International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 395 - 400
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0058

Monetary Smurfing Analytics Model using ML

Aaron Alex.X[1], Barathram.B[2], Anandhu.B[3], Chithrakumar T[4]*
[1, 2, 3] UG - Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore, India
[4]Assistant Professor (Sr. Gr), Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore, India
Emails: *aaronalex.2101001@srec.ac.in[1], barathram.2101026@srec.ac.in[2], anandhu.2101009@srec.ac.in[3], chithrakumar.thangaraj@srec.ac.in[4]*
***Corresponding Author Orcid ID:** *https://orcid.org/0009-0006-9700-4228*

Abstract
The act of taking money obtained from illegal operations, such as drug trafficking, and disguising it as profits from a legitimate business venture is known as money laundering. The money obtained through illegal action is regarded as dirty, and to make it appear clean, it is "laundered." Money laundering is a major concern to the country these days, as well as financial institutions. The sophistication of this illegal activity is growing; it appears to have beyond the tired old trope of drug trafficking to include financing terrorism and, of course, personal benefit. Research on anti-money laundering is crucial since money laundering is a global issue that seriously jeopardizes international security and financial stability. Furthermore, it's estimated that the banking system seizes only 0.2% of the money that has been laundered. The crime itself is growing more sophisticated and intricate, and banks are becoming more vulnerable as a result of its ongoing volume amplification. Considering the role that banking institutions play in the money-laundering industry, practitioners and researchers are becoming more interested in creative ways to solve problems and enhance anti-money-laundering efforts. Researchers are starting to look into the viability of artificial intelligence methods in this setting. A systematic knowledge deficit concerning a thorough study that meticulously examines and combines artificial intelligence methods for countering money laundering in the banking industry was discovered, though. To combat investment fraud, the majority of global financial institutions have been putting anti-money laundering measures into place. Data mining tools have emerged recently and are thought to be effective methods for identifying instances of money laundering.

Keywords: *Monetary Smurfing Analytics, Machine Learning, Logistic Regression, Random Forest*

1. Introduction

Money laundering has been impacting the world economy for a few decades now, spreading like a spider web. Massive sums of money are utilized in money laundering to transform illegally obtained monies into ones that are both legal and legitimately connected to criminal activity. Today, banking institutions are required to make significant investments in the fulfillment [1] of Anti-Money Laundering (AML). One of the most important responsibilities for many nations is combating money laundering. Money launderers typically split up the illicit funds into smaller amounts and then Legalize them through a series of tiny bank transfers or business deals. Therefore, it is an extremely difficult process to manually discover money laundering [2] operations. No organization supports the money laundering or small-scale financing of terrorist or criminal groups. Money obtained unlawfully through a variety of means, including drug sales and theft, must be cleaned up for tax purposes. This has a detrimental impact on a country's capital, savings, and investments while, on the other hand, providing financial support for illicit activity. [3] Money laundering detection, commonly referred to as Money laundering is just one of the numerous types of banking fraud that include fraud with credit or

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 395 - 400
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0058

debit cards, internet transactions, and money laundering, among other things. This is also known as an action that aims to prevent laundering of funds from an incident. The term "illegal income," which refers to judicial few activities, varies from nation to nation. AML's main objectives are to uncover planned offences, reduce drug sales, stop terror attacks, and maintain the financial services industry's standing. Every day, AML regulations [4] change and become more costly, complex, and challenging to follow. The burden of Anti Money Laundering Agreement reporting and regulations is falling on banking organizations. Figure 1 Flow Diagram for Monetary Smurfing Analytics process.

2. Proposed System

In this project, we demonstrate the solution that was created as a tool and some initial experiment findings using actual transaction datasets. Our strategy involves employing supervised machine learning techniques to categories transactions as either fraudulent or not, by utilizing data such as balance changes and [5] inbound and outgoing transactions both domestically and internationally. For every experiment, we divided the data into successive train and test datasets. Using the top 15 features from the train set, we train every supervised model, and then we evaluate them on the whole test set. to track one's performance over time. We employ the Cat boost Classifier and logistic regression (LR) using the scikit-learn implementation. [6]

