Bank Marketing Analysis

T. Dhikhi [1], Vikrant Kumar [2]
[1] Assistant Professor (O.G.), Department of Computer Science Engineering, SRM Institute of Science & Technology, Chennai, Tamil Nadu, India.
[2] Department of Computer Science Engineering, SRM Institute of Science & Technology, Chennai, Tamil Nadu, India.

Abstract – All bank advertising efforts are subject to clients' colossal electronic information. The measure of these information sources is unthinkable for a human examiner to concoct fascinating data that will help in the basic leadership process. Information mining models are totally helping in the execution of these crusades. This paper presents examination and utilizations of the most essential procedures in information mining; multilayer recognition neural system (MLPNN), tree expanded Naïve Bayes (TAN) known as Bayesian systems, Nominal relapse or calculated relapse (LR), and Ross Quinlan new choice tree demonstrate (C5.0). The goal is to analyze the execution of MLPNN, TAN, LR and C5.0 strategies on a genuine information of bank store membership. The design is expanding the crusade adequacy by recognizing the primary attributes that influence a win (the store bought in by the customer) in light of MLPNN, TAN, LR and C5.0. The exploratory outcomes illustrate, with higher correctnesses, the accomplishment of these models in foreseeing the best crusade contact with the customers for buying in store. The exhibitions are computed by three factual measures; characterization precision, affectability, and specificity.

Index Terms – Bank Marketing, Naïve Bayes, Nominal Regression, Neural Network, C5.0.

1. INTRODUCTION

In banks, enormous information records data about their clients. This information can be utilized to make and keep clear relationship and association with the clients so as to target them independently for unmistakable items or saving money offers. Typically, the chose clients are reached straightforwardly through: individual contact, phone cell, mail, and email or some other contacts to publicize the new item/sIn banks, gigantic information records data about their clients. This information can be utilized to make and keep clear relationship and association with the clients so as to target them exclusively for positive items or saving money offers. As a rule, the chose clients are reached specifically through: individual contact, phone cell, mail, and email or some other contacts to promote the new item/administration or give an offer, this sort of advertising is called coordinate showcasing. Truth be told, coordinate showcasing is in the primary a system of a significant number of the banks and insurance agencies for cooperating with their clients.

Truly, the name and distinguishing proof of the term coordinate promoting recommended first time in 1967 by Lester Wunderman, which he is thought to be the dad of direct showcasing. Also, a portion of the banks and monetary administrations organizations may depend just on procedure of mass showcasing for advancing another administration or item to their clients. In this system, a solitary correspondence message is communicated to all clients through media, for example, TV, radio or publicizing firm, and so forth In this methodology, organizations don't set up an immediate relationship to their clients for new-item offers. Actually, a significant number of the clients are not intriguing or react to this sort of offers advancement.

Appropriately, banks, money related administrations organizations and different organizations are moving far from mass advertising technique since its insufficiency, and they are currently focusing on a large portion of their clients by direct showcasing for particular item and administration offers. Because of the positive outcomes plainly estimated; numerous advertisers appealing to the immediate promoting. For instance, if an advertiser conveys 1,000 offers via mail and 100 react to the advancement, the advertiser can state with certainty that the battle drove quickly to 10% direct reactions. This metric is known as the 'Reaction Rate', and it is one of many clear quantifiable achievement measurements utilized by direct advertisers. In divergence, general promoting utilizes roundabout estimations, for example, mindfulness or commitment, since there is no immediate reaction from a shopper. From the writing, the immediate advertising is turning into a vital application in information mining nowadays. The information mining has been utilized generally in direct showcasing to recognize imminent clients for new items, by utilizing acquiring information, a prescient model to quantify that a client will react to the advancement or an offer.

