International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 7 ǁ 2021 ǁ PP. 54-57

Detection of Licit and Illicit Transaction in Bitcoin Network using Machine Learning

Siddhartha B S[1], Chaithra R[2], Meghana K[3], Navyashree H D[4], Priyanka V L[5]

ABSTRACT: *Every year, criminals capture billions of dollars worth of trivial crimes (e.g., terrorism, drug trafficking) and harm many people and their economy. Cryptocurrencies, in particular, have developed as a platform for money laundering. Machine learning can be used to detect these illegal methods First, we point out that existing state-of-the-art solutions that use unsupervised access methods are not enough to detect illegal patterns in real Bitcoin transaction databases. After that, we demonstrate that our supervised learning solution is capable of match performance.*

Keywords: Logistic Regression, Support Vector Machine, K-nearest Neighbor.

Date of Submission: 06-07-2021 Date of acceptance: 19-07-2021

I. INTRODUCTION:

Illegal money laundering is a major problem worldwide. Criminals are illegally obtaining money in large cases and then investing it in a financial system that is considered legitimate. Money laundering is an illegal way to make big money as a result of a criminal act, such as drug trafficking or terrorist financing, which seems to come from a legitimate source. Money from criminal activity is considered dirty, and the process is "washed" to make it look clean, Governments and international organizations have developed stricter regulations on the use of cryptocurrencies and are increasing their investment in cryptocurrencies, where criminals benefit from anonymity. In the financial sector, Anti-Money Laundering (AML) efforts often rely on law-based programs. However, the risk arises from the limited reduction of publicly available sets-sets, leading to higher positive positive (FPR) levels and lower detection rates. Machine learning strategies overcome the rigidity of legal-based programs by incorporating complex patterns from historical data, and can increase adoption rates and reduce FPR. Recently, Weber et al. extract data, elliptical data Elliptic Data Set Bitcoin transactions to real organizations under the licensing category as opposed to illegal ones (scams, malware, terrorist organizations etc.). The function in the database is to illegally separate and license the graph. This anonymous data set is a transaction graph collected on the Bitcoin blockchain. The node in the graph represents the transaction, the edge can be viewed as the flow of Bitcoins between transactions and so on. Each node has 166 features and is labeled as "Licit", "Illicit" or "Unknown" business.

Containing a sample of 200k labeled Bitcoin Transactions, Supervised reading is a machine learning activity that marks input output based on the model input output in pairs. [1] Provides work from the details of a training label that contains a set of training examples. [2] In supervised learning, an example is a pair that includes an input object and the required output value. Particularly, we show that: (1) Detecting money laundering cases in the Bitcoin network without any labels is impossible since illicit transactions hide within clusters of licit behavior.

1.1 Problem Statement: Machine learning can be used to find patterns. Labels are so frightening that traditional unsupervised algorithms do not work. Two percent (4,545) nodes are classified as class1 (licit). Twenty-one percent (42,019) are classified as class2 (illicit). The remaining transactions are not registered in respect of the license as opposed to being illegal. So most of the data doesn't have labels so label shortages are higher. Financial crime is extremely rare in licensing operations and the database is not very accurate.

II. SYSTEM ANALYSIS:

2.1 Existing System: To detect licit and illicit activity on the bitcoin transaction datasets they were using unsupervised anomaly detection method. AL has also successfully been applied in other fraud-related use-cases such as Bitcoins transactions are transfers from one Bitcoin address to another, represented as nodes in the graph.

2.2 Proposed System: In our project we have more imbalanced data, The class 3 is not considering, here we consider only a known things that is licit and illicit, Because unknown is uge if do any prediction regarding that it will show more on that , The train set includes all labeled samples up to the 34th time-step (16670 transactions), and the test set includes all labeled samples from the 35th time-step, in (29894= transactions). We used supervised algorithm to detect the data transactions as licit or illicit We use the logistic regression (LR)

,Support Vector Machine (SVM),K-nearest Neighbor to find accuracy.

2.3 System Architecture: In the figure shows that block diagram we collect a huge amount of data on the use of Bitcoin transaction. Here we implements the logistic regression and support vector machine(SVM) and K-nearest Neighbor Algorithms to classify problems. we are providing graphical output of licit and illicit types by using python on ML domain.

