International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958, Volume-2, Issue-6, August 2013

Bank Direct Marketing Based on Neural Network

Hany. A. Elsalamony, Alaa. M. Elsayad

Abstract*— *All bank marketing campaigns are dependent on customers' huge electronic data. The size of these data source is impossible for a human analyst to come up with interesting information that will help in the decision-making process. Data mining models are completely helping in performance of these campaigns. This paper introduces applications of recent and important models of data mining; Multilayer perceptron neural network (MLPNN) and Ross Quinlan new decision tree model (C5.0). The objective is to examine the performance of MLPNN and C5.0 models on a real-world data of bank deposit subscription. The purpose is increasing the campaign effectiveness by identifying the main characteristics that affect a success (the deposit subscribed by the client) based on MLPNN and C5.0. The experimental results demonstrate, with higher accuracies, the success of these models in predicting the best campaign contact with the clients for subscribing deposit. The performances are measured by three statistical measures; classification accuracy, sensitivity, and specificity.

Index Terms*—*Bank Marketing; Data Mining; Neural Network; C5.0.

I. INTRODUCTION

In banks, a huge data recorded about their customers. This data can be used to create and keep direct relationship and connection with the customers in order to target them individually for definite products or banking offers. Usually, the selected customers are contacted directly through personal contact, telephone cellular, mail, and e-mail or any other contacts to advertise the new product/service or give an offer, this kind of marketing is called direct marketing. In fact, direct marketing is in the main a strategy of many of the banks and insurance companies for interacting with their customers.

Historically, the name and identification of the term direct marketing suggested first time at 1967 by Lester Wunderman, which he considered the father of direct marketing [5].
In addition, some of the banks and financial-services companies may depend only on strategy of mass marketing for promoting a new service or product to their customers. In this strategy, a single communication message broadcasted to all customers through media such as television, radio or advertising firm, etc. [6]. In this approach, companies do not set up a direct relationship with their customers for new-product offers. In fact, many of the customers are not interesting or respond to this kind of sales promotion.

Accordingly, banks, financial-services companies and other companies are shifting away from mass marketing strategy because its ineffectiveness and they are now targeting most of their customers by direct marketing for specific product and service offers [3, 4]. Due to the positive results clearly measured; many marketers attractive to the direct marketing. For example, if a marketer sends out 1,000 offers by mail and 100 respond to the promotion, the marketer can say with confidence that campaign led directly to 10% direct responses. This metric known as the 'Response Rate', and it is one of much clearly quantifiable success metrics employed by direct marketers. In dissimilarity, general advertising uses indirect measurements, such as awareness or engagement, since there is no straight response from a consumer [9]. From the literature, the direct marketing is becoming very important application in data mining these days [7]. The data mining widely has been used in direct marketing to identify prospective customers for new products, by using history-purchasing data, a predictive model to measure that a customer is going to respond to the promotion or an offer [12].

Data mining has gained popularity for illustrative and predictive applications in banking processes. The Multilayer perceptron neural network (MLPNN) is one from its techniques, which have their roots in the artificial intelligence. MLPNN is a mutually dependent group of artificial neurons that uses a mathematical or computational model for information processing using a connection approach to computation (Freeman et al., 1991) [17].

Another technique of data mining is the decision tree approach. Decision tree provides powerful techniques for classification and prediction. There are many algorithms to build a decision tree model [21, 22]. Decision tree is the most widely used data mining methods for several reasons. It can generate understandable rules, and to handle both continuous and categorical variables [23].

This paper investigates the effectiveness of the recent and famous efficient decision tree model (C5.0), and back propagation of neural network (MLPNN) on the bank direct marketing. The dataset well known as bank marketing from the University of California at Irvine (UCI) [1]. The remainder of the paper organized as follows: Section (II) focuses on the definition and features of data mining. Section (III) reviews on the neural network. The recent model in the decision tree C5.0 it will discuss in section (IV). The discussion of data and interpretation of it will be in section (V). Section (VI) demonstrates the experimental results to show the effectiveness of each model. Finally, conclusions will introduced in Section (VII).

II. DATA MINING OVERVIEW

The definition of Data Mining or Knowledge Discovery in Databases is the action that extracting some new important information contained in large databases. The target of data

Manuscript received August, 2013.
Hany. A. Elsalamony, Math Department, Faculty of Science, Helwan University, Cairo, Egypt.
Alaa. M. Elsayad, Computers and Systems Department, Electronic Research Institute, Cairo, Egypt. Electrical Engineering Department, Engineering College, Salman Bin Abdul Aziz University, Wadi Addwasir, Saudi Arabia.

mining is to find unexpected characteristics, hidden features or other unclear relationships inside the data based on techniques' combination. Today, many applications in a wide and various range of business founded and worked in this regulation.

In 1996, U. Fayyad, and G. Shapiro defined the general knowledge discovery process as an interactive and iterative process involving more or less the following steps: understanding the application field, data selecting, preprocessing and cleaning data, integration of data, data reduction and transformation, selecting algorithms of data mining, data mining, interpretation and description of the results and using the discovered knowledge [12]. In fact, the data mining can classified into two categories: descriptive and predictive as in [13].

Actually, in the recent years data mining occupying great position of attention area in the society of business or banking because its elasticity in working with a large amount of data, and turning such data into cleared information and knowledge [13]. Most of the people may be confused in understanding between terms "knowledge discovery" and "data mining" in different areas. Knowledge discovery in databases is the process of identifying valid, novel, probably useful, and finally understandable patterns/models in data. On the other hand, data mining is a step in the knowledge discovery process consisting of particular data mining algorithms that under some acceptable computational efficiency limitations, finds patterns or models in data [14]. Submit your manuscript electronically for review.

