International Journal of Engineering and Applied Sciences (IJEAS)
ISSN: 2394-3661, Volume-5, Issue-2, February 2018

Prediction of Bank Investors using Neural Network in Direct Marketing

Rutu S Patel, Himanshu S Mazumdar

Abstract— **Direct marketing in banking is one of the most effective methods of predicting potential investors. Effectiveness of direct marketing is being analyzed using different methods like feature correlation, dataset balancing, neural network (NN) etc. Usually sixteen to twenty parameters are collected for training database to evaluate the potential client. A fully connected multilayer NN is developed that gradually optimizes the connection based on training dataset. This NN is used to predict the customer willingness for long term deposit with accuracy hire then 95% which corresponds to Accuracy, Sensitivity and Specificity of 95.19%, 92.32% and 95.42% respectively. One of the important parameter is false negative prediction which is 0.63% for above accuracy. Result of false negative indicates incorrectly predicting unwilling clients. With our algorithm, analyzing UCI test benchmark dataset gives 276 true prediction out of 451 records of customers who buy the bank product and only 23 false prediction out of 3668 records of customers who did not buy the bank product. This may be noted that false negative to true negative ratio increases rapidly with small decrease of accuracy. 2% decrease from 95% increases the false negative value from 23 to 379. Such increase leads to several fold non productive persuasion effort. On the other hand decrease in true positive reduces the true buyer but do not reduce the productivity due to false prediction. However it is seen that increase of network size do not increase the accuracy even after several hours of training. Hence an optimum size of the network needs to be achieved with automatic iterative pruning.**

Index Terms— **Banking, Data Mining, Neural Network, Prediction.**

I. INTRODUCTION

Since 1967, Lester Wunderman is considered the father of "Direct Marketing" and was responsible for important innovation in direct marketing projects associated with the financial services like banking [9]. Researchers in this area are addressing the issue of optimization of direct marketing input parameters to pin point the potential investors.

Most of the banks are now targeting their customers by direct marketing for their specific product and service because of the ineffectiveness of mass marketing [2, 8]. Various media like, television, radio or advertising are used in mass marketing to communicate with all customers [10]. For promotion of any new-product, the direct relationship with prospective customers is not set up in this approach. Many of the customers are found reluctant and do not respond to the sales campaign. Educating customers through expensive ad campaign about any specific bank product without one is to one contact may help competitors sale resulting in negative marketing.

Ms. Rutu Patel, Teaching Associate, Ahmedabad University, Ahmedabad, India

Dr. Himanshu S Mazumdar, Head R&D center and Professor EC department, Dharmsinh Desai University, Nadiad, India

Banks are now adopting the various intelligent data mining techniques of direct marketing to predict the customer behavior. Various predictive models are proposed by different researchers. Anirut Suebsing [2] presents a novel method by applied Cosine similarity to feature selection method. His algorithm selects a robust feature subset using the C5.0, CART and Neural Networks classifiers. He has shown that his method improves the performance of client detection error rate compared to other methods. George H John [3] describes a method for feature subset selection using cross-validation that is applicable to any induction algorithm, and discusses experiments conducted with ID3 and C4.5 on artificial and real datasets. Mark A. Hall [4] describes a new filter approach to feature selection that uses a correlation based heuristic to evaluate the worth of feature subsets. Ravinder Singh's [5] work reveals that support vector machine and genetic programming are superior tools for the purpose of classifying the loan applicant as their misclassification rates were least as compared to others. Xinguo Lu1 [6] in his paper, presents a novel feature selection method based on correlation-based feature selection (CFS). These models emphasize on the quality of acquired data and predict the expected customers that have a higher probability to use the offered service. Direct marketing of these services are becoming more and more important in intelligent data mining techniques [11].

