

# STUDY OF DATA MINING ON BANKING DATABASE IN FRAUD DETECTION TECHNIQUES

**Sayali Kishor Rodge**

Student, of Master of Computer Application ,YMT College of Management, Maharashtra, India

-----\*\*\*-----  
**Abstract:**

*At present, banking industry in a very huge sector where numerous operations are carried out of the different departments. Hence, banks have a very large database to handle for various purposes such as maintaining personal details of customers to large industries and their respective daily transactions. To retrieve the specific information from this large database and to take important decision, and also to secure data and prevent the frauds data Mining technique is used. This involves the decisions like customer relation management, investment decisions, credit decisions, manual transaction decisions, financial decisions and many more. This paper includes the overall Data Mining technique to overcome the conflicts of bank database, fraud detection, database security and to make the secure transactions from the database.*

understand the issues of data. Data mining is well known as knowledge discovery in database (KDD) which associates with Artificial Intelligence (AI), Decision Tree Approach, Visualization, and Genetic Algorithm.

Data mining technique works by following these major steps:

- 1) Predictive modeling
- 2) Clustering/Segmentation
- 3) Visualization
- 4) Link analysis
- 5) Deviation detection
- 6) Summarization

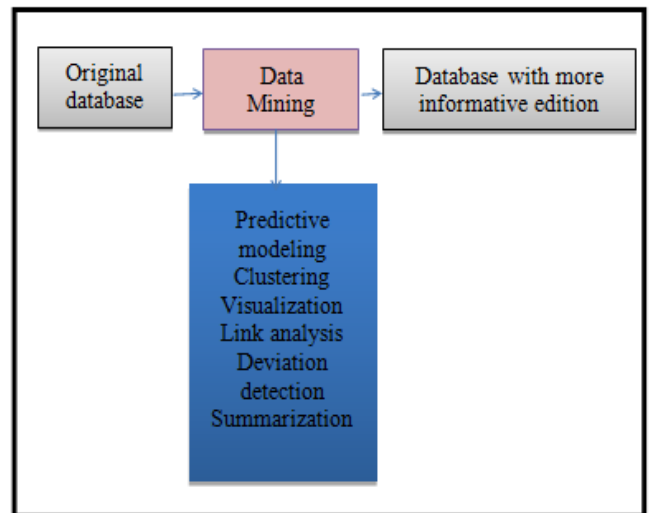
**Keywords:** Data mining operations, Database security, and Decision making, Fraud detection technique.

## 1.1 INTRODUCTION

Banking industry is spreading its branches vastly over a multiple area of various industries with providing more functionality to its customers. Implementing such a large industry requires to stores its huge information in a secured and proper format. Hence in addition it is developing its services rapidly in various operations like providing online transaction or offering various policies to their customer also giving access to multiple other functions. So currently banks have huge data to store which is increasing rapidly.

Handling such a large data leads to security if data which is the most important task of Data mining. Data mining have introduced various techniques and algorithms that will help to focus on important pattern of data from the database. And also helps to take important decisions.

Data mining is important tool which helps to take important decisions by drawing or fetching the information and convert them in some patterns to



## 2. Operations in data mining

### 2.1. Predictive Modeling

Predictive modeling in data mining refers to predicting a particular pattern which will be formed by collecting specific data and which will give additional data

or information about the existing database. Hence it is helpful in detecting frauds as, in bank database the relevant information of frauds which has been done in past records are collected and a similar pattern or statistic model is developed which will predict the future fraud, hence bank can prevent it by analyzing the report generated by predictive model. Predictive modeling can be done in various techniques as Decision Tree Algorithm or Artificial Neural Network Algorithm or Naïve Bayes Algorithm [1].

### 2.2. Clustering/Segmentation

Clustering refers to making a cluster of data from the provided database. It is required when we want to find the same group of data or recognize the same pattern for analysis. Clustering can be performed with number of algorithms such as K-means, K-Medoids, etc. This algorithm will make group of data which is similar to the structure or characteristics called as clusters [2]. Clustering will help in bank database for finding same preference of customers or same type of transactional account holder; hence it can help to draw similar kind of fraud techniques used in fraud detection.

Segmentation is performed to produce finer data patterns. Segmentation can be performed using three different algorithms that is, Sliding window approach, top-down approach, or bottom-up approach. Hence segmentation is performed because it produces finer and clearer clusters than clustering algorithm [3].

