IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Study of Analytical Methods in Banking Transactions

Ishan Padalkar[*], Pranav Kulkarni[†], Mahesh Kulkarni[‡] and Sukrut Pendharkar[§] Prof. Anita Shinde[¶]

Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Savitribai Phule Pune University Pune

Email: [*]ishandeva@gmail.com, [†]pkulkarni0410@gmail.com,

***Abstract*—In this day and age of digital transactions, massive transaction records are generated everyday. This poses a big problem in the sense that there is no robust system for extensive analysis of the same. With the introduction of every new technology there comes it's misuse. There are thousands of fraudulent transactions done everyday and it becomes impossible to identify them among the huge data that is collected. This paper outlines a survey framework for a new Transaction Analysis System. The proposed approach meets custom auditing needs with respect to categorization of narrations and applying basic validations. We intend to further extend this work to utilize anomaly detection techniques which can be applied for transaction authentication.**

***Index Terms*—transaction analysis, categorization of records, suspicious transaction detection, profiling**

I. INTRODUCTION

Enterprise risk management is increasingly important in all aspects of business. Numerous researchers and practitioners have attempted to identify methods to monitor and control enterprise risks. Hazard risk, financial risk, operational risk and strategic risk are the typical risks [5] [23]. Because of the importance of financial statement fraud, identification and management of the risks associated with financial statement fraud help to reduce enterprise risks.

The key contribution of this paper is that it provides a structured and broad overview of prerequisites that need to be satisfied before performing data mining operations on large financial records.

Firstly, accuracy of financial records. The two main sources of financial statement inaccuracy are deliberate dishonesty and human error. The system needs to understand the context for reporting these suspicious transactions. Secondly, knowing that all balance sheets are reconciled. Here we are concerned with reconciliation of amounts meaning deposited and withdrawal amount must reconcile with balance amount.

A. CLOUD

Cloud computing is the on-demand delivery of IT resources over the Internet with pay-as-you-go pricing. Rather than purchasing, owning, and keeping up physical server farms and servers, you can get to innovation administrations, for exam- ple, processing force, stockpiling, and databases, dependent upon the situation from a cloud supplier like Amazon Web Services (AWS).

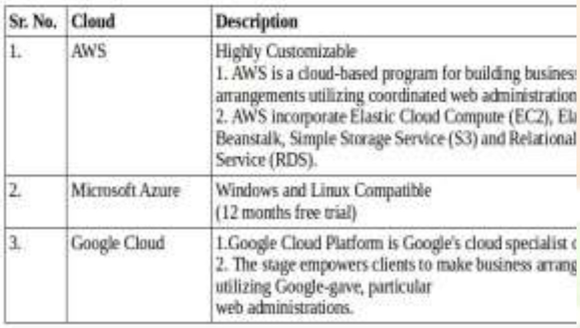
Amazon Web Services

Amazon Web Services (AWS) is the world's most compre- hensive and broadly adopted cloud platform, offering over 175 fully featured services from data centers globally.

Most secure- AWS is architected to be the most flexible and secure cloud computing environment available today. Our core infrastructure is built to satisfy the security requirements for the military, global banks, and other high sensitivity organizations.

Amazon Web Services includes functionalities like:

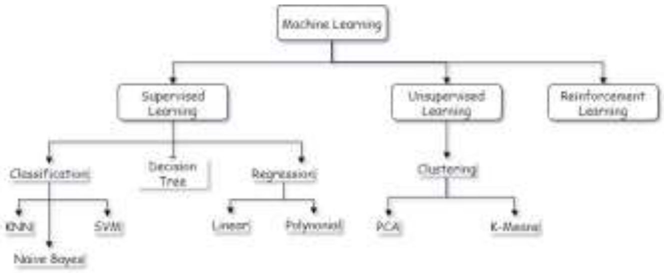
1) Amazon Elastic Compute Cloud
2) Amazon Elastic Block Store
3) Serverless Computing
4) Lambda functions
5) Batch Computing
6) Amazon Simple Storage Service
7) Web Server
8) Docker and Container



Sr. No.	Cloud	Description
1.	AWS	Highly Customizable 1. AWS is a cloud-based program for building busines arrangements utilizing coordinated web administratio 2. AWS incorporate Elastic Cloud Compute (EC2), Ela Beanstalk, Simple Storage Service (S3) and Relational Service (RDS).
2.	Microsoft Azure	Windows and Linux Compatible (12 months free trial)
3.	Google Cloud	1.Google Cloud Platform is Google's cloud specialist 2. The stage empowers clients to make business arrang utilizing Google-gave, particular web administrations.

B. *DATA MINING*

There are several data mining techniques, and most have been used in data mining research projects.



