MARKETING METHODS SUBSTANTIALLY ENHANCED BY USING DATA MINING TECHNIQUES FOR FINANCIAL DATA ANALYSIS

Ritesh N Varma
Research Scholar
Dept of CSE
Shri JJT University
Rajasthan
varmaritesh4@gmail.com

Dr. Vinod M Vaze
Research Guide
Dept of CSE
Shri JJT University
Rajasthan

Abstract:

Data mining is becoming important area for many corporate firms including banking industry. It is a process of analysing the data from numerous perspectives and finally summarize it into meaningful information, so data mining assists the bankers to take concrete decision. Objective of the project is to predict the fraud and fraud less transaction with respect to the time and amount of the transaction using classification machine learning algorithms and statistics and calculus (differentiation, chain rule etc) and linear algebra in building of the complex machine learning models for prediction and understanding of the data set. we have achieved accuracy of 94.84% using logistic regression and 91.62% using naive Bayes and 92.88% using decision tree and we step into deep learning to the continuous advancements in data mining, which is an interdisciplinary subfield of computer science and statistics, it has become significantly easier for the retail industry to improve customer service, acquire and retain new consumers by developing products in demand, and compete successfully in the marketplace. As a reason, data mining allows retail companies to analyse product sales trends, customer buying habits as well as preferences, delivery performance, customer peak traffic, and supplier lead times.

Key words: Data mining, Marketing Methods

1.0 INTRODUCTION:

The innovation of Technology improves the working of banking industry and the services provided by them. This is the gift of technology that banks enables sophisticated product development, better market infrastructure, etc. Internet has emerged as important medium for delivery of banking product and services since the early nineties each Indian bank has done some improvement on technology front by use of telemarketing, ATM, internet banking, mobile banking. This technology driven approach reach to maximum number of customers at low possible cost in efficient manner. The beauty of these banking innovation is possible only through the effective use of technology which has positive effect on growth and development Banking industry became highly competitive now a days. To be able to survive and grow in changing market environment and banks are going to adopt latest and updated technologies, it is also viewed as a tool of cost reduction and effective communication with clients and institution concern with banking transaction. With the advancement of computer technology and wide application of database, the banks gathered a huge amount of data which is stored in different form but identifying valuable information or knowledge for taking a concrete decision becomes very difficult. At this point of time a new concept of data mining came in existence which makes the ability to recognize the true value of data and decision-making process became easier. Data mining is very helpful in better standing of huge volume of data collected by CRM software system. In recent years many industries particularly banks as well as financial institution have recognized the true importance of information they have on their customer It is data mining which

minimize the purchasing cost, identifying the most effective and efficient promotion and provide an effective solution to numerous other organizations need. As is evident data mining solution have plenty of applications across industries.

2.0 LITERATURE REVIEW

M.B.Suvarchala [1] The study depicts that customer experience management is essential for enhancing customer satisfaction on the banking and financial services. Banks willing to improve their relationships with customers need constant monitoring of customer behavior and internal processes. An attempt has been made in this study to examine the customer experience management in largest public sector bank, State Bank of India and biggest private sector bank HDFC.

Neelam Verma, et al [2] The main reason for introducing biometric systems is to increase overall security. Biometrics offers greater security and convenience than traditional methods of personal recognition. In the world today, fingerprint is one of the essential variables used for enforcing security and maintaining a reliable identification of any individual.

Phakade, et al [3] The proposed work "Red-Tacton HAN Based ATM Machine Control System" has been successfully implemented and tested and verified using the hardware results. It has been one of newest technologies in this world which can grow very vast in upcoming years as it uses the human body as communication interface to talk to the machines and also it provides high security and decreases theft rate in ATM's.

Mayuri Bhosle [4] In this way, the implementation of ATM security by using fingerprint recognition and GSM model took advantage of the stability and reliability of finger print characteristics. Additionally, the system also contains the original verifying method which was inputting owner's password which is send by the controller. The security features enhanced largely for the stability and reliability of owner recognition. The whole system was built on the finger print technology which make the system more safe, reliable and easy to use.

Prashant P. Pittalia [5] According to latest scenario ATM fraud is a very grave problem for banks. This project leads to establish authentication results which can be used by banks as well as various organizations.

Patrick Enu [6] The study seek to examine the awareness of Information and Communication Technology banking among customers; to analyse the usage of Information and Communication Technology in banking and to identify the benefits of Information and Communication Technology in banking.

