International IT Journal of Research (IITJR), ISSN: 3007-6706
Volume 3, Issue 2, April- June, 2025
Available online at: https://itjournal.org/index.php

Approaches to Hybrid Data Mining for the Determination of Financial Transaction System Fraud

Fasi Ahmed Parvez Mohammad[1], Dr. Manisha[2]

[1]Assistant Professor Supervisor , Department of Computer Science & Engineering, JS University, Shikohabad, UP
[2]Research Scholar , Department of Computer Science & Engineering, JS University, Shikohabad, UP

ABSTRACT

The detection of fraudulent activity has become an important issue for financial institutions due to the rise in the risk of fraud brought about by the exponential expansion of online financial transactions. One of the biggest problems with using typical fraud detection algorithms is handling large volumes of skewed, multi-dimensional transaction data. Finding ways to successfully detect false information in financial transaction systems is the driving force behind this study. One approach is to use hybrid data mining approaches. The suggested solution employs a number of data mining techniques, including clustering, classification, and anomaly detection. Using these methods, we can increase detection accuracy and decrease false positives all at once. Experts in the field use a wide range of supervised and unsupervised machine learning algorithms, including as support vector machines, decision trees, and neural networks, to detect common and unusual forms of fraud. To improve the model's efficacy and performance, two processes are performed: preprocessing the data and feature selection. If you compare the hybrid method to the single-model approaches, you'll see that the latter are less reliable, precise, and easy to remember. When contrasted with more traditional approaches, this becomes clear. Financial institutions may benefit from the suggested method's ability to help them reduce losses and increase transaction security by providing a scalable solution for real-time fraud detection.

Keywords: Fraud Detection, Hybrid Data Mining, Financial Transactions, Machine Learning

INTRODUCTION

The term "data mining" refers to the practice of extracting useful insights from large datasets that were previously inaccessible. We can only hope that this novel strategy for extracting vital information from data warehouses pans out. Data mining allows companies to look forward with greater precision and become more knowledge-driven and strategic. Supplementing the event analysis generated by decision support system demonstration devices was the initial goal of using data mining to provide erroneous predictions about the future. It may take a lot more time to discover answers to company problems without data mining technologies. Record searches for analytical data and unexpected patterns are now much easier thanks to this. A primary objective of data mining is to discover trustworthy patterns in data that have been overlooked. Data mining research mostly aims to discover similarities in large datasets. The main goal is to find patterns in the data that weren't there before. The use of data mining methods is widespread across many different sectors. Methods such as regression models, advanced neural networks, classification, grouping, and prediction are examples. Financial fraud detection (FFD) relies on data mining to unearth previously unknown insights inside large datasets. The goal of data mining is to find actionable insights hidden in massive databases by means of systematic pattern recognition and statistical analysis. The term "data mining" describes the process of gleaning useful information from massive databases by the use of mathematical, statistical, and machine learning algorithms. Data mining is a technique for discovering new and valuable information in large datasets by extracting patterns and correlations. Data mining has several applications, one of which is the development of new models that might identify new dangers before people do. Detecting fraudulent actions is one of the most major uses of data mining, which can be seen in both the public and commercial sectors. A wide variety of approaches are used by the FFD inside its data mining architecture.

One of data mining's primary functions is to uncover and foil money laundering activities. Using a data mining tool to find signs of fraudulent activity in bank accounts is the most important part of detecting fraud. Disadvantages of fraud monitoring I would be very grateful if you could provide some light on the reasonable expectations of the banks about the transfer. A user's unique account details feed the anti-money-laundering strategy. Most people use their bank accounts when they shop, whether it's at a brick-and-mortar store or online.

Since money laundering is a severe danger to financial institutions and a criminal offense, the national government should pay attention to this matter. Although most banks do have damage prevention measures in place, regulatory bodies have found them to be insufficient. We have implemented message padding and other security measures to strengthen protection, however we have not yet implemented failure detection based on money laundering. When it comes to data mining, the usual steps that follow feature selection are modeling, data collection, management, and success evaluation. Many cases of fraud and money laundering have been discovered lately using data mining.

