Ensemble Methods for Classification of Direct Marketing

M. Govindarajan
Assistant Professor,
Department of Computer Science and Engineering,
Annamalai University, Annamalai Nagar – 608002,
Tamil Nadu, India.
e-mail: govind_aucse@yahoo.com

Abstract

One of the major developments in machine learning in the past decade is the ensemble method, which finds highly accurate classifier by combining many moderately accurate component classifiers. In this research work, new ensemble classification methods are proposed using both homogeneous ensemble classifiers with bagging and heterogeneous ensemble classifiers with arcing. Classifier ensembles are designed using Radial Basis Function (RBF) and Support Vector Machine (SVM), k-Nearest Neighbor (k-NN) and Multilayer Perceptron (MLP) as base classifiers. The feasibility and the benefits of the proposed approaches are demonstrated by the means of standard data sets of direct marketing. A wide range of comparative experiments are conducted for standard data sets of direct marketing and the classification performance of the proposed homogeneous and heterogeneous ensemble classifiers are evaluated in terms of accuracy. The objective is to examine the performance of proposed ensemble methods on a real-world data of bank deposit subscription and labor relations data. The purpose is increasing the campaign effectiveness by identifying the main characteristics that affect a success. The experimental results demonstrate, with higher accuracies, the success of these models in predicting the best campaign contact with the clients for subscribing deposit and also labor resources management practices.

Keywords: Data Mining, Ensemble, Radial Basis Function, Support Vector Machine, Accuracy, k-Nearest Neighbor, Multilayer Perceptron.

1. Introduction

Data mining methods may be distinguished by either supervised or unsupervised learning methods. One of the most active areas of research in supervised learning has been to study methods for constructing good ensembles of classifiers. It has been observed that when certain classifiers are ensembled, the performance is phenomenal compared to the performance of the individual classifiers. Recently, advances in knowledge extraction techniques have made it possible to transform various kinds of raw data into high level knowledge. However, the classification results of these techniques are affected by the limitations associated with individual techniques. Hence, hybrid approach is widely recognized by the data mining research community. Hybrid models have been suggested to overcome the defects of using a single supervised learning method. Hybrid models combine different methods to improve classification accuracy. The term combined model is usually used to refer to a concept similar to a hybrid model. Combined models apply the same algorithm repeatedly through partitioning and weighting of a training data set. Combined models also have been called Ensembles. Ensemble improves classification performance by the combined use of two effects: reduction of errors due to bias and variance (Haykin, 1999).

1.2 Direct Marketing

In general, businesses worldwide use mass marketing as their marketing strategy for offering and promoting a new product or service to their customers. The idea of mass marketing is to broadcast a single communication message to all customers so that maximum exposure is ensured. However; since this approach neglects the difference among customers it has several drawbacks. In fact a single product offering cannot fully satisfy different needs of all customers in a market and unsatisfied customers with unsatisfied needs expose businesses to challenges by competitors who are able to identify and fulfill the diverse needs of their customers more accurately. Thus in today's world where mass marketing has become less effective, businesses choose other approaches such as direct marketing as their main marketing strategy (Hossein Javaheri 2007).

Direct marketing is concerned with identifying which customers are more likely to respond to specific promotional offers. A response model predicts the probability that a customer is responsive/non-responsive to an offer for a product or service. A response modeling is usually the first type of target modeling that a business develops as its marketing strategy. If no marketing promotion has been done in the past, a response model can make the marketing campaign more efficient and might bring in more profit to the company by reducing mail expenses and absorbing more customers (Parr Rud 2001). Response model can be formulated in to a binary classification problem in which customers are divided in to two groups of respondents and non-respondents. In direct marketing a desirable response model should contain more respondents and fewer non-respondents. By doing so, one can significantly reduce the overall marketing cost without sacrificing opportunities (Shin 2006). In direct marketing, data mining has been extensively used to identify potential customers for new products. Various classification methods have been used for response modeling such as statistically and machine learning techniques: Neural Networks, Decision trees and Support Vector Machine.

The main purpose of this paper is to apply homogeneous and heterogeneous ensemble classifiers for standard datasets of direct marketing problem to improve classification accuracy. As a result they can improve Return on Investment (ROI), also improve customer relationships and retention (Shin and Cho, 2006). Organization of this paper is as follows. Section 2 describes the related work. Section 3 presents proposed methodology and Section 4 explains the performance evaluation measures. Section 5 focuses on the experimental results and discussion. Finally, results are summarized and concluded in section 6.

2. Related work

Various data mining techniques have been used to model customer response to catalogue advertising. Traditionally statistical methods such as discriminant analysis, least squares and logistic regression have been applied to response modeling.

Bounds and Ross showed that neural networks could improve the response rate from 2% up to 95% (Bounds 1997). Suh, Noh, and Suh (1999) and Zahavi and Levin (1997a, 1997b) found that neural network did not outperform other statistical methods. They suggested combining the neural network response model and the statistical method. On the other hand, Bentz and Merunkay (2000) reported that neural networks outperformed multinomial logistic regression. Bentz and Merunkay also showed that neural networks did better than multinomial logistic regression (Bentz 2000). Viaene et al have also used neural networks to select input variables in response modeling (Viaene, Baesens et al. 2001).

Potharst, Kaymak, and Pijls (2001) applied neural networks to direct mailing campaigns of a large Dutch charity organization. According to their results, the performance of neural networks surpassed that of CHAID or logistic regression. Ha, Cho, and MacLachlan (2005) proposed a response model using bagging neural networks. The experiments over a publicly available DMEF4 dataset showed that bagging neural networks give more improved and stabilized prediction accuracies than single neural networks and logistic regression.

Cheung, Kwok, Law, and Tsui (2003) used SVM for content-based recommender systems. The system is definitely a form of direct marketing that has emerged by virtue of recent advances in the World Wide Web, e-business, and on-line companies. They compared Naive Bayes, C4.5 and 1-nearest neighbor rule with SVM. The SVM yielded the best results among them.

Shin and Cho applied Support Vector Machine (SVM) to response modeling. In their study, they introduced practical difficulties such as large training data and class imbalance problem when applying SVM to response modeling. They proposed a neighborhood property based pattern selection algorithm (NPPS) that reduces the training set without accuracy loss. For the other remaining problem they employed different misclassification costs to different class errors in the objective function (Shin 2006).