Richa Tuli,[7] It can be concluded on some key points there is a difference in attitude of customer of ICICI and SBI bank towards use of ATM . Purpose of this study was to know the factors influencing the use of ATM. This study finds that most important factor which influence customer to use the ATM services is it's convenience in use in case of both ICICI and SBI bank.

Saurabh Kumar [8] In this project a new Framework is built which will be very useful for the current world. Financial transactions of bank customers via ATM are facing huge amount of thefts. Among all these thefts one is related to with the existence of more than one person in ATM room and another threat is forced illicit activity performed by the user.

3.0 DATA MINING TECHNIQUES

Data mining a field at the intersection of computer science and statistics is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Mostly use Data mining techniques are as follows:

A. Bayes Classification

A Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong independent assumptions and is particularly suited when the dimensionality of the inputs is high. A naive Bayes classifier assumes that the existence (or nonexistence) of a specific feature of a class is unrelated to the existence (or nonexistence) of any other feature. Classification is a form of data analysis which can be used to extract models describing important data classes. Classification predicts categorical labels (or discrete values).Data classification is a twostep process. In the first step, a model is built describing a predetermined set of data classes or concepts. The model is constructed by analyzing database tuples described by attributes. Each tuple is assumed to belong to a predefined class, as determined by one of the attributes, called the class label attribute. In the second step, the model is used for classification. First,

the predictive accuracy of the model is estimated.

Bayesian Algorithm:

1. Order the nodes according to their topological order.
2. Initiate importance function $Pr0 (X\backslash E)$, the desired number of samples m, the updating interval l, and the score arrays for every node.
3. $k \Downarrow 0$, $T \Downarrow \varnothing$ 4. for i $\Downarrow 1$ to m do
4. if (i mod l == 0) then
5. $k \Downarrow k+1$
6. update importance function $Prk (X\backslash E)$ based on T end if
7. $si \Downarrow$ generate a sample according to $Prk (X\backslash E)$
8. $T \Downarrow T \cup \{si\}$
9. Calculate Score
10. (si, $Pr(X\backslash E,e)$, $Prk (X|E)$ and add it to the corresponding entry of every array according to the instantiated states.
11. Normalize the score arrays for every node.

The major disadvantage of this model is that the predictive accuracy is highly correlated with this assumption. An advantage of this method is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification.

B. Decision Tree

A classification (decision) tree is a tree-like graph of decisions and their possible consequences. Topmost node in this tree is the root node on which a decision is supposed to be taken. In each inner node, it is done as a test on an attribute or input variable. Specifically each branch of the tree is a classification question and the leaves of the tree are partitions of the dataset with their classification. The

processes in decision tree algorithms are very similar when they build trees.
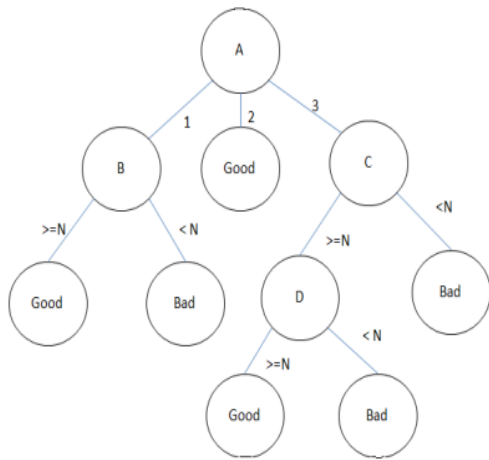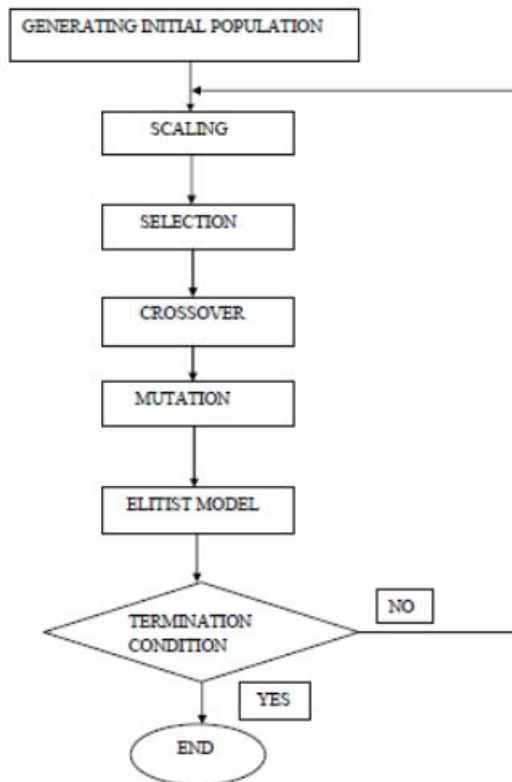