A. IDENTIFYING DECEIT IN THE MONEY-LAUNDERING PROCESS

Fraud is defined as telling a falsehood with the aim to deceive another person into parting with their money, damaging their reputation, or committing a crime. Systematic methods for identifying and avoiding fraudulent actions and losses are the two most potent weapons in this fight. One sensible way to prevent fraudulent transactions is to stay away from fraudulent activities. In order to detect any potential instances of unlawful activity, fraud monitoring software examine each and every transaction. When it comes to avoiding fraud, this doesn't matter. Furthermore, it identifies the ones that are not genuine even before the con artist has a chance to complete their deception.

The term "fraud identification" refers to the steps used to spot fraudulent activity where none existed before. First, we check whether the data trend is legitimate. In order to handle early scam data, a method called supervised learning is used. The focus of unsupervised learning focuses on data that may be used for criminal or fraudulent purposes but isn't really scam data. Unsupervised learning relies on this data. A lot of jargon is required for this task, and some of the terminology used are indicators of dishonest or unlawful behavior.

An increase in the frequency of fraudulent schemes, resulting in substantial financial losses, has accompanied the development of better information networks and IT. But there are a lot of different ways fraud may happen, such over the phone or online. Due to its universal accessibility, users' ability to hide their location, and the anonymity it provides for transactions, the Internet is ripe with opportunities for fraud. The proliferation of lightning-fast internetworking channels has made it easier for scam artists to plot their schemes. Through these mediums, they are able to communicate with people from all over the globe and share knowledge with them.

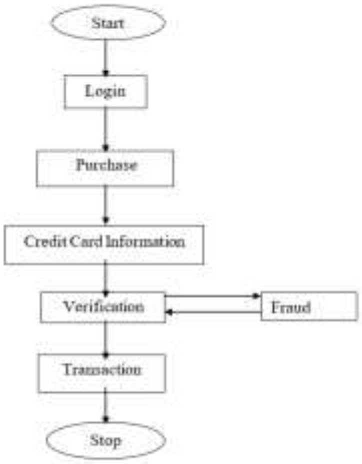


Figure: Flowchart for Credit Card Fraudulent Detection

LITERATURE REVIEW

Money laundering (ML) activities might be better detected and prevented with the use of data mining tools. An essential aspect of the ML method is investigating account user characteristics. Some odd behavior from the bank account indicates that anything is wrong. Methods for detecting fraud do not include using realistic concepts from machine learning banking. In order to identify individuals involved in money transfers, this research suggests a technique known as Probabilistic Relational Model and Audit Sequential Pattern Mining (PRM-ASP). For many-to-one and one-to-many file conversions, you may use association mapping (AM) files, which are a kind of data set creation tool. Nevertheless, the characteristic is said to become apparent when it is unable to provide scalability and flexibility in the offender identification process. To determine the degree of adaptation risk in money laundering, one might use the Bitmap Index-based Decision Tree (BIDT).

The BIDT learning method allows for the creation of knowledge trees, the identification of potential risks associated with machine learning, and the facilitation of progress. It is possible to utilize a BIDT to compromise the security of large financial institutions. As an alternative to a list of rowids, a BIDT bitmap index makes use of a collection of bits known as a bitmap. Within the context of this index type, a distinct sequential number is given to each key value (such an account number) kept in a database. An approach to efficiently detect money laundering based on association rule patterns (EARM-MLD) was developed in response to the challenge of handling highly dimensional data that is structured in several clusters. There are three main parts to the association rule pattern mining system in the EARM-MLD design. under order to back up and safeguard the data several times, it is necessary to first identify the common major item groupings under banking laws. One day, we may be able to use these enormous datasets to construct association rules based on spatiotemporal models. This will simplify discovery and decrease the quantity of false positives when used with a multi-clustering strategy. Lastly, the multi-clustering method makes advantage of a plethora of qualifying money transfer groups.