III. MULTILAYER PERCEPTRON NEURAL NETWORK

Multilayer perceptron neural network (MLPNN) with back-propagation is the most popular artificial neural network architecture [15, 16]. The MLPNN known as a powerful function approximation for prediction and classification problems. Historically, this direction field started when neurophysiologist Warren McCulloch and mathematician Walter Pitts introduced a paper on how neurons might work in 1943. They found a model for simple neural network using electrical circuits [8]. They named this model 'threshold logic'. The model opened the door on the way for research in neural network to divide into two distinct approaches. One approach concentrated on biological processes in the brain, and the other focused on the application of neural networks to artificial intelligence [10]. The interesting in the field renewed in 1982. John Hopfield introduced an approach to construct machines that are more useful that using bidirectional lines. In 1986, with multiple layered neural networks appeared three independent groups of researchers, one of which included David Rumelhart, presented similar ideas, which are calling now as back propagation networks because it distributes pattern recognition errors throughout the network. Hybrid networks used just two layers; these back-propagation networks use many. Neural networks applied to data mining in Craven and Sahvlik (1997) [8, 10].

Figure 1 shows that the MLPNN structure organized into layers of neuron's input, output and hidden layers. There is at least one hidden layer, where the actual computations of the network are processed. Each neuron in the hidden layer sums its input attributes x_i after multiplying them by the strengths of the respective connection weights w_{ij} and computes its output y_j using activation function (AF) of this sum. AF may range from a simple threshold function, or a sigmoid, hyperbolic tangent, or radial basis function.

$$y_i = f\left(\sum w_{ij} x_i\right) \quad (1)$$

Back-propagation (BP) is a common training technique for MLPNN. The available dataset is divided normally into training and test subsets. BP works by presenting each input sample to the network where the estimated output is computed by performing weighted sums and transfer functions. The sum of squared differences between the desired and actual asset value of the output neuron's E is defined as:

$$E = \frac{1}{2} \sum_j (y_{dj} - y_j)^2 \quad (2)$$

where y_{dj} is the desired value of an output neurons j, and y_j is the actual output of that neuron.

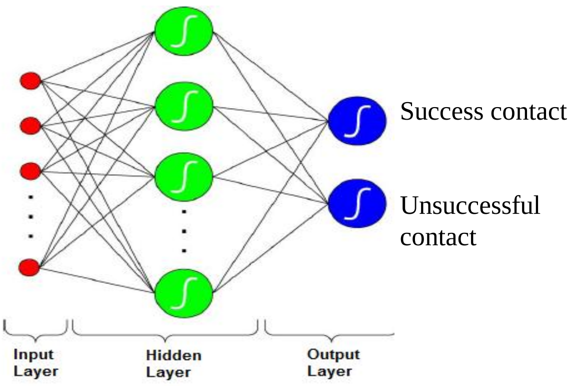

Figure.1. the structure of multilayer perceptron neural network.

Weights w_{ij} in Equation (1), are adjusted to reducing the error E of Equation (2) as fast, quickly as possible. BP applies a weight correction to reduce the difference between the network outputs and the desired ones; i.e., the neural network can learn, and can thus reduce the future errors. The performance of MLPNN depends on network parameters, the network weights and the type of transfer functions used [17].

When using MLPNN, three important issues need to be addressed; the selection of data samples for network training, the selection of an appropriate and efficient training algorithm and determination of network size. New algorithms for data portioning and effective training with faster convergence properties and fewer computational requirements are being developed [18, 19]. However, the third issue is a more difficult problem to solve. It is necessary to find a network structure small enough to meet certain performance specifications. Pruning methods for improving the input-side redundant connections were also developed that resulted in smaller networks without degrading or compromising their performance [20].

Finally, MLPNN has many advantages, such as good learning ability, less memory demand, suitable generalization, fast real-time operating, simple and convenient to utilize, suited to analyze complex patterns, and so on. Therefore, it has become a research hotspot in past few years [11]. On the other hand, there are some disadvantages like: the neural network requires high-quality data; variables must be carefully selected a priori, risk of over-fitting, and requires a definition of architecture. f you are using *Word*, use either the Microsoft Equation Editor or the *MathType* add-on (http://www.mathtype.com) for equations in your paper

(Insert | Object | Create New | Microsoft Equation *or* MathType Equation). “Float over text” should *not* be selected.

IV. DECISION TREE MODEL ALGORITHM

Data mining techniques include many that should be in the circle of interest for financial people dealing with huge and complicated data sets. One of the most popular of the data mining techniques, decision trees, originated in the statistics' discipline [24].

Decision tree algorithm partitions the data samples into two subsets so that the samples within each subset are more homogeneous than in the previous subset. This is a recursive process, the resulting two subsets are then split again, and the process repeats until the homogeneity criterion is reached or until some other stopping, criterion is satisfied [21, 22].

As the name implies, this model recursively separates data samples into branches to construct a tree structure for improving the prediction accuracy. Each tree node is either a leaf node or decision node. All decision nodes have splits, testing the values of some functions of data attributes. Each branch from the decision node corresponds to a different outcome of the test as in Figure 2.

Historically, the seminal book by Brieman et al. (1993) provided an introduction to decision trees that is still considered the standard resource on the topic. Two reasons for the popularity of decision tree techniques are (1) the procedures are relatively straightforward to understand and explain, and (2) the procedures address a number of data complexities, such as nonlinearly and interactions, that commonly occur in real data [24].

The famous model in decision trees is C5.0, which is a recently invented modeling algorithm, and it is an improved version of C4.5 and ID3 algorithms. C5.0 is a commercial product designed by Rule Quest Research Ltd Pty to analyze huge datasets and is implemented in SPSS Clementine workbench data mining software [2].