Figure: Decision Tree

Fig: Genetic Algorithm Flowchart

Selection deals with the probabilistic survival of the fittest, in those more fit chromosomes are chosen to survive. Where fitness is a comparable measure of how well a chromosome solves the problem at hand. Crossover takes individual chromosomes from P combines

them to form new ones. Mutation alters the new solutions so as to add stochasticity in the search for better solutions. In general, the main motivation for using GAs in the discovery of high-level prediction rules is that they perform a global search and cope better with attribute interaction than the greedy rule induction algorithms often used in data mining.

Marketing methods can be substantially enhanced by using data mining

many retailers is "Market Basket Analysis", which discovers a particular location where products can be effectively promoted. Using information collected from this method improves promotion effectiveness, the process of determining which products are likely to be sold to customers, which results in an increase in store sales. Most importantly, having exploited data mining for collecting and analyzing information related to Market Basket Analysis, detecting loyal customers becomes significantly faster.

4.0 RESULTS AND DISCUSSION

The procedure which we followed to predict the result are understanding problem statement and data by performing statistical analysis and visualization then checking whether the data is balance or not, In this data set the data is imbalanced, balanced by using oversampling, then scaling the data using standardization and normalization and testing data with different ML algorithms For any data science project some package are very important such as Numpy that is numeric python And pandas and for visualization of the data, matplotlib and seaborn is used which build on matplotlib with some extra features. Dataset contains the transaction

from Europe card owners during September 2021. In this 492 out of 2,84,807 are fraud transactions. Data is not balanced because less amount of fraud cases as compared to huge transaction data. Dataset is converted PCA transformation and contains only numeric values. Due to privacy and confidentiality many backgrounds information is not provided only PCA transformed data is given The figure shows the user interface for test and train the data. Train and Test buttons are given to the user where using train the algorithms are trained and then o predict the fraud by clicking predict button it will take to another window where the input is given and output is seen as fraud or nonfraud when predict button is clicked it will take to another window where it asks for data which is input to the machine learning algorithms and in the predict it will give output as fraud or nonfraud. comma separated 30 values are given including amount and time. Predicted result is displayed as fraud after providing the data. These results along with the classification report for each algorithm is given in the output as follows, where class 0 means the transaction was determined to be valid and 1 means it was determined as a fraud transaction.
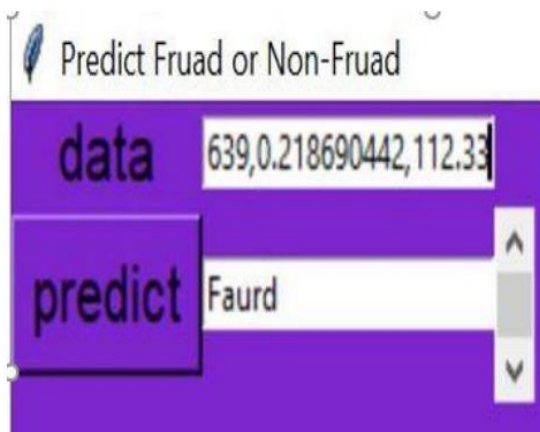

Fig: Detection of fraud or normal transaction

Confusion matrix for Logistic regression Algorithm:

represents confusion matrix for Logistic regression algorithm. Which contains True Positive, True Negative, False Positive, False Negative. False positive value is lesser which shows fraud not detected cases are low. For logistic regression algorithm accuracy, recall, precision achieved are 94.84, 92.00, 97.58 respectively**.**
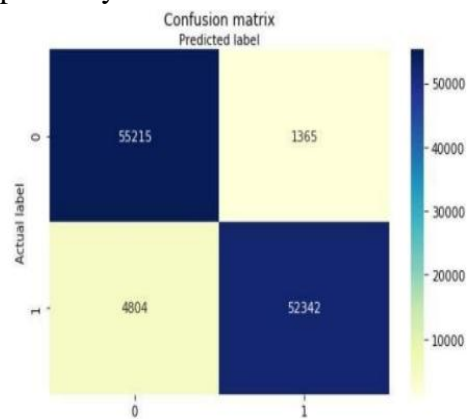

Figure: Confusion matrix for Logistic regression

Confusion matrix for Naive Bayes Algorithm:

represents confusion matrix for Naive Bayes algorithm. Which contains True Positive, True Negative, False Positive, False Negative. False positive value is lesser which shows fraud not detected cases are low. For Naive Bayes algorithm accuracy, recall, precision achieved are 91.62, 84.82, 97.09 respectively.

