International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Bank Fraud Detection Using Neo4j and Machine Learning Algorithms

Riyaz. A. Jamadarr[1], Atharva. A. Chavan[2]

[1]Assistant Professor, Department of Information Technology, AISSMS's Institute of Information Technology, Pune-411001, INDIA
[2]TE. (Information Technology), AISSMS's Institute of Information Technology, Pune-411001, INDIA

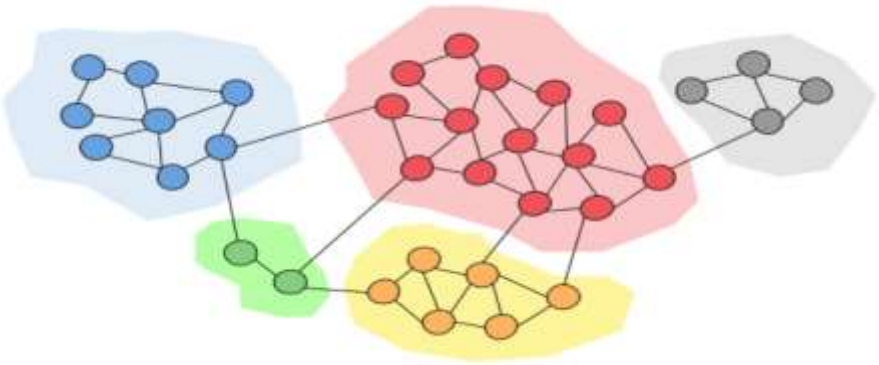
A B S T R A C T

Following the identification of trends, a degree of verification/authentication can be applied to banking operations. In today's world, almost everyone needs to do business with a bank, whether in person or online. When working with banks, both clients and ba Fraud has progressed from being committed by unorganised fraudsters to being committed by organised crime and fraud rings that techniques to take over..Bank fraud is a federal crime that involves cheating by accessing online transactions and credit card information. Banks and financial institutions lose billions annually because of fraud. Scammers attempt bankers with scams to obtain financial assets. The most common types of bank fraud include debit and credit card fraud, account fraud, insurance fraud, money laundering fraud, etc. Bankers needs to make sure that their assests are safe to armor the global financial system.Anti-Fraud the existing systems are easily cracked by fraudsters. This. The paper has proposed an application to detect bank fraud using a community-based detection algorithm that identifies certain behaviours or trends that could lead to fraud. Detecting fraud in the first place. software in banking has the tools and processes that banks use to monitor transactions and payments for suspicious activity. When a transaction or behavior pattern casts a red error, the bank's fraud team can step in. Banks have. advanced their fraud detection capabilities for years.There is a There is a clear need for better fraud detection applications and effective fraud management systems in the banking sector because most transactions are now digital. database used for database creation and representation, and the Cypher request was used as the graphic query language.

Keywords: fraud, community detection algorithm.

1. Introduction To Community Detection Algorithm

Communities are a property of many networks in which a particular network may have multiple communities such that nodes inside a community are densely connected. When analyzing different networks, it may be important to discover communities inside them. Community detection techniques are useful for social media algorithms to discover people with common interests and keep them tightly connected. Community detection can be used in machine learning to detect groups with similar behaviours.For example customers purchasing behaviour



Motivation

The selection of data sources and determination of community detection approaches can enhance the accuracy, efficiency and scalability of community. Social networks portray interactions among the interconnected nodes represented as graph. The community detection process (CDP), provide common relations between users and analyze each related part of a network. Communities in social networks can be performed by different methodologies which have high importance for understanding the types, detecting and analyzing useful and hidden patterns in aforesaid network. Community detection techniques are useful for social media algorithms to discover people with common interests and keep them tightly connected. Community detection can be used in machine learning to detect groups with similar properties and extract groups for various reasons.

1.Aim and Objective of the work

To detect the fraudsters in banking systems.

Project objectives:

To develop a successful scam detection model to detect online banking scams and classification. Additionally, the goal was to clarify how it would affect the prompt and reliable detection of any "odd" transactions.These abnormal transactions were considered outliers for the banking system and treated as fraud activities.To detect bank fraud using a community detection algorithm that identifies the patterns that can lead to fraud occurrences.

Risk:

Address cannot be taken as unique attribute because fraud can happen from same address also if more than one person is living in same house.

2. Features

2.1 Training Module

A training module was used for making step-by-step processes to present a more realistic view of the information flow. The training module consists of raw data given by the users to detect if there are any fraudsters exist or not.

2.2 Storage Module

The storage module is a system module to store the data from the training dataset.It stores the data in the system and sends the data to the training module while necessary

2.3 Detection Module

The core elements of the module are test data, cipher query builder, community detection algorithm, and response builder. Raw data were taken by the test data as a training datasetThe raw data were searched by the cypher query builder. Also the data passed through a community detection algorithm to check the connection between each node and showed the result using the response builder present in Neo4j database.

