Cryptocurrency Transaction Fraud Detection: An Advanced Machine Learning And Deep Learning Approach

Mr. K. MAHANTHI [1], Ms. ENDLA MOUNIKA [2]

#1 Assistant professor in the Department of IT at DVR & DR. HS MIC College of Technology (Autonomous), Kanchikacherla, NTR District.

#2 MCA student in the Department of Computer Applications (DCA) at DVR & DR. HS MIC COLLEGE OF TECHNOLOGY, Kanchikacherla, NTR District

ABSTRACT: Fraudulent transactions have a huge impact on the economy and trust of a cryptocurrency network. Consensus algorithms like proof of work or proof of stake can verify the validity of the transaction but not the nature of the users involved in the transactions or those who verify the transactions. This makes a cryptocurrency network still vulnerable to fraudulent activities. One of the ways to eliminate fraud is by using machine learning techniques. Machine learning can be of supervised or unsupervised nature. In this paper, we use various supervised machine learning techniques to check for fraudulent and legitimate transactions. We also provide an extensive comparative study of various supervised machine learning techniques like decision trees, Naive Bayes, logistic regression, multilayer perceptron, and so on for the above task.

1.INTRODUCTION

For a very long time, research has focused on the issue of identifying fraudulent transactions. Fake exchanges are hurtful to the economy and deter individuals from putting resources into bitcoins or in any event, confiding in other digital money based arrangements. Transactions that are fraudulent typically raise questions about the participants or the nature of the transaction. Individuals from a cryptographic money network need to recognize Fake exchanges as quickly as time permits to keep them from hurting the digital currency organization's local area and honesty. Many AI procedures have been proposed to manage this issue, a few outcomes give off an impression of being very encouraging, however there is no conspicuous unrivaled technique. This

paper looks at the exhibition of different managed AI models like SVM, Choice Tree, Gullible Bayes, Calculated Relapse, and barely any profound learning models in recognizing false exchanges in a digital money organization. Based on the trade-off between accuracy and computational speed, this comparative study will assist in selecting the best algorithm. We want to see which clients and exchanges have the most elevated likelihood of being associated with false exchanges.

2.LITERATURE SURVEY

2.1 S. Su, Y. Sun, X. Gao, J. Qiu* and Z. Tian*. A Correlation change based Feature Selection Method for IoT Equipment Anomaly Detection. Applied Sciences.

In the era of the fourth industrial revolution, there is a growing trend to deploy sensors on industrial equipment, and analyze the industrial equipment's running status according to the sensor data. Thanks to the rapid development of IoT technologies , sensor data could be easily fetched from industrial equipment, and analyzed to produce further value for industrial control at the edge of the network or at data centers. Due to the considerable development of deep learning in recent years, a common practice of such analysis is to conduct deep learning. Such methods select a subset of all fetched sensor data stream as the input features, and generate equipment predictions. As a result, the performance of the learning model was seriously impacted by the features selected, thus feature selection plays a critical role for such methods.

To select an appropriate set of features for the learning model, researchers aim to select the most relevant features to the prediction model to improve the prediction performance, or to select the most informative features to conduct data reduction. Unfortunately, both kinds of methods have intrinsic drawbacks when applied in the online scenarios. The former kind of methods seriously depends on predefined evaluation criteria, such as feature relevance metrics or a predefined learning model . Thus, such method are limited to certain dataset, and are not suitable for online scenarios which involve dynamical and unsupervised feature selection. The later kind of methods right fits in the online scenarios. However, data reduction mainly aims to improve the efficiency (but not accuracy) of the prediction model, which is not the most concerning factor of online industrial equipment status analysis.

To relieve the dependency of predefined evaluation criteria, researchers switch to

select the features which can indicate the online sensor data's characters, such as features which are smoothest on the graph , or the features with highest cluster ability. In this paper, we focus on the features with correlation changes such as smoothness and clusterability, which are important characters for traditional pattern recognition fields like image processing and voice recognition . We believe that correlation changes can significantly pinpoint status changes in industrial environment. As far as we know, this is the first work focusing on correlation changes for online feature selection.

2.2.X. Yu, Z. Tian, J. Qiu, F. Jiang. A Data Leakage Prevention Method Based on the Reduction of Confidential and Context Terms for Smart Mobile Devices. Wireless Communications and Mobile Computing, https://doi.org/10.1155/2018/5823439.