Confusion matrix for Decision Tree Algorithm:

To represents confusion matrix for Decision Tree algorithm. Which contains True Positive, True Negative, False Positive, False Negative. False positive value is lesser which shows fraud not detected cases are low. For Decision Tree

algorithm accuracy, recall, precision achieved are 92.88, 98.98, 99.48 respectively.

To understanding the behavior of customers, data mining helps retail companies to achieve more potential customers as well as retain them for a long period of time, control the process of manufacturing new products, and stay competitive in the market.
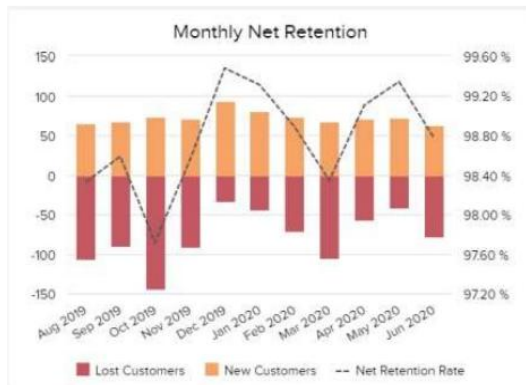

Figure: Data mining to analyze monthly net retention churn rate

Figure: Data mining to analyze monthly customer

Analyzing which specific products are frequently chosen by customers makes it easier to create new strategies for targeting new customers and retain previous ones, to bring a significant amount of revenues that can be redirected for the development of well selling products and, of course, to avoid producing goods that are vulnerable at competition and no longer profitable.

CONCLUSION

Credit card fraud is most common problem resulting in loss of lot money for peoples and loss for some banks and credit card company. This project want to help the peoples from their wealth loss and also for the banked company and trying to develop the model which more efficiently separate the fraud and fraud less transaction by using the time and amount feature in data set given in the Kegel. first, we build the model using some machine learning algorithms such as logistic regression, decision tree, support vector machine, this all are supervised machine learning algorithm in machine learning. Having taken into consideration above facts, it may be concluded that data mining has more positives than its negatives. Rightly used in retail companies, and not only, this powerful tool can be extremely beneficial when it comes to collecting and analyzing data. As we live in the century of big data and continuous technological advancements, the value of having proper information will rise significantly, resulting in a substantial use of data mining in the upcoming future. The only thing that has to be done by the companies is to maintain the harmony in using data mining technology and ensure its safety to avoid negative consequences.

References:

1. *Neelam Verma, Rakesh Patel, Priya Bag (2014), Global Journal of Engineering Science and Researches ATM Security Improvement Using Finger Print, ISSN 2348 – 8034*

2. *Prof. S.V. Phakade, Miss.Susmita Gopal Jadhav (2017), Red-Tacton HAN Based ATM Machine Control System, e-ISSN: 2395 -0056, Volume: 04*

3. *Prof. Mayuri Bhosle. Miss. Archana Takawale etc..(2019), A Survey On Cardless Automated Teller Machine (ATM), e-ISSN: 2395-0056, Volume: 06*

4. *Prof. (Dr.) Prashant P. Pittalia (2015), Enhancemnet of ATM Security and Theft Protection with the Use of One Time Password, ISSN: 2277 128X, Volume 5*

5. *Prof.Shinde S.P, Chingale R.R. etc.. (2017) , ATM Machine Security System Using GSM And MEMS Sensor, e-ISSN: 2395 -0056, Volume: 04*

6. *Patrick Enu , James Tetteh Gberbi (2015), Effect of Information and Communication Technology (ICT) on the Delivery of Banking Services in Ghana: A Case Study of Zenith Bank Ghana Limited, ISSN 2345-6086, Pages: 60-82*

7. *Shailesh N. Sisat, Vishwajit K. Barbudhe (2014), Secured Automatic Teller Machine (ATM) and Cash Deposit Machine (CDM), ISSN: 2321-7782, Volume 2*

8. *Richa Tuli,Abhijeet Khatri, Anita Yadav (2012) , A comparative study of customer attitude towards ATM of SBI and ICICI bank, ISSN: 2249-1058,*

9. *Saurabh Kumar, Ankit Mohan, Ashish Gupta, Paras Gupta (2019), ATM Security using Machine Learning, e-ISSN: 2395-0056, Volume: 06*