Classification

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. The data classification process involves learning and classification. Classification refers to identification of categories where the new data items belong.

Clustering

Clustering mainly focuses on grouping of similar data items. It is a type of Unsupervised Learning in machine learning. Clustering can be said to be the identification of similar classes of objects. In this technique, transactions with similar behavior are combined into one group. [10]

Prediction

Prediction as its name implies is one of the data mining techniques that discover relationship between independent variables and relationship between dependent variables. based on historical data, we can draw a fitted regression curve that is used for attempted fraud prediction. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data min- ing, independent variables are attributes already known and response variables are what we want to predict.

II. MOTIVATION

The financial services sector is on the eve of a major trans- formation, and the driving force behind it is AI. Innovative applications for AI have already been found across areas such as credit scoring, regulatory compliance, customer experience, and portfolio management. Thanks to rapid advancements in technology, tasks that once took employees hours to com- plete manually, can now be done in a matter of seconds. Traditionally, banks and financial institutions have approached fraud detection with manual procedures, or rule-based solu- tions, which have been limited in their success. A rule-based approach means that a complex set of criteria for flagging suspicious transactions has to be established and reviewed manually. While this can be effective in discovering anomalies which conform to known patterns, it is not capable of detecting fraud which follows new, or unknown patterns. This gives criminals the incentive to develop ever more sophisticated techniques to circumnavigate the rules, and they themselves are leveraging new technologies to achieve that is helping banks and financial institutions get one step ahead, is machine learning. Applying machine learning to fraud detection enables financial firms to identify genuine transactions versus fraudulent transactions in real time, and with greater accuracy. Through a combination of supervised and unsupervised methods, models are capable of learning and recognizing new patterns that may have been missed by other approaches to fraud management. This will also be useful to generate visual representations in the form of graphs, charts etc.

this. The solution

III. RELATED WORK

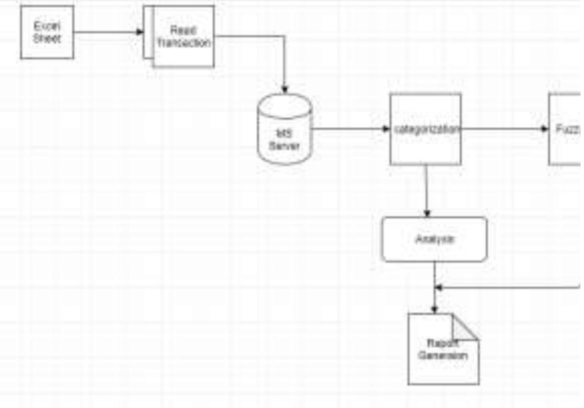
In order to maintain the high standards of security amidst the overwhelming flow of big banking data and the rapidly growing scale and complexity of cyber crimes, researchers have been exploring advanced DM techniques for effectively identifying unusual fraud behavior. Note that an existing review that targeted credit card processing can be found in [22]. From an internal perspective, the survey data of bank employees in India are collected in [3] to analyse their perceptions with regard to fraud.

Many researchers worked with transaction data, seeking better approaches to distinguish between patterns from genuine behavior with higher efficiency and accuracy [6], [21]. Among these, Wei et al. [21] proposed a framework named i-Alertor for major Australian banks; a semi-supervised decision support system named BankSealer was proposed in [4] for an Italian bank; authors in [13] proposed a hybrid DM method to predict network intrusions and detect fraud activities; FraudMiner model that integrated frequent itemset mining was introduced in [18] and verified with the data set from UCSD DM contest 2009; a comparative study [24] addressed the ensemble approach to build classifiers; in terms of a recent advancement in FraudMiner, the authors in [8] introduced the LINGO clustering technique [14] for the pattern matching process, and this enhancement helped maintain a satisfying performance in terms of accuracy while further reducing the false alarm rate; Behera and Panigrahi [2], [3] demonstrated the hybrid approach for credit card fraud detection by combining Fuzzy Clustering and NN techniques, and achieved over 93in [15]; APATE was proposed in [20] for automated fraud detection within a large credit card issuer in Belgium; both Luhn's and Hunt's algorithms were employed in [17] for proposing a novel system of credit card fraud detection; the authors in [10] illustrated the use of DM techniques on customer data to add a higher level of authentication to banking processes for real time fraud detection; a framework named FDiBC was developed in [31] for fraud detection within the Saman Bank in Iran; an e-banking security system employing Cryptogra- phy and Steganography was introduced in [6] for preventing online banking fraud. Apart from the main implementations on transaction data, the authors in [1] focused on phishing detection from official banking websites and applied a multi- label classifier based associative classification DM for effective detection of phishing in websites with high levels of accuracy. In order to improve the customer credit card churning prediction for a Latin-American bank, the authors in [19] adopted improved DM techniques that are based on K-means clustering and support vector machines (SVM). Blog mining (text mining and cluster analysis) was applied in [7], where security risks, protection strategy and security trends of mobile banking were summarized from more than 200,000 results of the Google blog search engine. There are also researchers who paid extra attention to money laundering detection. For instance, a DM model is presented in [11] that applied K-means clustering and Association Rule Mining for identifying suspected sequence of money laundering processes. A novel technique named Bitmap Index-based DT was proposed in [9] for evaluating the risk factor of money laundering with Statlog German credit data.