With the development of Internet and information technology, smart mobile devices appear in our daily lives, and the problem of information leakage on smart mobile devices will follow which has become more and more serious . All kinds of private or sensitive information, such as intellectual property and financial data, might be distributed to unauthorized entity intentionally or accidentally. And that it is impossible to prevent from spreading once the confidential information has leaked.

According to survey reports, most of the threats to information security are caused by internal data leakage. These internal threats consist of approximate 29% private or sensitive accidental data leakage, approximate 16% theft of intellectual property, and approximate 15% other thefts including customer information, and financial data. Further, the consensus of approximate 67% organizations shows that the damage caused from internal threats is more serious than those form outside.

Although laws and regulations have been passed to punish various behaviors of intentional data leakage, it is still hard to prevent data inleakage effectively. Confidential data can be easily disguised by rephrasing confidential contents or embedding confidential contents in nonconfidential contents . In order to avoid the problems arising from data leakage, lots of software and hardware solutions have been developed which are discussed in the following chapter.

In this paper, we present CBDLP, a data leakage prevention model based on confidential terms and their context terms, which can detect the rephrased confidential contents effectively. In

CBDLP, a graph structure with confidential terms and their context involved is adopted to represent documents of the same class, and then the confidentiality score of the document to be detected is calculated to justify whether confidential contents is involved or not. Based on the attribute reduction method from rough set theory, we further propose a pruning method. According to the importance of the confidential terms and their context, the graph structure of each cluster is updated after pruning. The motivation of the paper is to develop a solution which can prevent intentional or accidental data leakage from insider effectively. As mixed-confidential documents are very common, it is very important to accurately detect the documents containing confidential contents even when most of the confidential contents have been rephrased.

2.3 Y. Sun, M. Li, S. Su, Z. Tian, W. Shi, M. Han. Secure Data Sharing Framework via Hierarchical Greedy Embedding in Darknets. ACM/Springer Mobile Networks an

Geometric routing, which combines greedy embedding and greedy forwarding, is a promising approach for efficient data sharing in darknets. However, the security of data sharing using geometric routing in darknets is still an issue that has not been fully studied. In this paper, we propose a Secure Data Sharing framework (SeDS) for future darknets via hierarchical greedy embedding. SeDS adopts a hierarchical topology and uses a set of secure nodes to protect the whole topology. To support geometric routing in the hierarchical topology, a two-level bit-string prefix embedding approach (Prefix-T) is first proposed, and then a greedy forwarding strategy and a data mapping approach are combined with Prefix-T for data sharing. SeDS guarantees that the publication or request of a data item can always pass through the corresponding secure node, such that security strategies can be performed. The experimental results show that SeDS provides scalable and efficient end-to-end communication and data sharing

3.PROPOSED SYSTEM

Machine learning can be of supervised or unsupervised nature. In this paper, we use various supervised machine learning techniques to check for fraudulent and legitimate transactions. We also provide an extensive comparative study of various supervised machine learning techniques like decision trees, Naive Bayes, logistic regression, multilayer perceptron, and so on for the above task.

3.1 IMPLEMENTATION

Gathering the datasets: We gather all the r data from the kaggle website and upload to the proposed model

Generate Train & Test Model: We have to preprocess the gathered data and then we have to split the data into two parts training data with 80% and test data with 20%

Run Algorithms: For prediction apply the machine learning models on the dataset by splitting the datasets in to 70 to 80 % of training with these models and 30 to 20 % of testing for predicting

