JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Comparative Analysis & Implementation of AI Algorithms for Money Laundering Detection

#1 K.Ramya , Asst Prof , Mohamed Sathak Engineering College /
#2 S . Shenija, Final MCA, Mohamed Sathak Engineering College /
#3 Dr.S Sajitha Banu, Asst Prof , Mohamed Sathak Engineering College , Kilakarai
#4 M Sabari Ramachandran , Asst Prof , Mohamed Sathak Engineering College /
#5 G Balamurugan , Asst Prof , Mohamed Sathak Engineering College , Kilakarai

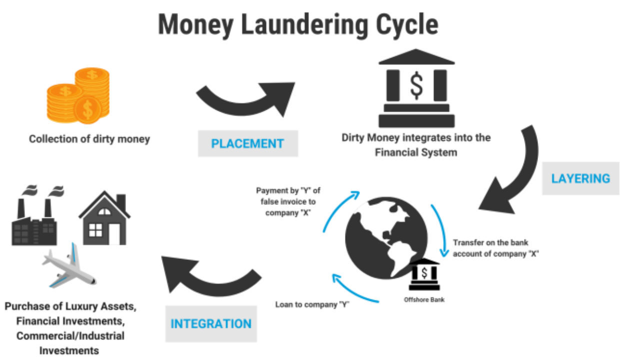
Abstract: With billions of dollars and Rupees worth of assets being laundered every year, the problem of money laundering is growing. In order to detect this illegal activity and prevent it from happening, a network of banks needs to work together. However, most existing approaches have failed to achieve this goal primarily because they do not allow any third-party access to information connecting one bank with another. The paper proposes a novel solution which follows three essential principles: preserving privacy, and enabling ML detection at massive scale. This paper also implements the comparative analysis of the ML Algorithms and depicts the same on graphical illustration with the accuracy level of each of the algorithm in a specialized approach to detect Money Laundering using the Latest ML Applications

Keywords *– Money Laundering Detection, Machine Learning, Malicious Application, Python, Random Forest, Laundering*

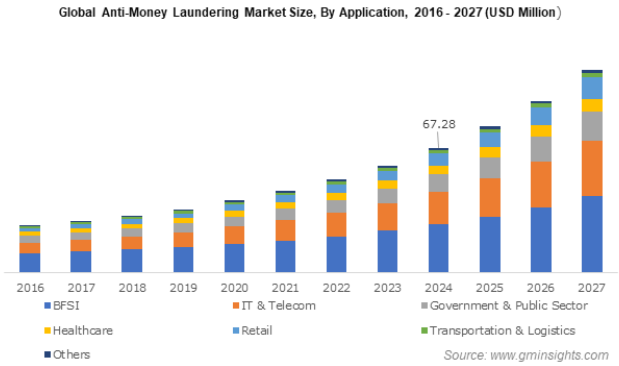
I.INTRODUCTION

Increasing the world's population, the volume of economic transactions and the global struggle against regional instability have resulted in increased money laundering activities so that the illegal movement of funds can be achieved in an efficient way. The money laundering process is achieved by a series of steps which include laundering, washing, preparing and making counterfeit notary documents. The most common disguise methods are cash exchange, mail transfer, counterfeit documents and counterfeit seals. All these methods are relied on by criminals to transport their ill-gotten gains undetected to other countries where they can be used to boost parallel economic activity or hide the proceeds from prosecution. rom these malicious Android apps. This is a severe threat and necessity to detect the money laundering initiatives and to protect the authentic data of the users posed under threat. This paper implements the same using the latest Machine Learning algorithms to prevent and predict the money laundering initiative from the scratch. Updated algorithmic deployment to the detection of the same can be accomplished by computing the accuracy levels of the predictions and the computational time required to process and detect the same with a minimum negligence

Fig 1 Money Laundering Process in detail


Money Laundering Cycle
Collection of dirty money
PLACEMENT
Dirty Money integrates into the Financial System
LAYERING
Payment by "Y" of false invoice to company "X"
Transfer on the bank account of company "X"
Offshore Bank
Loan to company "Y"
INTEGRATION
Purchase of Luxury Assets, Financial Investments, Commercial/Industrial Investments