Yong Seog Kim, (2009) provides insights on advantages and disadvantages of two ensemble models: ensembles based on sampling and feature selection. Experimental results confirm that both ensemble methods make robust ensembles and significantly improve the prediction performance of single classifiers at the cost of interpretability and additional computing resources. In particular, classifiers utilizing prior class distributions like support vector machine and naive Bayesian classifier only marginally benefit from ensembles, while classifiers with higher variance like neural networks and tree learners make a strong ensemble. Further, there seems to be an optimal ratio of selecting input variables that maximizes the performance of ensembles while minimizing computational costs when feature selection is used to create ensembles.

Ligang Zhou, et al., (2010) focuses several ensemble models based on least squares support vector machines (LSSVM) are brought forward for credit scoring. The models are tested on two real world datasets and the results show that ensemble strategies can help to improve the performance in some degree and are effective for building credit scoring models.

Nan-Chen Hsieh et al., (2010) focuses on predicting whether a credit applicant can be categorized as good, bad or borderline from information initially supplied. This is essentially a classification task for credit scoring. Given its importance, many researchers have recently worked on an ensemble of classifiers. However, to the best of our knowledge, unrepresentative samples drastically reduce the accuracy of the deployment classifier. Few have attempted to preprocess the input samples into more homogeneous cluster groups and then fit the ensemble classifier accordingly. For this reason, the concept of class-wise classification as a preprocessing step in order to obtain an efficient ensemble classifier is introduced. This strategy would work better than a direct ensemble of classifiers without the preprocessing step.

Tang applied feed forward neural network to maximize performance at desired mailing depth in direct marketing in cellular phone industry. He showed that neural networks show more balance outcome than statistical models such as logistic regression and least

squares regression, in terms of potential revenue and churn likelihood of a customer (Tang 2011).

A.I. Marqués et al., (2012) presented one step beyond by introducing composite ensembles that jointly use different strategies for diversity induction. Accordingly, the combination of data resampling algorithms (bagging and AdaBoost) and attribute subset selection methods (random subspace and rotation forest) for the construction of composite ensembles is explored with the aim of improving the prediction performance. The experimental results and statistical tests show that this new two-level classifier ensemble constitutes an appropriate solution for credit scoring problems, performing better than the traditional single ensembles and very significantly better than individual classifiers.

Recently, hybrid data mining approaches have gained much popularity; however, a few studies have been proposed to examine the performance of hybrid data mining techniques for response modeling (Maryam Daneshmandi et.al, 2013). A hybrid approach is built by combining two or more data mining techniques. A hybrid approach is commonly used to maximize the accuracy of a classifier.

Hany A. Elsalamony, (2014) introduces analysis and applications of the most important techniques in data mining; multilayer perception neural network (MLPNN), tree augmented Naïve Bayes (TAN) known as Bayesian networks, Nominal regression or logistic regression (LR), and Ross Quinlan new decision tree model (C5.0). The objective is to examine the performance of MLPNN, TAN, LR and C5.0 techniques 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, TAN, LR 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 calculated by three statistical measures; classification accuracy, sensitivity, and specificity.

Niharika Sharma et al., (2015) analyzes the performance of different classification techniques (i.e. Neural network, logistic regression, discriminant analysis, k-nearest neighbours, naïve bayes, support vector machine, decision trees and tree bagger) to select the one with the most accurate results for classification of bank direct marketing dataset. The performance of each classification model is evaluated using three statistical measures: classification accuracy, sensitivity and specificity. These measures are defined by a confusion matrix that contains information about actual and predicted classifications done by a classification system. Experimental results have shown the effectiveness of models. Ensemble method (Tree Bagging) achieves slightly better performance than rest of the models.

Ashkan Zakaryazad and Ekrem Duman (2016) introduced an ANN model with a new penalty function which gives variable penalties to the misclassification of instances considering their individual importance (profit of correctly classification and/or cost of misclassification) and then considered maximizing the total net profit. The effectiveness of the proposed models is appraised on two real-life data sets from fraud detection and a University of California Irvine (UCI) repository data set about bank direct marketing.

In this paper, a hybrid direct marketing system is proposed using radial basis function, support vector machine, k-Nearest Neighbor and Multilayer Perceptron and the effectiveness of the proposed bagged RBF, bagged SVM, bagged k-NN, bagged MLP and RBF-SVM, k-NN-MLP hybrid systems are evaluated by conducting several experiments on standard datasets of direct marketing. This paper analyzes the performance of the existing and proposed ensemble methods to achieve at good ensemble classifiers with the most accurate results for classification of bank marketing and labor relations datasets.

3. Proposed Methodology

3.1 Data Preprocessing

The dataset is related with direct marketing campaigns of a Portuguese banking institution and includes all collective agreements reached in the business and personal services sector for locals with at least 500 members (teachers, nurses, university staff, police, etc) in Canada in 87 and first quarter of 88. Before performing any classification method the data has to be preprocessed. In the data preprocessing stage it has been observed that the datasets consist of many missing value attributes. By eliminating the missing attribute records may lead to misclassification because the dropped records may contain some useful pattern for Classification. The dataset is preprocessed by removing missing values using supervised filters.

3.2 Existing Classification Methods

3.2.1 Radial Basis Function Neural Network

The Radial Basis Function Network (RBF) is in its simplest form a three layered feed forward neural network with one input layer, one hidden layer and one output layer (R. Callan, 1998). It differs from an MLP in the way the hidden layer performs its computation. The connection between the input layer and the output layer is nonlinear, while the connection between the hidden layer and the output layer is linear. RBF networks are instance based, meaning that it will compare and evaluate each training case to the previous examined training cases. In an MLP all instances are evaluated once while in an RBF network the instances are evaluated locally (Tom M, 1997). Instance based methods use nearest neighbor and locally weighted regression methods. An RBF network can be trained more efficiently than a neural net using backpropagation since the input and output layer are trained separately.

3.2.2 Support Vector Machine

Support Vector Machines has been introduced by Vapnik and his colleagues (C. Cortes and V. Vapnik, 1995), SVM models are very similar to classical multilayer perceptron

neural networks used for classification (R. Hua, Dai liankui, 2010), but recently they have been extended to solve regression problems (V. Vapnik et al., 1997).

SVM is very similar to an ANN since both receive input data and provide output data. For regression, the input and output of SVM are identical to the ANN. However, what makes the SVM primarily better is that the SVM does not suffer from over fitting like ANN does. So, the ANN memorizes the input data on the training stage and will not perform well at the testing data.