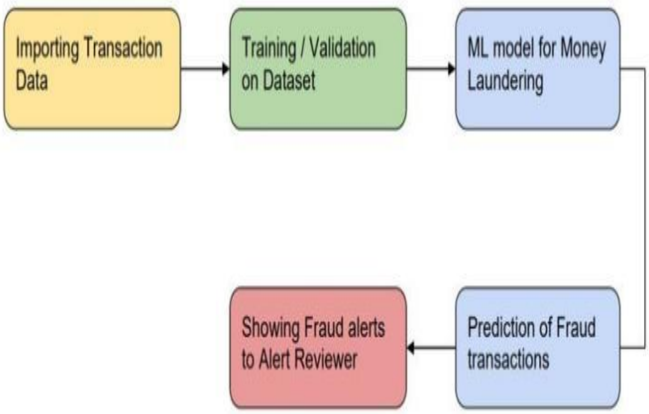


Figure 1 Flow Diagram for Monetary Smurfing Analytics

2.1 Procedure Monetary Smurfing Analytics

1. We start by using the transactions dataset.
2. Filter datasets with attributes based on requirements so that analysis may be performed.
3. Divide the dataset into test and training sets.
4. Use SMOTE to balance the data in the created resultant dataset.
5. After training, testing data can be analysed using Neural Networks, Random Forests, and Logistic Regression.
6. Lastly, data will be provided as accuracy metrics [7]. Figure 2 shows the Procedure Flow Process.

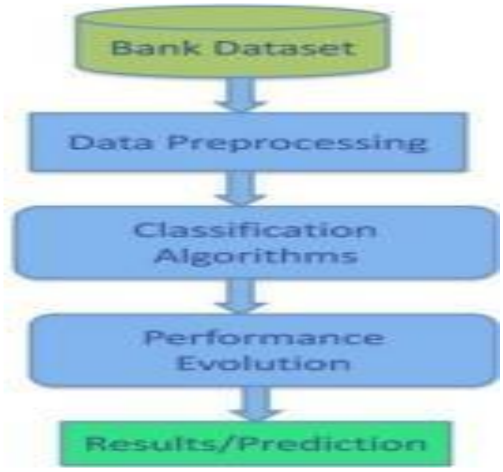


Figure 2 Procedure Flow

The prediction technique will make use of the following machine-learning algorithms: [8]

- Logistic Regression
- Random Forest
- Cat boost

2.1.1 Logistic Regression

To determine the probability of a categorical dependent variable, logistic regression, a machine learning classification technique, is employed.A binary variable with data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.) is the dependent variable in logistic regression. Stated differently, P(Y=1) is predicted by the logistic regression model as a function of X. [9]

2.1.2 Random Forest

Regression and classification challenges can be

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 395 - 400
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0058

handled by the ensemble methodology Random Forest. It accomplishes this by using a number of decision trees in conjunction with a method known as "bagging," or Bootstrap and Aggregation. Here, the fundamental idea is to use multiple decision trees to determine the outcome instead of depending only on one. Several decision trees are used by Random Forest as its fundamental learning models. For each model, we create sample data sets by randomly selecting rows and attributes from the dataset. Training Flow Diagrams are shown in Figure 3. [10]

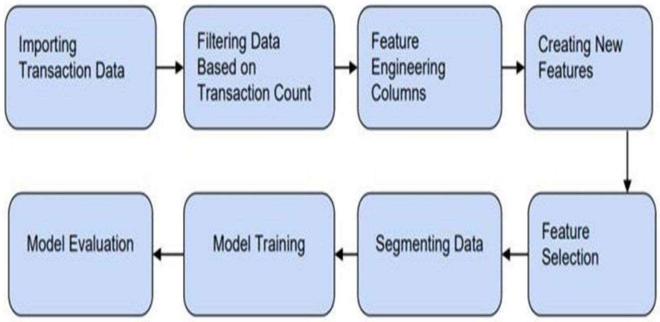


Figure 3 Training Flow Diagrams

2.1.3 Cat boost

Selecting the model by those best addresses the given problem is the goal of training. (Regression, classification, or multiclassification) depending on a set of features x_ {i} xi for any given input item. A training dataset, or collection of objects with known features and label [15] values, is used to find this model... Utilizing data in the same format as the training dataset, the validation dataset to verify accuracy; however, it is not utilized for actual training. Instead, it is used to assess the quality of training. Cat Boost relies on gradient enhanced decision trees as its base. During training, a succession of decision trees is built one after the other. Every new tree is built with less loss as compared to the previous ones. [11]

