_____

Customer Behaviour Prediction using Machine Learning Techniques

K.S.Puujhitha	J.Priyadharshini	S.Poovendran	S.V.Kogilavani
UG Student	UG Student	UG Student	Assistant Professor(SrG)
Department of CSE	Department of CSE	Department of CSE	Department of CSE
Kongu Engineering College	Kongu Engineering College	Kongu EngineeringCollege	Kongu Engineering College
Erode, Tamilnadu	Erode, Tamilnadu	Erode, Tamilnadu	Erode, Tamilnadu
+919943562288	+918526724123	+919095733563	+919486153223
puujhithaks@gmail.com	_priyagopal600@gmail.com_	_poovendranvijay143@gmail.com_	_kogilavani.sv@gmail.com_

Abstract:-In today's advanced world of business, Customer Relationship Management is a data mining framework which establishes customer relationship and manages the relation between organizations and customers. Data mining plays an important role and gained popularity in various CRM related applications. To extract useful information from the available data, data mining is used in this project, and machine learning techniques are employed to train the system. An important data mining technique called classification is mostly used in this field. The classification model is used to predict the behaviour of customers. This paper discusses about the usage of various classification models like Naive Bayes, Decision Tree Induction, Random Forest, Multilayer Perceptron Neural Networks, Support Vector Machine to predict the behaviour of customers been proposed to show the accuracy of J48 is comparatively better. The performance of these methods are measured on the basis of evaluation parameters like accuracy, specificity and sensitivity. The result shows that Decision Tree Induction provides better performance compared to other models.

Keywords:-_classification; naive bayes; customer behaviour;decision tree induction ;support vector machine_

_____*****_____

1. Introduction

Data mining[3][4] is the process of discovering interesting patterns and knowledge in large datasets involving methods at the intersection of machine learning, statistics and

database system. Machine learning is the field of computer science that gives computer systems the ability to learn with data, without being explicitly programmed. Customer Relationship Management is a strategy for managing the relationship and interaction between the organization and the customers. In the competitive world, it is necessary to identify and predict customer needs by using past customer behaviour. For this purpose data mining and machine learning techniques have been employed in Customer Relationship Management(CRM).In CRM the major task is analysing and understanding the customer behaviour, which further helps in retaining potential customers and acquiring new customers. The tasks involved in CRM are customer acquisition, customer loyalty, customer retention and

customer profitability. For each task in CRM various data mining models can be applied.

Data mining can be used for decision making and forecasting in organizations. Classification is the widely

used learning model in data mining that predicts the customer behaviour. The classification models used in our work are Naive Bayes, Decision Tree Induction, Random Forest ,Support Vector Machine and Multilayer Perceptron Neural Networks.

2. Literature Review

In this section, recent research work carried in the field of Costumer Relationship Management is discussed. E.W.T. Ngai [1] referred that data mining is the process of analyzing data from different prospective and summarizing it into useful information. A combination of data mining models is often required to support or forecast the effects of a CRM strategy. The direct marketing program would be classified as being supported by classification as prediction of customers' behaviour is the major concern. The fields of data mining and machine learning arose due to the advances of information technology (IT), leading to an exponential growth of business . The aim of DM is to analyze raw data and extract high-level knowledge for the domain user or decision-maker.

R tool includes a high variety of DM algorithms and it is currently used by a large number of DM/BI analysts.

R presents the advantage of being more flexible and extensible by design, and its open source availability and

_____

_____

users' activity, novel DM methods are in general more quickly encoded into R than into commercial tools.

Paulo Cortez[2] referred that while several DM algorithms are available for these tasks, the library is particularly suited for using neural networks (NNs) and support vector machines (SVMs). Both are flexible models that can cope