One of the most effective techniques suggested by earlier researchers is prediction of prospective customer using NN. In this method, customer's profile is collected through calls while promoting a bank product like fixed deposit offers. Collected information along with customer's final response is preprocessed as normalized training dataset to an appropriately configured NN. Neural Network is used in supervised learning mode to correlate customer's response to his collected profile and predict important customers. Standard datasets are used as benchmarks for training and validating prediction algorithms. The dataset used in this paper is known as bank marketing from the University of California at Irvine (UCI) [7]. Design of any NN configuration depends on the input-output dataset. It is often tedious and time consuming to optimize NN configuration which is important for final accuracy and training period. An overdesigned NN often memorizes the training dataset during training by indicating low error but performs poorly in real application or with test dataset.

In this paper, we present a NN which automatically configures to the training dataset for optimal performance using pruning method developed for large multi-layered NNs. This also indicates the consequence of dropping weak parameters for prediction model. A large set of input parameter collection often annoy customers hence reduced dataset increases the efficiency of direct marketing.

The rest of the paper is organized as follows: Section I describes the different approaches of data mining and prediction. Section II explains dataset and data preprocessing. Section III describes the MLP Neural Network architecture and back propagation learning algorithm. Section IV explains proposed method. Section V compares the results of different NN configurations with proposed technique. Finally, the conclusion is presented in Section VI.

II. DATASET AND DATA PREPROCESSING

In Direct Marketing, usually sixteen to twenty parameters of client profile record are used for training database. In this work we have used one such benchmark training database from the University of California at Irvine (UCI) Machine Learning Repository [1].

The dataset consists of 41,188 training records and 4,119 independent records for testing the learning algorithm. The training set is usually 10 times larger than test set. Test set is exclusive and used to check the NN for its real time performance. Table 1 lists all the fields of record. Each record consists of the desired output of training as last binary parameter 'Outcome' that is used as desired output while training.

Table 1. UCI benchmark training and testing Dataset of Bank Telemarketing

No.	Attribute Name	Description	Type
1	Age	Age ranges from 17 to 98	Numeric
2	Job	Type of job (admin., blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, unknown)	Categorical
3	Marital	Marital status (divorced, married, single, unknown; note: divorced means divorced or widowed)	Categorical
4	Education	(basic.4y, basic.6y, basic.9y, high school, illiterate, professional course, university degree, unknown)	Categorical
5	Default	Has credit in default? (no, yes, unknown)	Categorical
6	Housing	Has housing loan? (no, yes, unknown)	Categorical
7	Loan	Has personal loan? (no, yes, unknown)	Categorical
8	Contact	Communication type (cellular, telephone)	Categorical
9	Month	Last contact month of year(mar, ..., nov, dec)	Categorical
10	Day of Week	Last contact day of the week (mon, tue, wed, thu, fri)	Categorical
11	Duration	Last contact duration, in seconds (0 – 4918 seconds)	Numeric
12	Campaign	Number of contacts performed during this campaign and for this client (includes last contact) (1 - 56)	Numeric
13	Pdays	Number of days that passed by after the client was last contacted from a previous campaign (999 means client was not previously contacted)	Numeric
14	Previous	Number of contacts performed before this campaign and for this client (0- 7)	Numeric
15	Poutcome	Outcome of the previous marketing campaign (failure, nonexistent, success)	Categorical
16	Emp.var.rate	Employment variation rate - quarterly indicator (932 - 94767)	Numeric
17	Cons.price.idx	Consumer price index - monthly indicator (33 - 508)	Numeric
18	Cons.conf.idx	Consumer confidence index - monthly indicator (1 - 5045)	Numeric
19	Euribor3m	Euribor 3 month rate - daily indicator (5191 - 52281)	Numeric
20	Nr.employed	Number of employees - quarterly indicator	Numeric
21	Outcome	Has the client subscribed a term deposit? (yes, no)	Categorical