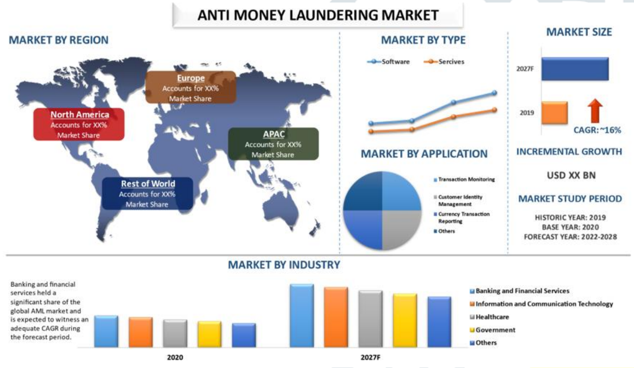
Money laundering is a criminal activity that can be used in the context of developing countries to launder money obtained through the illicit drug trade or other sources, and make it appear as though it has originated from legal activities. This procedure is often performed using legitimate companies located within more developed countries. It is is the process of converting money from criminal activity into a legitimate source of money, with the aim of avoiding detection by law enforcement authorities. With stricter laws being enacted as a result of the global financial crisis, criminals are looking for ways to launder their money in order to move it around the world and out of the reach of investigators.



Source: www.gminsights.com

Fig 2. Value wise Market Share of Money Laundering

The growth malicious applications in android over a period of ten years has been on the high rise from 85M to 1334M apps. The jump is so high and so is the demand to create a system to detect and classify malicious apps from the rest so as to protect user safety.



This paper proposes or mainly contributes the following

a. Detect & Classify Malignant & Benign Apps using novel implementation strategies such as ML Algorithms
b. Provide Illustrative result of the percentage of accuracy of the ML Algorithms on Malware Detection

II. LITERATURE SURVEY WORKS

Existing research papers of projects have three major issues

1. Data Quality Issues : We will face stringent data Quality issues on the process of processing of the financial data.An alternative approach is to replace missing values with those from a nearby observation. This can be achieved by smoothing values of the available data. For example, the missing date for a particular money transfer can be estimated by the mean of that day of month for neighboring transactions. A number of data smoothing methods have been proposed in the literature. In an organization like Banks, there are many transactions happening continuously and having a data of all the related records of such transactions helps credit risk management to understand the trend and then predict and take action before any default.

2. Corrupted Input Data: In case of any corrupted data from the users or banks, the same cannot be legitimated as that may be ponderable to more high authority issues.

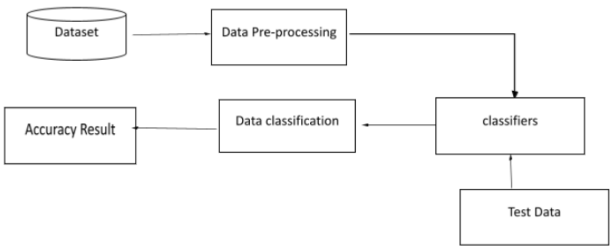
3. Conventional detection tools may not able to detect newly created or devised plan of money laundering. Detecting of large number of money laundering tasks over millions of fraudulent applications is still a challenging task using traditional way.

Alternate existing systems propose checking for anti-money laundering systems is time consuming and the results are not reliable. Moreover, no confusion matrices are applied to cross check the correctness of the results. Moreover, Non-Machine Learning techniques has been deployed in the existing systems. This can render to reduce the possible accuracy of the detection system.

III PROPOSED SYSTEM

The proposed system has been implemented with the help of Artificial Intelligence Packages, Python and NumPy packages. Our algorithm is based on Support Vector Machine and Decision Tree for classification of Android applications. We also have applied Support Vector Machine Algorithm on our dataset which is compared with the training dataset to validate our system. The performance of this system has been evaluated using one-class classifier, six-class classifier and seven-class classifier as we have used datasets from selected sources to test our performance while making up comparison between training data and test data.