Input data: In this module we will give dataset as our input

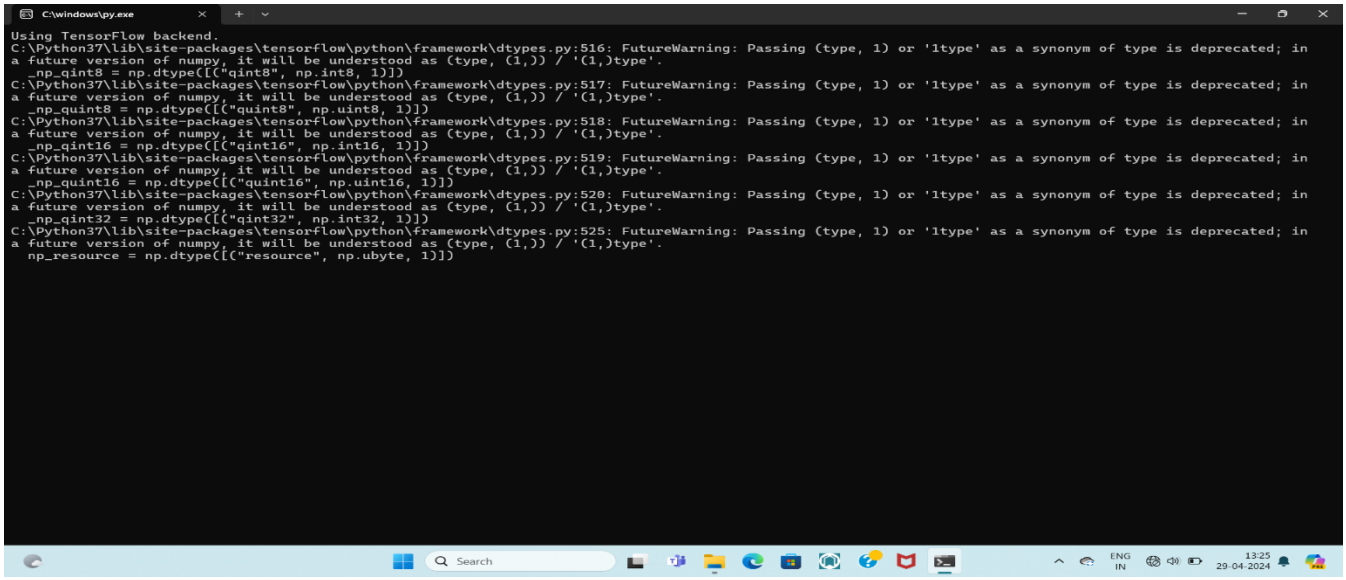
Predict output: in this module we will get output based on input dataset and find frauds

4.RESULTS AND DISCUSSION

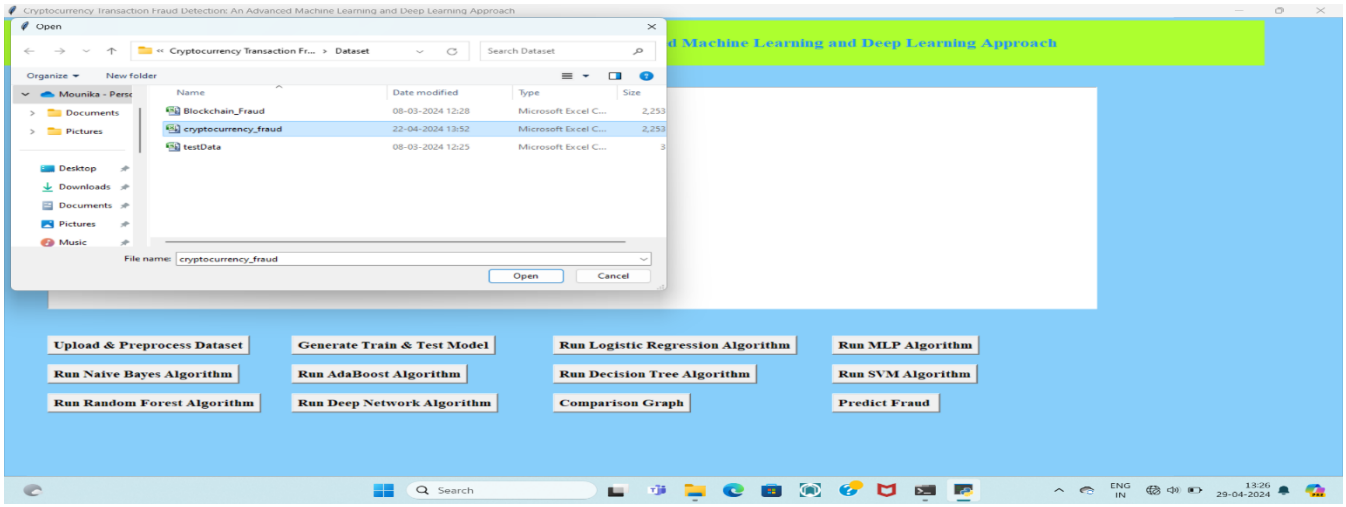
Fraud transaction effect economy of any country in the world and Cryptocurrency consider secured against any attack due to its proof of work and transaction validation via hash code but the user who is involving Cryptocurrency transaction cannot be trusted and he may perform fraud transaction and to predict such user transaction author of this paper is using various machine learning algorithms called Logistic Regression, MLP, SVM, Decision Tree and many more than evaluating performance of this algorithms in terms of accuracy.

To implement this paper author has used Cryptocurrency fraud transaction dataset which contains user and transaction details and then we extracted all transaction details and then process dataset to normalize value and then replace missing values with 0 and then remove all non-numeric data.