3. THE USE OF PRASP IN THE FIGHT AGAINST MONEY LAUNDERING

The identification of money laundering (ML) presents a significant challenge to data mining techniques that seek to streamline transactions. Among data mining's many applications in financial accounting is the identification of possible instances of misconduct. Important machine learning banking factors may not get enough attention in the fraud detection method. There is a financial database agreement with a capital "K" that ML cannot comment on. Log data kept by the user's account is used by the ML. Whether they do it online or off, most people who engage in commercial activities utilize bank accounts. The underlying workings of machine learning are a crucial influence in the challenges affecting the financial system.

This approach to financial fraud detection uses common traits to classify data mining jobs and tackles fraud detection-specific difficulties. On the other hand, the feedback provided by ML banks is not documented. For the same reason, the Joint Threshold Administration (JTA) Model key controls kernel function dependent financial systems. The production of transactions and replies is based on irrelevant database information in an effort to circumvent machine learning.

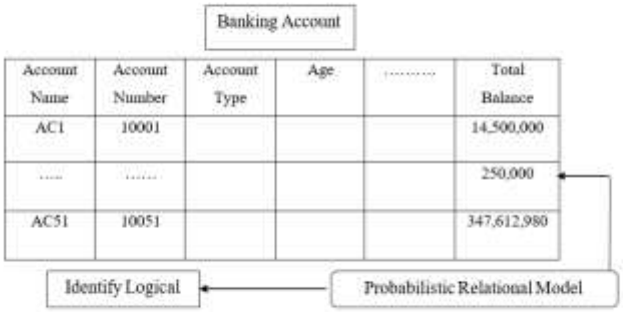
A probabilistic relational model with audit sequential pattern mining, or PRM-ASP, is what we provide to make machine learning discovery more effective. Dividing the work into many-to-one and one-to-many accounts is possible using the Association Mapping (AM) technique. An overarching objective of PRM-ASP mining is the discovery of new machine learning methods for use with time series data. Several kinds of link accounts between activities are identified to ensure that weak accounts are recognized. A PRM is necessary for the identification of possibly susceptible bank accounts as well as for the assembling of machine learning accounts. Both relationship thinking and ASP are used by the PRM-ASP to find trends in accounts that are vulnerable to hacking.

A. RELATIONAL PROBABILITY MODEL

The money transport amongst dissimilar bank clients are addressed in PRM-ASP mining and relational logic is analyzed. In the data mining step, AM are used for discovering the PRM logic. The PRM is represented in figure 3.3.

Figure 3.3 explains the money transaction to the different banking accounts with the different time frame. Transaction relational logic of ACC_15, ACC_12 and ACC_54 are tested using the PRM. Data mining step predicts susceptible account and amount of data transferred is calculated.

B. PRMUSINGTHEASPMININGONMONEY LAUNDERINGDETECTION



Account Name	Account Number	Account Type	Age	....	Total Balance
AC1	10001				14,500,000
.....	....				250,000
AC51	10051				347,612,980

PRM-ASP Mining determines the ML accounts in the bank dataset. ML is an illegal action for financial institutions and hence become a major risk to the entirenation, so that PRM-ASP mining is used to discover the faulty bank accounts.

From the figure 3.1, logical relationships among client information is employed to recognize the ML by using PRM-ASP.

PRM-ASP mining is achieved based on the personal information of the clients. The relational logic and ASP are extracted from the client and banking companies. PRM is used to explain the associations among the objects.

4. ESTIMATION OF THE RISK OF MONEY LAUNDERING

Finding many-to-one and one-to-many correlations between transactions is the goal of money laundering detection (ML) using time series data. This makes it possible to identify potentially exploitable accounts. By drawing on similar logic, the audit sequential pattern (ASP) may identify potentially susceptible account transactions. In this specific instance, ASP also makes good use of a Probabilistic Relational Model to provide an acceptable machine learning identification. Banks and other firms dealing with money are gravely threatened by criminals' use of ML. Many financial institutions do not adhere to the criteria set by regulators, despite the fact that the majority of them have security safeguards in place. Machine learning has the potential to uncover security holes, even after security measures have been implemented, making approaches like message padding more safe. A sensible trade-off between performance and security comes at the cost of a flawed financial structure. This property is considered vulnerable since it does not have the scalability and flexibility needed for ML crime detection. The word "ML" is used to describe the process of making money that seems to be legitimate but is really gained illegally. The term "ML" has expanded beyond its initial definition to include a wide range of financial crimes handled by different legal and judicial systems. Its original intent was to depict misuse of the financial system. Offenders now have additional options to conceal their identities because to the abundance of data accessible online, which has considerably improved the accuracy of crime scene identification.