Cat Boost is compatible with the subsequent feature types:

a. Quantitative. The height (182, 173) and any binary characteristic (0, 1) are two examples. [12]
b. Classifiable (cat). These attributes have a finite set of possible values. Usually, these numbers are set. Two such musical categories include styles (dancing, classical), and musical genres (rock, indie, pop).
c. Regular text is included in these features (Music to hear, why hearst thou music sadly?) [13]

The following phases are often involved in the process of converting categorical features to numerical: Deployment Flow Diagrams are shown in Figure 4.

a. Randomly ordering the collection of input objects by permutation.
b. Changing the floating-point label value to an integer [14]

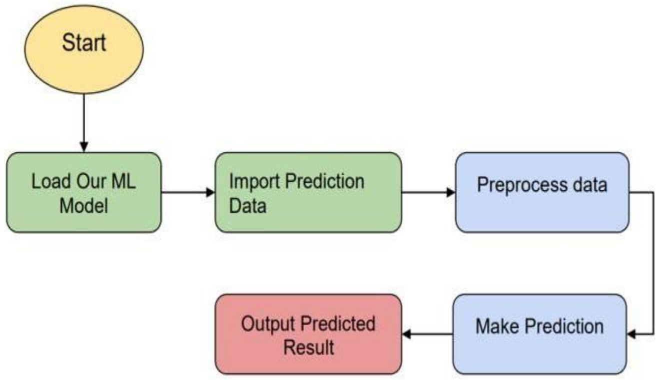


Figure 4 Deployment Flow Diagrams

3. Result

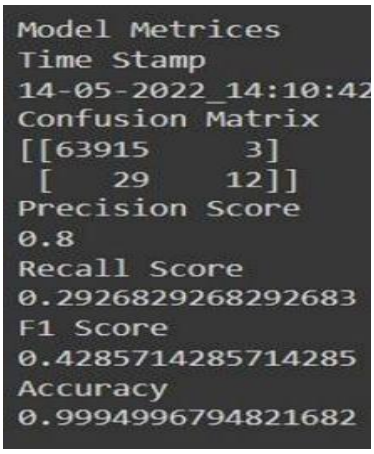
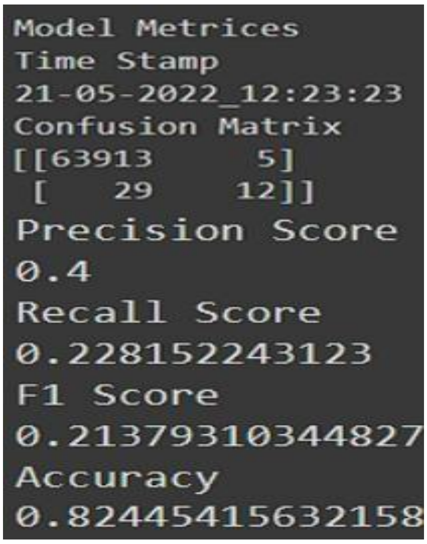


Figure 5 Model Metrices Values of the Processed Model

Figure 5 shows the model metrices values of the processed Model [16].

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 395 - 400
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0058



```
Model Metrices
Time Stamp
21-05-2022_12:23:23
Confusion Matrix
[[63913      5]
 [   29     12]]
Precision Score
0.4
Recall Score
0.228152243123
F1 Score
0.21379310344827
Accuracy
0.82445415632158
```

Figure 6 Comparison Result

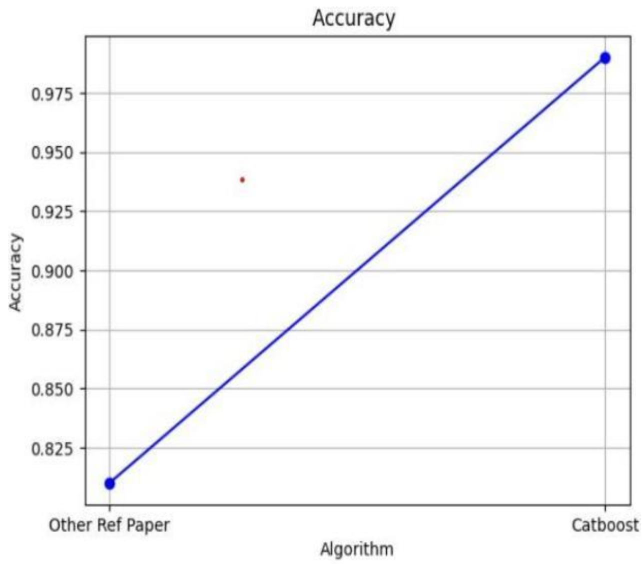
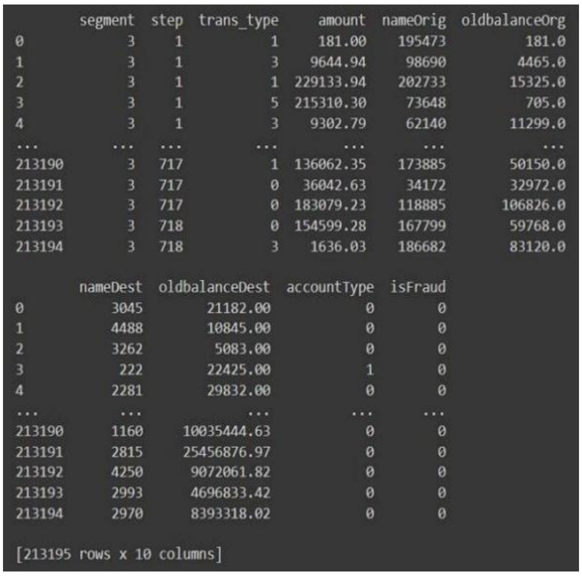