There are basically two kinds of parameters- Numeric and Categorical. Numeric parameters are like, Age, Duration, campaign, Pdays, Previous, Cons.conf.idx, Cons.conf.idx, Euribor3m and Nr.employed. There are basically two kinds of parameters- Numeric and Categorical. Numeric parameters are like, Age, Duration, campaign, Pdays, and Previous etc. Neural network needs normalized input. Each of these parameters has range which needs to be normalized. It is important to note that training and test data has different ranges. Both the dataset is normalized with common maximum range values. Example of a normalized record is shown in Table 2. Categorical parameters are like, Job, Marital, Education, Default, Housing, Loan, Contact, Month, Day of Week and Poutcome. The ranges of categorical parameters are same in both the training and testing dataset. These are normalized with corresponding range value. All field of the dataset is converted to numeric value and normalized in the range of 0 to 1. The output parameter in this study has only two categories: non-client and potential client that are normalized to 0 and 1. This output parameter is used as desired output in supervised back propagation algorithm. Experiments are carried out to validate the importance of each of the above parameters. Experimental Network-2 and 3 shows the elimination of different inputs through pruning and corresponding impact on result.

Table 2. Original with normalized record of each attribute

S.No.	Attributes	Original Record	Normalized Record
1	Age	56	0.506
2	Job	housemaid	0.272
3	Marital	married	0.333
4	Education	basic.4y	0
5	Default	no	0
6	Housing	no	0
7	Loan	no	0
8	Contact	telephone	1
9	Month	may	0.6666667
10	Day of Week	mon	0.25
11	Duration	261	0.4581983
12	Campaign	1	0.0
13	Pdays	999	1.0
14	Previous	0	0.0
15	Poutcome	nonexistent	0.5
16	Emp.var.rate	1.1	0.25
17	Cons.price.idx	93.994	0.72
18	Cons.conf.idx	;-36.4	0.36
19	Euribor3m	4.857	0.911
20	Nr.employed	5191	0.8
21	Outcome	no	0

III. Multi-Layer Neural Network architecture and Back Propagation Learning Algorithm

Neural network is computer algorithm that is influenced from way of human learning. Often an experienced banker can judge a potential investor from his profile. There are many factors which are modeled in brain that can predict new situations not just by remembering but by wisdom. Similarly, an efficient NN can model observed data and its outcome in training process. However, an oversized NN can also produce low training error by memorizing the input dataset. This is known as Over-fitting. An over-fitted NN performs poorly on new dataset. When a NN builds the process model, it predicts the outcome of new dataset with astonishing accuracy. We will see one such case study of model based banking in this paper where past profile records of clients can be used to train a multi-layer NN and it predicts the clients who are likely to buy the product with 95% accuracy.

A. *Multi-Layer Neural Network architecture*

Multi-Layer NN is also known as Multi-Layer Perceptron (MLP) consisting of more than one layer is capable of mapping nonlinear boundaries. This was first demonstrated by David Rumelhart in 1986[14]. The inner layer in MLP is called hidden layer. More than one hidden layer is often used in deep learning NN for complex modeling. MLP consist of two components: 1- Artificial Neuron with input and output functions and 2- Connections linking the neurons with weights. The choice of weight of connections makes a NN to perform complex mapping function. This is achieved by training the NN with real input-output data.

Fig. 1 shows a multi-layer NN architecture. Input neurons propagate the input through connection weights Wji to the hidden neuron inputs Xj. All the neurons except input neurons are consist of input summing function and output activation function. The activation functions are usually bounded functions like, sigmoid, hyperbolic tangent, or radial basis function. Each of these neurons is connected to bias neurons with constant output. All the connection weights are reinforced using Back Propagation Algorithm.

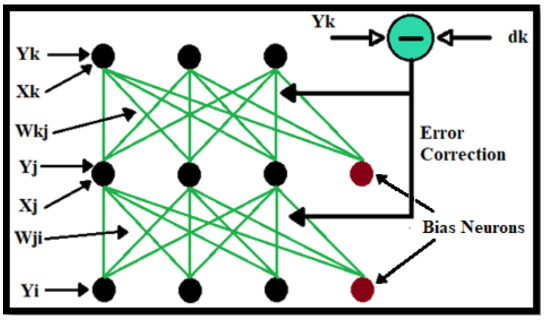


Fig. 1 shows a two-layer neural network with input neuron Y_i, hidden neuron Y_j and output neuron Y_k. The desired output d_k is compared with output Y_k and error is back propagated to correct the connection weights W_{kj} and W_{ji} using gradient descendent error correction algorithm.