IV. PROPOSED SYSTEM

In this scenario we propose a web based solution that serves two purposes namely, analysis and flagging suspicious narrations. This will be based on a machine learning model that will operate on the uploaded datasets.

There are different data mining tools and techniques available which are used as one of the domains of data mining which can excel at, suspicious transaction monitoring. Principle component analysis proposed in [16] aimed to represent each sample of transaction with few number of values so that attributes could be reduced by determine attributes contains major information and facilitate faster fraud detection for credit card transactions. Another algorithm which is LINGO produces reasonably described and meaningful clusters when implemented into the Carrot2 framework that significantly influence the quality of clustering. LINGO algorithm implemented in [12] and proposed this algorithm as an approach for clustering could be used for outlier detection as the proposed algorithm idea is to first find meaningful descriptions of clusters.



The system has following stages:

1) **Input Data Acquisition**: The data input is in the form of excel documents.
2) **Preprocessing**: The preprocessing Consists of a series of operations performed on the input data, which includes background noise reduction, empty data restoration, filtering etc. Various ETL tools and
3) **Attributes Mapping**: This step deals with matching of data attributes from input database with the dynamically generating database. This module involves the identification of the similar attributes used for further steps.
4) **Fuzzy name matching**:

- Fuzzy matching is the implementation of algorith- mic processes to determine the similarity between elements of data such as business name, personal name or address information.
- It is an approach to computing based on "degrees of truth".
- The fuzzy logic feature allows the algorithm to detect and evaluate near matches rather than require exact matching. Depending on the algorithm, it may consider alternate nicknames, such as "Mike" or "Mickey."

5) **Levenshtein Distance**: The Levenshtein distance is a string metric for measuring the difference between two sequences. Informally, the Levenshtein distance between two words is the minimum number of single-character edits (i.e. insertions, deletions, or substitutions) required to change one word into the other.

6) **Data Mining**: This stage focuses on the analysis part where visual representation is done on the basis of processed data attributes for better understanding.

A. *Advantages*

- Easy visualization for extensive analysis of large datasets.
- Easy to store and access information.
- Time as well as cost efficient.
- Helping users to reduce the inaccuracy caused due to manual entries

techniques may be used for this purpose.

B. *Limitations*

- Multiple users cannot log in at a time.
- Retrieval time may increase in case of large volume.

V. CONCLUSION

Though most of the fraud detection systems show good results in detecting fraudulent transactions, they also lead to the generation of too many false alarms. This assumes significance especially in the domain of credit card fraud detection where a credit card company needs to minimize its losses but, at the same time, does not wish the cardholder to feel restricted too often.

Thus We propose a novel system based on the integration of two approaches that is fuzzy clustering and neural network. Moreover, other learning techniques needs to be experimented and comparison between them are required to be done in future research.

VI. ACKNOWLEDGMENT

We take this to express our deep sense of gratitude towards our esteemed guide **Prof. Anita Shinde** for giving us this splendid opportunity to select and present this project and also providing facilities for successful completion. We would like to thank our project sponsor **India Forensic** for their valuable support and guidance in building this project for implementing and exploring new technologies.

We also thank **Dr. H. K. Khanuja**, Head, Department of Computer Engineering, for opening the doors of the department towards the realization of the project, all the staff members, for their timely support, priceless suggestions and for most valuable time lent as and when required. With respect and gratitude, we would like to thank all the people, who have helped us directly or indirectly.

VII. REFERENCES

[1] Neda Abdelhamid, Aladdin Ayesh, and Fadi Thabtah. Phishing detection based associative classification data mining. *Expert Systems with Applications*, 41(13):5948–5959, 2014.

[2] Tanmay Kumar Behera and Suvasini Panigrahi. Credit card fraud detec- tion using a neuro-fuzzy expert system. In *Computational intelligence in data mining*, pages 835–843. Springer, 2017.

[3] Madan Lal Bhasin. Menace of frauds in the indian banking industry: An empirical study. *Australian Journal of Business and Management Research*, 4(12), 2015.