with complex nonlinear mappings, potentially leading to more accurate predictions
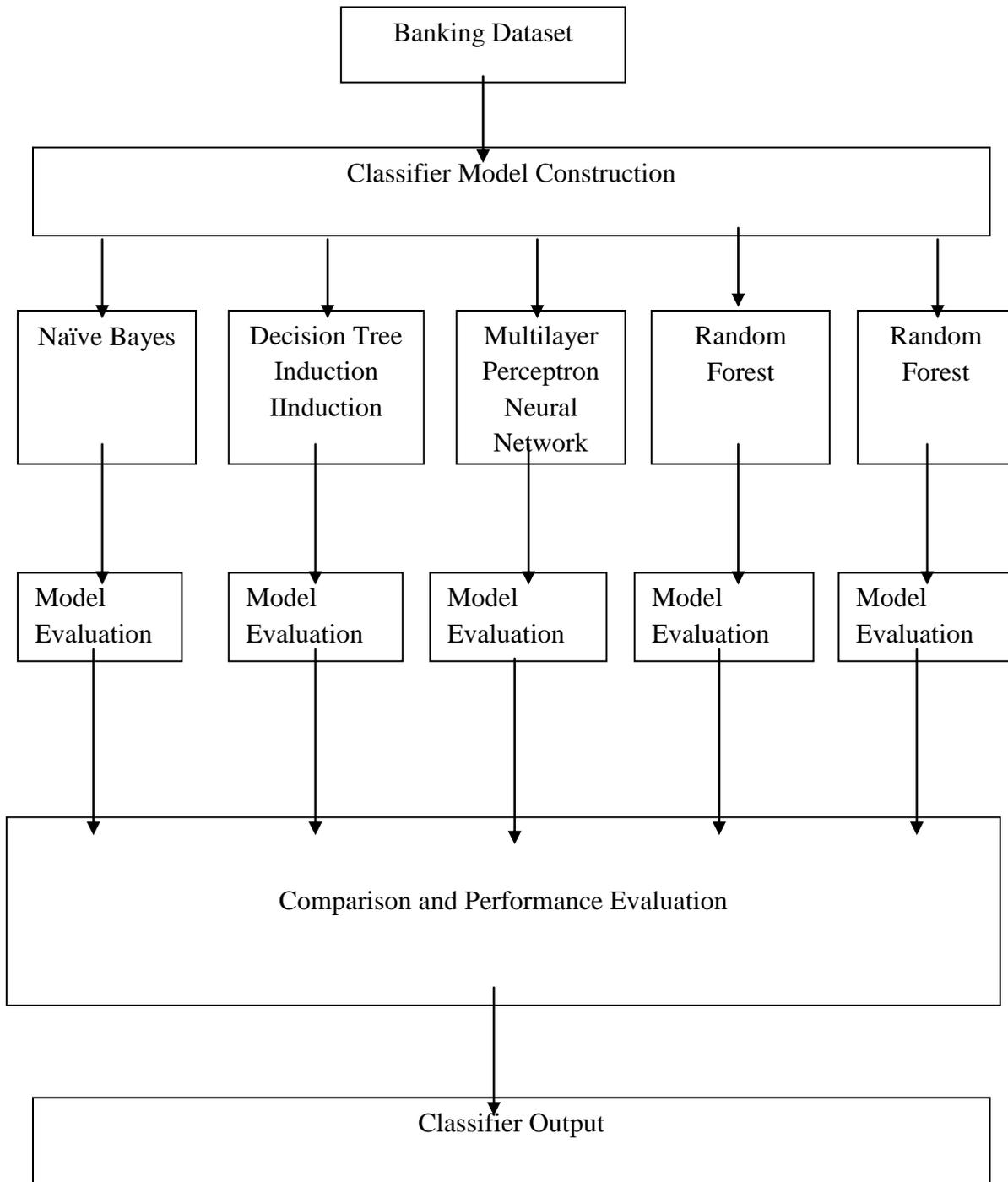
3. Proposed System

Figure 3.1 proposed system design

Instead of direct marketing campaigns performed by banks to improve the relationship between the customer and the organisation. But this project uses data mining and machine

learning techniques for CRM. Data mining is used to discover and extract productive information from the available customer details. Machine learning techniques are

_____

_____

employed to train the system based on the information discovered.

An efficient CRM framework in the domain of banking application for the purpose of customer behaviour prediction is shown in figure 3.1

3.1.Machine Learning Techniques

3.1.1.Naive Bayes

Naive bayes[3] classifier belongs to probabilistic classifiers based on applying bayes theorem. Bayesian classifiers are helpful in predicting the probability that a sample belongs to a particular class. To estimate the parameters necessary for classification, the classifier requires only small amount of training data. It also handles real and discrete data. Bayes theorem provides a way to calculate the posterior probability of each hypothesis, we can use it as the basis for learning algorithm[12] that calculates the probability for each hypothesis, then outputs the most probable. Because of the high accuracy and fastness to train, this technique is used for large databases.

The bayes theorem is represented in equation (1)

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \qquad (1)$$

where P(h) is probability of hypothesis h, $P(D)$ is prior probability of training data D, $P(h|D)$ is probability of h given D, $P(D|h)$ is probability of D given h.

The difficulty in Bayesian learning are it requires prior knowledge and computational cost is high to determine optimal hypothesis.

3.1.2 J48

Decision tree J48 is the implementation of ID3(Iterative Dichotomiser 3) developed by WEKA project team. It is one way to display an algorithm that contains conditional control statements. It helps in identifying a strategy most likely to reach the good. This algorithm is applicable for minimal

class, missing class values, binary class. J48[14] is used when decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm. It is very simple to understand and interpret. But it is unstable, complex if many values are uncertain.

