TEM Journal. Volume 8, Issue 3, Pages 938-944, ISSN 2217-8309, DOI: 10.18421/TEM83-35, August 2019

Marketing Intelligence in the Era of Big Data

Shorouq F. Eletter, Tahira Yasmin, Ghaleb A. El Refae

Al Ain University of Science and Technology, P.O. Box: 64141, Al Ain, UAE

Abstract **– Telemarketing is a cost-effective instrument that offers products or services to customers through two-way and one-to-one communications over the phone. This research will use a real data collected from a Portuguese retail bank dataset during (May/ 2008 to June/ 2013) period, in order to sell long term deposits. The dataset comprises a total of 20 variables in a total of 41188 telephone contacts, and a binary ("yes" or "no") response variable that describes whether or not the customer will purchase the long-term deposit. Logistic regression and Naïve base models are built for benchmarking. Based on the results, the predictive performance of the multilayer perceptron artificial neural networks is better than the other models.**

Keywords **– Big Data, Artificial neural network, Logistic regression, Naïve base, Receiver Operating Characteristic (ROC).**

1. Introduction

The revolution regarding information and communication technologies have greatly facilitated the value co-creation between firms and customers [1]. Nowadays, business data is generated and captured at exponential rate. This vast amount of gathered data, known as Big Data, is transforming the way businesses operate by opening a new customers, firms and markets in the way that every party is visible to all. Therefore, the movement of window into customers' behaviors, preferences and wants. Big Data acts as glass window for all parties: Big Data brought many opportunities and challenges to all parties.

Business Intelligence (BI) is an umbrella name that comprises numerous hardware and software capabilities, technologies, systems, and applications to make better timely business decisions. In the era of Big Data, businesses became more and more relying on BI and data analytics to discover and gain crucial insights from raw data gathered through various transactional enterprise systems as well as external sources. This led to unprecedented intelligence on buyer opinion and consumer wants to identify new business opportunities. Moreover, it offers a large extent data-related solutions to contemporary business organizations [2].

The Big Data era emanated along with both big opportunities and challenges for marketing in terms of data collection, storage, and management. This also demanded new marketing strategies to keep competitive advantage. In his research, [3] it is stated that firms need to be aware of, and respond to the demands of this era in order to avoid getting lost by competitors. Nowadays, businesses are able to collect data from social media, transactions, surveys, sensor networks, and many different sources. The ability to integrate heterogeneous sources of information will provide a holistic view of the domain and generates more accurate marketing intelligence [4]. Businesses are utilizing Data Mining (DM) techniques to reveal the hidden information in Big Data [5]. Moreover, data analysis models are used to capture, visualize, and analyze the underlying patterns in the data to make better decision. Therefore, firms need to classify and utilize Big Data based on business value for better decision making [6]. This puts consumer analytics at the epicentre of a Big Data revolution. Businesses are using analytics to extract and utilize consumer insight from Big Data to enhance marketing capabilities [7]. Hence, it is important to leverage the full potential offered by Big Bata and business intelligence analytics in order to gain competitive advantage [8]. Recently, businesses are using Big Data and BI capabilities to identify and target worthy customers directly.

Many industries are using direct marketing to provide an interactive contact with customers,

DOI: 10.18421/TEM83-35
https://dx.doi.org/10.18421/TEM83-35

Corresponding author: Shorouq F. Eletter,
Al Ain University of Science and Technology, Al Ain, UAE
Email: *shorouq.eletter@aau.ac.ae*

Received: 27 May 2019.
Revised: 04 August 2019.
Accepted: 10 August 2019.
Published: 28 August 2019.