The tree of C5.0 uses common splitting algorithms includes entropy based on information gain. The gain ratio is a robust and consistently gives a better choice of tests than the gain criterion (ID3) for large dataset. The model works by splitting the sample based on the attribute that provides the maximum information gain. Each subsample defined by the first split then split again, usually based on a different attribute, and the process repeats until the subsamples cannot be split any further. Finally, the lowest-level splits are reexamined, and those that do not contribute significantly to the value of the model are removed or pruned.

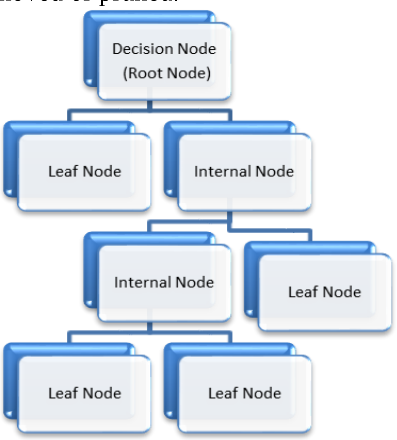

Figure.2. illustrated example of a binary decision tree

C5.0 model is quite robust in the presence of problems such as missing data and large numbers of input fields. It usually does not require long training times to estimate. In addition, C5.0 models tend to be easier to understand than some other model types, since the rules derived from the model have a very straightforward interpretation. Furthermore, C5.0 offers the powerful boosting method to increase accuracy of classification [25].

C5.0 uses information gain as a measure of purity, which is based on the notion of entropy. If the training subset consists of n samples $(x_1, y_1),\ldots, (x_n, y_n)$, $x_i \in R_p$ is the independent attributes of the sample, i and y_i is a predefined class $Y=\{c_1,c_2,\ldots,c_k\}$. Then the entropy, *entropy*(X), of the set X relative to this n-wise classification is defined as:

$$entropy(X) = (\sum_{i=0}^{n} -p_i \log_2 p_i) \quad (3)$$

where p_i is the ratio of X fitting in class c_i.

The gain (X, A) is simply expected reduction in entropy caused by partitioning the set of samples, X, based on an attribute A:

$$gain(X, A) = entropy(X) \sum_{v \in values(A)} \frac{|X_v|}{|X|} entropy(X_v) \quad (4)$$

where *values*(A) is the set of all possible values of attribute A, and X_v is the subset of X for which attribute A has the attribute value v, i.e., $X_v = \{x \in X \mid A(x) = v\}$.

Boosting, winnowing and pruning, three methods, be used in the C5.0 tree construction; they propose to build the tree with the right size [26]. They increase the generalization and reduce the over fitting of the decision tree model.

V.DATA SET DISCRIPTION

This paper employed the bank marketing dataset from the University of California at Irvine (UCI) Machine Learning Repository have been used to evaluate the performances of C5.0 decision tree classification model and multilayer perceptron neural network MPLNN. The bank direct marketing dataset used here was collected by S. Moro, R. Laureano and P. Cortez, (2011) at the European Simulation and Modeling Conference [1]. The data is related with direct marketing campaigns of a Portuguese banking institution. 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) were (or not) subscribed. The bank direct marketing dataset contains (45211) number of samples with (17) attributes without missing values [1, 27].

The characteristics of dataset composed in two kinds: nominal and numeral attributes, as shown in Table 1. This table shows that three kinds of attributes; Numerical, which are in range type for all of them like (Age, Balance, Day, Duration, campaign, Pdays, and Previous), Categorical are in set type as the attributes (Job, Marital, Education, Contact, Month, Poutcome), and Binary categorical are all of attributes that represented as yes or no in their classes; for example, the attributes (Default, Housing, Loan, Output).

The column headed Attributes illustration is presenting the number classes for each attribute and the relation with its name. In the second attribute named Job, there exist many kinds of jobs belonging to this attribute as (admin, unknown, unemployed, management, housemaid, entrepreneur, student, blue-collar, self-employed, retired, technician, and services). The attribute Marital can be illustrated in classes as (married, divorced, and single) where the class divorce means divorced or widowed. The Education classes is divided into unknown, secondary, primary, and tertiary; however, in attributes Default, Housing, Loan, and the output attribute have only two classes (yes, and no). The contact communication classes in the Contact attribute are unknown, telephone, and cellular. Clearly, in the attribute Month the classes are month's names Jan, Feb, etc. The attribute Poutcome presents the outcome of the previous marketing campaign like: unknown, other, failure, and success. Last column in table 1 introduces the duration for each range in numerical kind of attributes; for example, Age attribute has (18:95) in duration; that means all ages for customers or samples are ranged between 18 and 95 years, also; the average yearly balance is in between -8019 and 102127.

In the same way, the month's days of course are ranged from 1 to 31, and the last contact duration in seconds in the attribute of Duration is in between 0 to 4918 seconds. The attribute Campaign shows in its domain the number of contacts performed during this campaign, and for this client (includes last contact) is in the interval from 1 to 63; however, the domain ranged between -1 to 871 is representing the number of days that passed by after the client was last contacted from a previous campaign (-1 means client was not previously contacted) in the attribute Pdays. Last but not least, the attribute Previous presents the number of contacts performed before this campaign and for this client, its domain from 0 to 275.

In fact, methods for analyzing and modeling data can be splits into two groups: supervised learning and unsupervised learning. The supervised learning requires input data that has both predictor (independent) attributes and a target (dependent) attribute whose value is to be estimated. In addition, the process learns how to model (predict) the value of the target attribute based on predictor attributes. The famous examples of supervised learning are decision trees, and neural networks. Actually, the supervised learning is suitable for analysis dealing with the prediction of some attribute [7].