3.2.3 k-Nearest Neighbor

K-nearest neighbor (Margaret H.Dunham**,** 2003) is a supervised learning algorithm where the result of new instance query is classified based on majority of k-nearest neighbor category. The purpose of this algorithm is to classify a new object based on attributes and training samples. The classifiers do not use any model to fit and only based on memory. Given a query point, k number of objects (k=1) are found closest to the query point. The classification is using majority vote among the classification of the k objects. Any ties can be broken at random. K-Nearest neighbor algorithm used neighborhood classification as the prediction value of the new query instance.

K-nearest neighbor algorithm is very simple. It works based on minimum distance from the query instance to the training samples to determine the k-nearest neighbors. After gathered k nearest neighbors, simple majority of these k-nearest neighbors are taken to be the prediction of the query instance.

3.2.4 Multilayer Perceptron

The simplest neural network is called a perceptron. A perceptron is a single neuron with multiple inputs and one output. The original perceptron proposed the use of a step activation function, but it is more common to see another type of function such as a sigmoidal function. A simple perceptron can be used to classify into two classes. Using a unipolar activation function, an output of 1 would be used to classify into one class, while an output of 0 would be used to pass in the other class.

An MLP is a network of perceptrons. The neurons are placed in layers with outputs always flowing toward the output layer. If only layer exists, it is called a perceptron. If multiple layers exist, it is an MLP. Although the back propagation algorithm can be used very generally to train neural networks, it is most famous for applications to layered feed forward networks, or multilayer perceptrons.

3.3 Homogeneous Ensemble Classifiers

3.3.1 Dagging

This meta classifier creates a number of disjoint, stratified folds out of the data and feeds each chunk of data to a copy of the supplied base classifier. Predictions are made via majority vote, since all the generated base classifiers are put into the Vote meta classifier. It is useful for base classifiers that are quadratic or worse in time behavior, regarding number of instances in the training data.

3.3.2 Proposed Bagged Classifiers

Given a set D, of d tuples, bagging (Breiman, L. 1996a) works as follows. For iteration i (i =1, 2,…..k), a training set, D_i, of d tuples is sampled with replacement from the original set of tuples, D. The bootstrap sample, D_i, by sampling D with replacement, from the given training data set D repeatedly. Each example in the given training set D may appear repeated times or not at all in any particular replicate training data set D_i. A classifier model, M_i, is learned for each training set, D_i. To classify an unknown tuple, X, each classifier, M_i, returns its class prediction, which counts as one vote. The bagged classifiers (RBF, SVM, k-NN, MLP), M^*, counts the votes and assigns the class with the most votes to X.

Algorithm: Homogeneous ensemble classifiers using bagging

Input:

- D, a set of d tuples.
- $k = 4$, the number of models in the ensemble.
- Base Classifiers (RBF, SVM, k-NN, MLP)

Output: Bagged Classifiers (RBF, SVM, k-NN, MLP), M^*

Method:

1. for i = 1 to k do // create k models
2. Create a bootstrap sample, D_i, by sampling D with replacement, from the given training data set D repeatedly. Each example in the given training set D may appear repeated times or not at all in any particular replicate training data set D_i
3. Use D_i to derive a model, M_i;
4. Classify each example d in training data D_i and initialized the weight, W_i for the model, M_i, based on the accuracies of percentage of correctly classified example in training data D_i.
5. endfor

To use the bagged RBF and SVM models on a tuple, X:

1. if classification then
2. let each of the k models classify X and return the majority vote;
3. if prediction then
4. let each of the k models predict a value for X and return the average predicted value;

3.4 Heterogeneous Ensemble Classifiers

3.4.1 Weighted Majority Algorithm

The weighted majority algorithm corrects the trivial algorithm. It maintains a weighting of the experts. Initially all have equal weight. As time goes on, some experts are seen as making better predictions than others, and the algorithm increases their weight proportionately. The algorithm's prediction of up/down for each day is computed by going with the opinion of the weighted majority of the experts for that day.

Weighted majority algorithm

Initialization: Fix an $\eta \leq 1/2$. For each expert i, associate the weight $w_i^{(1)} := 1$.

For t = 1,2,…,T:

1. Make the prediction that is the weighted majority of the experts' predictions based on the weights $w_1^{(t)}$,….., $w_n^{(t)}$. That is, predict "up" or "down" depending on which prediction has a higher total weight of experts advising it (breaking ties arbitrarily).
2. For every expert i who predicts wrongly, decrease his weight for the next round by multiplying it by a factor of (1 - η):

$$w_i^{(t+1)} = (1 - \eta) w_i^{(t)} \qquad \text{(update rule).}$$

3.4.2 Proposed Hybrid System

Given a set D, of d tuples, arcing (Breiman. L, 1996) works as follows; For iteration i (i =1, 2,…..k), a training set, D_i, of d tuples is sampled with replacement from the original set of tuples, D. some of the examples from the dataset *D* will occur more than once in the training dataset D_i. The examples that did not make it into the training dataset end up forming the test dataset. Then a classifier model, M_i, is learned for each training examples *d* from training dataset D_i. A classifier model, M_i, is learned for each training set, D_i. To classify an unknown tuple, X, each classifier, M_i, returns its class prediction, which counts as one vote. The hybrid classifiers (RBF-SVM and k-NN-MLP), M^*, counts the votes and assigns the class with the most votes to X.

Algorithm: Hybrid System using Arcing

Input:

- D, a set of d tuples.
- *k* = 4, the number of models in the ensemble.
- Base Classifiers (RBF, SVM, k-NN, MLP)

Output: Hybrid Models (RBF-SVM and k-NN-MLP), M^*.

Procedure:

1. For i = 1 to k do // Create k models
2. Create a new training dataset, D_i, by sampling D with replacement. Same example from given dataset *D* may occur more than once in the training dataset D_i.

3. Use D_i to derive a model, M_i
4. Classify each example d in training data D_i and initialized the weight, W_i for the model, M_i, based on the accuracies of percentage of correctly classified example in training data D_i.
5. endfor

To use the hybrid model on a tuple, X:

1. if classification then
2. let each of the k models classify X and return the majority vote;
3. if prediction then
4. let each of the k models predict a value for X and return the average predicted value;

The basic idea in Arcing is like bagging, but some of the original tuples of D may not be included in Di, where as others may occur more than once.