receiving a direct response from them, hence, get to know their needs and fulfil them. Direct marketing instruments are usually used to obtain new customers and to generate additional return from current customers and therefore, enhancing the economic rewards for the organization [9]. Moreover, direct marketing instruments eventually aim to create cost-effective channels through two-way and one-to-one communications [10]. Contacting customers through the phone to sell products or services (known as telemarketing) is a powerful interactive direct marketing technique [11]. Organizations use telemarketing to contact their customers in order to improve the marketing campaigns through an integrated and systematic use of telecommunications and information processing technologies. Unlike other sales techniques, it is believed that telemarketing offers many benefits to organizations through reducing costs even by half, and increasing the amount of sales significantly. Telemarketing allows targeting a selected set of customers, in order to choose those who are extremely enthusiastic to buy the offered product or service. Telemarketing enable organizations to understand their customers' expectations and needs; to offer products/services as a result of directly pointing out the advantages of purchasing such products/services while replying to customer objections or fears [12]. Targeting only worthy customers after identifying their characteristics can improve response rate, reduce expenses, and hence, increase profits.

Data mining is a promising set of business intelligence techniques and procedures that improve the quality of raw data. Currently, in the era of big data, it is necessary to have such tools to analyse the enormous amount of data that organization need. Generally, classification is a data mining task that discriminates and distinguishes data classes. The goal of classification is to utilize a training dataset to understand how attribute values relate to class labels. Then , the classifier is able to predict class labels of given cases [13]. Building and establishing automated protocols for selecting customers in advance using available customers' information and metrics might help to decrease the time and costs of campaigns, and perform fewer and more effective phone calls, which will diminish client stress and intrusiveness [14].

Many banks are using data mining techniques prior screening the projected customers to understand their behaviour by classifying the customers before offering special services [11]. This research uses and compares three data mining techniques models: artificial neural networks (ANNs), Naïve Bayes (NB), and logistic regression (LR). ANNs represent promising data mining tools that have been used successfully in classification problems. The NB model is widely used for classification problems in data mining and machine learning fields because of its simplicity and remarkable classification accuracy [13]. In addition, Logistic regression is a simple model to interpret classical model for comparison.

2. Methodology

2.1. Data

This research will use a real phone call marketing campaign dataset collected from a Portuguese retail bank dataset during (May/ 2008 to June/ 2013) period to sell long term deposits. Typically, clients were contacted more than once to determine whether the client will purchase the product. The dataset comprises a total of 20 variables in a total of 41188 telephone contacts, and a binary ("yes" or "no") response variable that describes whether or not the customer will purchase the long- term deposit (Table 1). This dataset was used by [15], [10], [14].

During the telemarketing campaign a list of clients is either targeted with phone call and offered to buy the deposit or clients call the contact-center for any other reason, and they are asked to subscribe the deposit. Hence, the response will be either "yes" or "no" (binary variable) depending on whether or not the client is going to purchase the long- term deposit [14].

Table 1: Telemarketing Dataset Variable Description

	Number	Variable Description	Type	Role
Client information	1	age	numeric	input
	2	job: type of job	categorical	input
	3	Marital: marital status	categorical	input
	4	Education	categorical	input
	5	default: has credit in default?	categorical	input
	6	Housing: has housing loan?	categorical	input
	7	loan: has personal loan?	categorical	input
Attributes of Last contact of Current campaign	8	contact: contact communication type	categorical	Input
	9	month: last contact month of year	categorical	Input
	10	Day-of-week: last contact day of the week	categorical	Input
	11	duration: last contact duration, in seconds	numeric	Input
other attributes	12	Campaign: number of contacts performed during	numeric	Input
	13	pdays: number of days that passed by after the client was	numeric	Input
	14	previous: number of contacts performed before this this	numeric	Input
	15	poutcome: outcome of the previous marketing campaign	numeric	Input
social and economic context attributes	16	emp.var. rate: employment variation rate - quarterly	numeric	Input
	17	cons.price.idx: consumer price index - monthly indicator	numeric	Input
	18	cons.conf.idx: consumer confidence index - monthly	numeric	Input
	19	euribor3m: euribor 3-month rate - daily indicator	numeric	Input
	20	nr. employed: number of employees - quarterly indicator	numeric	Input
Client	21	y - has the client subscribed a term deposit?	categorical	output

The dataset comprises 4640 (11.3%) successful calls and 36548 (88.7%) unsuccessful calls. The dataset was divided into training set 69.9% for model building, and a testing set of 30.1% for model evaluation. The training set was used to determine the classifier parameters. Then the testing set was applied on the trained classifier to make prediction. Cross-validation was used to evaluate predictive performance of the classifiers. SPSS 24 was used to develop the different data mining models.