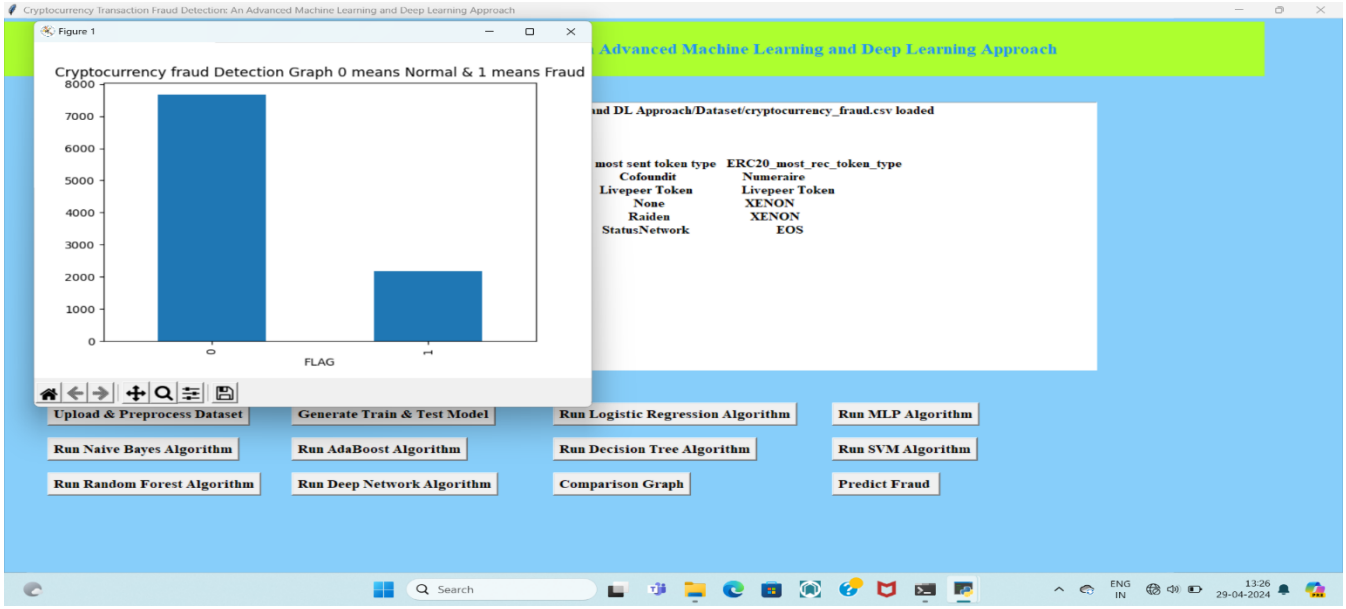
Below screen showing dataset details



In above screen first row contains dataset column names and remaining rows contains dataset values and in dataset we have column called FLAG which contains values as 0 and 1 where 0 means Normal transaction and 1 means fraud transaction

