International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Customer Retention and Fraud Detection for Credit Card

Ajinkya Pathak[1], Rohan Aney[2], Soham Baheti[3]

Under Guidance of : Mr. Manjitsing K Valvi at IT Department
K J Somaiya College of Engineering , Vidyanagar, Vidyavihar East, Mumbai, Maharashtra 400077.

Abstract: ***In today's digital world banking sector has reached every business, every small business and households in multiple forms from loans to credit cards to Fixed Deposits. It is very crucial for a bank to analyses its customer purchases and to interpret the trends to enhance the customer base and retain existing customers. There are various strategies for customer retention; the modern day banks use latest IT tools to capture and interpret the data for customer attributes. The present project deals with developing a decision tree program to gain a rough estimate on whether a customer with particular attribute would be profitable for the bank in the long term. For a manager Visualization tools are also provided to help facilitate a much better understanding of its customers. The project also addresses a very vital aspect of fraud detection and offers a suitable program for the same. Fraud detection is not only essential for the bank but lowest level of fraud will enhance customer's confidence in the transactions with the bank. The program is based on genetic algorithms for Fraud detection. This program can be incorporated dynamically (i.e. Real time) on any system thus making it more agile and effective.***

Keywords: Banking tools, Data Classification, Genetic algorithms, Visualization tools

1. Introduction

Data Mining is a process of abstracting potential and useful information, knowledge from plentiful, incomplete, noisy, fuzzy and stochastic data. This information and this knowledge can't be achieved relying on a simple data query or or database understanding. The key of data mining include three parts: data, information and business decisions. The ultimate task of data mining is not only to retrieve useful information. But in fact, it is to use that information to improve business decision-making efficiency and to develop more appropriate decisions. We have divided the project in 3 modules for simplicity.

2. Components

A. Customer Analysis

Every customer has certain set of attributes associated with him/her. The attributes provide more information about the individual which may provide vital information for the bank to find some patterns within the customer base in the bank. The manager can use information acquired by analysis and enrich the customer relationship base. In this section we use data mining approach called as decision trees. The decision tree is classification algorithm. The decision splits the set of attributes. For splitting we have to decide the splitting criteria. We make us of Gini index value to find the attribute to be split at each step.

B. Fraud detection

Fraud detection proves important in any money transaction operation happening online. Fraud detection requires a robust system to work every time a unusual request is sent to the server. The fraud detection mechanism uses the genetic algorithm in neural network which records the transactions done and works on it to find the Fraudulent ones. The usual pattern is known to the system. If multiple parameters are showing unusual then the system gives an alert. The threshold (point beyond which the transaction is declared fraudulent) that declares the transaction as potentially fraudulent can change based on the system requirement and its application.

C. Visualization

The managerial level application is hugely benefiting from the visual tools that help them with the understanding otherwise huge dataset with huge amount of records. The algorithms used help the manager visualize the output based on his/her customization.

3. Theoretical Review

A. Decision trees

Decision Tree splitting and probability value set Splitting is finding the best possible breakage of the data.

$$InformationGain(a_i, S) = Entropy(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i = v_{i,j}} S|}{|S|} \cdot Entropy(y, \sigma_{a_i = v_{i,j}} S)$$

where:

$$Entropy(y, S) = \sum_{c_j \in dom(y)} -\frac{|\sigma_{y=c_j} S|}{|S|} \cdot \log_2 \frac{|\sigma_{y=c_j} S|}{|S|}$$

Where,
S - Training Set
A - Input Feature Set
y - Target Feature
Create a new tree T with a single split and make it the rootnode.
IF One of the Stopping Criteria is fulfilled
THEN
Mark the root node in T as a leaf with the most common value of y in S as a split label.
ELSE
Find a discrete function f(A) of input attributes values such that splitting the S according to f(A)s outcomes (v1,...,vn) gains
IF best splitting solution threshold
THEN
the best splitting solution.
IF best splitting solution threshold
THEN
Label t with f(A)
FOR each outcome vi of f(A): Set Subtree i= TreeGrowing(f(A)=vi,S,A,y).

Connect the root node of tT to Subtree i with an edge that is split labelled as vi END
FOR Mark the first root node in T as a leaf with the most common value of y in S as a split label.
ELSE
END IF
END IF
RETURN T

B. Genetic Algorithms

Genetic Algorithm is a part of Neural networks. The initial entry is selected randomly from the sample datasets which has many entries.

The fitness value is calculated in each entry and is sorted out. In selection process is selected through tournament method. The Crossover is calculated using single point probability. Mutation mutates the new offspring formed by previous entries using uniform probability measure for given entry. The new population is generated and undergoes the same process it maximum number of generation is reached.

Pseudo code of genetic algorithm
1) Initialize the data Entry
2) Evaluate initial data entry
3) Repeat
4) Perform competitive selection
5) Apply genetic operators to generate new values
6) Evaluate solutions in the dataset
7) Until some criteria is achieved

Selection process

Selection is used for choosing the best data value for selecting higher fitness entities. The selection operation takes the current population and produces a merging pool which contains the individuals which are going for another cycle. There are several selection methods such as the biased selection, random selection, tournament selection.

Tournament Selection

it selects optimal Entities from diverse datasets. It selects n individuals from the current population uniformly at random, forms a tournament and the best entities of a group wins the tournament selection has been used in this as tournament and is put into the merging pool for recombination.