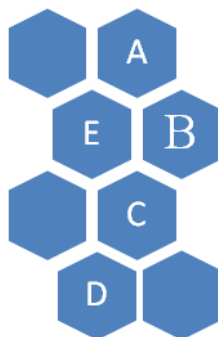


Fig2.2.1 Clustering/Segmentation of database

### 2.3. Visualization

Visualization technique in data mining is introduced for more effective presentation of formed data. Hence study reviews that human brain is more functional to remember visualizing the image rather than

remembering information in data format [8]. Visualization converts the any raw data of characters or numbers to the image; image is in the form of static graph or any kind of graphical representation. Visualization includes techniques such as tree map, scatter plot matrix, parallel coordinates, and spatial visualization [4]. Banking data base is very large and there is possibility of conflicting of data to do the study of banking database visualization technique is very effective. It can draw the tree map of or plot matrix to encounter in which area the actual fraud has been done or the type of customer or employee which have done the fraud also branch location where it has been done. Hence it is useful in locating the fraud area and to prevent it.

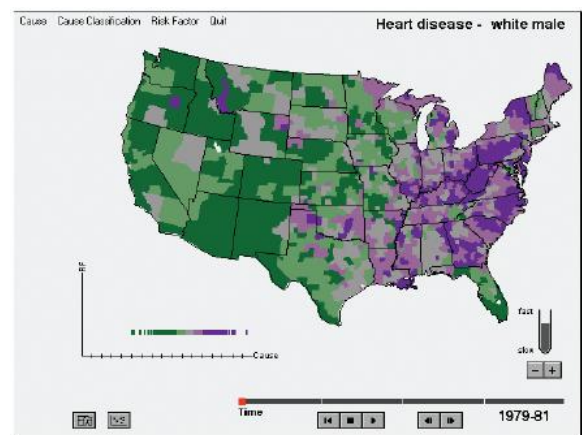


Fig 2.3.1 E.g. Special graph visualization

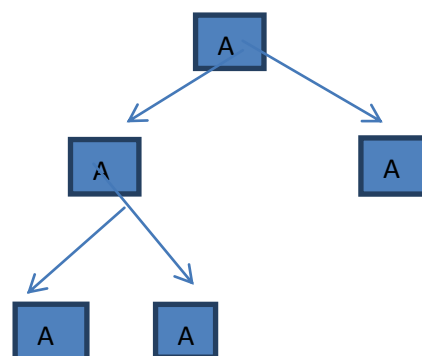


Fig 2.3.2 E.g. Tree Diagram

### 2.4 Link analysis

Link analysis is one of the most important operations of data mining. It works very efficiently to find out the related data to each other. Link analysis is found out the related data i.e. one part of database is linked with

other part and that connection is established by link analysis. In banking database customer linked with account then account linked with transactions further it link with type of transactions and this will continued.

Link analysis is based on part of mathematics called as graph theory where edges are connected to each other by some vectors to find out the correct path or some particular pattern of given data. It helps in fraud detection in many ways as bank employee can link the fraud detection area with each other and can summarize the data to find out the exact problem and their solution.

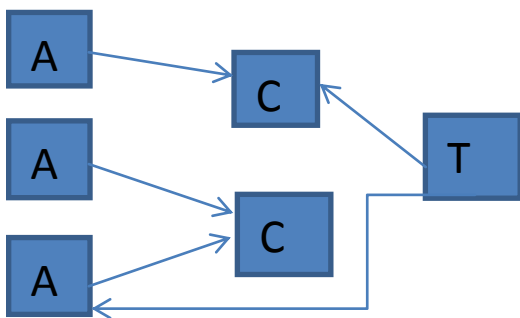


Fig 2.4.1 E.g. Link analysis

A-Account C-Customer T-Transaction

## 2.5. Deviation detection

Deviation detection is refers to identifying the errors or noise or exception which occurs in database. Clustering algorithm will only cluster the pattern of deviated data and non- deviated data, clustering algorithm focuses on eliminating the exception from the database where as deviation detection algorithm focuses on isolating those exception for better performance and error free data. Deviation detection is helpful while testing the many application of bank like credit card fraud can be isolates or rectify by deviation detection algorithms. Deviation detection algorithm considers the sequential exception problem to work on large databases; hence it discovers the all possible exceptions on every single dataset [5].

## 2.6. Data summarization

Data summarization is the major part of mining techniques at this will give the final outline to the data which we gathered from the previous operations i.e. relations between the clusters and the dependency of the

subsets of data are clarify to take the important decisions. Hence data summarization will give idea about which product should be brought together and which should be avoided for better performance also to avoid the conflicts of data and to prevent the data from fraud and provide security to bank database. Data summarization is process of generating the better and more informative version of original database [7].