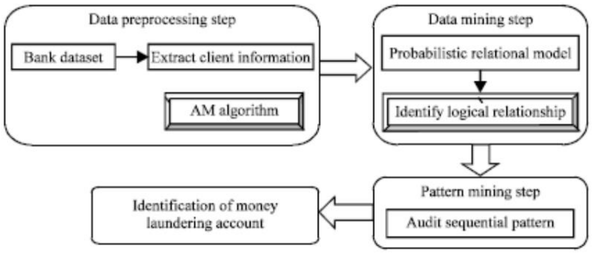
Illegit Money Laundering is very complex and difficult to detect, as they often use sophisticated techniques to hide any vulnerability. Advanced AI - Machine Learning technique helps in finding the same and its variants simply by using hidden features. This paper implements a new scheme of detection of the money laundering symptoms by implementing the AI CNN Process we can predict for 95% accuracy rate of malware on android phones.



A. APPLICATION STRUCTURE & DATASETS

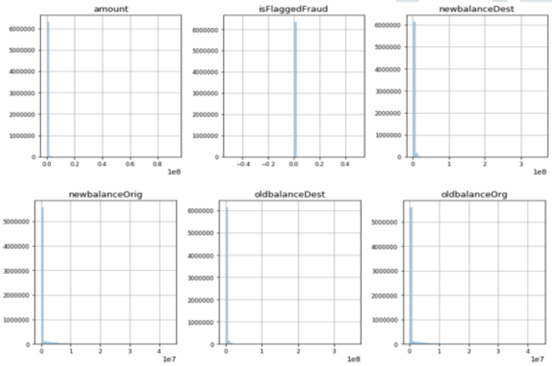
Detection of Money Laundering using AI domain is Machine Learning is a field which is heavily data dependent and you need good data before you can build anything. Noisy large data will typically yield garbage models (or super biased ones). But it’s generally way worse as compared to state of the art and will need a lot of time to become usable.

We have created an accurate smallish dataset and used few shot techniques. These techniques are still developing and are slowly becoming usable by transfer learning, meta learning etc. With the emergence of new Deep Learning techniques Computational resources and Data both have become prerequisites to work in AI - Machine Learning.


Data preprocessing step
Bank dataset
Extract client information
AM algorithm
Data mining step
Probabilistic relational model
Identify logical relationship
Pattern mining step
Audit sequential pattern
Identification of money laundering account

B. NEURAL NETWORKS - CONVOLUTIONAL

AI – CNN is applied in a variety of applications including medical, security and criminal justice systems. This essay provides a brief overview of machine learning techniques and applications in this application domain. Implemented smart mechanisms on the Internet of Things (IoT) devices are based on machine learning algorithms that can be used for identification and prediction of behaviors such as check-in, check-out, theft and vandalism among others. With the rapid growth in mobile network penetration via smart phones worldwide, theft or vandalism detection using IoT devices becomes critical due to security concerns.



Deep learning is a subset of machine learning and AI. Machine learning is a branch of artificial intelligence. Convolutional neural network (CNN) is a type of deep learning used in computer vision with images to extract and learn features from the image. In CNN, filters are applied pixel-by-pixel at each image location to obtain feature detections which each pixel belonging to. Although fully connected multilayer sequential neural networks can handle feature detection directly, CNN is more accurate and efficient for processing images than alternative methods such as Cascading Style Sheets (CSS) and Paths for HTML (PHP).

B. Proposed Method Flow Diagram

The algorithms extract the important relational logic data such as customer, bank, transaction and their relationship with each other by maximizing their probability to identify all the possible relationships that can be derived through PRM-ASP Mining. This algorithm is also popular throughout the data mining community as an efficient means of reasoning over uncertain data.

Feature Extraction expects to decrease the number of highlights in a dataset by making new features from the current ones. This is done by using the decision tree algorithm. The Datasets are then sorted out using several classifiers.

C.PREDICTING LAUNDERING using AI CLASSIFIERS

1. DECISION TREE ALGORITHM

Decision Tree Algorithms work on the Model that will work on bi polar model of answering. Certain questions on bipolar degree and put forth to the data and if the received answer is yes or no, the actions that has been decided or programmed to will take place. This will not be much significant in the changing of the complete accuracy levels or the programming scenarios as per the real-world prediction data.