On the other hand, unsupervised learning instead of identifying a target (dependent) attribute treats all the attributes equally. In this kind of methods, the goal is searching about patterns, groupings or other ways to distinguish the data, which may lead to understanding of data relations; not to predict the value of an attribute like the previous kind of analyzing method. The examples of unsupervised learning are correlation, statistical measures, and cluster analysis [7].

Table.1. Attributes description

#	Attributes	Kind	Type	Attributes illustration	Domain
1	Age	Numeric	Range		18:95
2	Job	Categorical	Set	('admin.','unknown','unemployed','management','housemaid','entrepreneur', 'student','blue-collar','self-employed','retired','technician','services')	
3	Marital	Categorical	Set	marital status ('married','divorced','single'; note: 'divorced' means divorced or widowed)	
4	Education	Categorical	Set	('unknown','secondary','primary','tertiary')	
5	Default	Binary (Categorical)	Flag	has credit in default? (binary: 'yes','no')	
6	Balance	Numeric	Range	average yearly balance, in euros	-8019: 102127
7	Housing	Binary (Categorical)	Flag	has housing loan? (binary: 'yes','no')	
8	Loan	Binary (Categorical)	Flag	has personal loan? (binary: 'yes','no') # related with the last contact of the current campaign	
9	Contact	Categorical	Set	contact communication type (categorical: 'unknown','telephone','cellular')	
10	Day	Numeric	Range	last contact day of the month	1:31
11	Month	Categorical	Set	last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')	
12	Duration	Numeric	Range	last contact duration, in seconds	0:4918
13	Campaign	Numeric	Range	number of contacts performed during this campaign and for this client (includes last contact)	1:63
14	Pdays	Numeric	Range	number of days that passed by after the client was last contacted from a previous campaign (-1 means client was not previously contacted)	-1:871
15	Previous	Numeric	Range	number of contacts performed before this campaign and for this client	0:275
16	Poutcome	Categorical	Set	outcome of the previous marketing campaign (categorical: 'unknown','other','failure','success')	
17	Output	Binary (Categorical)	Flag	Output variable (desired target): y - has the client subscribed a term deposit? (binary: 'yes','no')	

This paper is using supervised learning of data analysis to reach to the best prediction for the last attribute Y, which is the target. The objective is to examine the performance of MLPNN and C5.0 model on a real-world data of bank deposit subscription and increasing the campaign effectiveness by identifying the main characteristics that affect the success (the deposit subscribed by the client).

Table 2 shows that a classification for all attributes, they are divided into two parts; each one has some of the attributes with ranges (for numerical attributes) and classes (for categorical and binary categorical attributes), and also the percentages for every class or interval in the range of attributes are calculated. By these percentages, the most common age category for the customers in this dataset of the bank is inside the interval from 30 to 40 years in the attribute of Age by 40%, also; the public job in these samples is Blue-collar in the attribute of Job by 22.47%. In the same context, the highest percentage is 60% for married customers in the attribute of Marital, and most of them learnt to the secondary class by ratio 51% in the Education attribute.

The customers have no credit are the majority in the attribute Default by 98%; but whose their average yearly balance is between -8019 and 10,000 Euros take the highest percentage 98.17% in the attribute of Balance. In the attributes of Housing and Loan, which they are telling that the customer whose take housing or personal loans, 56% of customers they subscribe in housing loan and only 16% from them subscribe in personal loan; conversely, 84% they have not subscriptions in personal loan and 44% in housing loan. Cellular contact is the winner in the Contact attribute by 65% for communication with the customer. The last contacts with the customers whose are completed from 18 and 21 days before the contact campaign started to have the highest percentage 15% for the density of communication. In addition, the month May in the attribute Month is the most of the months that it has high ratio 30% with respect to others for the last month contact in the year. In the attribute Duration, which represents the last contact duration in seconds, 73% are contacted through 300 seconds as a maximum for contacting. The number of contacts performed during this campaign, and for these clients (includes last contact) in the attribute, Campaign is concentrated in the interval from 1 to 10 with percentage 96.7%. The Pdays attribute presents the number of days that passed by after the client was last contacted from a previous campaign, and -1 means client was not previously contacted; this value is presented in these samples with 81.7% as a higher percentage in this attribute. The number from 0 to 25 in the attribute Previous, that represents the number of contacts performed before this campaign and for this client, is the highest range in this attribute. Finally yet importantly, the class unknown in the attribute Poutcome, which is the outcome of the previous marketing campaign, is determined as the greatest one with the percentage 82%.

TABLE 2 PART (1) BANK DIRECT MARKETING ATTRIBUTES' VALUES PERCENTAGES

Age		Job		Marital		Education		Default		Balance		Housing		Loan	
Range	**%**	**Class**	**%**	**Class**	**%**	**Class**	**%**	**Class**	**%**	**Range**	**%**	**Class**	**%**	**Class**	**%**
[18, 30)	12%	Admin	11%	Married	60%	Primary	16%	Yes	2%	[-8019, 10,000)	98.17%	Yes	56%	Yes	16%
[30, 40)	40%	Entrepreneur	3%	Single	28%	Secondary	51%	No	98%	[10,000, 20,000)	1.4%	No	44%	No	84%
[40, 50)	25.5%	Blue-collar	22%	Divorced	12%	Tertiary	29%			[20,000, 30,000)	.312%				
[50, 60)	18.6%	Retired	5%			Unknown	4%			[30,000, 40,000)	0.051%				
[60, 70)	2.7%	Technician	17%							[40,000, 50,000)	.022%				
[70, 80)	0.94%	Student	2%							[50,000, 60,000)	.022%				
[80, 90)	0.27%	Management	21%							[60,000, 70,000)	.008%				
[90, 100)	0.02%	Self-employed	3.74%							[70,000, 80,000)	.0022%				
		Services	9%							[80,000, 90,000)	.009%				
		Unknown	0.64%							[90,000, 100,000)	.0022%				
		Housemaid	2.74%							[100,000, 110,000)	.0022%				
		Unemployed	2.88%												