METHODOLOGY:

The architecture in this paper consists of three modules such as a training module, a storage module, and a detection module. The input data were entered by the legitimate bankers or users and received by the training module. The training module passed test data to the storage module to store the data. The retrieval of data by the training module occurred when necessary. After then data was processed to the detection module to detect the fraud by using a cipher query language and community system algorithm.

Training Module

It gives more realistic view of information.Some unique entities like Aadhar number,Passport number,Mobile Numberare passed to the module and later stored using storage module and then into detection module

Storage Module

The storage module is a system module to store the data from the training dataset. It stores the data in the system and sends the data to the training module while necessary. In this system, neo4j is used to store the data and reprocess the data to the training module if necessary. Besides, the neo4j cluster was being used to process the data in the storage module

Detection Module

The main elements of this module are test data, cipher query builder, community detection algorithm, and response builder. Raw data were taken by the test data as a training dataset. The raw data were searched by the cypher query builder. Then the data is passed through a community detection algorithm to check the connection between each node of the raw data and showed the result set through the response builder.

3. Architecture

3.1 New Account

A application was developed where authorized person from bank can add the customers by clicking on ‘New Account” Button.Then the record will be stored in database using Neo4j. Later that authorized person can check the record using “Account List” option which will be made available in application.

3.2 Choose Search Type

Later it can be checked if the cutomer exists or not with some unique attributes like aadhar number or passport number,mobile number.Here we can take aadhar number for example

3.3 Result

In case of any one of the attributes gets matched then it is detected immediately by the application and the person is suspected as a fraud. The Bank may then institute other proceedings against them. Neo4j was used to create and depict the database graphically because it makes it easier to detect when looking at the connections.

Conclusion-

Machine learning is a method to extract key information from large amounts of data and enable better decision-making in the banking and retail sectors. They use statistical warehousing to combine a variety of database data within an operational framework from which facts can be exploited. Community detection is important for the science of multi-application networks. Community sensing is used extensively in the field of machine learning applications. It also allows researchers to understand the attributes of aggressive processes that take place in a network. As syntagmatic examples, the processes of epidemic spread are significantly affected by the EU structure of the graph. Data-based approaches are extremely useful for identifying trends across the data set. Safety. in the banking industry can be achieved by utilizing encryption and data based on the fraud detection system. Fraud. detection is a challenge and a smart task in today's digital age. Expert. Intelligent systems and software are used to fight fraud. This. The paper presented a system for detecting fraud through a community detection algorithm. was developed through a web app which acts as an intermodal hub between bankers and customers. Neo4j, A graphical database system has been implemented for finding and screening fraud cases. Chartered Banker input, transition from training module to storage and retrieval module via on-demand training module.

Advantages of Database used(Neo4j)-

- More Connected data can be retrieved, traversed, and navigated much more quickly and easily.
- It makes semi-structured data representation relatively simple.
- Neo4j CQL query language commands are presented in a readable manner for ease of use.
- It makes use of a strong yet simple data model.
- It does not require complicated joins to retrieve connected/associated information as it's far very smooth to retrieve it is adjoining node or courting information with out Joins or Indexes.

Challenges of Algorithm used(Community Algorithm)-

- The number of communities has to be pre-determined.
- Not applicable to overlapping communities. Does not suit dynamic communities.
- This algorithm requires heavy computing resources.
- While there is no single solution, many solutions are grouped together.

REFERENCES

[1] S. Daliri, "Using Harmony Search Algorithm in Neural Networks to Improve Fraud Detection in Banking System," (in English), *Computational Intelligence and Neuroscience,* Article vol. 2020, p. 5, Feb 2020, Art. no. 6503459.

[2] S. M. Darwish, "An intelligent credit card fraud detection approach based on semantic fusion of two classifiers," (in English), *Soft Computing,* Article vol. 24, no. 2, pp. 1243-1253, Jan 2020

[3] E. A. Minastireanu and G. Mesnita, "Methods of Handling Unbalanced Datasets in Credit Card Fraud Detection," (in English), *Brain-Broad Research in Artificial Intelligence and Neuroscience,* Article vol. 11, no. 1, pp. 131-143, Mar 2020.

[5] O. Ata and L. Hazim, "Comparative Analysis of Different Distributions Dataset by Using Data Mining Techniques on Credit Card Fraud Detection," (in English), *Tehnicki Vjesnik-Technical Gazette,* Article vol. 27, no. 2, pp. 618-626, Apr 2020.

[6] H. Alqahtani et al., "Cyber Intrusion Detection Using Machine Learning Classification Techniques," in Computing Science, Communication and Security, Singapore, 2020, pp. 121-131: Springer Singapore

[7] S. Arora and M. P. S. Bhatia, "Fingerprint Spoofing Detection to Improve Customer Security in Mobile Financial Applications Using Deep Learning," (in English), *Arabian Journal for Science and Engineering,* Article vol. 45, no. 4, pp. 2847-2863, Apr 2020.

[8] O. Ata and L. Hazim, "Comparative Analysis of Different Distributions Dataset by Using Data Mining Techniques on Credit Card Fraud Detection," (in English), *Tehnicki Vjesnik-Technical Gazette,* Article vol. 27, no. 2, pp. 618-626, Apr 2020