4. Performance Evaluation Measures

4.1 Cross Validation Technique

Cross-validation (Jiawei Han and Micheline Kamber**,** 2003) sometimes called rotation estimation, is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. 10-fold cross validation is commonly used. In stratified K-fold cross-validation, the folds are selected so that the mean response value is approximately equal in all the folds.

4.2 Criteria for Evaluation

The performance of classification models are evaluated using classification accuracy. Classification Accuracy is the percentage of test samples that the ability of a given classifier to correctly predict the label of new or previously unseen data (i.e. tuples without class label information). The confusion matrix is used to visualization the performance of the proposed ensemble classifiers. Predicted values are compared with actual values to compute the confusion matrix. The statistical measures are calculated using true positive (TP), true negative (TN), false positive (FP) and false negative (FN) of

confusion matrix. The percentage of Correct/Incorrect classification is the difference between the actual and predicted values of variables. Table 1 shows the confusion matrix for a two-class classifier.

Table1. The confusion matrix for a two-class classifier

Classifier / Predicted class	TRUE CLASS	
	p (positive)	n (negative)
Y	True Positives	False Positives
N	False Negatives	True Negatives
Total	P	N

5. Experimental Results and Discussion

5.1 Bank Marketing dataset Description

The dataset is obtained from UCI Machine Learning Repository. The data is related with direct marketing campaigns of a Portuguese banking institution, based on phone calls. Often, more than one contact of the same potential customer was required, in order to determine if the product (bank term deposit) would (or would not) be bought. The goal is to predict if the customer will subscribe or not. With a valid prediction, the marketing department can focus on the most promising leads and increase the overall ROI of the campaign. The Properties of bank marketing dataset are given in Table 2.

Table 2: Properties of Bank marketing dataset

Data Set Characteristics:	Multivariate	Number of Instances:	45211
Attribute Characteristics:	Real	Number of Attributes:	17
Associated Tasks:	Classification	Missing Values?	N/A

The source data used was related to direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls in order to attain the information if bank term deposit would be subscribed or not. This data was obtained from: http://archive.ics.uci.edu/ml/datasets/Bank+Marketing. It contains the following attributes:

1 - Age

2 - Job: type of job

3 - Marital: marital status4 - education

5 - Default: has credit in default?

6 - Balance: average yearly balance, in Euros

7 - Housing: has housing loan?

8 - Loan: has personal loan?

9 - Contact: contact communication type

10 - Day: last contact day of the month

11 - Month: last contact month of year

12 - Duration: last contact duration, in seconds

13 - Campaign: number of contacts performed during this campaign and for this client

14 - Pdays: number of days that passed by after the client was last contacted from a previous campaign

15 - Previous: number of contacts performed before this campaign and for this client

16 - Poutcome: outcome of the previous marketing campaign

17 – Subscription

Data related to the last contact of the current campaign are:

1- Contact: contact communication type (telephone, cellular
or unknown).

2- Day: last contact day of the month.

3- Month: last contact month of year.

4- Duration: last contact duration, in seconds.

Data related to the contact of the various marketing campaigns are:

1- Campaign: number of contacts performed during this campaign and for this customer.

2- Pdays: number of days that passed by after the customer was last contacted from a previous campaign.

3- Previous: number of contacts performed before this campaign and for this customer.

4- Poutcome: outcome of the previous marketing campaign.

The target is binary to predict if the customer will subscribe or not.

5.2 Labor Relations Dataset Description

The data includes all collective agreements reached in the business and personal services sector for locals with at least 500 members (teachers, nurses, university staff, police, etc) in Canada in 87 and first quarter of 88. Data was used to test 2 tier approach with learning from positive and negative examples. Each case concerns one contract, and the outcome is whether the contract is deemed acceptable or unacceptable. The acceptable contracts are ones in which agreements were accepted by both labor and management. The unacceptable ones are either known offers that fell through because one party would not accept them or acceptable contracts that had been significantly perturbed to the extent that, in the view of experts, they would not have been accepted. Table 3 shows the properties of labor relations dataset.

Table 3: Properties of Labor Relations dataset

Data Set Characteristics:	Multivariate	Number of Instances:	57
Attribute Characteristics:	Categorical, Integer, Real	Number of Attributes:	16
Associated Tasks:	N/A	Missing Values?	N/A

It contains the following attributes:

1. dur: duration of agreement
2. wage1.wage : wage increase in first year of contract
3. wage2.wage : wage increase in second year of contract
4. wage3.wage : wage increase in third year of contract
5. cola : cost of living allowance
6. hours.hrs : number of working hours during week

7. pension : employer contributions to pension plan
8. stby_pay : standby pay
9. shift_diff : shift differencial : supplement for work on II and III shift
10. educ_allw.boolean : education allowance
11. holidays : number of statutory holidays
12. vacation : number of paid vacation days
13. lngtrm_disabil.boolean : employer's help during employee longterm disabil
14. dntl_ins : employers contribution towards the dental plan
15. bereavement.boolean : employer's financial contribution towards the covering the costs of bereavement
16. empl_hplan : employer's contribution towards the health plan

5.3 Experiments and Analysis

In this section, evaluation of models and analysis are described in detail. All experiments have been performed using Intel Core 2 Duo 2.26 GHz processor with 2 GB of RAM and weka software (Weka: Data Mining Software in java).

5.3.1 Homogeneous Ensemble Classifiers Bagging

The bank marketing and labor relations datasets are taken to evaluate the proposed Bagged RBF, SVM, k-NN, MLP classifiers.

5.3.1.1 Proposed Bagged RBF, SVM, k-NN, MLP

Table 4. The Performance of Base and proposed bagged Classifiers for Bank Marketing dataset

Dataset	Classifiers	Classification Accuracy
Bank Marketing	RBF	71.16 %
	SVM	69.00 %
	k-NN	75.33 %
	MLP	77.33 %
	Proposed Bagged RBF	76.16 %
	Proposed Bagged SVM	73.33 %
	Proposed Bagged KNN	97.83 %
	Proposed Bagged MLP	95.50 %
	Dagged RBF	76.00 %
	Dagged SVM	73.66 %
	Dagged K-NN	83.50 %
	Dagged MLP	82.33 %