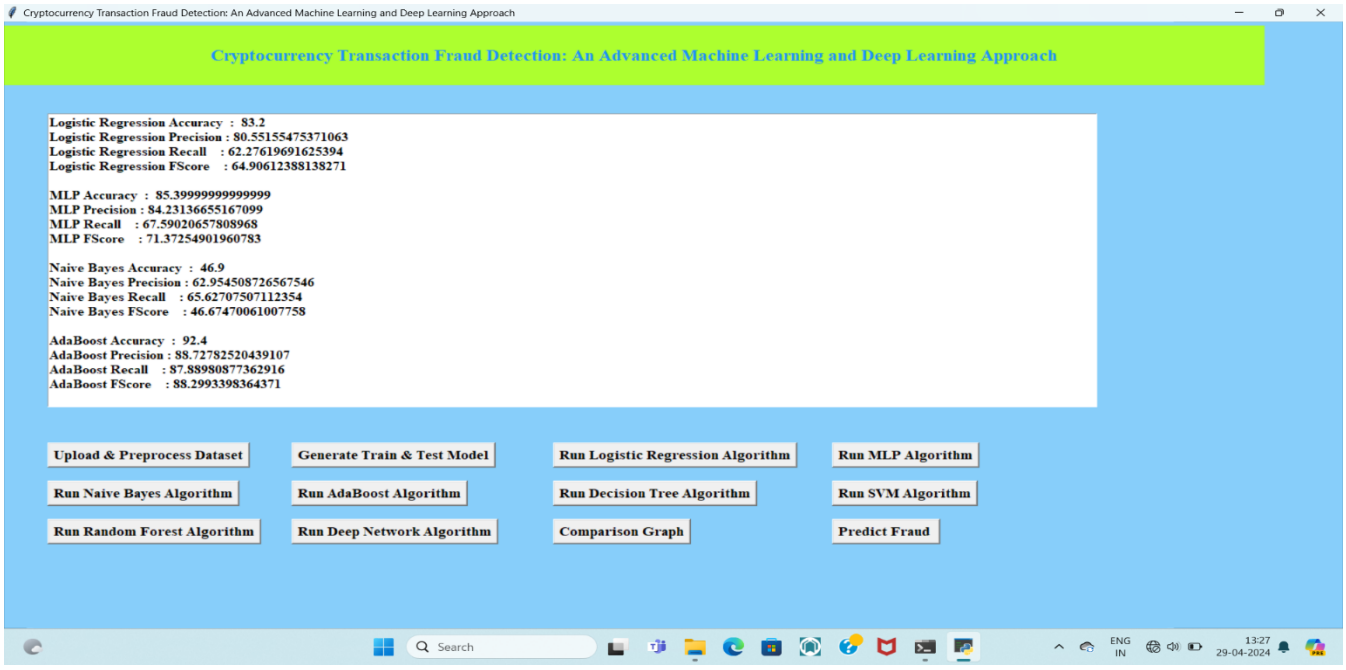


In above screen selecting and uploading dataset and then click on 'Open' button to load dataset and get below output



In above screen dataset loaded and dataset contains some non-numeric data and ML algorithms will not take such data so we need to remove and graph x-axis contains type of transaction and y-axis contains number of records and now close above graph and then click on 'Generate Train & Test Model' button to get below output

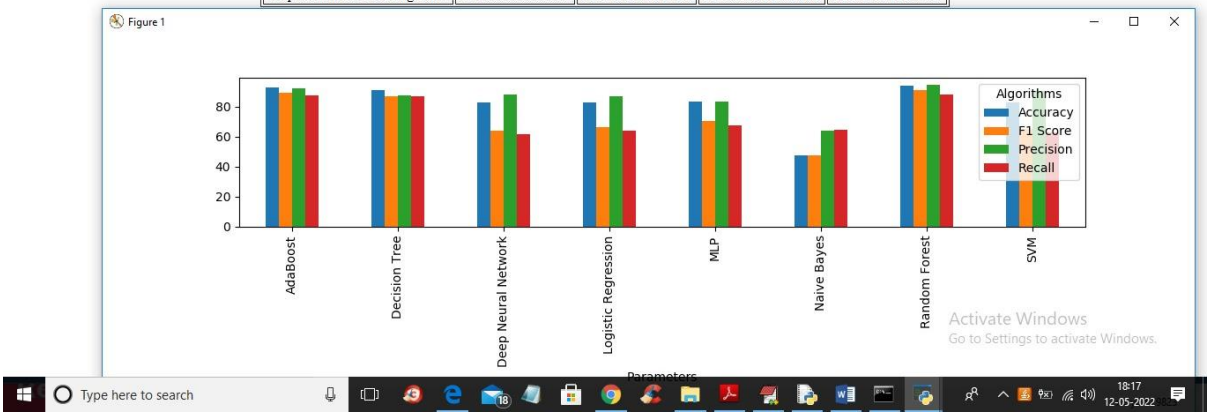
In above screen we can see all data converted to numeric format and we can see total records found in dataset with total columns and then split dataset into train and test and now train and test data is ready and now click on each button to run all algorithms and get below output



In above screen we can see the performance or accuracy of each algorithm and below is the remaining algorithm accuracy

In above screen we can see accuracy of AdaBoost, Decision Tree and SVM and below is the accuracy of remaining algorithms

Algorithm Name	Accuracy	Precision	Recall	FSCORE
Logistic Regression Algorithm	82.6	86.7816091954023	63.710919193947916	66.43285413338889
MLP Algorithm	83.5	83.216280925778 14	67.41480886389503	70.69682939694076
Naive Bayes Algorithm	47.5	63.652193494666 18	64.54386199817009	47.4722128005715
AdaBoost Algorithm	92.9	92.16902239035376	87.50027894936511	89.53591025053389
Decision Tree Algorithm	90.8	87.3966003090628	86.72312601816519	87.05239277410928
SVM Algorithm	82.5	90.701381509032 94	62.60683760683761	65.01056680882374
Random Forest Algorithm	93.899999999999 99	94.616026983474 58	88.15302046372541	90.8555461112835
Deep Neural Network Algorithm	93.899999999999 99	94.616026983474 58	88.15302046372541	90.8555461112835