To evaluate the adaptation risk linked to money laundering, the BIDT approach is recommended. In order to improve scalability and reduce machine learning risks, BIDT learning relies on building an information tree. With bitmap indexing, BIDT can quickly and easily access large volumes of financial data. The BIDT table descriptions are as follows:

Each key value (the account number) and row IDs are represented by a sequentially numbered bitmap, an array of bits, rather than a catalog. The BIDT approach then uses the "select" query performance to apply bit-wise and count-wise logical operations on AND variables. It is possible to acquire a better picture of the adaptation risk in ML methods if the query responses are combined to form a decision tree. All that's needed to get the population frequencies for the BIDT root node is to include the total number of "1" into the bitmap structure. One can predict occurrences of money laundering and evaluate the risk factor rate using this skill.

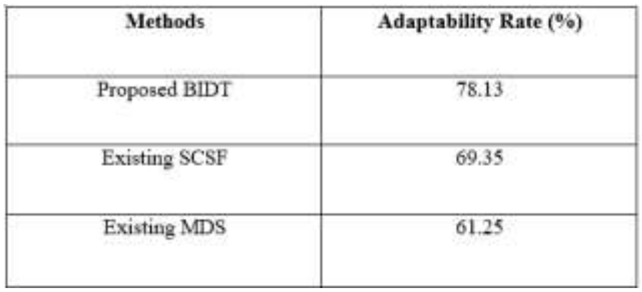
A. DECISION TREE FOR BITMAP INDEX-BASED RISK ASSESSMENT OF FINANCIAL MONEY LAUNDERING

The risk factors are estimated on financial organizations ML using the indexing scheme. The indexing scheme uses the rows and columns to store the information that improves the scalability rate. ML is the illegal amount transacted between different users, which are evaluated using the BIDT technique. The risk related to larger amount of illegal transaction is controlled in a financial organization by constructing a decision tree with mapping of the bit in fuzzy form ‘0’ and ‘1’. The decision tree contains the root and sub co-ordinate nodes to create the determination rules in the BIDT technique. The purpose of indexing in BIDT technique is to make available pointers to rows in a table consisting of given key values.

B. PERFORMANCEANALYSISOFADAPTABILITYRATE(AR)

AR on crime using BIDT technique is the capability of service provider to alter changes in services based on customers’ requests during ML operation. Adaptability of offense measures the time taken for ML changes or updates the service in higher level at less interval of time. Higher the adaptability rate, more quickly, the anti ML system is and therefore is said to be more efficient in handling the ML operations. It is measured in terms of percentage (%).

Table:TabulationforAdaptability Rate



Methods	Adaptability Rate (%)
Proposed BIDT	78.13
Existing SCSF	69.35
Existing MDS	61.25

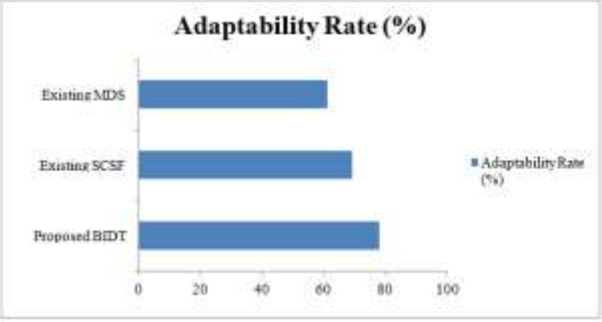


Figure 4.7 Measureof adaptability

EXPERIMENTAL EVALUATION

Efficient Association Rule Pattern based Money Laundering Detection (EARM- MLD) framework is experimented in JAVA language using Statlog (German Credit Data) Data Set. The Statlog German Credit Data classifies the people using a set of attributes list. To efficiently implement the algorithms in EARM-MLD framework, numerical attributes from Strathclyde University are added to make it effectivealgorithm for ML identification. The Statlog German Credit Data include 17 attributes and has been coded as integer type and 3 under the categorical type.

