iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: III Month of publication: March 2020

DOI:

Suspicious Financial Activities Detection using Machine Learning

Snehal Deshmukh[1], Ankita Dharurkar[2], Sudhanshu Tarale[3], Preeti Godabole[4], Niraj Borkar[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering, SIES Graduate School of Technology.*

Abstract: Money Laundering refers to the act of trying to pump in illegally generated money into the authentic source so as to utilise it for malicious intents that is terror funding etc. Though there are current anti-money laundering systems, they are highly inefficient and tend to give false alerts. We propose to develop a system using machine learning and current norms and rules, that chiefly aims to reduce these false positives and also provides a dynamic screening regularly so as to detect money laundering instances faster. We aim to account for the cumulative amount in a person's account and his/her net worth to map the ability of the person to receive or transfer or even facilitate the transaction, if found to be not possible that person's details will be thoroughly scanned to check if there is a history of such behaviour. Then the transaction will be either backtracked to get an idea about the source of the money or followed to audit the trail and to possibly detect infringement of given guidelines thus reporting suspicious and fraudulent activity.

Keywords: Machine learning, supervised learning, anti-money laundering.

I. INTRODUCTION

The top most priority of financial institutions has always been monitoring of transactions for the purpose of detection of money laundering practices. The efficiency of current Transaction Monitoring Systems is gauged through the false positive rates which can range from 93% to 99.9% as they all are statistical rule based model. For example, if 634,000 suspicious reports are generated over a period of two years then, according to the false positive rates mentioned above, firms would have to investigate between 8 to 124 million false positives. This requires manual investigation which eventually leads to a significant increase in the operational costs. Hence, a system is required that reduces the number of false positive alerts. Therefore saving time and resources which are required otherwise.[1] The main aim of our system is to bring down this rate of false positives by applying machine learning algorithms on these generated alerts. The machine learning model will act as a filter and filter out most of the false positives thus reducing the time and effort that would have been otherwise required to investigate the reports.

II. SURVEY

A. Money Laundering Detection using TFA (Transaction Flow Analysis) System [2]

Money laundering is not a single step act but in fact a combination of the three following acts:

1) *Placement:* The sum of money is broken down into parts and the cash is deposited in banks.
2) *Layering:* Money is transferred into overseas bank accounts where the banks have secrecy codes.
3) *Integration:* The money invested in the second step can now be used to purchase luxury assets and the illegal money appears legal.

Clustering algorithm is applied on the data of bank statements to form clusters which have a graph structure. The calculations were performed on transfers provided by the test data generator which creates random accounts and connects them creating money transfers. The elements of these clusters can be treated as suspicious operations. The frequent sets and sequences in the clusters are then mined to identify potential money laundering cases. Data visualisation is then applied on the frequent patterns and clusters in order to analyse the results. Thus, graphs with at least one account which has significantly a large number of incoming transfers assists in finding gatekeepers. The frequent mining algorithm finds those entities which always appear in the transactions. The time taken for the realisation of the transfer of money is also taken into consideration which eventually detects money laundering.

Some of the areas that we have to monitor to keep money laundering in check are:

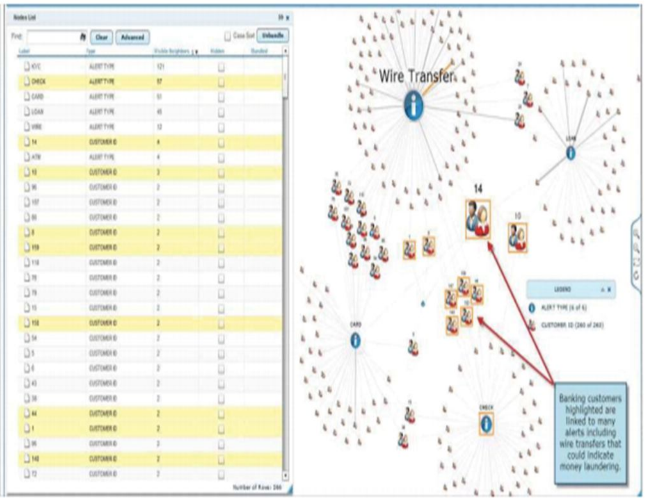
a) Transaction Type
b) Frequency
c) Amount
d) Geographical origin/destination
e) Account Signatories

Frequent patterns of transactions are:

Actual transaction usage	Percentage
ATM	40.2
EFTPoS	22.8
Branch	22.0
Cheques	8.4
Others	6.6 (includes internet banking at 0.2 percent)

Table 2.1 Frequent patterns of transaction

The customer transactions can be clustered as shown in the image. Clustering these transactions help us in finding the outliers in the dataset. These outliers can be processed further and according to the rules that we establish we can further look into these transactions.



Banking customers highlighted are linked to many alerts including wire transfers that could indicate money laundering.