B. *Back Propagation Learning Algorithm*

Back Propagation (BP) is most popular algorithm for estimating the connection weights of MLP. Initially, the algorithm starts with random set of weights. Output is estimated using equation 1a, 1b, 1c and 1d. A cost function shown in equation 2a is used to estimate the output error. In BP algorithm the sensitivity of each connection weight is calculated using first derivative of cost function. The sensitivity is proportional to the reinforcement factor ‘ η ’ which is also known as learning rate. The reinforcement equation for output layer and hidden layer are shown in equation 2b and 2c respectively. It uses iterative method till the error reduces with threshold limit.

MLP Equations-

$$X_j = \Sigma Y_i * W_{ji} \quad (1a)$$

$$Y_j = \frac{1}{1+e^{-xj}} \quad (1b)$$

$$X_k = \Sigma Y_j * W_{kj} \quad (1c)$$

$$Y_k = \frac{1}{1+e^{-xk}} \quad (1d)$$

Where, Y_k- Output, X_k- Input to Output Neurons, Y_j- Hidden Neuron Output, X_j- Input to Hidden Neurons, Y_i- Input, W_{kj} – Output layers’ connection weight and W_{ji} – Hidden layers’ connection weights.

Back-Prop Equations-

Cost Function – $E=\frac{1}{2}*(Y_k-d_k)^2$ (2a)

Error Sensitivity is calculated by taking partial derivative $\frac{\partial E}{\partial W}$ of the cost function with respect to output and hidden layer connections as shown in equation 2b and 2c respectively.

$$W_{kj}=W_{kj}+\eta*\Sigma\{(Y_k-d_k)*Y_k*(1-Y_k)\}*Y_j \quad (2b)$$

$$W_{ji}=W_{ji}+\eta*\Sigma\{(Y_k-d_k)*Y_k*(1-Y_k)*W_{kj}\}*Y_j*(1-Y_j)*Y_i \quad (2c)$$

Where, η = Learning rate and d_k - Desired output for training.

The algorithm described above uses three-layer architecture with single hidden layer. However, in actual practice we may need more than one hidden layer or a non-layered configuration. This situation arises after pruning an oversized fully connected network by removing neurons and connections. Section IV shows optimized network to overcome the risk of over-fitting.

IV. Proposed Method

A general purpose visual NN (VNN) framework shown in Fig. 2 is designed to solve different type of supervised learning problems. The framework is having user-friendly GUI frontend to define feed-forward type NN of desired configurations. The system allows interfacing user data in different format like; excel sheet, images and online data generators. The framework supports multiple network training in parallel and visually interconnects them for deep learning applications. The network configuration for randomly connected system is supported. It can be trained and pruned to optimize the network configuration for given dataset. The example in Fig. 2 shows three different nets that are copied from a fully connected and trained network. These three networks are independently pruned and further trained for optimal solution. The training and testing result of these networks are presented in Table 3. Framework supports parallel operation of all the networks for validating and optimizing multiple tasks at the same time. It also provides C++ source code output with network configuration in .xml and connection values for portability.

In this study, we used UCI benchmark dataset having 20 inputs and one output parameter through .csv files interface. The parameters are listed at the input of the network in Fig. 2. The training dataset consist of 41,188 records and test dataset consist of 4,119 independent records. Initially an oversized network consists of 20 input neurons, one output neuron and four hidden layers containing 10 neurons per layer is selected experimentally for training from different configurations in parallel. The network is trained for accuracy of 95.0146% with training dataset and examined using independent test dataset with 95.1828% accuracy. This confirmed the learning without over-fitting. It is further trained along with multiple pruning cycles to reduce the network size without considerably affecting the accuracy. The pruning also indicated the input parameters which has very less contribution in prediction model. The system provides detail error breakup in four different categories which are used to calculate the system performance of classification.