TABLE 2 PART (2) BANK DIRECT MARKETING ATTRIBUTES' VALUES PERCENTAGES

Contact		Day		Month		Duration		Campaign		Pdays		Previous		Poutcome	
Class	**%**	**Range**	**%**	**Class**	**%**	**Range**	**%**	**Range**	**%**	**Range**	**%**	**Range**	**%**	**Class**	**%**
Unknown	29%	[0, 3)	4%	Jan	3%	[0, 300)	73%	[1, 10)	96.7%	-1	81.7%	[0, 25)	99.93%	Unknown	82%
Cellular	65%	[3, 6)	10%	Feb	6%	[300, 600)	19%	[10, 20)	2.6%	[0, 100)	3.1%	[25, 50)	0.06%	Success	3%
Telephone	6%	[6, 9)	12%	Mar	1%	[600, 900)	5%	[20, 30)	0.43%	[100, 200)	6.4%	[50, 75)	0.01%	Failure	11%
		[9, 12)	8%	Apr	6%	[900, 1200)	2%	[30, 40)	0.2%	[200, 300)	3.3%	[75, 100)	0.00%	Other	4%
		[12, 15)	11%	May	30%	[1200, 1500)	1%	[40, 50)	0.04%	[300, 400)	5%	[100, 125)	0.00%		
		[15, 18)	11%	Jun	12%	[1500, 1800)	0%	[50, 60)	0.03%	[400, 500)	0.3%	[125, 150)	0.00%		
		[18, 21)	15%	Jul	15%	[1800, 2100)	0%	[60, 70)	.001%	[500, 600)	0.1%	[150, 175)	0.00%		
		[21, 24)	9%	Aug	14%	[2100, 2400)	0%			[600, 700)	0.02%	[175, 200)	0.00%		
		[24, 27)	5%	Sept	1%	[2400, 2700)	0%			[700, 800)	0.06%	[200, 225)	0.00%		
		[27, 30)	10%	Oct	2%	[2700, 3000)	0%			[800, 900)	0.02%	[225, 250)	0.00%		
		[30, 31]	5%	Nov	9%	[3000, 3300)	0%					[250, 275)	0.00%		
				Des	1%	[3300, 5100)	0%					[275, 300)	0.00%		

VI. THE EXPERIMENTAL RESULTS

The performance of each classification model is evaluated using three statistical measures; classification accuracy, sensitivity and specificity. These measures are defined using true positive (*TP*), true negative (*TN*), false positive (*FP*) and false negative (*FN*). The percentage of Correct/Incorrect classification is the difference between the actual and predicted values of variables. True Positive (*TP*) is the number of correct predictions that an instance is true, or in other words; it is occurring when the positive prediction of the classifier coincided with a positive prediction of target attribute. True Negative (*TN)* is presenting a number of correct predictions that an instance is false, (i.e.) it occurs when both the classifier, and the target attribute suggest the absence of a positive prediction. The False Positive (*FP*) is the number of incorrect predictions that an instance is true. Finally, False Negative (*FN*) is the number of incorrect predictions that an instance is false. Classification accuracy is defined as the ratio of the number of correctly classified cases and is equal to the sum of *TP* and *TN* divided by the total number of cases *N* [7, 28].

$$Accuracy = \frac{TP+TN}{N} \tag{5}$$

Sensitivity refers to the rate of correctly classified positive and is equal to *TP* divided by the sum of *TP* and *FN*. Sensitivity may be referred as a *True Positive Rate.*

$$Sensitivity = \frac{TP}{TP+FN} \tag{6}$$

Specificity refers to the rate of correctly classified negative and is equal to the ratio of *TN* to the sum of *TN* and *FP*. *False-Positive Rate* equals (100-*specificity)* [28].

$$Specificity = \frac{TN}{TN+FP} \tag{7}$$

Figure 3 shows the component nodes of the proposed stream. The stream is implemented in SPSS Clementine data mining workbench using Intel ® core ™ 2 Duos, CPU with 1.83 GHz. Clementine uses client/server architecture to distribute requests for resource-intensive operations to powerful server software, resulting in faster performance on larger datasets [2]. The software offers many modeling techniques, such as prediction, classification, segmentation, and association detection algorithms.

Bank direct marketing dataset node is connected directly to EXCEL sheet file that contains the source data. The dataset was explored as ordinal data types.

Type node specifies the field metadata and properties that are important for modeling and other work in Clementine. These properties include specifying a usage type, setting options for handling missing values (the used dataset in this paper has not been missing values to handle), as well as setting the role of an attribute for modeling purposes; input or output. As previously stated, the first 16 attributes in table 1 are defined as input (predictive) attributes and the output attribute (y) is defined as a target.

MLPNN classifier node is trained using the pruning method. It begins with a large network and removes the weakest neurons in the hidden and input layers as training proceeds. To prevent overtraining, the portion of the training subset has been used as a validation subset. The network is trained on the rest of training subset, and accuracy is estimated based on the validation subset. The stopping criterion is set based on time; the network is trained for one minute.