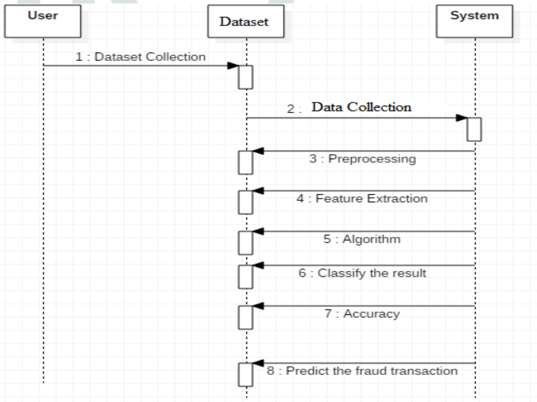
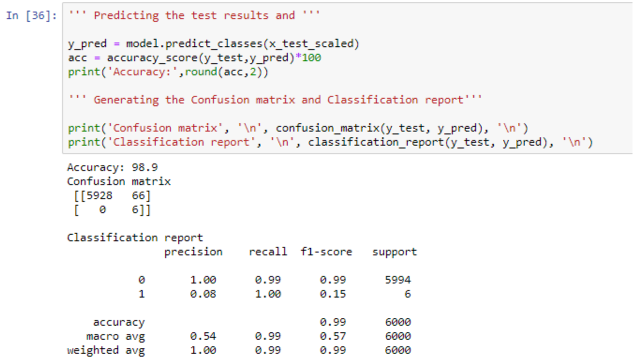


Figure 6. Flowchart of the Proposed Method

II. NAIVE BAYES ALGORITHM

Naïve Bayes is a simple algorithm used to classify given data points into classes. It works in such a way that it assumes the features extracted from the given dataset have no relationship with each other, or in other words - feature independence. In addition, it uses probability values to assign class labels using Bayes theorem and concludes that the evidences which support any hypotheses are more significant and informative than those contradicting them.
processing of the application.

Applying statistical analysis to a set of historical data and observed correlations between subsets of variables. It aims to assign a probability to every possible hypothesis in the form of "samples" with the help of conditional statements such as: If (Z_i =1), then P(X = x). The assumptions on which this algorithm can be built upon depend upon a set of attributes that one needs to be considered – Currency in our paper implementation.



```
In [36]: ''' Predicting the test results and '''

         y_pred = model.predict_classes(x_test_scaled)
         acc = accuracy_score(y_test,y_pred)*100
         print('Accuracy:',round(acc,2))

         ''' Generating the Confusion matrix and Classification report'''

         print('Confusion matrix', '\n', confusion_matrix(y_test, y_pred), '\n')
         print('Classification report', '\n', classification_report(y_test, y_pred), '\n')
```

```
Accuracy: 98.9
Confusion matrix
 [[5928   66]
 [   0    6]]

Classification report
               precision    recall  f1-score   support

           0       1.00      0.99      0.99      5994
           1       0.08      1.00      0.15         6

    accuracy                           0.99      6000
   macro avg       0.54      0.99      0.57      6000
weighted avg       1.00      0.99      0.99      6000
```

Fig 7. Classification Model of the ML Implementation

E. CLASSIFICATION

The trained datasets are used for classification, a type of supervised learning in which the ML algorithms learn from data input given to them and then utilize this learning to classify new data for accurate observation and prediction. Classification of money laundering data means that the dataset may be classified into various classes for more in-depth observation. The attributes that comprehend to the liabilities of the applications may vary depending upon the app dataset and chosen application. Closer to their domain, accurate results are obtained. The trained datasets are to be done classification, which is a kind of technique used to categorize the data into a desired and distinct number of classes where we can assign labels to each class. This is an example of supervised learning in which the ML algorithms learn from the data given to it and then use this learning to classify new datasets for more accurate observation and prediction. Money Laundering data classification means that we can classify certain datasets into various classes depending upon their proximity to the application domain.