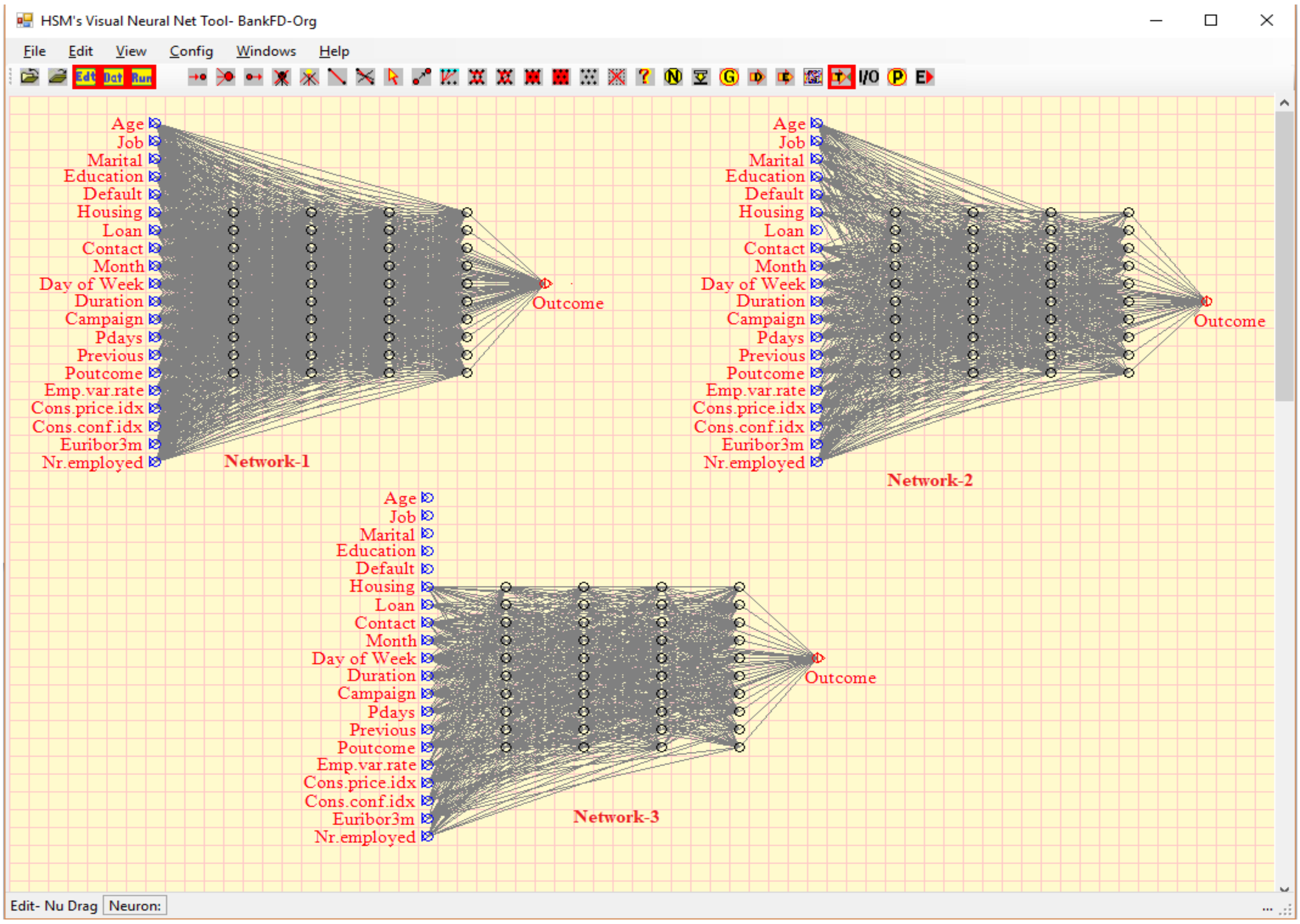


Fig. 2 shows Visual NN (VNN) Framework developed to solve complex mapping problem using supervised learning. In this study we predict the willingness of potential client to purchase bank product from parameters collected through telephone calls.