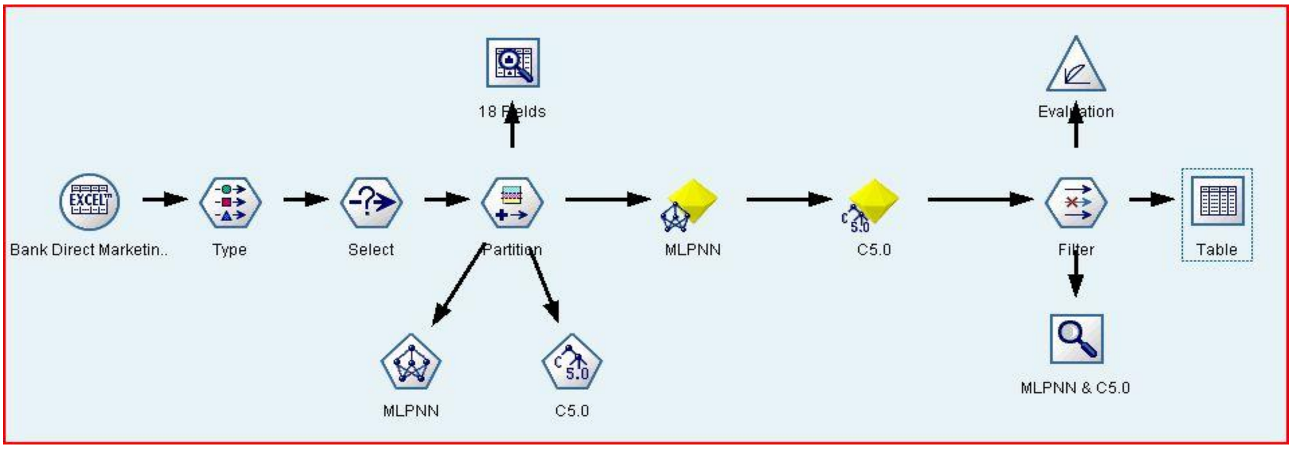

Figure.3. Data mining stream for the prediction of the bank deposit subscription using MLPNN and C5.0 respectively.

However, the training process may be interrupted at any point to save the network model with the best accuracy achieved so far. By using the direct-marketing campaigns of a Portuguese banking institution dataset, the resulting structure consists of four layers; one input with 48 neurons, two hidden layer: first hidden layer has 20 neurons and the second hidden layer has 15 neurons, and the output layer with one neuron. The prediction accuracies of training and test samples are 90.84% and 90.32% respectively. *C5.0 node* is trained and tested using a simple model with the partitioned data. The minimum number of samples per node is set to be 2 and the decision tree with 12 in depth. It will examine the importance rate of the predictors before starting to build the model by winnow attribute's option. Predictors that are found to be irrelevant are then removed from the model-building process [2]. The prediction accuracies of training and test samples are 93.23% and 90.09% respectively.

Filter, Analysis and Evaluation nodes are used to select and rename the classifier outputs in order to compute the performance statistical measures and to graph the evaluation charts.

Table 3 shows the numerical illustration for the importance of the attributes with respect to models MLPNN and C5.0. The table illustrates that the attribute Duration is the most important for the two examined models. In MLPNN, the ratio is 0.326 and in C5.0 is 0.722 that they are highest ratios among all the other attributes. The attribute Default is removed by C5.0 because it is trivial or not has any degree of importance with respect this model; however, some attributes are not removed, and they measured by zero because their importance is very low and rounded to zero. These all ratios are illustrated in a graph in Figure 4.

TABLE 3 THE IMPORTANCE OF ATTRIBUTES RELATED TO THE MLPNN AND C5.0

Models	**Importance**															
Attributes	**Age**	**Job**	**Marital**	**Education**	**Default**	**Balance**	**Housing**	**Loan**	**Contact**	**Day**	**Month**	**Duration**	**Campaign**	**Pdays**	**Previous**	**Poutcome**
MLPNN	0.0223	0.0504	0.021	0.0247	0.0062	0.0083	0.0226	0.0174	0.0555	0.0478	0.1762	0.326	0.0198	0.0376	0.0142	0.15
C5.0	0	0.0186	0.0301	0	-----	0.0237	0	0.0152	0	0	0	0.722	0.0206	0.0095	0.0072	0.153

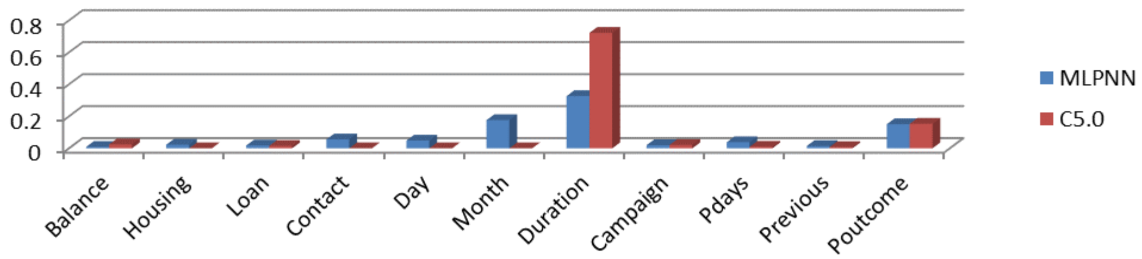

Figure.4. the most importance of attributes based on MLPNN and C5.0.

Figure 5 shows the cumulative charts of the two models for training and test subsets. The higher lines indicate better models, especially on the left side of the chart. The two curves are identical for test subset and almost identical with the training one. This figure shows that MLPNN line is the best in the training subsets in some positions; however, in the testing subset, the success is observed in C5.0. Actually, all of them coincide together to the end of the curve.

The predictions of all models are compared to the original classes to identify the values of true positives, true negatives, false positives and false negative. These values have been computed to construct the confusion matrix as tabulated in Table 4 where each cell contains the raw number of cases classified for the corresponding combination of desired and actual classifier outputs.