[4] Michele Carminati, Roberto Caron, Federico Maggi, Ilenia Epifani, and Stefano Zanero. Banksealer: An online banking fraud analysis and decision support system. In *IFIP International Information Security Conference*, pages 380–394. Springer, 2014.

[5] Yongrok Choi, Xiaoxia Ye, Lu Zhao, and Amanda C Luo. Optimizing enterprise risk management: a literature review and critical analysis of the work of wu and olson. *Annals of Operations Research*, 237(1- 2):281–300, 2016.

[6] Namrata Devadiga, Harshad Kothari, Hardik Jain, and Smita Sankhe. E-banking security using cryptography, steganography and data mining. *International Journal of Computer Applications*, 164(9):0975–8887, 2017.

[7] Wu He, Xin Tian, and Jiancheng Shen. Examining security risks of mobile banking applications through blog mining. In *MAICS*, pages 103–108, 2015.

[8] Mohamed Hegazy, Ahmed Madian, and Mohamed Ragaie. Enhanced fraud miner: credit card fraud detection using clustering data mining techniques. *Egyptian Computer Science Journal (ISSN: 1110–2586)*, 40(03), 2016.

[9] Vikas Jayasree and RV Siva Balan. Money laundering regulatory risk evaluation using bitmap index-based decision tree. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, 23(1):96–102, 2017.

[10] SN John, C Anele, O Okokpujie Kennedy, Funminiyi Olajide, and Chinyere Grace Kennedy. Realtime fraud detection in the banking sector using data mining techniques/algorithm. In *2016 international conference on computational science and computational intelligence (CSCI)*, pages 1186–1191. IEEE, 2016.

[11] Mahesh Kharote and VP Kshirsagar. Data mining model for money laundering detection in financial domain. *International Journal of Computer Applications*, 85(16), 2014.

[12] SOZAN SULAIMAN MAGHDID. *INTELLIGENT SYSTEM FOR IDENTIFICATION HEART DISEASES*. PhD thesis, NEAR EAST UNIVERSITY, 2019.

[13] Milad Malekpour, Maryam Khademi, and Behrouz Minae-Bidgoli. A hybrid data mining method for intrusion and fraud detection in e-banking systems. *J. Comput. Intell. Electron. Syst*, 3:1–6, 2014.

[14] Stanisław Osiński, Jerzy Stefanowski, and Dawid Weiss. Lingo: Search results clustering algorithm based on singular value decomposition. In *Intelligent information processing and web mining*, pages 359–368. Springer, 2004.

[15] Suvasini Panigrahi, Amlan Kundu, Shamik Sural, and Arun K Majum- dar. Credit card fraud detection: A fusion approach using dempster– shafer theory and bayesian learning. *Information Fusion*, 10(4):354–363, 2009.

[16] Amruta D Pawar, Prakash N Kalavadekar, and Swapnali N Tambe. A survey on outlier detection techniques for credit card fraud detection. *IOSR Journal of Computer Engineering*, 16(2):44–48, 2014.

[17] Prajal Save, Pranali Tiwarekar, Ketan N Jain, and Neha Mahyavanshi. A novel idea for credit card fraud detection using decision tree. *Inter- national Journal of Computer Applications*, 161(13), 2017.

[18] KR Seeja and Masoumeh Zareapoor. Fraudminer: A novel credit card fraud detection model based on frequent itemset mining. *The Scientific World Journal*, 2014, 2014.

[19] G Ganesh Sundarkumar and Vadlamani Ravi. A novel hybrid undersam- pling method for mining unbalanced datasets in banking and insurance. *Engineering Applications of Artificial Intelligence*, 37:368–377, 2015.

[20] Véronique Van Vlasselaer, Cristián Bravo, Olivier Caelen, Tina Eliassi- Rad, Leman Akoglu, Monique Snoeck, and Bart Baesens. Apate: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, 75:38–48, 2015.

[21] Wei Wei, Jinjiu Li, Longbing Cao, Yuming Ou, and Jiahang Chen. Effective detection of sophisticated online banking fraud on extremely imbalanced data. *World Wide Web*, 16(4):449–475, 2013.

[22] Pornwatthana Wongchinsri and Werasak Kuratach. A survey-data mining frameworks in credit card processing. In *2016 13th Inter- national Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 1–6. IEEE, 2016.

[23] Desheng Dash Wu and David L Olson. Introduction to special section on “risk and technology”, 2010.

[24] Masoumeh Zareapoor, Pourya Shamsolmoali, et al. Application of credit card fraud detection: Based on bagging ensemble classifier. *Procedia computer science*, 48(2015):679–685, 2015.