Prediction accuracy, sensitivity and specificity are three important factors to evaluate the performance of the model. The model groups the predicted output in four categories as explained below:

- True Positive (TP) is number of correctly predicted clients who are potential buyers of the product.

- False Positive (FP) is number of incorrectly predicted clients as potential buyers of the product.
- True Negative (TN) is number of correctly predicted clients who are not potential buyers of the product.
- False Negative (FN) is number of incorrectly predicted clients who are not potential buyers of the product.

The above statistical factors are required to compute the prediction accuracy, sensitivity and specificity as stated below:

- Accuracy = $\frac{TP+TN}{N}$ (3)
- Sensitivity = $\frac{TP}{TP+FN}$ (4)
- Specificity = $\frac{TN}{TN+FP}$ (5)

Section V describes the prediction result using above equations in Table 5

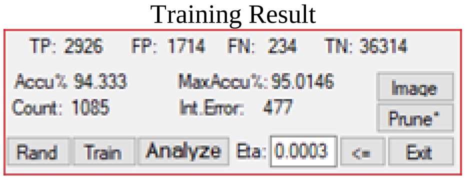
Table 3. Results of Pruned Network

Net No.	Connections	Input Neurons	Hidden Neurons	Accuracy	
				Training	Testing
1	1460	20	40	95.015%	95.183%
2	740	19	40	94.151%	94.306%
3	626	15	40	93.054%	92.945%

V. Results

All results shown in this section is calculated using UCI benchmark dataset [7] downloaded as 21 columns excel data sheet containing numeric and text parameter attributes in two batches for training and testing. These raw data are preprocessed and normalized as described in Table 2 from section II. The preprocessed normalized data is store in 2 csv files that are interfaced with VNN framework for training and testing. Initially training is performed with an oversized fully connected network till 95.0146% learning accuracy. A copy of this network is made for further optimization experiments. These networks are further pruned to reduce the inputs in different combinations. The results of pruned network are shown in Table 3. The Fig. 3 shows the training and testing operations. The final values are analyzed to compute TP, FP, TN, and FN as shown in Table 4 and accuracy, sensitivity, specificity as shown in Table 5.

Training Result



Testing Result

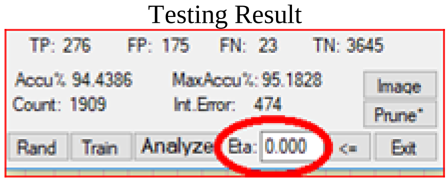


Fig. 3 shows the result dialogue boxes of Network-1, Figure 2 for Training and Testing operation. Training and Testing operation differ in dataset and learning coefficient η .

Table 4. The confusion matrices of MLPNN model for Training and Testing subsets

Neural Network Configuration	Model	Training Data			Testing Data		
		Desired output	Yes	No	Desired output	Yes	No
Conn:1460 Neuron:61	MLPNN	Yes	TP=2,926	FP=1,714	Yes	TP=276	FP= 175
		No	FN=234	TN=36,314	No	FN=23	TN=3645

Table 5. Percentage of Statistical Measures of MLPNN for Training and Testing Subsets calculated using equation 3, 4 and 5

Network Configuration	Model	Partition	Accuracy	Sensitivity	Specificity
Conn:1460 Neuron:61	MLPNN	Training	95.27%	92.59%	95.49%
		Testing	95.19%	92.31%	95.42%

VI. Conclusion

This paper presents a neural network based investor predictive model by utilizing the network and input parameter optimizing technique for the bank telemarketing prediction. Different pruned network configuration along with training results are shown. Importance of input parameters and corresponding accuracy is shown in tabular form. The experimental results showed that false negative to true negative ratio increases rapidly with small decrease of accuracy. Such increase leads to several fold non productive persuasion effort. The algorithm can be applied to similar applications with minor modification to optimize data collection, computational performance and decision making accuracy.