Figure 7 Accuracy Graph Compared to Other Ref Paper vs. Cat Boost

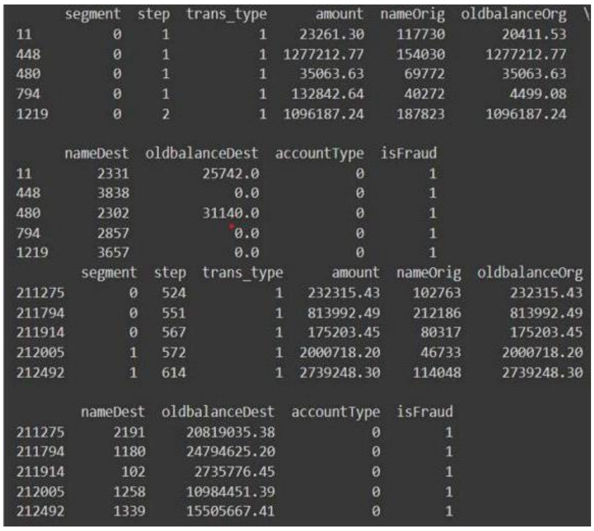


	segment	step	trans_type	amount	nameOrig	oldbalanceOrg
0	3	1	1	181.00	195473	181.0
1	3	1	3	9644.94	98690	4465.0
2	3	1	1	229133.94	202733	15325.0
3	3	1	5	215310.30	73648	705.0
4	3	1	3	9302.79	62140	11299.0
...	...	...	...	...	...	...
213190	3	717	1	136062.35	173885	50150.0
213191	3	717	0	36042.63	34172	32972.0
213192	3	717	0	183079.23	118885	106826.0
213193	3	718	0	154599.28	167799	59768.0
213194	3	718	3	1636.03	186682	83120.0

	nameDest	oldbalanceDest	accountType	isFraud
0	3045	21182.00	0	0
1	4488	10845.00	0	0
2	3262	5083.00	0	0
3	222	22425.00	1	0
4	2281	29832.00	0	0
...	...	...	...	...
213190	1160	10035444.63	0	0
213191	2815	25456876.97	0	0
213192	4250	9072061.82	0	0
213193	2993	4696833.42	0	0
213194	2970	8393318.02	0	0

[213195 rows x 10 columns]

Figure 8 Filtered Final Data



	segment	step	trans_type	amount	nameOrig	oldbalanceOrg
11	0	1	1	23261.30	117730	20411.53
448	0	1	1	1277212.77	154030	1277212.77
480	0	1	1	35063.63	69772	35063.63
794	0	1	1	132842.64	40272	4499.08
1219	0	2	1	1096187.24	187823	1096187.24

	nameDest	oldbalanceDest	accountType	isFraud
11	2331	25742.0	0	1
448	3838	0.0	0	1
480	2302	31140.0	0	1
794	2857	0.0	0	1
1219	3657	0.0	0	1

	segment	step	trans_type	amount	nameOrig	oldbalanceOrg
211275	0	524	1	232315.43	102763	232315.43
211794	0	551	1	813992.49	212186	813992.49
211914	0	567	1	175203.45	80317	175203.45
212005	1	572	1	2000718.20	46733	2000718.20
212492	1	614	1	2739248.30	114048	2739248.30

	nameDest	oldbalanceDest	accountType	isFraud
211275	2191	20819035.38	0	1
211794	1180	24794625.20	0	1
211914	102	2735776.45	0	1
212005	1258	10984451.39	0	1
212492	1339	15505667.41	0	1