3.1.3 Random Forest

Random Forest[11] are an ensemble learning method for classification , regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting that is the mode of the classes or mean prediction. In general, the more trees in the forest the robust the forest looks like. In the same way in the random forest classifier the higher the number of trees in the forest gives the high accuracy. Random Forest[7] is like bootstrapping algorithm with decision free model. Random Forest tries to build multiple CART model with different sample and different initial variables. Final prediction can simply be the mean of each prediction.

3.1.4 MultiLayer Perceptron Neural Network(MLNN)

MultiLayer Perceptron Neural Network is organized as a layered set of neurons. Among the input , output and hidden layer of neurons the actual computational of the network are performed in the hidden layer. Back propagation is the most common training technique used for MLPNN[13]. The difference between the network outputs and the desired ones is reduced by the application of weight correction by Back Propagation. The neural networks helps in learning and reducing the future errors. Advantages may include good learning activity, fast real-time operation, less memory demand, analysis of complex patterns. High quality data requirement of the network, careful selection of variables a priori are some of the disadvantages of MLPNN.

3.1.5.Support Vector Machine(SVM)

_____

_____

Support Vector Machine [6] (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well SVM [5] is based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. Support Vector Machine (SVM) is primarily a classier method that performs classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables.

3.2.TOOLS USED

3.2.1 Weka

The Waikato Environment for knowledge Analysis is introduced by Waikato University, New Zealand. It is a popular machine learning toolkit used extensively for research, education and projects. It consist of collection of machine learning algorithms and tools for data mining tasks such as data pre-processing or data preparation, classification, association rules, clustering, regression, forecasting and visualization and is well suited for developing new machine learning schemes[8]. It is an open source software. It consists of algorithms data pre processing, classification, association rules, clustering, regression, forecasting and visualization. Weka 3.74 is used for experimentation in this project.

3.2.2 R-Studio

RStudio [9] is a free and open-source integrated development environment (ide) for r, a programming language for statistical computing and graphics. RStudio was founded by JJ Allaire, creator of the programming language ColdFusion. Hadley Wickham is the Chief Scientist at RStudio. R Studio is available in two edition:Rstudio Desktop and R Studio Server.We used RStudio Desktop for our work.

4. Performance Evaluation

4.1 Dataset Description

To illustrate the performance of classification models we use the dataset that contains results of direct bank marketing campaigns[10].The customer was approached for a long-term deposit over phone. Our dataset has 17 variables which includes 16 input variables and a target variable with two possible outcomes either the client signed for long-term deposit or not .This is shown in Table 1 . Our dataset contains 45211 instances. For this work 10% of the pre processed dataset[10] from University of California at Irvine(UCI) is taken. Among sixteen input variables eight are related to client, four are related to last contact of the current campaign and the remaining four related to campaign.

_____

_____

Table1.Dataset Description

Variable Name	Type	Description
Age	Numeric	-
Job	Categorical	admin,unknown,unemployed,management,housemaid,entrepreneur,student,bluecollar,selfemployed,retired,techinician,services.
Martial	Categorical	married, single, divorced, note: divorced means divorced or widowed
Education	Categorical	unknown, secondary, primary, tertiary
Default	Binary	has credit in default?
Balance	Numeric	average yearly balance, in euros
Housing	Binary	has housing loan?
Loan	Binary	has personal loan?
Contact	Categorical	unknown, telephone, cellular, note: contact communication type
Day	Numeric	last contact day of the month
Month	Categorical	jan, feb, mar,.... nov, dec, note: last contact month of year
Duration	Numeric	in seconds, note: last contact duration
Campaign	Numeric	number of contacts performed during this campaign and for this client which includes last contact
Pdays	Numeric	number of days that passed by after the client was last contacted from a previous campaign, note: -1means client was not previously contacted
Previous	Numeric	number of contacts performed before this campaign and for this client
Poutcome	Categorical	unknown, other, failure, success, note: outcome of the previous marketing campaign
Y	Binary	has the client subscribed a term deposit?

4.2 Evaluation Metrics

For each method the classification accuracy(amount of correctly classified instances), true positive rate(the proportion of actual positives which are correctly identified as such), false positive rate(incorrectly classified positive), ROC area(area under the ROC curve) and time taken to build the classifier model is shown.

The results of our experiments for automatically classifying our dataset is summarized in Table 2.

Table 2.Result of comparison of classifiers.

Classifier	Classification Accuracy (%)	True Positive Rate	False Positive Rate	ROC	Time Taken (in sec)
Naive Bayes	88	0.880	0.426	0.86	0.48
J48	90.3	0.903	0.463	0.84	7.12
Random Forest	90	0.904	0.509	0.92	32.6

_____

3rd **National Conference on Innovative Research Trends in Computer Science and Technology (NCIRCST 2018)**
Volume: 4 Issue: 3

ISSN: 2454-4248
73 – 79

_____

MLPNN	89.6	0.896	0.461	0.88	1427.65
SVM	88.7	0.887	0.481	0.82	40.7

J48 classifier shows better accuracy(90.31%) among the models experimented. But the time taken is less for Naive Bayes(NB) compared to J48.