Information digging has picked up fame for illustrative and prescient applications in managing an account forms. Four systems will apply to the informational index on the bank coordinate showcasing. The Multilayer recognition neural system (MLPNN) is one of these procedures, which have their underlying foundations in the man-made reasoning. MLPNN is a commonly subordinate gathering of counterfeit neurons that applying a scientific or computational model for data preparing utilizing an associated way to deal with calculation (Freeman Another strategy of information mining is the choice tree

approach. Choice tree gives ground-breaking strategies to arrangement and forecast. There are numerous calculations to fabricate a choice tree demonstrate. It can produce reasonable standards, and to deal with both nonstop and straight out factors . One of the popular ongoing procedures of the choice tree is C5.0, which will be connected in this paper.

A guileless Bayes classifier (TAN) is a simple and straightforward probabilistic classifier in view of applying Bayes' hypothesis with solid (gullible) autonomy presumptions. It can foresee class participation probabilities, for example, the likelihood that a given example has a place with a specific class. The supposition is called class contingent autonomy. It is made to streamline the calculation included and, in this sense, is considered "credulous".

The fourth system will utilize is Logistic relapse investigation (LR). Cornfield was the first to utilize strategic relapse in the mid 1960s and with the wide accessibility of refined measurable programming for rapid PCs; the utilization of calculated relapse is expanding. LR ponders the relationship between a downright reliant and an arrangement of free (expressive) fields. The name calculated relapse is frequently utilized when the needy variable has just two qualities. The name various gathering calculated relapse (MGLR) is normally held for the situation when the needy variable has at least three extraordinary qualities. Various gathering calculated relapse is now and again called multinomial, polytomous, polychotomous, or ostensible strategic relapse ervice or give an offer, this sort of advertising is called coordinate showcasing. Actually, coordinate showcasing is in the fundamental a procedure of a large number of the banks and insurance agencies for communicating with their clients.

Table 1: ATTRIBUTES

Attributes	Age	Job	Marital	Education	Defult	Balance	Housing	Loan	Contact	Day	Month	Duration	Campagn	Pdays	Previous	Poutcome	Outnut
Kind	Numeric	Categorical	Categorical	Categorical	Binary	Numeric	Binary	Binary	Categorical	Numeric	Categorical	Numeric	Numeric	Numeric	Numeric	Categorical	Binary

2. DATA MINING OVERVIEW

The definition of Data Mining or Knowledge Discovery in Databases is the action that extracts some new important information contained in large databases. The objective of information mining is to discover sudden attributes, shrouded highlights or other hazy connections in the information in light of systems' blend. Today, numerous applications in a wide and different scopes of business established and worked in this direction.

In 1996, U. Fayyad, G. Shapiro characterized the general learning revelation process as an intelligent and iterative process including pretty much the accompanying advances: understanding the application field, information choosing, preprocessing and cleaning information, mix of information, information decrease and change, choosing calculations of information mining, translation and portrayal of the outcomes and utilizing the found information . Truth be told, the information mining can be arranged into two classes graphic and prescient.

In reality, in the ongoing years, information mining is involving incredible position of consideration zone in the general public of business or managing an account since its flexibility in working with a lot of information, and transforming such information into clear data and learning. A large portion of the general population might be befuddled in comprehension between the expressions "information disclosure" and "information mining" in various territories. Information disclosure in databases is the way toward distinguishing substantial, novel, likely valuable, lastly reasonable examples/models with information. Then again, information mining is a stage in the learning revelation process comprising of specific information mining calculations that under some satisfactory computational effectiveness constraints, discovers examples or models in information.