The Statlog German Credit Data contains the 1000 instances on financial area for performing the experimental work to identify the vulnerable accounts. The proposed Efficient Association Rule Pattern based Money Laundering Detection framework is compared with existing methods namelyAnomalyDetection usingPrincipal Component Analysis and Detecting and Investigating crime using Data Mining. To evaluate the Efficient Association Rule Pattern based Money Laundering Detection framework, the following metrics are evaluated.

i) Timefordetectingmoneylaundering
ii) Falsepositiverate
iii) Fraudidentificationaccuracy
iv) Systemefficiencyratio

RESULTS

Finding the accounts that are engaged in money laundering (ML) is the reason for establishing the PRM-ASP Mining model. The preprocessed data set is run through the AM algorithm to decouple the transactions from the different account types. By identifying different kinds of accounts in transactions, machine learning identification may utilize time series data to locate susceptible accounts. To identify potentially insecure bank accounts, it compiles ML accounts and applies PRM to transaction classification. The PRM, which ASP employs to provide a logical ML identity, is also useful in this case.

In machine learning, the adaptation risk is evaluated using the Bitmap Index-based Decision Tree (BIDT) technique. Before BIDT learning inducts a knowledge tree, it boosts the scalability and determines the ML risk to the enterprise. Utilizing a bitmap index, BIDT efficiently gains access to large financial datasets. A bitmap array of bits represents each key value (e.g., account number) in a table's description in a BIDT instead of a list of rowdies. Then, using the "select" query performance, the BIDT approach applies AND with count and bit-wise logical operations. Using the query results to construct the decision tree improves the accuracy of assessing the adaptation risk in ML operations. The decision tree's main node, or root node, can do full "1" computations using bitmap architecture, and it may utilize population frequencies to predict money laundering and evaluate the risk factor rate.

We developed a powerful money-laundering detection system using association rule patterns to handle high-dimensional data with a multi-clustering structure. In order to mine association rule patterns, the EARM-MLD framework mainly uses three parts. We begin by searching for large item sets with confidence and support values over a certain level that are often used in banking rules. The time needed to detect ML is therefore reduced. Using the spatio-temporal model, we can then construct an association rule from those large collections of items. Integrating with a multi-clustering algorithm, it performs the detection operation with ease while aiming to reduce the false positive rate. Last but not least, the multi clustering technique makes use of the set of qualified money transfer groups. The combination of the EARM-MLD framework with multi-cluster components is seen as questionable behavior while doing ML detection tasks.

CONCLUSION

Successfully determining money laundering accounts with a minimal false positive rate is achieved using the PRM-ASP Mining model. The first step is to use the AM algorithm to partition the transaction process. The mapping technique effectively identifies the transactions involving many-to-one and one-to-many accounts. The Probabilistic Relational Model is a collection of relational logic transactions used to classify accounts as susceptible. The PRM-ASP Mining model improves the audit sequential pattern to classify the route of monetary transfers. The PRM-ASP mining model also provides a logical structure for use in a wide variety of practical settings. Improving the accuracy of fraud detection with little time is achieved by performance analysis of the PRM-ASP mining model. Last but not least, the PRM-ASP Mining model improves account monitoring processing time while decreasing false positive rates.