2.2. The Data Mining Models

This research compares classification ability of three data mining models: naïve base (NB), neural network (NN) and logistic regression (LR). The NB classifiers are useful, efficient and commonly used for solving classification problems in data mining. NB is enjoying a renaissance because it is easy to use and its distinct stability. NB classifier discriminates between classes by calculating the probability membership for each class while removing the corresponding probabilities for missing attribute. Furthermore, NB classifier assumes class conditional independence, i.e., the effect of all attributes on a particular class are independent of each other [13]. NB classifier is a probabilistic model based on the Bayes rule along with a strong independence assumption. According to the Bayes Rule, the probability that a particular customer belongs to a class is given by equation (1):

$$P(c_i|d) = \frac{P(d|c_i)*P(c_i)}{P(d)} \qquad (1)$$

Multilayer perceptron is a feedforward neural network model composed of a number of fully connected units called neurons, which are comparable to the biological neurons in the brain [16], [17]. A typical feedforward neural network comprises of three layers: an input layer, hidden layer, and an output layer. A neuron is the basic processing unit of the network. Neurons cooperate across different layers through several weighted connections. A transfer function is associated with each neuron in order to describe how the weighted sum of its input is converted into an output [18]. A neuron receives information from other neurons, or sometimes from external stimulus. Then, it generates an output that is transferred into destination neurons for further processing. The network learns the relationship between input and output by adjusting the connection weights repeatedly until its output matches the actual output [19]. After training, the ANN uses the acquired knowledge to respond to new input, and make prediction with good accuracy while the input-output relationship remains unknown [20].

For a given input x_k the state of the i^{th} neuron (o_i) is computed by equation (2):

$$o_i = f(w_{i0} + \sum_{j \epsilon P_i} w_{ij} * o_j) \quad (2)$$

in which P_i denotes the set of neurons reaching the i^{th} neuron, f is the transfer function, w_{ij} is the connection weight between neuron j and i, and o_j represents the output of the j^{th} neuron. Before applying MLP model, by default, covariates are standardized before training to a zero mean and one standard deviation. The automatic architecture selection was used to set the number of hidden neurons in the hidden layer.

Logistic regression (LR) is a predictive model broadly used in classification. LR is a linear regression in which the dependent variable is categorical (binary). The LR model is used to estimate the probability of a binary response using a set of predictor variables. A multiple LR model was used with the predictor variables such as in equation (3):

$$P(y_i = 1|xi) = 1/(1 + \exp(-\alpha - x_i\theta)) \quad (3)$$

in which xi is a vector of independent variables, θ is the corresponding regression coefficients and α is the intercept [14]. LR is easy to interpret because of the additive linear combination of the independent variables. The current study uses LR model to determine those clients who purchased the long- term deposit.

2.3. Performance Evaluation

At this stage, the generated models needed to be evaluated to search for the best model. Model evaluation techniques are used to explore and highlight the best classification model. However, the average correct classification rate (ACC) and the receiver operating characteristic (ROC) are used to evaluate the performance of the developed models in order to promote the optimal classification model. ACC is a widely used criterion that measures predictive power of a classifier. When a model exhibits a high classification power, it can differentiate between those customers who will purchase the long- term deposit and those who will not. ACC measures the percentage of the accurately classified cases in a particular data set. Furthermore, ACC equals the number of cases correctly classified and divided by total number of cases. The output in a binary classification problem that is summarized by the following table, in which yes represents the variable of interest as in Figure (1).