3. MULTILAYER PERCEPTION NEURAL NETWORK

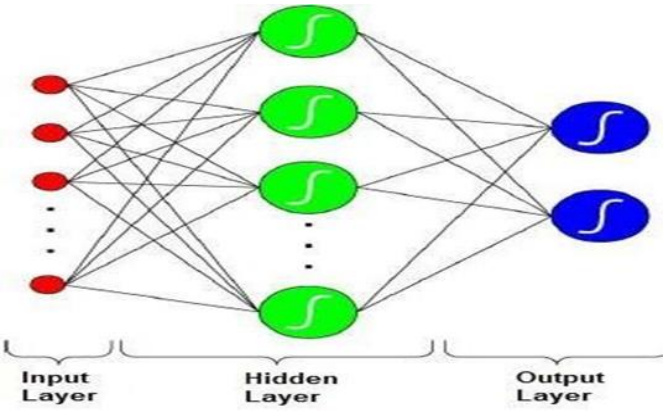
Multilayer perception neural network (MLPNN) with back-propagation is the most popular artificial neural network architecture . The MLPNN is known to be a powerful function approximation for prediction and classification problems. Historically, this direction field started when neurophysiologist Warren 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. They named this model ‘threshold logic’. The model opened the door on the way for research into 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.

The most interesting in the field was renewed in 1982. John Hopfield introduced an approach to construct more useful machines 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 now, called 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 are applied to data mining in Craven and Sahvlik (1997).

Figure 1 shows that the MLPNN structure is 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 xi after multiplying them by the strengths of the respective connection weights wij and computes its output yj using activation function (AF) of this sum. AF may range from a simple threshold function, or a sigmoid, hyperbolic tangent, or radial basis function.

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



Success contact Unsuccessful contact

Fig 1. The structure of multilayer perceptron neural Network

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. 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.

Finally, MLPNN has many advantages, such as the 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 the past few years. On the other hand, there are some disadvantages like: the neural network requires high-quality data; variables must be carefully selected a priori, the risk of over- fitting, and requires a definition of architecture.

1. NAÏVE BAYES CLASSIFIER

Naïve Bayes (TAN) is one of the most effective and efficient classification algorithms. It is a one special case of a Bayesian network. The structure and parameters of the unconstrained Bayesian network would appear to be a logical means of improvement. However, (TAN) was found by Friedman (1997) as an easily outperforms such an unconstrained Bayesian network classifier on a huge sample of benchmark data sets. Bayesian classifiers are helpful in predicting the probability that a sample belongs to a particular class or grouping. This technique is useful for large databases because it is highly accurate and quickly in classification and fast to train with simple models and intuitive. It requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification, handles real and discrete data.

amount of training data to estimate the parameters (means and variances of the variables) necessary for classificatIn addition, to improve the conditional independence assumption there exists one way, it is to enlarge the structure of naïve Bayes to represent explicitly attribute dependencies by adding arcs between attributes.

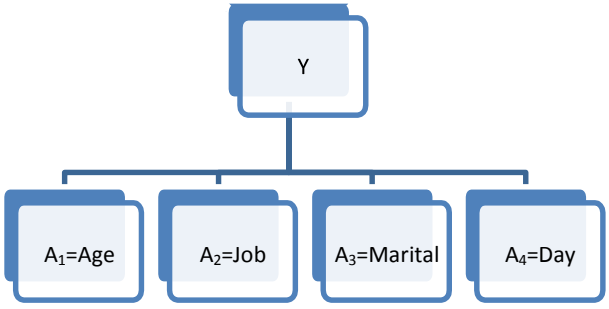


Fig 2. An example of naïve Bayes

Tree augmented naïve Bayes (TAN) is an extended tree such that a class node directly walks to all attribute nodes, also an attribute node can have only one parent from another attribute node. . Figure 3 shows an example of TAN from the direct bank marketing database.