In above screen we can see random forest and Deep neural accuracy and in all algorithms Random Forest is giving better accuracy. Now click on 'Comparison

Graph' button to get below output

In above screen we can see the accuracy, precision, recall and FSCORE of each algorithm in graph and tabular format and in all algorithms Random Forest giving better result

5.CONCLUSION

A machine learning-based approach for identifying fraudulent transactions in a bitcoin network has been presented. Several supervised learning techniques, including logistic regression, support

vector machines, decision trees, and dense neural networks, were examined in this method. Accuracy is used to conduct a complete comparison examination of all the approaches. For the comparative analysis of unsupervised algorithms like as clustering, this work can be expanded. We also want to conduct a thorough investigation into fraudulent activity in a private cryptocurrency in the future.

REFERENCES

[1] ai, Y., Zhu, D. Fraud detections for online businesses:a perspectivefromcryptocurrencytechnology.Finance Innov 2,20 (2016). https://doi.org/10.1186/s40854-016-0039-4

[2] yvarinen, H., Risius, M. & Friis, G. A Cryptocurrency-Based Approach ¨ Towards Overcoming Financial Fraud in Public Sector Services. Bus Inf Syst Eng 59, 441–456 (2017). https://doi.org/10.1007/s12599-017-0502-4

[3] u, J.J. Are cryptocurrency's immune to all malicious attacks?. Finance Innov 2, 25 (2016). https://doi.org/10.1186/s40854-016-0046-5

[4] stapowicz M., Zbikowski K. (2019) Detecting Fraudulent Accounts on ˙ Cryptocurrency: A Supervised Approach. In: Cheng R., Mamoulis N., Sun Y., Huang X. (eds) Web Information Systems Engineering – WISE 2019. WISE 2020. Lecture Notes in Computer Science, vol 11881. Springer, Cham. https://doi.org/10.1007/978-3-030-34223-4 2

[5] C odgorelec, B., Turkanovic, M. and Karakati ´ c, S., 2020. A Machine ˇ Learning-Based Method for Automated Cryptocurrency Transaction Signing Including Personalized Anomaly Detection. Sensors, 20(1), p.147.

[6] H arrugia S, Ellul J, Azzopardi G. Detection of illicit accounts over the Ethereum cryptocurrency. Expert Systems with Applications. 2020 Jul 15;150:113318.

[7] ham, Thai, and Steven Lee. "Anomaly detection in bitcoin network using unsupervised learning methods." arXivX preprint arXiv:1611.03941 (2016).

[8] onamo, Patrick, Vukosi Marivate, and Bheki Twala. "Unsupervised learning for robust Bitcoin fraud detection." 2016

Information Security for South Africa (ISSA). IEEE, 2016.

[9]
hi, Fa-Bin, et al. ”Anomaly detection in Bitcoin market via price return analysis.” PloS one 14.6 (2019): e0218341.

[10]
i, Ji, et al. ”A Survey on Cryptocurrency Anomaly Detection Using Data Mining Techniques.” International Conference on Cryptocurrency and Trustworthy Systems. Springer, Singapore, 2019.

[11]
. N. Sureshbhai, P. Bhattacharya and S. Tanwar, ”KaRuNa: A Cryptocurrency-Based Sentiment Analysis Framework for Fraud Cryptocurrency Schemes,” 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland, 2020, pp. 1-6, doi: 10.1109/ICCWorkshops49005.2020.9145151.

[12]
[12] Brenig, Christian, and Gunter M ¨ uller. ”Economic analysis of cryptocur- ¨ rency backed money laundering.” (2015).

AUTHOR PROFILE

Mr. K. MAHANTHI Assistant professor in the department of IT at DVR & DR. HS MIC COLLEGE OF TECHNOLOGY (Autonomous), Kanchikacherla, NTR (DT). His areas of interest include C language, Data science and Python, Web technologies.

Ms. ENDLA MOUNIKA, as MCA student in the department of Computer Application at DVR & DR. HS MIC COLLEGE OF TECHNOLOGY, Kanchikacherla, NTR (DT). She has completed B. Sc (MSCs) in Maris Stella College From KRISHNA UNIVERSITY. His areas of interests are Frontend developer, Full Stack Developer.