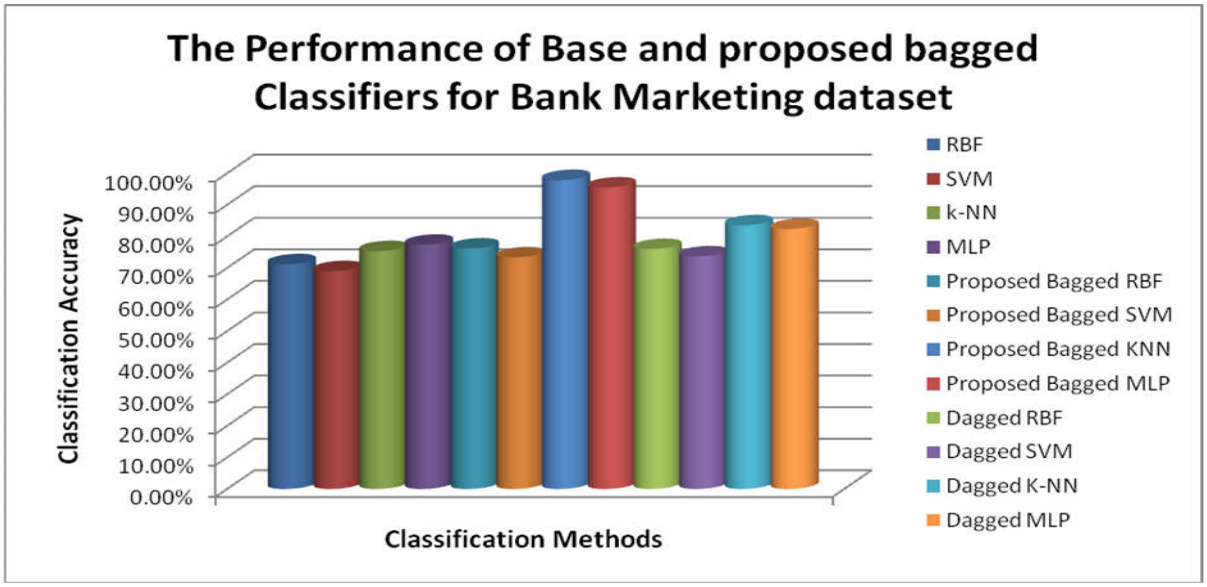

Figure 1: Classification Accuracy of Base and Proposed Bagged Classifiers Using Bank marketing dataset

Table 5.Confusion Matrices of Base and proposed bagged Classifiers for Bank Marketing dataset (A) RBF, (B) SVM, (C) K-NN, (D) MLP, (E) Proposed Bagged RBF, (F) Proposed Bagged SVM, (G) Proposed Bagged K-NN, (H) Proposed Bagged MLP, (I) Dagged RBF, (J) Dagged SVM, (K) Dagged K-NN, (L) Dagged MLP

(A)			**(B)**			**(C)**			**(D)**		
a	b	classified as	a	b	classified as	a	b	classified as	a	b	classified as
174	100	a = YES	156	118	a = YES	214	60	a = YES	210	64	a = YES
73	253	b = NO	68	258	b = NO	88	238	b = NO	72	254	b = NO
(E)			**(F)**			**(G)**			**(H)**		
a	b	classified as	a	b	classified as	a	b	classified as	a	b	classified as
180	94	a = YES	168	106	a = YES	267	7	a = YES	256	18	a = YES
49	277	b = NO	54	272	b = NO	6	320	b = NO	9	317	b = NO
(I)			**(J)**			**(K)**			**(L)**		
a	b	classified as	a	b	classified as	a	b	classified as	a	b	classified as
194	80	a = YES	190	84	a = YES	210	64	a = YES	215	59	a = YES
64	262	b = NO	74	252	b = NO	35	291	b = NO	47	279	b = NO

Table 6. The Performance of Base and proposed bagged Classifiers for Labor Relations Dataset

Dataset	Classifiers	Classification Accuracy
Labor Relations	RBF	94.73 %
	SVM	89.47 %
	k-NN	82.45 %
	MLP	85.96 %
	Proposed Bagged RBF	96.49 %
	Proposed Bagged SVM	96.49 %
	Proposed Bagged KNN	100 %
	Proposed Bagged MLP	98.24 %
	Dagged RBF	68.42 %
	Dagged SVM	92.98 %
	Dagged KNN	94.73 %
	Dagged MLP	94.73 %

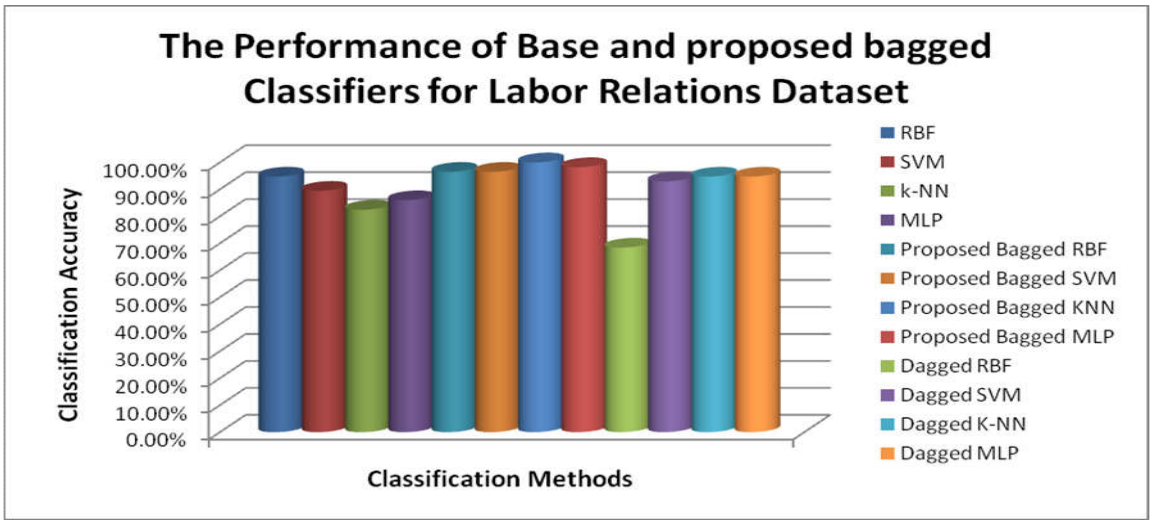

Figure 2: Classification Accuracy of Base and Proposed Bagged Classifiers Using Labor Relations Dataset

Table 7.Confusion Matrices of Base and proposed bagged Classifiers for Labor Relations dataset (A) RBF, (B) SVM, (C) K-NN, (D) MLP, (E) Proposed Bagged RBF, (F) Proposed Bagged SVM, (G) Proposed Bagged K-NN, (H) Proposed Bagged MLP, (I) Dagged RBF, (J) Dagged SVM, (K) Dagged K-NN, (L) Dagged MLP