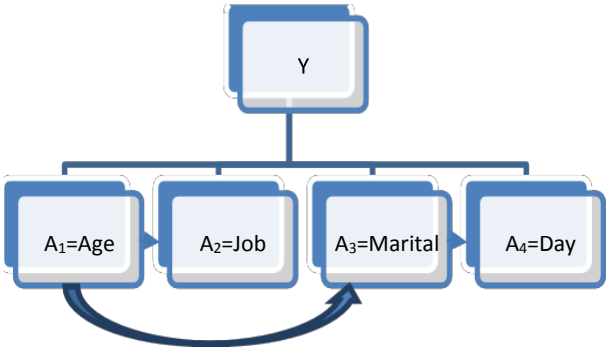


Fig 3. An example of TAN

2. LOGISTIC REGRESSION ANALYSIS

The logistic regression (LR) model is very suitable for addressing issues of many kinds of data sets; it is provided sufficiently several and well-distributed samples. In addition, it is well suited for describing and testing hypotheses about relationships between a categorical outcome variable and one or more categories or continuous predictor attributes.

Furthermore, LR uses maximum probability estimation rather than the least squares estimation used in traditional multiple regression. Starting values of the predicted parameters are used and the probability that the sample came from a population with those parameters is computed. The values of the estimated parameters are adjusted iteratively until the greatest probability value of them is obtained. That is, maximum probability approaches try to find estimates of parameters that make the data observed "most likely"

3. DECISION TREE MODEL TECHNIQUE

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.

Decision tree algorithm partitions the data samples into two or more 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 (in binary decision tree) are then split again, and the process repeats until the homogeneity criterion is reached or until some other stopping, criterion is satisfied.

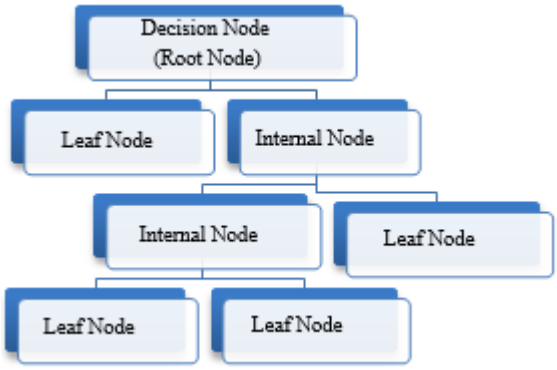


Fig 4. Illustrated example of a binary decision tree

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 to split, testing the values of some functions of data attributes. Each branch of the decision node corresponds to a different outcome of the test as in Figure 4. Historically, the book by Bremen 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 the procedures are relatively straightforward to understand and explain, and the procedures address a number of data complexities, such as nonlinearly and interactions, that commonly occur in real data.

The popular run of the mill in choice trees is C5.0, which is an as of late designed displaying calculation, and it is an enhanced adaptation of C4.5 and ID3 calculations. C5.0 is a business item outlined by Rule Quest Research Ltd Pty to investigate tremendous informational collections and is executed in SPSS Clementine workbench information mining programming.

The tree of C5.0 utilizes regular part calculations incorporates entropy in light of data gain. The model works by part the example in light of the property that gives the greatest data gain. Each sub test characterized by the main split is then part once more, generally in light of an alternate quality, and the procedure rehashes until the subsamples can't be part any further. At long last, the low-level parts are rethought, and those that don't contribute essentially to the estimation of the model are evacuated or pruned.

C5.0 demonstrate is very vigorous within the sight of issues, for example, missing information and expansive quantities of info fields. It for the most part does not require long preparing occasions to appraise. What's more, C5.0 models have a tendency to be less demanding to comprehend than some other model composes, since the tenets got from the model have an extremely direct elucidation. Besides, C5.0 offers the intense boosting strategy to build precision of arrangement. C5.0 utilizes entropy as a proportion of immaculateness, which depends on a data gain.

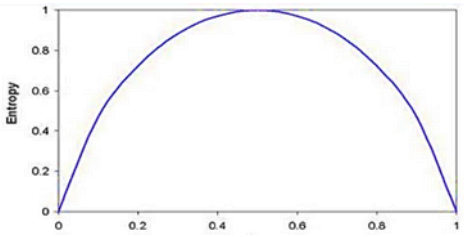


Fig 5: The entropy function relative to a binary classification, as the proportion of yes

4. DATA SET DISCRIPTION

This paper employed the bank direct marketing data set from the University of California at Irvine (UCI) Machine Learning Repository have been used to evaluate the performances of the multilayer perception neural network (MPLNN), Naïve Bayes (TAN), logistic regression (LR), and C5.0 decision tree classification model. The bank direct marketing data set used here was collected by S. Moro, R. Laureano and P. Cortez. The data is related to direct marketing campaigns of a Portuguese banking institution.