Acknowledgment

We would like to thank Chakarin Vajiramedhin and Anirut Suebsing for their article which inspired this work. We are also thankful to UCI repository, for providing the datasets.

References

[1] A. Asuncion, D. Newman, CA: University of California, School of Information and Computer Science, UCI machine learning repository. Irvine, 2012.
[2] A. Suebsing, N. Hiransakolwong, A novel technique for feature subset selection based on cosine similarity, Applied Mathematical Sciences, vol.6, no. 133, 2012, 6627-6655.
[3] G. John, R. Kohavi, and Pfleger. Irrelevant features and the subset selection problem. Int. Conf. on Machine Learning, Morgan Kaufman, San Francisco. 1994, 121-129.

[4] Mark A. Hall, Lloyd A. Smith, Feature Selection for Machine Learning: Comparing a Correlation-based Filter Approach to the Wrapper, American Association for Artificial Intelligence, 1998.

[5] R. Singh, R. R. Aggarwal, Comparative Evaluation of Predictive Modeling Techniques on Credit Card Data. International Journal of Computer Theory and Engineering, 3 (2011), 598-603.

[6] Xinguo Lo and et al., A Novel Feature Selection Method Based on CFS in Cancer Recognition, IEEE 6th International Conference on System Biology IISB), 2012.

[7] http://archive.ics.uci.edu/ml/index.html (last accessed: November 2012).

[8] Petrison, L. A., Blattberg, R. C. and Wang, P. „Database marketing: Past present, and future", Journal of Direct Marketing, 11, 4, 109–125, 1997.

[9] Brandweek 50, no.36.D1-D4 "The Next Generation of DIRECT MARKETING." Academic Search Complete, EBSCOhost, P.6, 2009.

[10] O'guinn, Thomas." Advertising and Integrated Brand Promotion". Oxford Oxfordshire: Oxford University Press. p. 625. ISBN 978-0-324-56862-2. , 2008.

[11] Eniafe Festus Ayetiran, "A Data Mining-Based Response Model for Target Selection in Direct Marketing", I.J.Information Technology and Computer Science, 2012, 1, 9-18.

[12] Hany. A. Elsalamony, Alaa. M. Elsayad," Bank Direct Marketing Based on Neural Network", International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-2, Issue-6, August 2013 https://www.researchgate.net/publication/272170309_Bank_Direct_Marketing_Based_on_Neural_Network

[13] Chakarin Vajiramedhin, Anirut Suebsing, "Feature Selection with Data Balancing for Prediction of Bank Telemarketing", Applied Mathematical Sciences, Vol. 8, 2014, no. 114, 5667 – 5672

[14] Rumelhart, D. E. and J. L. McClelland, 1986, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, MIT Press, Cambridge, MA

Rutu S. Patel is BE. Information Technology, Gujarat Technological University in 2014, MBA in Marketing from Ahmedabad University in 2017. Presently working as a Teaching Associate in Ahmedabad University.

Himanshu S. Mazumdar is BE. Electronics and Telecom, Jadavpur University in 1968, PhD in Computer Engineering from DDU, Nadiad in 2004.

He has worked in international space missions in University College London, NASA's Space Shuttle and Indian Space Research Organization. He is recipient of Hari Om Ashram Prerit Vikram Sarabhai Award in the year 1991. He has research work published in the area of Bio-Informatics, Computational Intelligence and Space Instrumentation. Presently, he is working as Professor, EC and Head of R&D Center at Dharmsinh Ddesai University, Nadiad.