(A)			**(B)**			**(C)**			**(D)**		
a	b	classified as	a	b	classified as	a	b	classified as	a	b	classified as
19	1	a = bad	16	4	a = bad	16	4	a = bad	16	4	a = bad
2	35	b = good	2	35	b = good	6	31	b = good	4	33	b = good
(E)			**(F)**			**(G)**			**(H)**		
a	b	classified as	a	b	classified as	a	b	classified as	a	b	classified as
21	1	a = bad	18	2	a = bad	20	0	a = bad	19	1	a = bad
1	34	b = good	0	37	b = good	0	37	b = good	0	37	b = good
(I)			**(J)**			**(K)**			**(L)**		
a	b	classified as	a	b	classified as	a	b	classified as	a	b	classified as
2	18	a = bad	18	2	a = bad	18	2	a = bad	19	1	a = bad
0	37	b = good	2	35	b = good	1	36	b = good	2	35	b = good

Here, the base classifiers are constructed using RBF, SVM, k-NN and MLP. 10-fold cross validation (Kohavi, R, 1995) technique is applied to the base classifiers and evaluated classification accuracy. Bagging is performed with RBF, SVM, k-NN and MLP to obtain a very good classification performance. Tables 5 and 7 show the confusion matrices generated from base and proposed bagged classifiers for bank marketing and labor relations datasets. Figures 1 and 2 show classification performance for standard datasets of direct marketing using existing and proposed bagged RBF, SVM, k-NN and MLP models. The analysis of results shows that the proposed bagged RBF, SVM, k-NN and MLP classifies are shown to be superior to individual approaches for standard datasets of direct marketing problem in terms of classification accuracy and confusion

matrix. This means that the combined methods are more accurate than the individual methods in the field of direct marketing.

5.3.2 Heterogeneous Ensemble Classifiers using Arcing

The bank marketing and labor relations datasets are taken to evaluate the proposed hybrid (RBF-SVM and k-NN-MLP) classifiers.

5.3.2.1 Proposed Hybrid (RBF-SVM and k-NN-MLP) System

Table 8. The Performance of Base and Proposed Hybrid (RBF-SVM and k-NN-MLP) Classifiers for Bank marketing dataset

Dataset	Classifiers	Classification Accuracy
Bank Marketing dataset	**RBF**	71.16 %
	SVM	69.00 %
	k-NN	75.33 %
	MLP	77.33 %
	Proposed Hybrid RBF-SVM	88.33 %
	Proposed Hybrid KNN-MLP	98.50 %
	Voted RBF-SVM	70.50 %
	Voted KNN-MLP	96.50 %

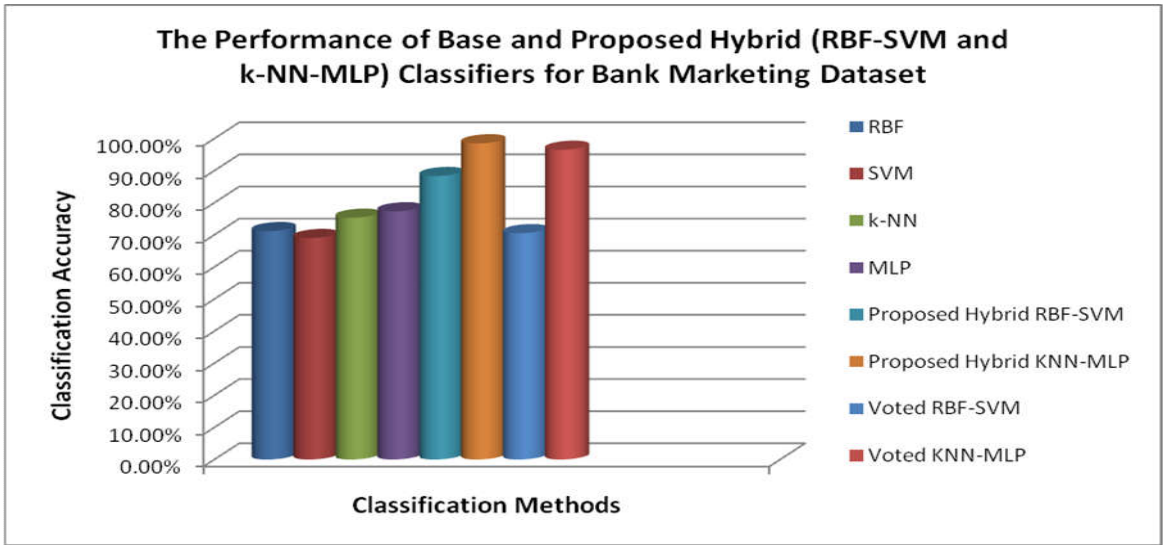

Figure 3: Classification Accuracy of Base and Proposed Hybrid (RBF-SVM and k-NN-MLP) Classifiers Using Bank marketing Dataset

Table 9.Confusion Matrices of Base and Proposed Hybrid Classifiers for Bank Marketing dataset (A) RBF, (B) SVM, (C) K-NN, (D) MLP, (E) Proposed Hybrid RBF-SVM (F) Proposed Hybrid KNN-MLP, (G) Voted RBF-SVM, (H) Voted KNN-MLP

(A)			**(B)**			**(C)**			**(D)**		
a	b	classified as	a	b	classified as	a	b	classified as	a	b	classified as
174	100	a = YES	156	118	a = YES	214	60	a = YES	210	64	a = YES
73	253	b = NO	68	258	b = NO	88	238	b = NO	72	254	b = NO
(E)			**(F)**			**(G)**			**(H)**		
a	b	classified as	a	b	classified as	a	b	classified as	a	b	classified as
447	16	a = YES	455	0	a = YES	148	126	a = YES	260	14	a = YES
54	83	b = NO	9	136	b = NO	51	275	b = NO	7	319	b = NO

Table 10. The Performance of Base and Proposed Hybrid (RBF-SVM and k-NN-MLP) Classifiers for Labor Relations Dataset

Dataset	**Classifiers**	**Classification Accuracy**
Labor Relations	**RBF**	94.73 %
	SVM	89.47 %
	KNN	82.45 %
	MLP	85.96 %
	Proposed Hybrid RBF-SVM	98.24 %
	Proposed Hybrid KNN-MLP	100 %
	Voted RBF-SVM	98.24 %
	Voted KNN-MLP	100 %