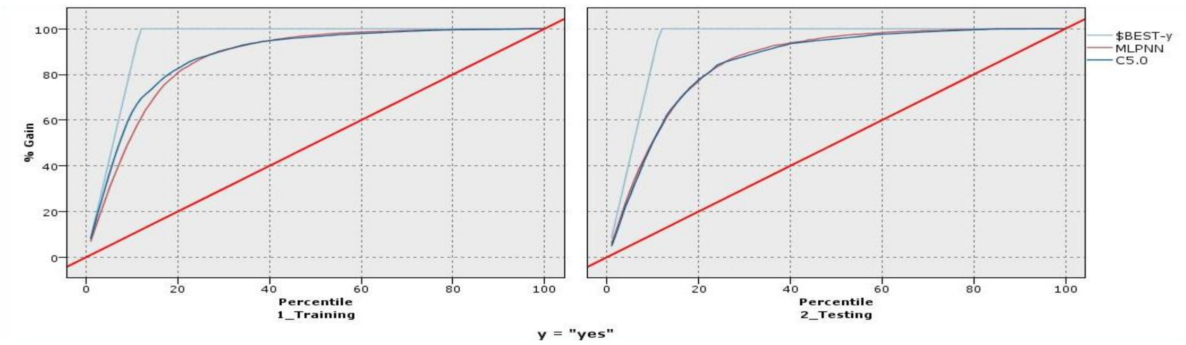

Figure.5. the cumulative gains charts of the four models for training and test subsets.

The values of the statistical parameters (sensitivity, specificity and total classification accuracy) of the four models were computed and presented in Table 5. Accuracy, Sensitivity and Specificity approximate the probability of the positive and negative labels being true. They assess the usefulness of the algorithm on a single model. By using the results shown in Table 5, it can be seen that the sensitivity, specificity and classification accuracy of all models has achieved 94.92% success of test samples.

TABLE 4 THE CONFUSION MATRICES OF MLPNN and C5.0 MODELS FOR TRAINING AND TESTING SUBSETS

Model	**Training Data**			**Testing Data**		
	Desired output	**Yes**	**No**	**Desired output**	**Yes**	**No**
MLPNN	**Yes**	TP=1,913	FP=1,799	**Yes**	TP=769	FP=808
	No	FN=1,093	TN=26,784	**No**	FN=510	TN=11,535
C5.0	**Yes**	2,258	1,454	**Yes**	740	837
	No	684	27,193	**No**	513	11,532

TABLE 5 PERCENTAGES OF THE STATISTICAL MEASURES OF MLPNN AND C5.0 FOR TRAINING AND TESTING SUBSETS

Model	**Partition**	**Accuracy**	**Sensitivity**	**Specificity**
MLPNN	**Training**	90.84%	63.64%	93.71%
	Testing	90.32%	60.12%	93.45%
C5.0	**Training**	93.23%	76.75%	94.92%
	Testing	90.09%	59.06%	93.23%

In this table, for MLPNN model the accuracy is 90.84% of training samples and 90.23% of testing samples. However, the classification accuracy of the C5.0 model is 93.23% of training samples and 90.09% of testing samples. Furthermore, the sensitivity analysis for it is 63.64% of training samples, and 60.12% of testing samples; but the specificity analysis percentages are 93.71% for training samples, and 93.45% for testing samples. On the other hand, the classification accuracy of C5.0 is 93.23% of training samples, and 90.09% of testing samples. Furthermore, the sensitivity analysis of C5.0 model is 76.75 % of training samples, and 59.067% for testing samples. For the sensitivity, analysis is 94.92% and 93.23% of testing samples.

Therefore, C5.0 is the best in accuracy analysis for training by 93.23%; however, in testing samples, MLPNN is better than it by 90.84% is. Furthermore, the same thing for sensitivity analysis; the training samples have 76.75% for C5.0 and in testing samples of MLPNN outperforming by 60.12%. Finally yet importantly, the best percentage 94.92% is achieved in specificity analysis by C5.0 in training samples; however, in testing samples the percentage 93.45% is achieved by MLPNN. From the previous, C5.0 is the best in accuracy, sensitivity, and specificity analysis for training samples; but the MLPNN is the best for accuracy, sensitivity, and specificity analysis for testing samples. Figure 6 shows that the comparison of accuracy, sensitivity, and specificity with respect to the two models MLPNN and C5.0.

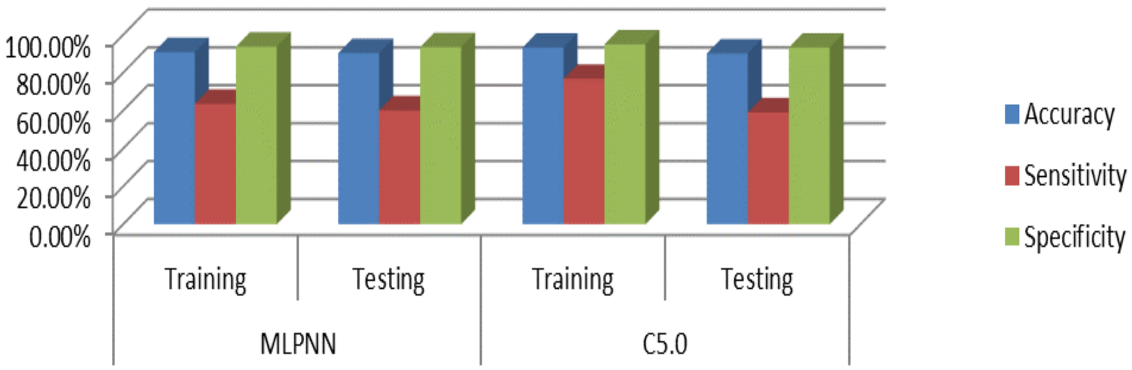

Figure.6. the comparison between MLPNN and C5.0 in accuracy, sensitivity, specificity

VII. CONCLUSION

Bank direct marketing and business decisions are more important than ever for preserve the relationship with the better customer. To success and survival, the business there is a need for customer care and marketing strategies. Data mining and predictive analytics can provide help in such marketing strategies. Its applications are Influential in almost every field containing complex data and large procedures. It has proven the ability to reduce the number of false positives and false-negative decisions.