Observed	**Predicted**	
	no	**yes**
no	TN	FN
yes	FP	TP

Figure 1: Observed and Predicted.

True negative (TN): number of correctly predicted "no" response
False negative (FN): number of incorrectly predicted "no" response
True positive (TP): number of correctly predicted "yes" response
False positive (FP): number of incorrectly predicted "yes" response (see Equation 4):

$$CC = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

ROC curve is a useful procedure to evaluate classifiers performance for a binary response variable. The ROC is generated by plotting sensitivity (the percentage of yes predicted correctly, i.e. true positive rate: $sensitivity = \frac{TP}{TP+FN}$) against '1-specificity' or false positive rate in which specificity is the percentage of customers who did not subscribe and predicted correctly (true negative rate): $specifity = \frac{TN}{TN+FP}$

This sensitivity rate shows the effect of previously determining customers that will respond positively to the campaign. On the other hand, the specificity rate shows the effect of previously determining customers that will respond negatively to the campaign [10]. The 45° diagonal line indicates that at each point on the line, the true positive rate equals the false positive rate, which means it has no predictive power. In addition, the upper the curve is from the 45°line, the better the classification accuracy of a model. ROC demonstrates the resulting of overall performance at all cut-off values. The ROC illustrates the behavior of the classification model without considering the misclassification cost. Additionally, the area under the ROC (AUC) is another graphical assessment technique that determines the accuracy measure of a classifier. It is a quantitative measure ranges between 0.5 (which indicates no better than random chance) and 1 (indicates a perfect classifier) [14], [21]. AUC measures the discriminative power of the classifier. A 50% AUC corresponds to a random classifier (i.e. area under the 45° line). The curve that has a larger AUC is superior.

3. Analysis of Results

After building the classification models, their predictive power was assessed using a number of evaluation criteria, namely ACC, ROC and AUC techniques. This helped to highlight the optimal model that predicts those customers who will subscribe. Table 2 shows that although MLP model slightly outperformed the NB and LR; models however, it is criticized because of its black box nature.

Table 2. Classification Results.

Observed response	Predicted	MLP	NB	LR
No	success	96.4%	95.5%	97.1%
	failure	3.6%	4.5%	2.9%
Yes	success	51.7%	57.9%	45.1%
	failure	48.3%	42.1%	54.9%
ACC (%)		91.5%	91.4%	91.4%
AUC		94.4%	92.1%	93.6%

Additionally, the ROC was used to evaluate the predictive power of the created models. The higher the curve and the closer to the upper left corner, the better the predictive power is. Figure 2 shows that the ROC of the MLP, NB and LR are close although ROC for MLP is a little higher than the other two curves. This also can be supported from the MLP AUC of 94.4%, which was the highest among the three classifiers indicating that ROC of MLP is the highest.

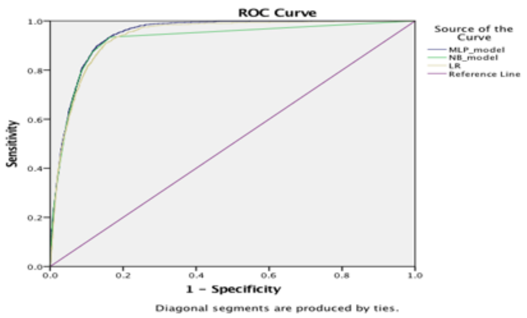

Figure 2: ROC Curve.

The MLP, NB and LR models accuracy prediction is higher than the model of [10]. The findings of [10] showed that ANN achieved 88.31%, NB achieved 69.14%, and LR achieved 81.18% prediction accuracy. In addition, the findings of this research regarding the AUC are higher than the findings of [14] of AUC for ANN and LR are 92.9% and 71.5% respectively. On the other hand, the results of this research for all models are lower than the results of [11] which achieved AUC was 96.5%.