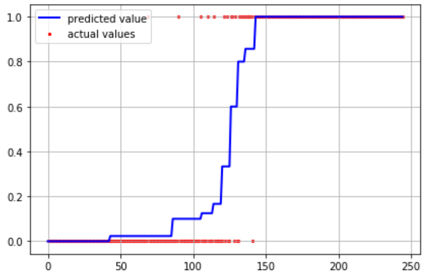


Figure 8. Malware Liklihood using DT Algorithm

IV. SOFTWARE SYSTEM

A. Python 2.5 / 3.5

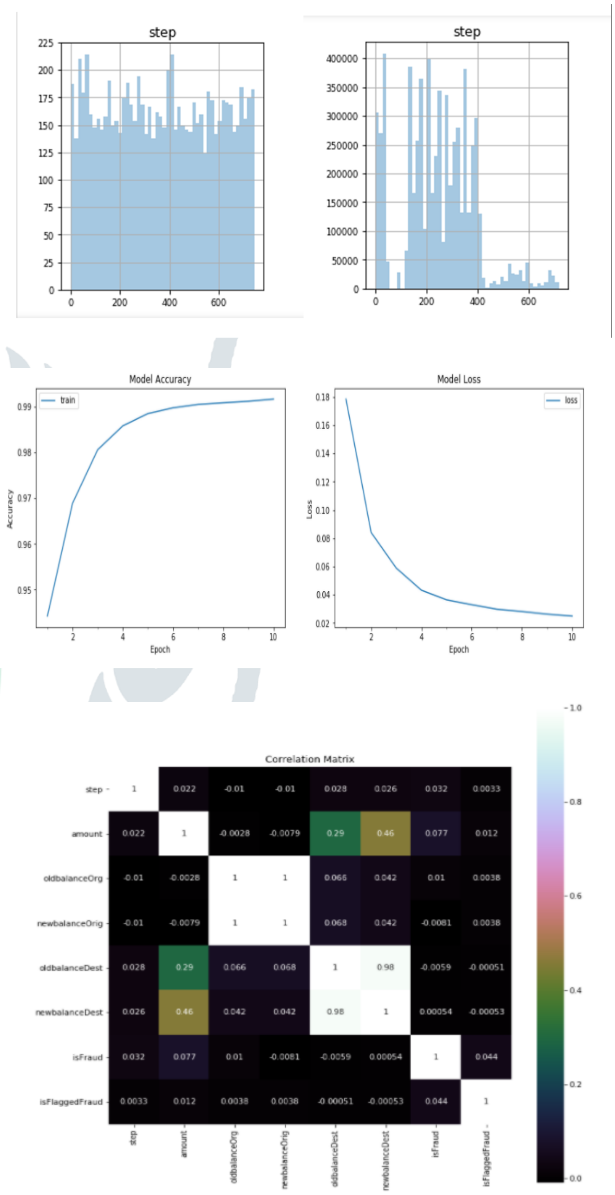
Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable Python is Object-Oriented − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

B. Anaconda Navigator

Anaconda Navigator is the GUI that helps in visualizing the results in the form of graphics. Output predictions can be well visualized using the navigator

V RESULTS AND DISCUSSION

Various results can be categorized from the total list of applications that has been tested. The trained dataset of applications are checked using a confusion matrix as listed below.



	step	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
step	1	0.022	-0.01	-0.01	0.028	0.026	0.032	0.0033
amount	0.022	1	-0.0028	-0.0079	0.29	0.46	0.077	0.012
oldbalanceOrg	-0.01	-0.0028	1	1	0.066	0.042	0.01	0.0038
newbalanceOrig	-0.01	-0.0079	1	1	0.068	0.042	-0.0081	0.0038
oldbalanceDest	0.028	0.29	0.066	0.068	1	0.98	-0.0059	-0.00051
newbalanceDest	0.026	0.46	0.042	0.042	0.98	1	0.00054	-0.00053
isFraud	0.032	0.077	0.01	-0.0081	-0.0059	0.00054	1	0.044
isFlaggedFraud	0.0033	0.012	0.0038	0.0038	-0.00051	-0.00053	0.044	1

The above fig shows a comparative plot of the various algorithms that has been employed in the proposed system

