PREDICTIVE ANALYSIS APPLICATION IN BANKING SECTOR USING MINING TECHNIQUE ALGORITHMS

Prof. Poonam Sadafal, Akanksha Bairagi, Mamta Jadhav, Chetna Rathod, Sakshi Hajare

Sinhgad Institute of Technology And Science, Narhe Pune-41

ABSTRACT

Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. It is an essential process where intelligent methods are applied to extract data patterns.

Nowadays, there are many risks related to bank loans, health loan, car loan, for the bank and for those who get the loans. This paper describes data mining with predictive analytics for banking policy applications and explores methodologies and techniques in data mining area combined with predictive analytics for application driven results for interested customers. The basic idea is to apply patterns on available data and generate new assumptions and anticipated behavior using predictive analysis. Data mining methods used in these applications naive Bayes data analysis. Data Mining is one of the most motivating and vital area of research with the aim of extracting information from tremendous amount of accumulated data sets. The model has been built using data from banking sector to predict the status of loans particular user if they want. The model has been built using data from banking sector.
Here we find out the interested user who wants the service.

Key words—*Predictive Analysis; Loan; data mining; algorithm*

INTRODUCTION

In this research paper, how problems faced in banking system for any activity for analysis. In the banking industry, various business models associated with it and the problems faced by the banking industry in adopting this technology. So we minimized the some problems for accurate analysis for better output through Mining technology. The main goal and objective of the project is to develop a Machine Learning model to perform predictive analytics on the banking data. The banking data set consists of details about customers like and whether the customer will buy a product provided by the bank or not.

This project describes data mining with predictive analytics for banking policy applications and explores methodologies and techniques in data mining area combined with predictive analytics for application driven results for interested customers. The basic idea is to apply patterns on available data and generate new assumptions and anticipated behavior using predictive analysis. Data mining methods used in these applications naive Bayes data analysis.

LITERATURE SURVEY

Sr. No.	Title of paper	Author	year	Advantages	Limitations
1.	The motivations for Big data Mining Technologies adoption in Saudi Banks	Nuha Almoqren , Dr. Mohammed Altayar	2016	Big data mining technology provides high of information management system that contributes to the achievement of the three quality factors: system quality, information quality, and service quality in Saudi bank.	This research is restricted to one country, Saudi Arabia; to know more about the issues that influence the implementation of big data mining technology in the banking sector.
02.	Data mining Application in banking sector with clustering & Classification methods	Asli calis, Ahmet Boyaci, Kasim Baynal	2015	In the study, it was aimed to analysis of existing personal loan customers and estimate potential customers' repayment performances.	For the banks to attain competitive advantage in the sector and to stay operating for long time periods, they must understand their customers correctly and must separate risky customers from others.
03.	Analysis of banking E-Data Downloading	Fuxin,Fu Shuai	2013	The scheme can solve security demand of users at present in E-banking Server.	E-banking server can provide more flexible security mechanism with current client decryption tool.

(IJRST) 2018, Vol. No. 8, Issue No. I, Jan-Mar **e-ISSN: 2249-0604, p-ISSN: 2454-180X**

EXISTING SYSTEM APPROACH

Some of the banks and financial-services companies may depend only on strategy of mass marketing for promoting a new service or product to their customers. In this strategy, a single communication message is broadcasted to all customers through media such as television, radio or advertising firm, etc. In this approach, companies do not set up a direct relationship to their customers for new-product offers. In fact, many of the customers are not interesting or respond to this kind of sales promotion.

Accordingly, banks, financial-services companies and other companies are shifting away from mass marketing strategy because its ineffectiveness and they are now targeting most of their customers by direct marketing for specific product and service offers. Due to the positive results clearly measured; many marketers attractive to the direct marketing.

Disadvantages:

1. It was restricted to only one country.
2. It was not possible to stay with the peoples for long time.
3. Not secure.
4. They must understand their customers correctly to provide accurate services.

PROPOSED SYSTEM APPROACH

This project describes data mining with predictive analytics for banking policy applications and explores methodologies and techniques in data mining area combined with predictive analytics for application driven results for interested customers. The basic idea is to apply patterns on available data and generate new assumptions and anticipated behavior using predictive analysis.

Data mining methods used in these applications naive Bayes data analysis. It is applied to all bank and other services company's except the exempted Bank policy and services and on transactions below the threshold limit. This system is used to any organization or banking services to analysis the future predication. User can access the web application from remote location.

Standard internet connection is required .TCP/UDP connection will be required.

Advantages:

1. Fast to train (single scan). Fast to classify
2. Not sensitive to irrelevant features
3. Handles real and discrete data
4. Handles streaming data well

SYSTEM ARCHITECTURE

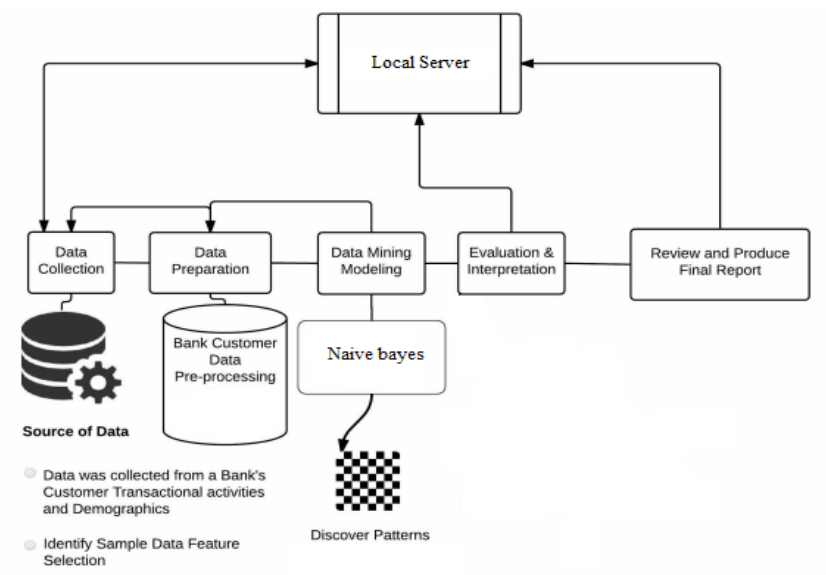


Fig No 01 System Architecture
(source:

MATHEMATICAL MODEL

Mathematical Model

Let M be the proposed system,

M= { S , E , I , O , F , Success , Failure }

where,

S = Start state of system
E = End state of system
I = Set of input
O = Set of output
F = Set of functions
Success = Desired output is generated.
Failure = Desired output is not generated.

I = {I1, I2}

Where, I1=Password (number, symbols, alphabet)
I2=Create Banking services.

O = {O1, O2, O3}

Where,

O1=add feedback.
O2=Sort the user which is interested to Banking services.
O3=Display list to send the notification

F = {f1, f2}

Where,

f1=Send_SMS()

where, if user interested to the banking system send SMS through pop up block.

f2=Transmit ()

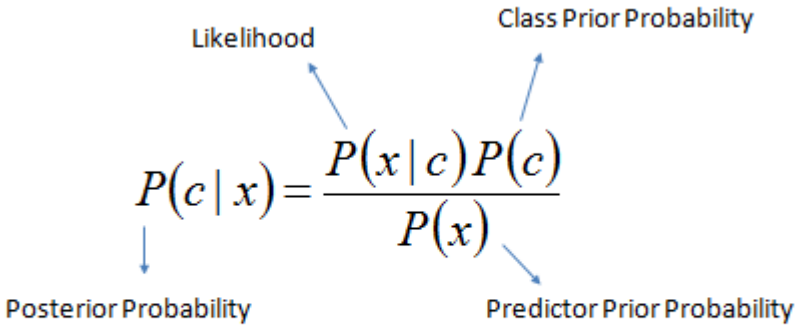
Where, It is used for transmit for security purpose

- Success= {Success state is when administrator should view the Bank user list.}
- Failure= {failure state is load kept on increasing.}

Algorithm Analysis

Naïve Bayes Algorithm

- Step 1: Convert the data set into a frequency table
- Step 2: Create Likelihood table by finding the probabilities like Overcast probability and probability of playing .
- Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Complexity of Naive Bayes Algorithm:

Time complexity The theoretical time complexity for learning a naive Bayes classifier is O(Np), where N is the number of training examples and p is the number of features. To empirically test the implemented algorithm's time complexity, we used an artificially created data set. We varied the number of samples with a fixed number of features and the maximum size of bins, shown in Figure 1. Also, we varied the numbers of features with a fixed number of samples and the maximum size of bins, shown in Figure 2, and tested varying number of bin sizes with a fix number of samples and features, shown in Figure 3. The figures show the time is linear in respect to the number of samples, features, and the maximum size of bins. In our algorithm, we first had an outer loop for each feature. Within this loop, we enumerated the number of samples, finding the correct bin (or discretized bin) based on whether it was a real valued or discrete feature. This led to a running time of O(Np).

In our first iteration of the algorithm, the running time was O(Npq), because all values of the bins had to be searched in order to find the correct bin for the value of the specific feature. Later, we realized that if a different data structure was used we would not need to search through all q values of each feature because an index could be used to do a fast lookup. We employed the use of a Java Hashtable through the Java Interface in Matlab in order to accomplish this task. This was able to reduce the overall time complexity to meet the theoretical bound. This same format was used in the testing of the data, which also resulted in an O(Np) running time.

Space complexity The theoretical space complexity for naive Bayes algorithm is O(pqr), where p is the number of features, q is values for each feature, and r is alternative values for the class. We were able to achieve the same theoretical space complexity, because in our implementation, we require the number of features (p) * the maximum number of bins (q) * the number of classes. Since the number of classes was two for the project, the space complexity of the implementation is O(2pq) = O(pq).

CONCLUSION

Data mining is a technique used to extract vital information from existing huge amount of data and enable better decision-making for the banking industries.

REFERENCES

- Sonntag D, Zillner S, van der Smagt P, et al. Overview of the CPS for Smart Factories Project: Deep Learning, Knowledge Acquisition, Anomaly Detection and Intelligent User Interfaces [M]// Industrial Internet of Things. Springer International Publishing, 2017: 487-504
- Nian K, Zhang H, Tayal A, et al. Auto insurance fraud detection using unsupervised spectral ranking for anomaly [J]. The Journal of Finance and Data Science, 2016, 2(1): 58-75
- Weller-Fahy D J, Borghetti B J, Sodemann A A. A survey of distance and similarity measures used within network intrusion anomaly detection [J]. IEEE Communications Surveys & Tutorials, 2015, 17(1): 70-91
- Pulakkazhy, Sreekumar, and R. V. S. Balan. "Data mining in banking and its applications-a review." Journal of computer science 9.10 (2013): 1252.
- Chitra, K., and B. Subashini. "Data Mining Techniques and its Applications in Banking Sector." International Journal of Emerging Technology and Advanced Engineering 3.8 (2013): 219-226.