In this project work, to evaluate the performance of the classification models we used three statistical measures namely classification accuracy, sensitivity and specificity. For obtaining these information we need true positive(TP) which are the , true negative(TN), false positive(FP), false negative(FP). This information about actual and predicted classification defines a confusion matrix. The confusion matrix for each classifier model is shown in Table 3.

Classification accuracy is obtained by adding TP and TN then dividing it by total number of instances N. It can be calculated using the equation (2).

$$classification\ accuracy = \frac{TP+TN}{N} \qquad (2)$$

Sensitivity is equal to the ration of TP to the sum of TP and FN and it refers to the rate of correctly classified positive(True positive rate). It can be calculated using the equation (3).

$$sensitivity = \frac{TP}{TP+FN} \qquad (3)$$

Table3 Confusion matrix for each classifier

Classifier	A	B	class
Naive Bayes	36988 2498	2924 2791	A=yes B=no
J48	38289 2747	1633 2542	A=yes B=no
MLPNN	37937 2526	1985 2563	A=yes B=no
Random Forest	38601 3024	1321 2265	A=yes B=no
SVM	38431 3125	1231 3524	A=yes B=no

Specificity is equal to the ratio of TN to the sum of TN and FP and it refers to the rate of correctly classified negative(true negative rate). It can be calculated using the equation (4).

$$specificity = \frac{TN}{TN+FP} \qquad (4)$$

The performance of the classifiers is compared in-terms of accuracy, sensitivity and specificity and the results of the comparison are shown in Table4.

Table4 performance comparison of classifiers in terms of accuracy, sensitivity, specificity

Classifier	Accuracy	Sensitivity	Specificity
Naive Bayes	88%	52.7	92.67
J48	90.3%	48.1	95.9
Random Forest	90%	42.8	96.69
MLPNN	89.6%	48.5	95
SVM	88.7%	45.9	94.5

Out of the 45211 instances in the dataset the instances classified correctly and instances classified incorrectly for each model is shown Table 5.J48 classified 40831 instances correctly where as other algorithms like Naive Bayes classified 39789 instances, Random Forest classified 40366 instances and MLPNN classified 40500 instances correctly.

Table5 classification of 45211 instances in the dataset

Classifier	Correctly classified instances	Incorrectly classified instances
Naive Bayes	39789	5422
J48	40831	4380
Random Forest	40366	4845
MLPNN	40500	4711
SVM	39678	5533

5. Conclusion

In this paper we prposed an efficient CRM for customer behaviour prediction using various data mining and machine learning techniques.Four classification models were used to predict the customer behaviour.The best model that achieves better accuracy is J48 of 90.3%.We also compared the performance of the classifiers in terms of accuracy, sensitivity and specificity.

References

[1] EWT Ngai, L Xiu, DCK.Chau. Application of Data Mining Techniques in Customer Relationship Management: A Literature Review on Classification, Expert Systems with Applications; 36- 2, 2009. p. 2592-2602.

[2] Cortez P. Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool, In

_____

_____

Proceedings of the 10th Industrial Conference on Data Mining, Germany: Springer; 2010. p. 572–583.

[3] https://www.techopedia.com/definiti on/1181/data-mining

[4] http://myweb.sabanciuniv.ed/rdehkharghani/files/2016/02/ The-Morgan-Kaufmann-Series-in-Data-Management- Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data- Mining.-Concepts-and-Techniques-3rd-Edition-Morgan- Kaufmann-2011.pdf

[5] http://www.statsoft.com/Textbook/Support-Vector- Machines

[6] https://www.analyticsvidhya.com/blog/2017/09/understa ing-support-vector-machine-example-code/

[7] http://dataapirant.com/2017/05/22/random-forest- algorithm-machine-learning/

[8] An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour Femina Bahari Ta,*, SudheepElayidomMb4465j0j7&sourceid=chrome&ie=U TF-8

[9] https://www.google.co.in/search?q=rstudio&rlz=1C1__ ___enN657IN657&oq=rstudio&aqs=chrome..69i57j69i 65j69i60l3[10] S Moro, R Laureano, P Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology Proceedings of the European Simulation and Modelling Conference; Portugal, 2011 . p. 117-121.

[10] https://www.analyticsvidhya.com/blog/2014/06/introduct ion-random-forest-simplified/

[11] https://rviews.rstudio.com/2016/11/02/naive-bayes-a- generative-model-and-big-data-classifier/

[12] https://www.uow.edu.au/~markus/teaching/CSCI323/Lec ture_MLP.pdf

[13] https://pdfs.semanticscholar.org/2456/a979fbe8eea47b90 d625c1a064162be5382e.pdf

_____