VI CONCLUSION

In this paper, a proposed AI based Laundering detection scheme to detect money laundering in the process of transforming the proceeds of crime and corruption into ostensibly 'legitimate' assets. It is a process that incorporates the movement, placement and integration of proceeds of crime and corruption in the legitimate economy by means of complex, sophisticated and professional financial transactions, finally obtaining the appearance of legitimate wealth or revenue from unlawful activities. In this paper, we propose a framework to detect fraudulent activities using machine learning algorithms but, the key is achieved when all banks send their results to central database. The framework explains on how banks collaborate with each other to achieve higher success rate by increasing the number of authorities that are provided the money laundering detection data. In this scenario, two banks which do not collaborate with each other will increase the risk involved in money laundering as the fraudster can choose a bank that does not share information with central database

VII FUTURE SCOPE

The scope of the project can be extended further with implementing additions of the same. Changes in algorithms and processes such as natural language processing, Enhanced visualization techniques will prove to be good. Further accuracy levels of the key factors in money laundering have been as currency as a general level. It may be then done for various factors such as fund allocation, stock calculation, asset allocation and more.

REFERENCES

[1] E. Fell, O. James, H. Dienes, N. Shah, and J. Grimshaw, “Understanding organised crime 2015/16,” Feb. 2019, pp. 11,18,31, 38.

[2] G. Leite, A. Albuquerque, and P. Pinheiro, “Application of Technological Solutions in the Fight Against Money Laundering — A Systematic Authorized licensed use limited to: University of Wollongong. Downloaded on August 11,2020 at 01:54:38 UTC from IEEE Xplore. Restrictions apply. Literature Review.” Multidisciplinary Digital Publishing Institute, Nov. 2019.

[3] M. Brown and J. Kros, “Data mining and the impact of missing data,” Industrial Management and Data Systems, vol. 103, pp. 611–621, 11 2003.

[4] S. Haider, “Clustering based anomalous transaction reporting,” Procedia CS, vol. 3, pp. 606–610, 12 2011.

[5] L.-T. Lv, N. Ji, and J.-L. Zhang, “A rbf neural network model for anti-money laundering,” in 2008 International Conference on Wavelet Analysis and Pattern Recognition, vol. 1. IEEE, 2008, pp. 209–215.

[6] C. Ju and L. Zheng, “Research on suspicious financial transactions recognition based on privacy-preserving of classification algorithm,” in 2009 First International Workshop on Education Technology and Computer Science, vol. 2. IEEE, 2009, pp. 525–528.

[7] A. Shamir, “How to share a secret.” New York, NY, USA: Association for Computing Machinery, Nov. 1979, vol. 22, no. 11, p. 612–613. [Online]. Available: https://doi.org/10.1145/359168.359176

[8] D. Flores, O. Angelopoulou, and R. Self, “Combining digital forensic practices and database analysis as an anti-money laundering strategy for financial institutions,” 09 2012, pp. 218–224.

[9] Z. Chen, L. Khoa, E. Teoh, A. Nazir, E. Karuppiah, and K. Lam, “Machine learning techniques for anti-money laundering (aml) solutions in suspicious transaction detection: a review,” Feb. 2018.

[10] M. Cooper, Y. Dzambasow, P. Hesse, S. Joseph, and R. Nicholas, “Internet X.509 Public Key Infrastructure: Certification Path Building.” Internet Engineering Task Force, 2005. [Online]. Available: https://tools.ietf.org/html/rfc4158

[11] E. Rescorla, “The Transport Layer Security (TLS) Protocol Version 1.3.” Internet Engineering Task Force, 2018. [Online]. Available: https://tools.ietf.org/html/rfc8446

[12] R. Rivest, A. Shamir, and L. Adleman, “Cryptographic communications system and method.” Google Patents, 1977. [Online]. Available: https://patents.google.com/patent/US4405829

[13] “Sec 1: Elliptic curve cryptography.” Certicom Corp., 2009. [Online]. Available: https://www.secg.org/sec1-v2.pdf

[14] J. Daemon and V. Rijmen, “AES Proposal: Rijndael.” National Institute of Standards and Technology, 1999. [Online]. Available: https://csrc.nist.gov/csrc/media/projects/cryptographic-standardsand-guidelines/documents/aes-development/rijndael-ammended.pdf

[15] D. Harkins, “Synthetic Initialization Vector (SIV) Authenticated Encryption Using the Advanced Encryption Standard (AES).” Internet Engineering Task Force, 2008. [Online]. Available: https://tools.ietf.org/html/rfc5297