

Chapter 8

K-Nearest Neighbors Algorithm (KNN): An Approach to Detect Illicit Transaction in the Bitcoin Network

Abdelaziz Elbaghdadi

Abdelmalek Essaadi University, Morocco

Soufiane Mezroui

Abdelmalek Essaadi University, Morocco

Ahmed El Oualkadi

 <https://orcid.org/0000-0002-4953-1000>

Abdelmalek Essaadi University, Morocco

ABSTRACT

The cryptocurrency is the first implementation of blockchain technology. This technology provides a set of tracks and innovation in scientific research, such as use of data either to detect anomalies either to predict price in the Bitcoin and the Ethereum. Furthermore, the blockchain technology provide a set of technique to automate the business process. This chapter presents a review of some research works related to cryptocurrency. A model with a KNN algorithm is proposed to detect illicit transaction. The proposed model uses both the elliptic dataset and KNN algorithm to detect illicit transaction. Furthermore, the elliptic dataset contains 203,769 nodes and 234,355 edges; it allows to classify the data into three classes: illicit, licit, or unknown. Each node has associated 166 features. The first 94 features represent local information about the transaction. The remaining 72 features are called aggregated features. The accuracy exceeded 90% with $k=2$ and $k=4$, the recall reaches 56% with $k=3$, and the precision reaches 78% with $k=4$.

DOI: 10.4018/978-1-7998-5781-5.ch008

I. INTRODUCTION

New Technologies are created to spur financial innovation and improve the financial inclusion. These technologies deviate to their main goals and give a new opportunity for criminals and terrorists to launder their proceeds or their illicit activities. The Financial Action Task Force international standards combating money laundering and the financing of terrorism members in 2012 adopted their standards to monitor the risks relating to new technology. In 2014, the Financial Action Task Force (FATF) published virtual currency key definitions and potential AML /CFT Risks regarding the revolution of the cryptocurrency and their mechanisms associated with payment for giving a new method for transmitting values over the internet.

The FATF defines cryptocurrency as a decentralized convertible virtual currency protected by cryptography. The FATF discover and analyze the concrete action taken by criminals to launder incriminated funds through cryptocurrencies, they offer recommendations for compliance officers and companies that deal with cryptocurrencies.

Today, data analysis can be used to detect the anomalies or predict the future results with the help of data in the different fields. Furthermore, others issues should be solved with data analysis in the cryptocurrencies such as the influence of the distance used in the performance of k-Nearest Neighbors (KNN) model and how to use deep learning methods to evaluate precision, recall, F1 and accuracy for this task. The aim of this study is the detection of illicit transaction with KNN algorithm using Elliptic dataset. Furthermore, the blockchain technology uses it for a set field such as Business Intelligence, this technology gives a set technique such as smart contract to automate the processes in the enterprise without a central authority.

The rest of this chapter is organized as follows. The related works in cryptocurrency are presented in section II. In the third section, the Bitcoin and Ethereum network overview are described. In the fourth section, the machine learning technique is described. The fifth section presents the proposed methodology in this study. In the sixth section, the obtained results are discussed. Finally, the conclusion is given.

II. RELATED WORKS

The blockchain technology creates potential innovations in the processing of the business activities in various sectors, which makes this technology face to a set of attacks and illicit activity. This section reviews some related work which target cryptocurrency, such as anomaly detection, data analysis and business intelligence

1. Detection of Fraud and Anomalies in Cryptocurrency

The openness is one of the main characteristics of the Bitcoin network. This technology is open to public at any time. In addition, the public key of the bitcoin is a 160 bits hash generated by the secp256k1 curve (Antonopoulos M. Andreas, 2014) (Joppe W. Bos and al, 2014). This address can change often (Antonopoulos M. Andreas, 2014), and this propriety removes the possibility to identify the Bitcoin users via the public key. This still removes the possibility of tracking identity by analyzing the use of public keys on the network. However the address bitcoin is the main identifier to make a transaction and any one can stipulate that if two addresses (public key) are used as entries in the same transaction, then

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/k-nearest-neighbors-algorithm-knn/267871

Related Content

A Timeline Optimization Approach of Green Requirement Engineering Framework for Efficient Categorized Natural Language Documents in Non-Functional Requirements

K. Mahalakshmi, Udayakumar Allimuthu, L Jayakumarand Ankur Dumka (2021). *International Journal of Business Analytics* (pp. 21-37).

www.irma-international.org/article/a-timeline-optimization-approach-of-green-requirement-engineering-framework-for-efficient-categorized-natural-language-documents-in-non-functional-requirements/269485

The Current State of Analytics in the Corporation: The View from Industry Leaders

Thomas Coghlan, George Diehl, Eric Karson, Matthew Liberatore, Wenhong Luo, Robert Nydick, Bruce Pollack-Johnsonand William Wagner (2010). *International Journal of Business Intelligence Research* (pp. 1-8).

www.irma-international.org/article/current-state-analytics-corporation/43677

Quality Management in Web Warehouses

Adriana Marotta, Laura González, Lorena Etcheverry, Bruno Rienzi, Raúl Ruggia, Flavia Serraand Elena Martirena (2012). *Business Intelligence Applications and the Web: Models, Systems and Technologies* (pp. 1-25).

www.irma-international.org/chapter/quality-management-web-warehouses/58409

Integrated QFD, Fuzzy Linear Regression and ZOGP: An Application of E-Store Design

Pelin Celikand Talha Ustasuleyman (2019). *International Journal of Business Analytics* (pp. 61-73).

www.irma-international.org/article/integrated-qfd-fuzzy-linear-regression-and-zogp/238066

A Framework for Feature Selection Using Natural Language Processing for User Profile Learning for Recommendations of Healthcare-Related Content

Mona Tanwar, Sunil Kumar Khatriand Ravi Pendse (2022). *International Journal of Business Analytics* (pp. 1-17).

www.irma-international.org/article/a-framework-for-feature-selection-using-natural-language-processing-for-user-profile-learning-for-recommendations-of-healthcare-related-content/292059