To find the flexibility risk in money laundering, the Bitmap Index-based Decision Tree (BIDT) method is suggested. The anti-money-laundering strategy now revolves on preserving regulatory risk rate and protecting financial institutions. Time spent detecting money laundering risks is decreased if the level of the real positive rate (i.e., regulatory risk rate) is improved. To assess the impact of regulatory risk rate on performance, Bitmap Index-based Decision Trees are used. By classifying the rows and columns according to the customer's account data, the bitmap indexing approach effectively decreases the time it takes to identify risks and substantially increases its adaptability rate. Bitmap Indexing was first used to enhance the regulatory risk rate; it easily handles big money laundering accounts and delivers results in fuzzy form. After that, in order to reduce the number of false positives, a Select Query Structure is created using many key-value databases that collaborate using a bitwise logical operator. Bitmap Index Frequency, which includes the identifiers for the rows and columns, was subsequently included as well.

Use Statlog's German Credit Data to improve the genuine positive rate using a low cardinality column.

To deal with high-dimensional data with a multi-clustering structure, we built an efficient money-laundering detection system based on association rule patterns. There are primarily three components to the EARM-MLD framework's association rule pattern mining. At first, we look for big itemsets that appear often in banking regulations and have confidence and support values that are greater than a certain threshold. As a result, the time needed to identify instances of money laundering is decreased. The next step is to use the spatio-temporal model to build an association rule from those massive item sets. With the goal of lowering the false positive rate, it combines with a multi-clustering algorithm and effortlessly conducts the detection operation. At last, the multi-clustering method incorporates a collection of transfer groups that meet the row criterion, collecting funds for a specific account with a minimal set size.

The suggested PRM-ASP mining model improves processing time for user account monitoring by 8% and decreases the false positive rate by 22% when utilizing Statlog German Credit Data for money laundering detection. The bitmap indexing approach significantly enhances the adaptation rate and decreases the risk identification time by 21%. Finally, when compared to state-of-the-art approaches for identifying money laundering, the EARM-MLD framework's use of a mapping algorithm improves performance, leading to a 15% increase in the system efficiency ratio and a 9% improvement in fraud detection accuracy.

REFERENCES

[1]. AashleshaBhingarde, AvnishBangar, Krutika Gupta and SnigdhaKarambe International Journal of Advanced Research in Computer and Communication Engineering, Volume 4, Issue 3, March 2015, Pages 169   170.

[2]. Andrei Sorin-122 Response System for Relational Databa Engineering, Volume 23, Issue 6, June 2011, Pages 875 888 International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 2, February 2015, Pages 997   1000.

[3]. Clifton Phua, Kate Smith-Miles, Vincent Cheng-Siong Lee, and Ross Gayler, Engineering, Volume 24, Issue. 3, March 2012, Pages 533-546

[4]. Data Mining in Money Laundering Detection Springer, Volume 7197, Pages 207-216

[5]. Gilbert Sebe- Performance of Agricultural Develo Accounting Auditing and Finance Research, Volume 2, Issue 1, March 2014, Pages 1-23

[6]. Mihaela A. Bornea, Vasilis Vassalos, Yannis Kotidis, and Antonios Deligiannakis, Transactions on Knowledge and Data Engineering, Volume 22, Issue 8, August 2010, Pages 1110   1125.

[7]. Roberto Cortinas, Felix C. Freiling, Marjan Ghajar-Azadanlou, Alberto Lafuente, Mikel Larrea, Lucia Draque Pe

[8]. Volume 9, Issue 4, July/August 2012, Pages 610-625

[9]. Rui Liu., Xiao-long Qian., Shu Mao., Shuai- ch on anti-money Conference (CCDC), 2011, Pages 4322 4325

[10]. SutapatThiprungsri, and Miklos A. Vasarhelyi Accounting Research, Volume 11, 2011, Pages 69-84

[11]. Tamer Hossam Eldin Helmy , Mohamed zaki Abd-ElMegied, Tarek S. Sobh,Laundering an Applications Volume 1, Issue 1, November - December 2014,

[12]. for Secure Computations with Two Non-Colluding Servers and Multiple Clients, Security, Volume 10, Issue 3, March 2015, Pages 445 457.

[13]. Mohammad Reza Keyvanpour, Mostafa Javideh and Mohammad Reza Ebrahimi, d investigating crime by means of data mining: a general crime matching 880.