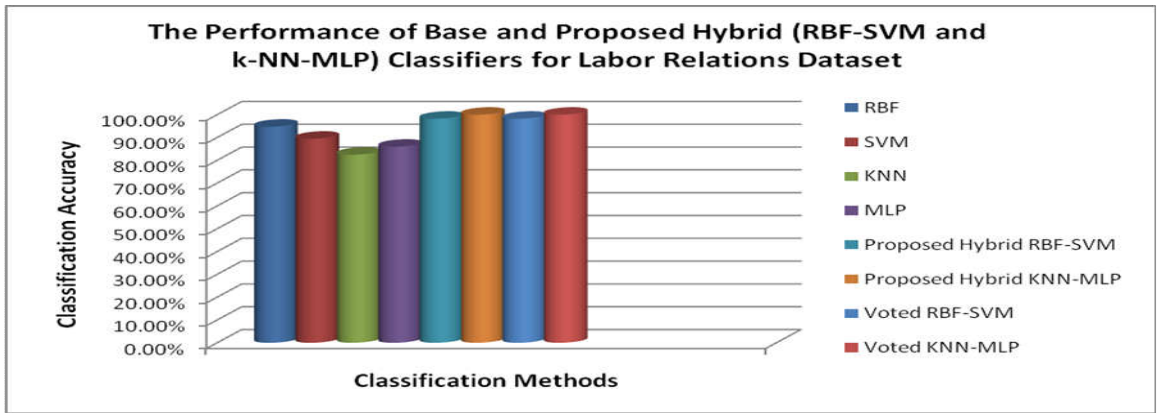

Figure 4: Classification Accuracy of Base and Proposed Hybrid (RBF-SVM and k-NN-MLP) Classifiers Using Labor Relations Dataset

Table 11.Confusion Matrices of Base and Proposed Hybrid Classifiers for Labor Relations dataset (A) RBF, (B) SVM, (C) K-NN, (D) MLP, (E) Proposed Hybrid RBF-SVM (F) Proposed Hybrid KNN-MLP, (G) Voted RBF-SVM, (H) Voted KNN-MLP

(A)			**(B)**			**(C)**			**(D)**		
a	b	classified as	a	b	classified as	a	b	classified as	a	b	classified as
19	1	a = bad	16	4	a = bad	16	4	a = bad	16	4	a = bad
2	35	b = good	2	35	b = good	6	31	b = good	4	33	b = good
(E)			**(F)**			**(G)**			**(H)**		
a	b	classified as	a	b	classified as	a	b	classified as	a	b	classified as
18	1	a = bad	12	0	a = bad	19	1	a = bad	20	0	a = bad
0	38	b = good	0	45	b = good	0	37	b = good	0	37	b = good

The data set described in section 5 is being used to test the performance of base classifiers and hybrid classifiers. In the proposed approach, first the base classifiers RBF, SVM, k-NN and MLP are constructed individually to obtain a very good generalization performance. Secondly, the ensemble of RBF and SVM, k-NN and MLP are designed. In the ensemble approach, the final output is decided as follows: base classifier's output is given a weight (0–1 scale) depending on the generalization performance. Table 9 and 11 exhibit the confusion matrices for base and proposed hybrid classifiers for bank marketing and labor relations datasets respectively. According to figure 3 and 4, the proposed hybrid models show significantly larger improvement of classification accuracy

than the base classifiers. The experimental results show that proposed hybrid (RBF-SVM and k-NN-MLP) models are superior to individual approaches for direct marketing problem in terms of classification accuracy. Tables 4 and 6; 8 and 10 gives classification accuracy of existing and the proposed homogeneous and heterogeneous models calculated on the results of various confusion matrices obtained which forms the basis of comparison. Though the accuracy for proposed hybrid k-NN-MLP is found to be 100% which is same as that of standard heterogeneous model, voted k-NN-MLP, the confusion matrices show that the proposed hybrid k-NN-MLP model produces a higher number of correctly identified respondents as against a lower number non-respondent incorrectly identified as respondents. This attests to the proposed hybrid k-NN-MLP model's higher accuracy compared to that of voted k-NN-MLP.

This paper aims at bringing an improvement in efficiency: making lesser contacts, but achieving a better number of successes (clients subscribing the deposit). The results of bank marketing dataset show that by using this model, bank can obtain a percentage as large as possible of the targeted customers responds to the product offer and also minimize the marketing cost and also can prevent to irritate the customers. Predictive response model helps companies (marketers) to identify a subset of customers who are more likely to respond than others and establish a direct relationship with them and increase the marketing efficiency. The results of labor dataset demonstrate a significant increase in the accuracy of estimating practices in the contractor's management systems.

From the results, the level of likely favorable response of customers is found to be well above the level of unlikely favorable response of customers. The implication is that the bank can plan effective marketing of its products/services through the guiding report obtained on each of the customers. This will enable the management increase its sales by targeting the respondents and prevent wasteful expenditure that will have been incurred as a result of sending promotion offers to the non respondents. These will go a long way in increasing the bank's Return on Investment. As a result of this framework, productivity and efficiency of management would increase. This model would improve labor resources management practices and forecasts. The improvement of labor

management practices could have a significant impact on reducing schedule delays and budget overruns.

6. Conclusion

Bank direct marketing and business decisions are more important than ever for preserving the relationship with the best customers. For success and survival in the business, there is a need for customer care and marketing strategies. Data mining and predictive analytics can provide help in such marketing strategies. This work aims on building a response model with ensemble classifiers for bank marketing dataset and improve resources management practices by using labor relations dataset. In this research work, new combined classification methods are proposed using homogeneous ensemble Classifiers with bagging and the performance comparisons have been demonstrated using standard datasets of direct marketing in terms of accuracy. The purpose is increasing the campaign effectiveness by identifying the main characteristics that affect the success (the deposit subscribed by the client). Here, the proposed bagged RBF, SVM, k-NN, MLP models combine the complementary features of the base classifiers. Similarly, new hybrid (RBF-SVM and k-NN-MLP) models are designed in heterogeneous ensembles involving RBF and SVM, k-NN, MLP models as base classifiers and their performances are analyzed in terms of accuracy. Experimental results have shown the effectiveness of models. The proposed ensemble models achieve slightly better performance than rest of the models.

The experiment results lead to the following observations.