Figure 9 Filtered Fraud Data

The accuracy of the naive Bayes classifier, according to the reference study, is 0.8125. Since the accuracy obtained [17] with the Cat boost classifier is 0.99, it can be concluded that greater accuracy can be obtained with it. Comparison Result is shown in Figure 6. Figure 7 shows the difference between the accuracy level of other ref papers and the cat boost algorithm [18]. Figure 8 Shows the Filtered Final Data.

The column labeled "Fraud" in Figure 9 FILTERED FRAUD DATA denotes [19] a fraudulent transaction that requires manual monitoring. False Positive is decreased by this model. The F1 score was 0.42 and the average test accuracy was 0.99. This high precision indicates that the model has retained the material effectively. The F1 Score for the testing process is shown in Figure 10. [20-22]

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 395 - 400
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0058

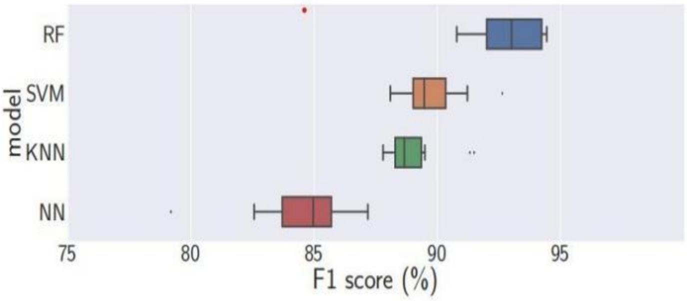


Figure 10 F1 Score for Testing

Conclusion

In this project, a suggested ML detection strategy lays out the conditions and steps that must be taken in order for numerous banks to work together on an AML project. Each supervised model is trained using a subset of characteristics from the train set, and it is subsequently assessed across the complete test set. to track one's performance over time. We employ the keras-based Neural Network (NN) implementation from scikit-learn. We classified money laundering activities into two groups—illegal and legal— using the data mining technique. Following the use of data mining, we clarified how real-time banking transaction detection is possible in terms of the anti-money laundering classification. It provides the optimal choice. Achieving the highest precision is also beneficial. This method could be useful in addressing important questions about anti-money laundering. We discovered the classification and probability by applying the Random Forest classification technique to a few dataset attributes. This investigation showed that, occasionally, this classification aids in the discovery of fraud control, including anti-money laundering procedures or money laundering detection.

References

[1]. Amr Ehab Muhammed Shokry1, Mohammed Abu Rizka2 and Nevine Makram Labib3 "Counter Terrorism Finance by Detecting Money Laundering Hidden Networks Using Unsupervised Machine Learning Algorithm", International Conferences ICT, Society, and Human Beings 2020.

[2]. NOVA IMS & Feedzai Maria Inês Silva, Feedzai David Aparício, Feedzai João Tiago "Machine learning methods to detect money laundering in the Bitcoin blockchain in the presence of label scarcity" Joana Lorenz, arXiv: 2005.14635v2 [cs.LG] 5 Oct 2021

[3]. 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016, pp. 954– 960. doi: 10.1109/ICMLA.2016.73. Camino, R. D. et al. (2017) "Finding suspicious activities in financial transactions and distributed ledgers," IEEE International Conference on Data Mining Workshops, ICDMW, 2017-November, pp. 787–796. doi:10.1109/ICDMW.2017.109.

[4]. Z. Chen, L. D. Van Khoa, E. N. Teoh, A. Nazir, E. K. Karuppiah, and

[5]. K. S. Lam, "Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review," Knowl. Inf. Syst., vol. 57, no. 2, pp. 245–285, 2018, doi: 10.1007/s10115-017- 1144-z

[6]. N. A. Le Khac and M. T. Kechadi, "Application of data mining for antimoney laundering detection: A case study," Proc. - IEEE Int. Conf. Data Mining, ICDM, pp. 577–584, 2010, doi: 10.1109/ICDMW.2010.66.