Figure (3) reveals the descending ranking of the importance and normalized importance values of all variables (the variable names are described in Table 1). The importance of an independent variable is referring to the measures, which is how much the network's predicted value varies for different values of the independent variable. Therefore, this figure will help to identify the characteristics of the customers who will subscribe. Considering the variables input ranking (Figure 3) the four best relevant variables are pdays, previous, duration and campaign. Pdays represents number of days that passed by after client was contacted from previous campaign. The fewer days pass the higher is a possibility that the client will subscribe. The logical explanation of that might be the following: the customer is still aware of the product, but after a longer period customers might forget the product and move to something else. Remarkably customers are targeted by too many products and services through social media and online shopping; therefore, they become attracted to shopping rather than savings. The second variable is previous: number of contacts performed before this campaign for this client indicates performing more contacts with clients and will make them good targets to buy long term deposits. Duration (*last contact duration in seconds*) is also important characteristic. This happens since long calls with clients will help them to ask questions and find more details about the product's benefits. What is more, campaign (*number of contacts performed during this campaign for this client*) is an important variable which indicates that during campaigns marketers try to contact clients as much as possible, which might help to keep the product in their mind, and enthusiastically they buy the product. This result is consistent to the findings of [22], and also to the findings of [14]. However, the way in which the independent variables are correlated to the predicted value of the clients' decision is not obvious. Based on commonsense, one could guess that the more the client is contacted the higher is probability that he will subscribe.

On the other hand, variables such as a day of the week, marital housing and default are in the bottom of the variable rank list (Figure 3). This might indicate that these variables might have the least influence on customer's decision. This contradicts the findings of [22], showing that married clients are more willing to subscribe as they need to secure their children. While the results in this research showed that a month has a moderate influence on client's decision, [14], it is found that the month (in which the client is contacted) played a significant role in the decision.

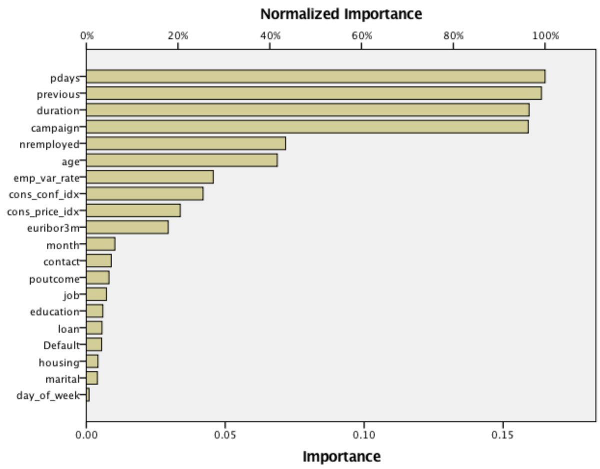

Figure 3: Normalized Importance.

4. Conclusion

Banks are using cost-effective instruments through targeting potential clients for telemarketing in order to increase profits while reducing cost. Simultaneously, if clients buy the long term deposit this will increase the bank's capital. This research used the predictive power of three classification models: MLP, LR and NB classifiers to select bank telemarketing clients. The objective was to compare the predictive power of the three models, and then to highlight the best model based on a number of evaluation criteria. The performances of the proposed algorithms were examined using the classification accuracy ACC, ROC and AUC.

The proposed methods showed good classification accuracy rates. Although the classification accuracy of the three models were close, the performance of MLP was slightly better than the NB and LR. The experimental results showed that the proposed models can be used successfully for the classification of real life problems. Additionally, this research has pointed to some variables that may increase the possibility of a successful call (the product is sold). The results revealed that the fewer days that passed by after client was contacted from previous campaign, the more contacts performed before this campaign for this client, longer call duration, and the more contacts performed during this campaign for this client are among the important characteristics for the clients who will subscribe.