## 3. Issues in bank database security

### 3.1. Loss of integrity

Maintaining data integrity is the superior task among all other task for bank. Loss of integrity leads to improper data maintenance and that leads to frauds in banking area. When improper modification made by some threat or by some person it results into loss of data integrity. For example, bank may have many credit card holder with same name in different branch hence to maintain the integrity in this case card holders name should not conflict and data should

Be preserved in a proper format.

### 3.2. Loss of confidentiality

Confidentiality is lost when some private data is handled by some unauthorized user or threats. Bank database holds all the personal as well as confidential information of their customers. Also it maintains the records of various industries, trust, companies with whom they have tie up. Hence, losing such valuable information leads to a very big loss of banking industries.

### 3.3. Maintaining data quality

Maintaining data quality involves securing data from outsiders and including the data which is informative and useful for the organization. Including information or data which is not required anymore will only results in conflict in future database.

### 3.4. Financial crime detection

Today most bank industries face the problem of financial crime i.e. customers get hacked by some unknown threat likewise credit card PIN is also hacked by some threats which will leads to financial crime.

## 4. DATA MINING TECHNIQUES IN FRAUD DETECTION

### 4.1. Limiting access

Using data mining algorithm confidentiality of bank database can be maintained by clustering the group of authorized employee of bank which will handle the bank database [6].

### 4.2. Eliminate unnecessary grouping

When some fraud happened in banking industries in a particular area to detect the actual theory behind it and to prevent it data mining can be used i.e. bank database will compare its own database patterns with other database where fraud has been detected and the similar patterns will generate using clustering algorithms. Those clusters will eliminate in deviation detection algorithms to avoid the future errors and exceptions [6].

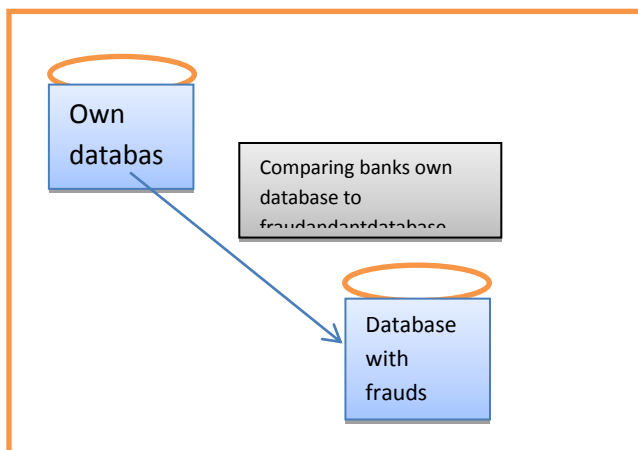


Fig 4.1 E.g. Detecting frauds

Fig.4.2 E.g. Comparing databases

### 4.3. Priority mechanism

To avoid the fraud bank database can be limited to some authorized people i.e. data mining help to cluster the database information into priority basis and that information will link to particular authorized employee using link analysis this will secured the data in that loop only and if any fraud is detected it will be easier to detect because of link analysis system.

## 5. CONCLUSION

Data mining is a very important tool to prevent fraud and detecting fraud activities in bank related database. Operations performed in data mining are used to give security to database and to enhance the decision making power. It fetches the important pattern from the large database which will help in improving database quality. This research paper includes almost all the issues related to banking database security and how data mining is used to overcome those techniques. Paper also focuses on important algorithms like K-means or K-Medoids etc. this will give proper way to extract the data.

## 6. REFERENCES

- [1] Evaluation of Predictive Data Mining Algorithms in Erythematous-Squamous Disease Diagnosis. Kwetishe Danjuma1, Adenike Osofisan2
- [2] CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling. George Karypis, Eui-Hong Han, Vipin Kumar, IEEE Computer 32(8): 68-75, 1999 (<http://glaros.dtc.umn.edu/gkhome/node/152>)
- [3] Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms Stan Salvador and Philip Chan Dept. of Computer Sciences Florida Institute of Technology Melbourne, FL 32901
- [4] Visualization Techniques for Data Mining in Business Context: A Comparative Analysis Ralph K. Yeh University of Texas at Arlington
- [5] A Linear Method for Deviation Detection in Large Databases Andreas Arning IBM Germany
- [6] Security and privacy implication of data mining By Chris Clifton and Don Marks.
- [7] Summarization - Compressing Data into an Informative Representation VarunChandola and Vipin Kumar Department of Computer Science, University of Minnesota, MN, USA.



[8] A Review of Data Mining Techniques ARTICLE in  
INDUSTRIAL MANAGEMENT & DATA SYSTEMS ·  
OCTOBER 2001 Impact Factor: 1.23 · DOI:  
10.1108/02635570110365989 · Source: CiteSeer