- RBF, MLP exhibits better performance than k-NN, SVM in the important respects of accuracy**.**
- The proposed bagged methods show significantly higher improvement of classification accuracy and confusion matrix than the base classifiers.
- The hybrid (RBF-SVM and k-NN-MLP) models show higher percentage of classification accuracy than the base classifiers.
- The proposed ensemble methods provide significant improvement of accuracy compared to individual classifiers and the proposed bagged RBF, SVM, k-NN,

MLP models perform significantly better than dagging and the proposed hybrid (RBF-SVM and k-NN-MLP) model perform significantly better than voting.

- The heterogeneous models exhibit better results than homogeneous models for standard data sets of direct marketing.
- The direct marketing dataset could be detected with high accuracy for homogeneous and heterogeneous models.

The future research will be directed towards developing more accurate base classifiers particularly for the direct marketing problem. With the development of data mining techniques and databases, some areas which are not covered in this study are interesting and need to be explored in the future work.

Acknowledgements

Author gratefully acknowledges the authorities of Annamalai University for the facilities offered and encouragement to carry out this work.

References

1. Ashkan Zakaryazad and Ekrem Duman. (2016). “A profit-driven Artificial Neural Network (ANN) with applications to fraud detection and direct marketing”, Neurocomputing, 175, Part A, 121-131.
2. Bentz, Y., & Merunkay, D. (2000). “Neural networks and the multinomial logit for brand choice modeling: A hybrid approach”, Journal of Forecasting, 19(3), 177–200.
3. Breiman. L, (1996), “Bias, Variance, and Arcing Classifiers”, Technical Report 460, Department of Statistics, University of California, Berkeley, CA.
4. Breiman, L. (1996a). Bagging predictors. Machine Learning, 24(2):123–140.
5. Bounds, D., Ross, D. (1997). "Forcasting Customer Response with Neural Network." Handbook of Neural Computation G6.2: 1-7.
6. R. Callan. (1998): Essence of neural networks. Prentice Hall PTR Upper Saddle River, NJ, USA.
7. Cheung, K.-W., Kwok, J. K., Law, M. H., & Tsui, K.-C. (2003). “Mining customer product rating for personalized marketing. Decision Support Systems”, 35: 231–243.

8. Ha, K., Cho, S., MacLachlan, D. (2005). "Response models based on bagging neural networks", Submitted for publication. Journal of Interactive Marketing 19(1), 17–30.
9. Hany A. Elsalamony, (2014), "Bank Direct Marketing Analysis of Data Mining Techniques", International Journal of Computer Applications, 85(7):12-22.
10. Haykin, S. (1999). "Neural networks: a comprehensive foundation" (second ed.), New Jersey: Prentice Hall.
11. HosseinJavaheri, S. (2007). "Response Modeling in Direct Marketing-A Data Mining Based Approach for Target Selection", http://www.directworks.org/. Retrieved 2013/03/15.
12. R. Hua, Dai liankui, (2010), support vector machine classification and regression based hybrid modeling method and its application in raman spectral analysis, Chinese Journal of Scientific Instrument, (11):2440-2446.
13. Jiawei Han, Micheline Kamber**,** (2003), "Data Mining – Concepts and Techniques", Elsevier Publications.
14. Kohavi, R. (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection", Proceedings of International Joint Conference on Artificial Intelligence, pp.1137–1143.
15. Ligang Zhou, Kin Keung Lai, Lean Yu, (2010), "Least squares support vector machines ensemble models for credit scoring", Expert Systems with Applications, 37(1):127-133.
16. Margaret H.Dunham, (2003), "Data Mining-Introductory and Advanced Topics", Pearson Education, 90-113.
17. A.I. Marqués, V. García, J.S. Sánchez, (2012), "Two-level classifier ensembles for credit risk assessment", Expert Systems with Applications, 39(12):10916-10922.
18. Maryam Daneshmandi, Marzieh Ahmadzadeh (2013), "A Hybrid Data Mining Model to Improve Customer Response Modeling in Direct Marketing, Indian Journal of Computer Science and Engineering, 3(6):844-855.
19. Nan-Chen Hsieh, Lun-Ping Hung, (2010), "A data driven ensemble classifier for credit scoring analysis", Expert Systems with Applications, 37(1):534-545.
20. Niharika Sharma, Arvinder Kaur, Sheetal Gandotra, Bhawna Sharma, (2015), "Evaluation and Comparison of Data Mining Techniques Over Bank Direct

Marketing", International Journal of Innovative Research in Science, Engineering and Technology, 4(8):7141 – 7147.

21. Parr Rud, O. (2001). "Data Mining Cook book : Modeling Data for Marketing, Risk, and Customer Relationship Management", John Wiley & Sons, Inc.
22. Potharst, R., Kaymak, U., Pijls W. (2001). "Neural networks for target selection in direct marketing", Erasmus Research Institute of Management (ERIM), Erasmus University Rotterdam in its series Discussion Paper with number 77, http://ideas.repec.org/s/dgr/eureri.html.
23. Shin, H., Cho, S. (2006). "Response Modeling with Support vector Machines", Expert Systems with Applications 30: 746-760.
24. Suh, E. H., Noh, K. C., & Suh, C. K. (1999). "Customer list segmentation using the combined response model", Expert Systems with Applications, 17(2):89–97.
25. Tang, Z. (2011). "Improving Direct Marketing Profitability with Neural Networks." International Journal of Computer Applications 29(5): 13-18.
26. Tom M. Mitchell, (1997), Machine Learning, McGraw-Hill, New York.
27. V. Vapnik, S. Golowich and A. Smola, (1997), Support vector method for function approximation, regression estimation, and signal processing, Advances in neural information processing systems, pp. 281-287.
28. Viaene, S., B. Baesens, et al. (2001). "Wrapped Input Selection Using Multilayer Perceptrons for Repeat-Purchase Modeling in Direct Marketing", International Journal of Intelligent Systems in Accounting, Finance and Management, 10(2): 115-126.
29. Yong Seog Kim, (2009), "Boosting and measuring the performance of ensembles for a successful database marketing", Expert Systems with Applications, 6(2): 2161-2176.
30. Weka: Data Mining Software in java http://www.cs.waikato.ac.nz/ml/weka/
31. Zahavi, J., & Levin, N. (1997a). "Issues and problems in applying neural computing to target marketing", Journal of Direct Marketing, 11(4):63–75.
32. Zahavi, J., & Levin, N. (1997b). "Applying neural computing to target marketing", Journal of Direct Marketing, 11(4):76–93.