[7]. S. Gao, D. Xu, H. Wang, and Y. Wang, "Intelligent anti-money laundering system," 2006 IEEE Int. Conf. Serv. Oper. Logist. Informatics, SOLI 2006, no. 7001805, pp. 851–856, 2006, doi: 10.1109/SOLI.2006.235721

[8]. C. H. Tai and T. J. Kan, "Identifying Money Laundering Accounts," Proc. 2019 Int. Conf. Syst. Sci. Eng. ICSSE 2019, pp. 379–382, 2019, doi: 10.1109/ICSSE.2019.8823264

[9]. Joao Paulo A. Andrade, Leonardo S. Paulucio,Rodrigo F. Berriel,Teresa Cristina James Carencro, "A Machine Learning-based System for Financial Fraud Detection", Universidad Federal do Esp´ırito Santo (UFES), Brazil,Instituto Federal do Espırito Santo (IFES), Brazil

[10]. CHUNG-CHIA HUANG AND ASHER TRANGLE,"Anti-Money Laundering and Block chain Technology", HARVARD Law School, January 26, 2020

[11]. Al-Suwaidi, N.A. and Nobanee, H. (2020),

"Anti-money laundering and anti-terrorism financing: a survey of the existing literature and a future research agenda", Journal of Money Laundering Control, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/JMLC-03-2020-0029

[12]. Joana Lorenz, Maria Inês Silva and David Aparício.”Machine learning methods to detect money laundering in the Bitcoin blockchain in the presence of label scarcity ”presented at ICAIF ’20,

[13]. October 15–16, 2020, New York, NY, USA

[14]. Amr Ehab Muhammed Shokry1 , Mohammed Abu Rizka2,* and Nevine Makram Labib3,”COUNTER TERRORISM FINANCE BY DETECTING MONEY LAUNDERING HIDDEN NETWORKS USING UNSUPERVISED MACHINE LEARNING ALGORITHM”, International Conferences ICT, Society, and Human Beings 2020)

[15]. Kidist Sintayehu1 & Hussien Seid (PhD)2,”Developing Anti Money Laundering Identification using Machine Learning Techniques “,published by Irish Interdisciplinary Journal of Science & Research (IIJSR),30 March 2023,Volume 7, Issue 1, Pages 64-74,

[16]. Todd Price, Jake Rigg, Shantell Goodwin, Jason Piazza, “Money Laundering: An Integrated Collaborative Framework and Compliance Strategy. “published by Southern Oregon University, March 20, 2019

[17]. Jiayi Liu 1 , Changchun Yin 1 , Hao Wang 1 , Xiaofei Wu 2 , Dongwan Lan 1 , Lu Zhou 1,* and Chunpeng Ge 3 ,”Graph Embedding-Based Money Laundering Detection for Ethereum”,published by electronics,21 July 2023

[18]. Floris Visser,”Detection of Money Laundering Transaction Network Structures and Typologies using Machine Learning Techniques”, ERASMUS SCHOOL OF ECONOMICS,03-05-2020

[19]. Michele Starnini1(B), Charalampos E. Tsourakakis1,4, Maryam Zamanipour1, André Panisson1, Walter Allasia2, Marco Fornasiero2, Laura Li Puma3, Valeria Ricci3, Silvia Ronchiadin3, Angela Ugrinoska2, Marco Varetto2, and Dario Moncalvo2.”Smurf-Based Anti-money Laundering in Time-Evolving Transaction Networks”,10 March 2022.

[20]. Jos´e-de-Jesus Rocha-Salazar a,* , Maria Jesus Segovia-Vargas b , Maria-del-Mar CamachoMi˜nano c ,”Money laundering and terrorism financing detection using neural networks and an abnormality indicator”, Expert Systems With Applications

[21]. Indurthy Meghana1 , Bitra Pavan Venkatesh2 , Gaddipati Keerthi Ganesh3 , Nadendla Sumanth4 , and Redrouthu Tarun Teja5 ,”Prediction of Financial Crime Using Machine Learning ”,published by International Journal of Innovative Research in Computer Science and Technology (IJIRCST) ,ISSN (online): 2347-5552, Volume-11, Issue-3, May 2023https://doi.org/10.55524/ijircst.2023.11.3.19 Article ID IJIRD-1257,Pages 96-100 www.ijircst.org

[22]. Kamlesh D. Rohit and Dharmesh B. Patel, “Review On Detection of Suspicious Transaction In Anti-Money Laundering Using Data Mining Framework”, published by IJIRST –International Journal for Innovative Research in Science & Technology| Volume 1 | Issue 8 | January 2015 ISSN (online): 2349-6010