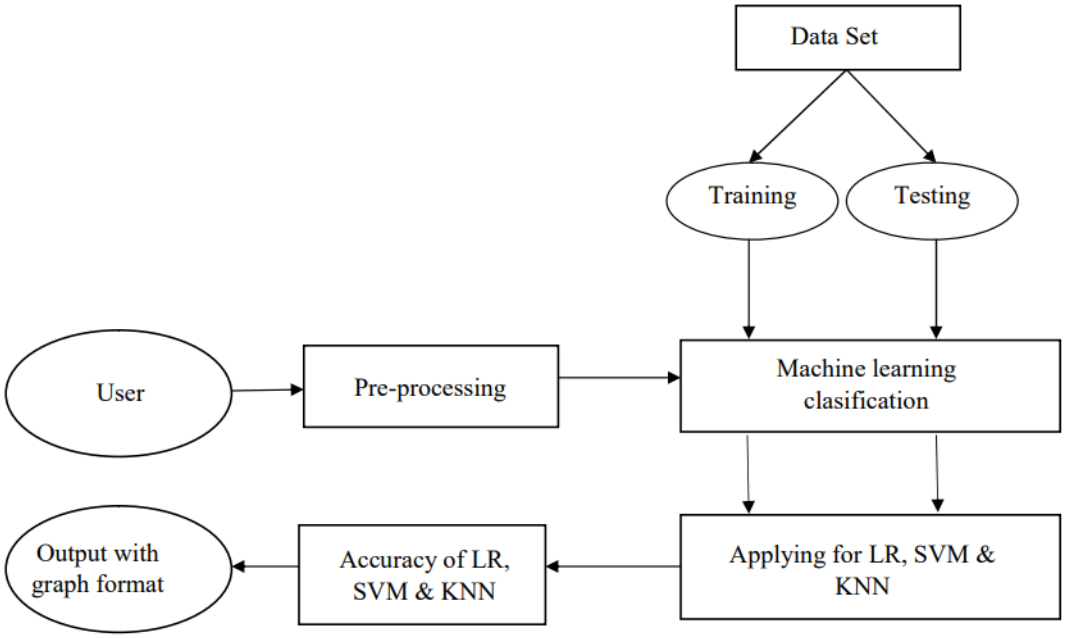

Figure 1: System Architecture

2.4 Data Flow Diagram:

in this data flow diagram here going to explain how the data will be process user will provide the data to the data set which will be classified as illicit illicit and unknown and legal transaction which will be converted as currency and which will be usually Known by the end user and the information about the transaction will be store if we come to the part of illicit transaction as by the name it is the illegal activity which is also known by the end-user another one is unknown which is known directly by the end user.

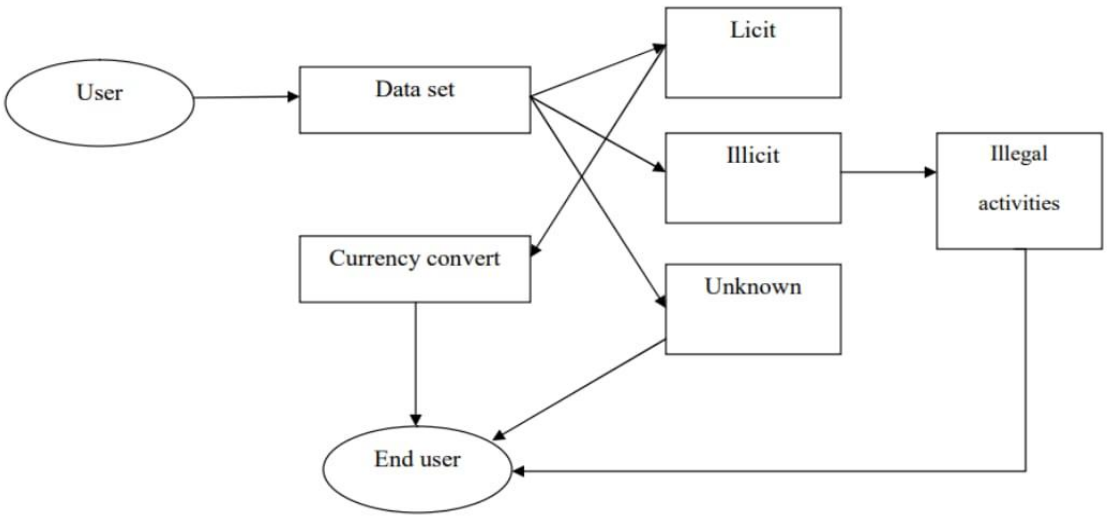

Figure 2: Data Flow Diagram

2.5 Methodology: in this methodology we have input data set which is called in.csv extension in this data visualisation consists of data which is visualised in the form of graph we are adding new data to the existing data set which will be used to feature extraction and it undergoes train are test by using ml classification finally we get output.

