of bank telemarketing prediction using data mining techniques," Int. J. Recent Technol. Eng., vol. 8, no. 2, pp. 5666–5674, 2019.
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. Y. O. Serrano-Silva, Y. Villuendas-Rey, and C. Yáñez-Márquez, "Automatic feature weighting for improving financial Decision Support Systems," Decis. Support Syst., vol. 107, pp. 78–87, 2018.
12. Y. Villuendas-Rey, C. F. Rey-Benguría, Á. Ferreira-Santiago, O. Camacho-Nieto, and C. Yáñez-Márquez, "The Naïve Associative Classifier (NAC): A novel, simple, transparent, and accurate classification model evaluated on financial data," Neurocomputing, vol. 265, pp. 105–115, 2017.
13. J. Asare-Frempong and M. Jayabalan, "Predicting customer response to bank direct telemarketing campaign," 2017 Int. Conf. Eng. Technol. Technopreneurship, ICE2T 2017, vol. 2017-Janua, pp. 1–4, 2017.
14. 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.
15. S. Moro, R. M. S. Laureano, and P. Cortez, "Using data mining for bank direct marketing: An application of the CRISP-DM methodology," ESM 2011 - 2011 Eur. Simul. Model. Conf. Model. Simul. 2011, no. October, pp. 117–121, 2011.
16. S. Lahmiri, "A two-step system for direct bank telemarketing outcome classification," Intell. Syst. Accounting, Financ. Manag., vol. 24, no. 1, pp. 49–55, 2017.
17. A. Lawi, A. A. Velayaty, and Z. Zainuddin, "On identifying potential direct marketing consumers using adaptive boosted support vector machine," Proc. 2017 4th Int. Conf. Comput. Appl. Inf. Process. Technol. CAIPT 2017, vol. 2018-Janua, pp. 1–4, 2018.
18. G. Marinakos and S. Daskalaki, "Imbalanced customer classification for bank direct marketing," J. Mark. Anal., vol. 5, no. 1, pp. 14–30, 2017.
19. M. M. Al-Rifaie and H. A. Alhakbani, "Handling class imbalance in direct marketing dataset using a hybrid data and algorithmic level solutions," Proc. 2016 SAI Comput. Conf. SAI 2016, pp. 446–451, 2016.
20. K. Kok-Chin and N. Keng-Hoong, "Evaluation of cost sensitive learning for imbalanced bank direct marketing data," Indian J. Sci. Technol., vol. 9, no. 42, pp. 1–8, 2016.
21. 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.
22. N. Barraza, S. Moro, M. Ferreyra, and A. de la Peña, "Mutual information and sensitivity analysis for feature selection in customer targeting: A comparative study," J. Inf. Sci., vol. 45, no. 1, pp. 53–67, 2019.
23. W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bull. Math. Biophys., vol. 5, no. 4, pp. 115–133, 1943.
24. F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," Psychol. Rev., vol. 65, no. 6, pp. 386–408, 1958.
25. H. Bourlard and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition," Biol. Cybern., vol. 59, no. 4–5, pp. 291–294, 1988.
26. R. M. Sadek et al., "Parkinson's Disease Prediction Using Artificial Neural Network," vol. 3, no. 1, pp. 1–8, 2019.
27. I. M. Nasser and S. S. Abu-Naser, "Predicting Tumor Category Using Artificial Neural Networks," Int. J. Acad. Heal. Med. Res., vol. 3, no. 2, pp. 1–7, 2019.
28. M. Fedrizzi, F. Antonucci, G. Sperandio, S. Figorilli, F. Pallottino, and C. Costa, "An artificial neural network model to predict the effective work time of different agricultural field shapes," Spanish J. Agric. Res., vol. 17, no. 1, pp. 1–9, 2019.
29. E. Ropelewska, "Classification of wheat kernels infected with fungi of the genus Fusarium using discriminative classifiers and neural networks," Chil. J. Agric. Res., vol. 79, no. 1, pp. 48–55, 2019.
30. E. C. Ang and S. A. Suandi, "Smart Manufacturing with An Artificial Neural Network to Predict Manufacturing Healthiness," Proc. - 2019 IEEE 15th Int. Colloq. Signal Process. its Appl. CSPA 2019, no. March, pp. 120–123, 2019.
31. S. Kursun Bahadir, U. K. Sahin, and A. Kiraz, "Modeling of surface temperature distributions on powered e-textile structures using an artificial neural network," Text. Res. J., vol. 89, no. 3, pp. 311–321, 2019.
32. S. Liu, C. W. Oosterlee, and S. M. Bohte, "Pricing options and computing implied volatilities using neural networks," Risks, vol. 7, no. 1, pp. 1–22, 2019.
33. D. Doyle and C. Groendyke, "Using neural networks to price and hedge variable annuity guarantees," Risks, vol. 7, no. 1, 2019.
34. S. García, J. Luengo, and F. Herrera, "Data Preprocessing in Data Mining," Intell. Syst. Ref. Libr., p. 72, 2015.
35. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., 2002.
36. I. Tomek, "TWO MODIFICATIONS OF CNN.," IEEE Trans. Syst. Man Cybern., 1976.
37. G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," ACM SIGKDD Explor. Newsl., 2004.
38. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., 2011.
39. G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," J. Mach. Learn. Res., 2017.
40. B. M. Shashidhara, S. Jain, V. D. Rao, N. Patil, and G. S. Raghavendra, "Evaluation of Machine Learning Frameworks on Bank Marketing and Higgs Datasets," Proc. - 2015 2nd IEEE Int. Conf. Adv. Comput. Commun. Eng. ICACCE 2015, pp. 551–555, 2015.
41. R. Ramesh, "Predictive analytics for banking user data using AWS Machine Learning cloud service," Proc. 2017 2nd Int. Conf. Comput. Commun. Technol. ICCCT 2017, no. November 2010, pp. 210–215, 2017.

AUTHORS PROFILE

Khairul Nizam Abd Halim is a postgraduate (Ph.D.) student at Faculty of Information and Communication at Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. He obtained his M.Sc. in Information Technology from Universiti Putra Malaysia in 2001. His current project focuses on Artificial Intelligent, Machine Learning and Data Mining.

Abdul Syukor Mohd Jaya is a senior lecturer in Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. He got his Bachelor of IT (Industrial Computing) from Universiti Kebangsaan Malaysia and Master of Mechanical Engineering (Advanced Manufacturing Technology) from Universiti Teknologi Malaysia. He received his Ph.D. in Computer Science also from Universiti Teknologi Malaysia in 2014. His research interest focuses on Intelligent Manufacturing and Data Analytics.

Ahmad Firdaus Ahmad Fadzil is a lecturer at Universiti Teknologi MARA, Melaka. He obtained his M.Sc. in Computer Science (Research) from Universiti Teknologi MARA in 2014. Current research interest includes parallel processing, image processing, Evolutionary Computing, and Machine Learning.