Critical Value Identification

Based on Credit card use
Float
credfreq=Float.valueOf(temp[3])/Float.valueOf(temp[6]);
if(credFreq ≥ 0.2
if(Float.valueOf(temp[7]) ≥ (5 ∗ credF req))
res[0]=1;
*res[1]=(Float.valueOf(temp[7])*credFreq);*
if(res[0] ≤ 1)
res[1]=(float)credFreq;
credFreq = Total number card used (CU) / CC age
If credFreq is less than 0.2 , it means that this criteria may not applicable for fraud critical value = credFreq
Otherwise, it check for condition of fraud (i.e) = Fraud condition = number of time Card used Today (CUT) ≥ (5 ∗ credFreq)
*If true, there may chance for fraud using this property and its critical value is CUT*credFreq*
If flase, no fraud occurance and critical value = credFreq

C. Visualization

Visualisation provides the ability to retrieve, evaluate, comprehend, and act on huge dataset much faster and more effectively. It encompasses various data sets quickly and efficiently and makes it accessible to the viewer. Different Visualization tools such as Jfree Charts provide a good platform for the efficient display of information.

In this project, we have used JFreeChart, a java library provided especially for creating charts, to visualize the data in form of Pie Charts, Bar Charts, line plots, etc.

JFreeChart is an open source and 100percent free java library. It provides a consistent and well-documented API, supporting a wide range of chart types. In this project, JFreeChart is implemented through NetBeans by importing the following.

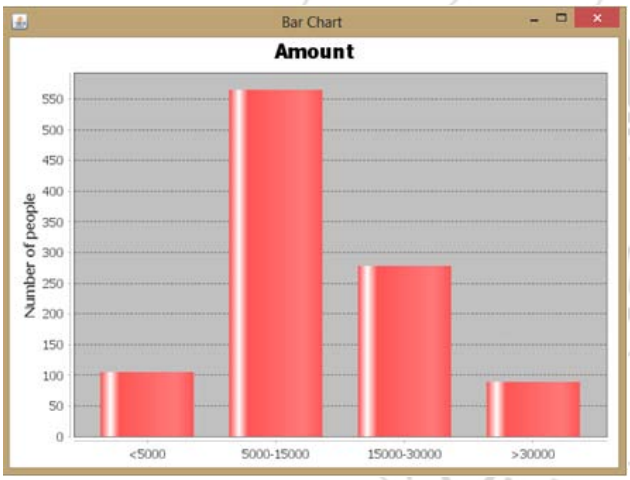
In this project,
JFreeChart is implemented through NetBeans by importing the following libraries:

import org.jfree.chart.ChartFactory;
import org.jfree.chart.ChartFrame;
import org.jfree.chart.JFreeChart;
import org.jfree.chart.plot.CategoryPlot;
import org.jfree.chart.plot.PiePlot;
import org.jfree.chart.plot.PlotOrientation;
import org.jfree.chart.renderer.category.BarRenderer;
import org.jfree.data.category.DefaultCategoryDataset;
import org.jfree.data.general.DefaultPieDataset;
import org.jfree.data.jdbc.JDBCCategoryDataset;

In the visualization module, the dynamic SQL queries provide the count for the specified category and the count is accordingly plotted on the Pie/Bar chart.
This is done for mainly 4 variables namely Job, Age, Amount and Location. Also, the visualization module also provides a line graph module for the purpose of displaying outliers in the database for the four decider variables for fraud as discussed in the above section.

A demo implementation of a Bar graph looks like below:

```
dataset.setValue(new
Integer(Amount_1),"Amount","<"+Amount1);
dataset.setValue(new  Integer(Amount_2),"Amount",+Amount2+"-
"+Amount3);
    dataset.setValue(new
Integer(Amount_3),"Amount",+Amount4+"-"+Amount5);
     dataset.setValue(new
Integer(Amount_4),"Amount",">"+Amount6);
    JFreeChart
chart=ChartFactory.createBarChart("Amount","","Number  of
people",
    dataset,PlotOrientation.VERTICAL,false,true, false );
CategoryPlot p=chart.getCategoryPlot();
p.setRangeGridlinePaint(Color.black);
ChartFrame frame=new ChartFrame("Bar Chart ",chart);
frame.setVisible(true);
frame.setSize(450,350);
```



4. Conclusion

From the training set we worked, we are assertive that the system can be integrated with a real time banking system. The dataset works perfectly on huge dataset and is adaptive and flexible to any kind of new attribute set. We have also made a database admin to operate the database of the system. In some cases the database may need as expert assistant for e.g. if the dataset is too small and not many entities are present to help the system generate and valuable information from the database. The 3 modules i.e. 1) Decision trees for customer analysis 2) Genetic Algorithm for Fraud Detection and 3) Visualization for customer patterns can be incorporated as managerial tools by most organization.

References

[1] Jiawei Han, Micheline Kamber, Data Mining : Concepts and Techniques, 2nd ed. Morgan Kaufmann, ISBN 1558609016, 2006.
[2] Marc Loy, Robert Eckstein, Dave Wood, James Elliott, Brian Cole, Java Swing, 2nd ed. O'Reilly Media, 2006.
[3] K.RamaKalyani, D.UmaDevi . "Fraud Detection of Credit Card Payment System by Genetic Algorithm" International Journal of Scientific Engineering Research Volume 3, Issue 7, July-2012 ..
[4] Amin, R.K. ; Comput. Sci. Study Program, Telkom Univ., Bandung, Indonesia ; Indwiarti ; Sibaroni, Y. "Implementation of decision tree using C4.5 algorithm in decision making of loan application by debtor (Case study: Bank pasar of Yogyakarta Special Region)."Information and Communication Technology (ICoICT), 2015 3rd International Conference on 27-29 May 2015..
[5] Zhao Li Ping; Shu Qi Liang " Data mining application in bankingcustomer relationship management." "IEEE Conference Publications.