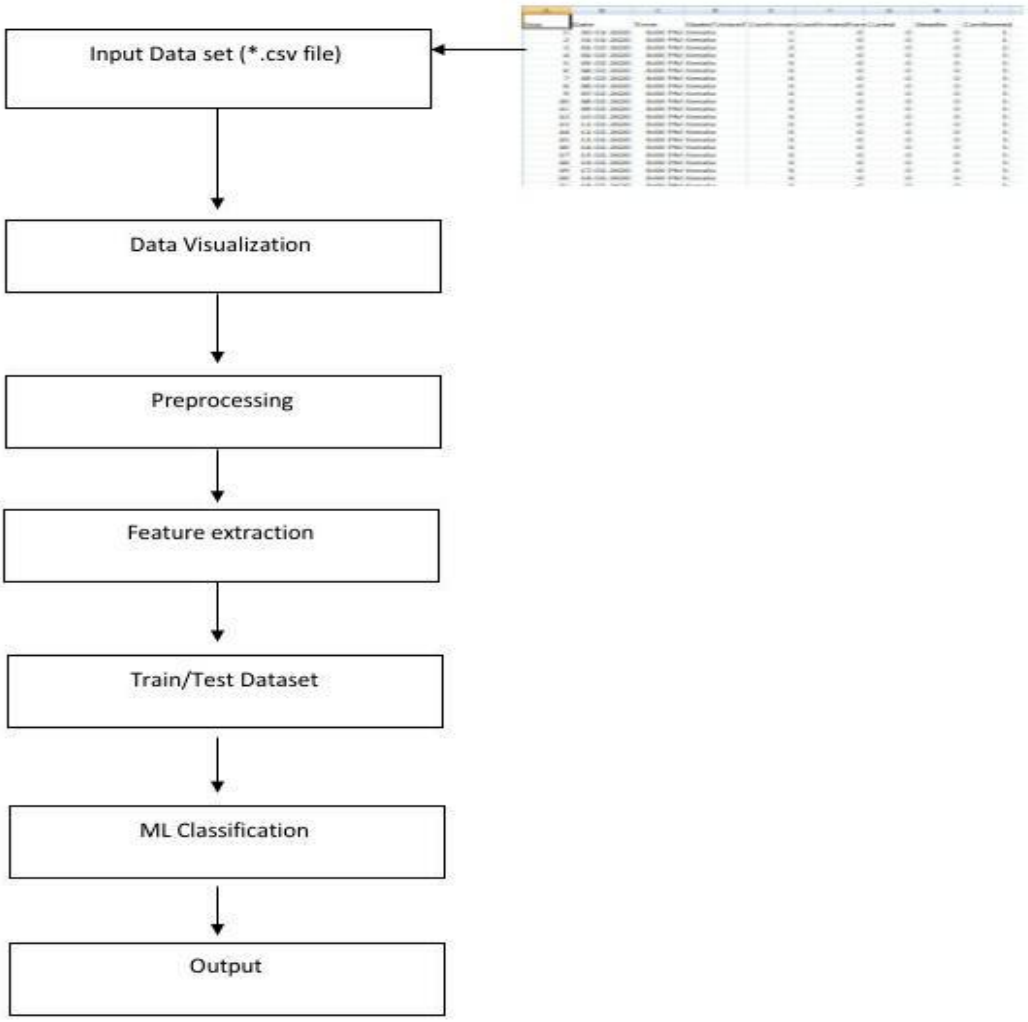

Figure 3: Methodology

III. IMPLEMENTATION:

3.1 Numpy library: NumPy represents a python number python package for calculating and processing the same components of multiple and one dimension. It makes good use of various layouts. NumPy provides built-in functions for straight algebra and random number generation. NumPy makes a computer targeted at the same members.

3.2 Pandas library: Pandas is described as an open library that provides high-level data fraud in Python. Data analysis requires a lot of processing, such as reorganization, cleaning or merging, etc. But we love pandas because working with pandas is faster, easier and more expressive than other tools.