2.1 Visualised cluster according to how transactions happen [3]

B. *Master Directions on Frauds – Classification and Reporting by commercial banks and select FIs [3]*

According to the RBI, there are some Early Warning Signals (EWS) which should alert the bank officials about some wrongdoings. Some of the EWS that we are looking into are sudden increase in turnover or net worth, non-production of KYC documents, frequent loan requests, not producing original bills on request and high value RTGS payments to unrelated parties.

Banks may choose to adopt or adapt the relevant signals from this list and also include other alerts/signals based on their experience, client profile and business models. The EWS so compiled by a bank would form the basis for classifying an account as a Red Flagged Account (RFA).

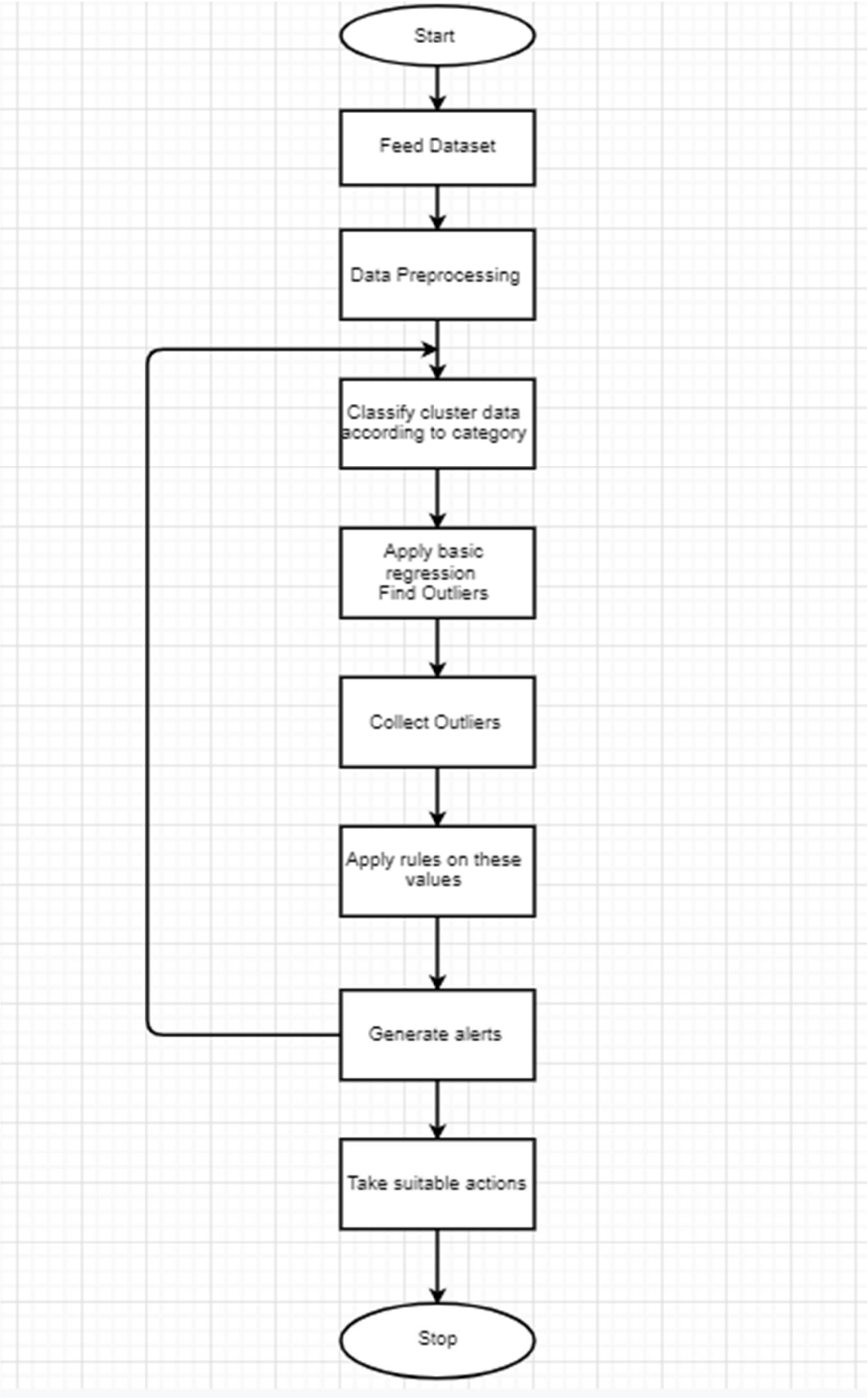
Frauds are broadly classified into:

1) Misappropriation and criminal breach of trust.
2) Fraudulent encashment through forged instruments, manipulation of books of account or through fictitious accounts and conversion of property.
3) Unauthorised credit facilities extended for reward or for illegal gratification.
4) Negligence and cash shortages.
5) Cheating and forgery.
6) Irregularities in foreign exchange transactions.
7) Any other type of fraud not coming under the specific heads as above.