Additionally, Multilayer perceptron neural network (MLPNN) is a mathematical model based on biological neural networks; it is a simulation of a biological neural system. MLPNN is one of the famous data mining techniques for classification. It is developed and organized to be able for dealing with large data.

This paper has been evaluating and comparing the classification performance of the two different data mining techniques' models MPLNN and C5.0 on the bank direct marketing dataset to classify for bank deposit subscription. The purpose is increasing the campaign effectiveness by identifying the main characteristics that affect the success (the deposit subscribed by the client). The classification performances of the two models have been using three statistical measures; Classification accuracy, sensitivity and specificity. This dataset has partitioned into training and test by the ratio 70%:30% respectively. Experimental results show the effectiveness of both models. However, C5.0 has achieved slightly better performance than MLPNN. Importance analysis has shown that attribute "Duration" in both models has achieved the most important attribute.

REFERENCES

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

[2] http://www-01.ibm.com/software/analytics/spss (last accessed: November 2012).

[3] Ou, C., Liu, C., Huang, J. and Zhong, N. 'One Data mining for direct marketing', *Springer-Verlag Berlin Heidelberg*, pp. 491–498., 2003.

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

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

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

[7] 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.

[8] Adem Karahoca, Dilek Karahoca and Mert Şanver, "Data Mining Applications in Engineering and Medicine", ISBN 978-953-51-0720-0, InTech, August 8, 2012".

[9] http://en.wikipedia.org/wiki/Direct_marketing.Feb, 2011

[10] http://en.wikipedia.org/wiki/Neural_network#History_of_the_neural network analogy. Oct 2010.

[11] TIAN YuBo, ZHANG XiaoQiu, and ZHU RenJie. "Design of Waveguide Matched Load Based on Multilayer Perceptron Neural Network". Proceedings of ISAP, Niigata, Japan 2007.

[12] Usama M. Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From Data Mining to knowledge Discovery: An Overview, Advances in Knowledge Discovery and Data Mining, AAAI Press, pp 1-34. 1996.

[13] Han, J., & Kamber, M. "Data mining: Concepts and Techniques". San Francisco: Morgan Kaufmann Publishers., 2001.

[14] Ho, T.B. (nd). Knowledge Discovery and Data Mining Techniques and Practice Available on. http://www.netnam.vn/unescocourse/knowlegde/knowfrm.htm 2006.

[15] T. Munkata, "Fundamentals of new artificial intelligence," 2nd edition, London, Springer-Verlag, 2008.

[16] B. Chaudhuri and U. Bhattacharya. Efficient training and improved performance of multilayer perceptron in pattern classification. Neuro computing, 34, 11–27, 2000.

[17] G. Thimm, E. Furuhashi and Takeshi. "Neural network pruning and pruning parameters". Proceeding of the 1st Workshop on Soft Computing, 1996.

[18] Tetko IV. "Efficient partition of learning data sets for neural network training". Neural Networks, 10, 1361-1374, 1997.

[19] L. Ma and K. Khorasani. "New training strategy for constructive neural networks with application to regression problems". Neural Networks, 17,589-609, 2004.

[20] Thimm, G. and E. Fiesler, Furuhashi and Takeshi. "Neural Network Pruning and Pruning Parameters". In proceeding of the 1st Workshop on Soft Computing, 1996.

[21] J. W. Han and M. Kamber. Data mining concepts and techniques, The 2nd edition, Morgan Kaufmann Publishers, San Francisco, CA, 2006.

[22] R. Nisbet, J. Elder and G. Miner. Handbook of statistical analysis and data mining applications. Academic Press, Burlington, MA, 2009.

[23] A. Floares., A. Birlutiu. "Decision Tree Models for Developing Molecular Classifiers for Cancer Diagnosis". WCCI 2012 IEEE World Congress on Computational Intelligence June, 10-15, 2012 - Brisbane, Australia.

[24] Derrig, Richard A., and Louise A. Francis, "Distinguishing the Forest from the TREES: A Comparison of Tree-Based Data Mining Methods," Variance 2:2, 2008, pp. 184-208.

[25] Su-lin PANG, Ji-zhang GONG, C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks, Systems Engineering - Theory & Practice, Volume 29, Issue 12, Pages 94–104, December 2009.

[26] M. Karabatak and M. Cevdet. An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications 36: 3465–3469, 2009.

[27] S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimarães, Portugal, October, 2011.

[28] K. Hornik, Stinchcombe and H. White. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward network. Neural Networks, 3, 359-66, 1990.

Hany A. Elsalamony received a Ph.D degree in mathematics from Helwan University in 2006. Currently, He works as a lecturer of computer science and mathematics at Helwan University, Cairo, Egypt. Before working in different universities and institutes in the Arab region, he has collected different experiences in information technology by working in many concerned projects. Dr. Hany published and reviewed several papers; all are focusing in image processing, and data mining. His current research interests cover the theory of data mining algorithms and their applications to image understanding and signal processing.

Alaa M. Elsayad received a PhD degree in image processing from Cairo University in 1998. He has completed his postdoctoral research studying the telemedicine system and image analysis in the Imaging Sciences and Information System (ISIS) unit at the Medical Center of Georgetown University. Currently, He works as an associate professor of computer science at the Electronics Research Institute, Cairo, Egypt where he heads the information and decision support center. Before working in different universities in the Arab region, he has collected different experiences in information technology while working in different governmental organizations in Egypt, which include Ministry of Administrative Development, Ministry of Communication and Ministry of International Cooperation. Dr. Alaa published several papers; all are focusing in image analysis and data mining. His current research interests cover the theory of data mining algorithms and their applications to image understanding and signal processing. Special research topics include medical image analysis and telemedicine systems.