3.3 Matplotlib library: People's minds work better on data display than text data. We can easily grasp things from a figurative perspective. It is better to represent the data using a graph where we can analyze the data well and make a certain decision based on the data analysis. Before reading matplotlib, we need to understand data recognition and why data recognition is important.

3.4 Seaborn library: Seaborn is an amazing viewing library of mathematical graphics editing in Python. It offers beautiful automated styles and color palettes to make mathematical sites more attractive. It is built on top of the matplotlib library and is also heavily integrated into data structures from pandas. Seaborn aims to make

visibility an integral part of data analysis and understanding. Provides database-based APIs, so we can switch between different presentations of the same variable views to better understand the database.

3.5 Scikit-learn library: Skikit-learn (Sklearn) is a very useful and robust library of machine learning in Python. It offers a selection of effective machine learning tools and mathematical modeling including division, deceleration, merging and reduction of size using the Compatible interface in Python. This library, mostly written in Python, is built on NumPy, SciPy and Matplotlib.

3.6 Training: The training data set is the common name for the samples used to create the model while we are training and measuring our model equally to match the parameters of the training data available for modeling.

3.7 Testing: Test data is used only to assess performance of model, testing data is the unseen data for which predictions have to be made.

3.8 Logistic Regression: Logistic regression that is used to divide and produce an effect in binary format the result should be divided as 0 0r 1, Yes or No, Used for multiple divisions. collecting the data we are importing the libraries after that analyzing the data here we explore the data, in data wrangling cleaning the data and checking whether their is a null values then, splitting dataset in to training and testing build model on the train data and predict output on the test data then accuracy check.

3.9 Support Vector Machine: SVM is a supervised learning method that looks at data and integrates it into one of two categories, we have past labeled information and model training that will be done after the new data will be available and predictions will be made in the end the result will be predict.

3.10 K-nearest Neighbor: The KNN algorithm is one of the easiest segmentation algorithms and is one of the most widely used learning algorithms. Here we process the data and extract the data and apply the KNN split. If the data is identified then we get the exact output otherwise we do the data processing and then we get the result.

IV. FUTURE ENHANCEMENT AND CONCLUSION

In future development we will try to use active learning to label new data points with the o/p you want in order to achieve the best test design. We have conducted research to find a licensed and illegal activity in the Bitcoin transaction database issued by elliptic. Our results show that the uncontrolled acquisition method has malfunction, to improve unchecked performance, to study the case where fewer labels can be obtained through use and to determine the minimum number of label conditions required to achieve the closest operation of the most monitored base.

REFERENCES

[1]. Jesse Crawford, Yong Guan. 2020. Knowing your bitcoin customer: Money laundering in the Bitcoin Economy.
[2]. Mohamed Florida. 2018. Bitcoin Concepts, Threats, and Machine-Learning Security Solution.
[3]. Sagwadi Mabunda. 2018. Crypto currency: The New Face of Cyber Money Laundering.
[4]. Reza Soltani, Uyen Trang Nguyen, Yang Yang,. 2016. A New Algorithm For Money Laundering Detection Based On Structural Similarity.
[5]. Haohua Sun Yin, Ravi Vatrapu. 2017. A First Estimation of the Proportion of Cybercriminal Entities in the Bit coin Ecosystem using Supervised Machine Learning.
[6]. Yining Hu,Suranga Seneviratne,Kanchana Thilakarathn. 2019. Characterizing and Detecting Money Laundering Activities on the Bitcoin Network.
[7]. Junwoo Seo,Mookyu Park,Haengrok Oh,Kyungho Lee. 2018. Money Laundering in the Bitcoin Network: Perspective of Mixing Services.
[8]. Sidharth Samanta,Bhabendu Kumar Mohanta, Siba Parsad Pati, Debasisih jena 2019(ICIT). A Framework to Build User Profile on Cryptocurrency Data for Detection of Money Laundering Activities.
[9]. Valeriia Dynut, Oleh Dykyi. (2018), Cryptocurency in The System of Money Laundering.
[10]. Linogxiao Yang, Xuewen Dong, Siyu Xing, jiawei Zheng, Xinyu Gu,Xiongfel Song. 2019. An Abnormal Transaction Detection Mechanism on Bitcoin.