C. Detection of Money Laundering Groups: Supervised Learning on Small Networks [4]

They have in detail explained an automated system for detecting money laundering operations in transaction networks. This system boosts the current state-of-the-art by its analysis of both explicit transaction relationships and implicit relationships which have been derived from supplementary information contained in the transactions. Evaluating the system shows that there is a permissible level of accuracy that is being achieved at high levels of precise analysis. The most important characteristic for their system is it uses a live environment to cut down rates of false positives.

III. PROPOSED SYSTEM


Start
Feed Dataset
Data Preprocessing
Classify cluster data according to category
Apply basic regression Find Outliers
Collect Outliers
Apply rules on these values
Generate alerts
Take suitable actions
Stop

3.1 Flowchart

IV. METHODOLOGY

A. Data Mining

Data sets consisting of the following three tables are to be mined which will be utilized by the model.

1) Customer Details

a) Customer number
b) Gender
c) Type of customer (Individual/ Firm / Company)
d) Occupation (Business/ Service/ Student/ Pensioner/ Others)
e) Address/ Location/ District
f) Net Worth
g) Annual Income
h) No. of deposit accounts
i) No. of loan accounts
j) Liability/ Total loan amount
k) Cumulative credit in accounts
l) Cumulative debit in accounts

2) Account Details

a) Account number
b) Type (Deposit/ Loan)
c) Subtype (FD/ Savings/ Current/ Term Loan)
d) Account open date
e) Status (Open/ Inoperative/ Close)
f) Present balance
g) Cumulative credit transactions over a period (1 Month/1 Quarter/ 1 Year)
h) Cumulative debit transactions over a period (1 Month/1 Quarter/ 1 Year)
i) Average credit balance
j) Average debit balance in loan amount
k) Cumulative cash deposit over a period (1 Month/1 Quarter/ 1 Year)
l) Cumulative cash withdrawal over a period (1 Month/1 Quarter/ 1 Year)
m) Average monthly transactions in previous period (1 Month/1 Quarter/ 1 Year)

3) Transaction Details

a) Account No
b) Date of transaction
c) Transaction sequence no.
d) Time of transaction
e) Type(Cash/ Transfer/ Clearing/ Electronic mode)
f) Type of transaction signature(Credit/ Deposit)
g) Electronic mode (ATM/ UPI/ NEFT/ RTGS)
h) Amount
i) Reference no. to identify
j) Balance

B. Data Pre-processing

The data stored in the above tables may not follow a defined format. Hence, all the data will be converted into a specific format for the ease of developing the model. For e.g. data will be cleansed and all non ASCII values will be discarded and then will be replaced under suitable headers. If there are null values in the key attributes then the transaction will be side lined and will be reported to the respective bank for further clarifications.

C. Clustering

Clustering will be applied on the transactions to form clusters based on the general rules according to the guidelines given by the RBI under the Prevention of Money Laundering Act. These rules are based on parameters like the amount, frequency and the date range of the transactions.

Eg: We can have a basic rule that all accounts which exceed their net transactions by 50,000 should be out under observation.

D. Finding Outliers

Every cluster will be taken into consideration to apply regression on each one of them. The purpose of applying regression is to find the outliers which represent the most probable suspicious transactions. The output of this step will contain a number of false positives. Finding outliers is basically applying some basic rule to get transactions which fall outside of the rule and then gathering those transactions and applying AML rules on it.

If the transaction seems unusual or suspicious, we check the cumulative debit/credit and net worth of the account holder.

E. Applying Rules

Another set of rules are applied on the collection of outliers obtained in the above step. These rules will be more specified depending on the characteristics of the respective cluster. This step filter outs some of the false positives obtained in the above step.

Further methods of backtracking will be used to know the source of the money.

F. Generation of Alerts

Alerts are generated as a result of the previous step, every alert is scored in terms of number of rules violated. A manual input on these results is taken and fed back to the model so that the machine can learn based on the executed algorithm as well as human feedback. These alerts then function as additional rules and the other inputs again go through classification of clusters, regression to find outliers and then application of rules to finally obtain more precise suspicious alerts which contain comparatively few number of false positives. These reduced numbers of reports are taken into account for further investigation by the respective authorities.

V. REFERENCES

[1] https://blogs.sas.com/content/hiddeninsights/2018/03/12/measures-for-anti-money-laundering-transaction-monitoring-systems/

[2] P. Umadevi and E. Divya, "Money laundering detection using TFA system," *International Conference on Software Engineering and Mobile Application Modelling and Development (ICSEMA 2012)*, Chennai, 2012, pp. 1-8.
doi: 10.1049/ic.2012.0150.

[3] https://m.rbi.org.in/Scripts/BS_ViewMasDirections.aspx?id=10477

[4] David Savage; Xiuzhen Zhang; Qingmai Wang; Xinghuo Yu & Pauline Chou, “Detection of Money Laundering Groups: Supervised Learning on Small Networks”, The AAAI-17 Workshop on AI and Operations Research for Social Good WS-17-01