The above discussion highlights some policy implications including identifying, and then targeting a group of clients with the potential characteristics rather than the whole population facilitate to achieve the marketing campaign objective more quickly, regarding the reduced cost and increase profit. Additionally, some organizational changes are demand, such as having data analytics department to make better decisions by utilizing the value of Big Data. It is necessary for systems that address the explicit use of knowledge in decision making. For future work, other classification algorithms, such as decision trees and genetic algorithms can be used to deal with real-time multi-class classification problems.

References

[1]. Bettencourt, L. A., Lusch, R. F., & Vargo, S. L. (2014). A service lens on value creation: marketing's role in achieving strategic advantage. *California management review*, *57*(1), 44-66.

[2]. Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, *36*(4): 1165-1188.

[3]. Hu, Y. (2018). Marketing and Business Analysis in the Era of Big Data. *American Journal of Industrial and Business Management*, *8*(07), 1747.

[4]. Fan, S., Lau, R. Y., & Zhao, J. L. (2015). Demystifying big data analytics for business intelligence through the lens of marketing mix. *Big Data Research*, *2*(1), 28-32.

[5]. Hassani, H., Huang, X., & Silva, E. (2018). Digitalization and big data mining in banking. *Big Data and Cognitive Computing*, *2*(3), 18.

[6]. Xie, K., Wu, Y., Xiao, J., & Hu, Q. (2016). Value co-creation between firms and customers: The role of big data-based cooperative assets. *Information & Management*, *53*(8), 1034-1048.

[7]. Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897-904.

[8]. Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management*, *16*(3), 547-578.

[9]. Kotler, P.G.A. (2006) Principles of Marketing. Pearson Prentice Hall.

[10]. Keles, A., & Keles, A. (2015). IBMMS Decision Support Tool for Management of Bank Telemarketing Campaigns. *arXiv preprint arXiv:1511.03532*.

[11]. Vajiramedhin, C., & Suebsing, A. (2014). Feature selection with data balancing for prediction of bank telemarketing. *Applied Mathematical Sciences*, *8*(114), 5667-5672.

[12]. Gherman, C., & Marcu, M. (2010). The Role And Implications Of Telemarketing In Media. Case Study: Transilvania Expres Daily Publication Of Brasov. *Annals of the University of Petroşani*, *10*(4), 97-104.

[13]. Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., & Strachan, R. (2014). Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems with Applications*, *41*(4), 1937-1946.

[14]. Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, *62*, 22-31.

[15]. Yang, C. (2016). *Predicting success of bank telemarketing with classification trees and logistic regression* (Doctoral dissertation).

[16]. Kankal, M., Akpınar, A., Kömürcü, M. İ., & Özşahin, T. Ş. (2011). Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. *Applied Energy*, *88*(5), 1927-1939.

[17]. Eletter, S. F., Yaseen, S. G., & Elrefae, G. A. (2010). Neuro-based artificial intelligence model for loan decisions. *American Journal of Economics and Business Administration*, *2*(1), 27-34.

[18]. Bekhet, H. A., & Eletter, S. F. K. (2014). Credit risk assessment model for Jordanian commercial banks: neural scoring approach. *Review of Development Finance*, *4*(1), 20-28.

[19]. Ekonomou, L., Christodoulou, C., & Mladenov, V. (2016). A short-term load forecasting method using artificial neural networks and wavelet analysis. *Int. J. Power Syst*, *1*, 64-68.

[20]. Turban, E., Sharda, R. & Delen, D. (2011). Decision support and business intelligence systems (9th ed.). London: Pearson.

[21]. Asare-Frempong, J., & Jayabalan, M. (2017, September). Predicting customer response to bank direct telemarketing campaign. In *2017 International Conference on Engineering Technology and Technopreneurship (ICE2T)* (pp. 1-4). IEEE.

[22]. Ahn, J. H., & Ezawa, K. J. (1997). Decision support for real-time telemarketing operations through Bayesian network learning. *Decision Support Systems*, *21*(1), 17-27.