#	Attributes	Kind	Type	Attributes illustration	Domain
1	Age	Numeric	Range	NaN	18:95
2	Job	Categorical	Set	('admin.','unknown','unemployed','management','housemaid','entrepreneur','student','blue-collar','self-employed','retired','technician','services')	NaN
3	Marital	Categorical	Set	marital status ('married','divorced','single'; note: 'divorced' means divorced or widowed)	NaN
4	Education	Categorical	Set	('unknown','secondary','primary','tertiary')	NaN
5	Default	Binary (Categorical)	Flag	has credit in default? (binary: 'yes','no')	NaN
6	Balance	Numeric	Range	average yearly balance, in euros	-8019: 102127
7	Housing	Binary (Categorical)	Flag	has housing loan? (binary: 'yes','no')	NaN
8	Loan	Binary (Categorical)	Flag	has personal loan? (binary: 'yes','no') # related with the last contact of the current campaign	NaN
9	Contact	Categorical	Set	contact communication type (categorical: 'unknown','telephone','cellular')	NaN
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')	NaN
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	Categoricasl	Set	outcome of the previous marketing campaign (categorical:'unknown','other','failure','success')	NaN
17	Output	Binary (Categorical)	Flag	Output variable (desired target):y-has the client subscribed a term deposit? (binary: 'yes','no')	NaN

The marketing campaigns were based on phone calls. Often, more than one contact with the same client was required, in order to access if the product (bank term deposit) were (or not) subscribed. The bank direct marketing data set contains (45211) number of samples with attributes without missing values.

The characteristics of data set composed of two kinds: nominal and numeral attributes, as shown in Table 2. 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 categories are all the attributes that represented as yes or no in their classes; for example, the attributes (Default, Housing, Loan, Output).

5. EXPERIMENTAL RESULTS

It is 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 suggests 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. Table 4 shows the confusion matrix for a two-class classifier.

TABLE OF CONFUSION MATRIX

		Predicted	
		Positive (yes)	Negative (no)
Actual	Positive (yes)	TP	FP
	Negative (no)	TN	FN

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.

6. CONCLUSION

Bank direct marketing and business decisions are more important than ever for preserving the relationship with the best 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. This paper has been evaluating and comparing the classification performance of four different data mining techniques' models MPLNN, TAN, LR and C5.0 on the bank direct marketing data set 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 four models have been using three statistical measures; Classification accuracy, sensitivity and specificity. This data set has partitioned into training and test by the ratio 70% and 30%, respectively. Experimental results have shown the effectiveness of models. C5.0 has achieved slightly better performance than MLPNN, LR and TAN. Importance analysis has shown that attribute "Duration" in C5.0, LR, and MLPNN models have achieved the most important attribute; however, the attribute Age is the only assessed as more important than the other attributes by TAN.

REFERENCES

[1] 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.
[2] Adem Karahoca, Dilek Karahoca and Mert Şanver, "Data Mining Applications in Engineering and Medicine", ISBN 978-953-51-0720-0, In Tech, August 8, 2012.
[3] B. Chaudhuri and U. Bhattacharya." Efficient training and improved performance of multilayer perceptron in pattern classification". Neuro computing, 34, pp11–27, September 2000
[4] D.G. Kleinbaum and M. Klein, Logistic Regression, Statistics for Biology and Health, DOI 10.1007/978-1- 4419-1742-31, Springer Science Business Media, LLC 2010
[5] 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.