IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.1, January 2021 E-ISSN: 2347-2693

Performance of Machine Learning Techniques in the Prevention of Financial Frauds

Saleha Farheen[1*], Monika Raghuwanshi[2]

[1,2]Dept. Computer Science Engineering, Bhopal, India

[*]*Corresponding Author: salehafarheen9@gmail.com, Tel.: +91-88307-31700*

DOI: https://doi.org/10.26438/ijcse/v9i1.2729 | Available online at: www.ijcseonline.org

Received: 14/Jan/2021, Accepted: 20/Jan/2021, Published: 31/Jan/2021

Abstract—Financial fraud presents more and more threat that has serious consequences in the financial sector. As a result, financial institutions are forced to continually improve their fraud detection systems. In recent years, several studies have used machine learning and data mining techniques to provide solutions to this problem. In this paper, we propose a state of art on various fraud techniques, as well as detection and prevention techniques proposed in the literature such as classification, clustering, And regression. The aim of this study is to identify the techniques and methods that give the best results that have been perfected so far. Financial fraud presents more and more threat that has serious consequences in the financial sector. As a result, financial institutions are forced to continually improve their fraud detection systems. In recent years, several studies have used machine learning and data mining techniques to provide solutions to this problem. In this paper, we propose a state of art on various fraud techniques, as well as detection and prevention techniques proposed in the literature such as classification, clustering, and regression. The aim of this study is to identify the techniques and methods that give the best results that have been perfected so far.

Keywords— Financial fraud, clustering, regression, machine learning

I. INTRODUCTION

A special method has to be considered to detect the fraud. Addition to that a expert system is to be incorporated to achieve the targets like finding the frauds. The fraud detecting system is basically on information retrieval from the existing database. The core concept of information retrieval is based on knowledge discovery in database (KDD), data mining, machine learning system and statistical analysis. The application of information retrieval techniques placed a major role in the area of fraud detections . Based on the applications of information retrieval techniques in fraud detection, the fraud detection techniques are divided into two different types namely, statistical techniques and artificial intelligence. Statistical techniques include . Data pre-processing for detection, corroboration, error handling and managing missing values or wrong data. Statistical parameters are considered like quantities, performance measure, average and probability. Different models and probability distribution. User profile constrains. Time series and time dependency ration is calculated. Clustering, classification and association techniques were considered towards result. Different matching algorithms are used to detect anomalies.

II. LITERATURE SURVEY

We reviewed the latest techniques to detect anomaly and trust relaying in IoT environment. Also, we focused on reviewing the methods and algorithms to detect financial fraud from traditional methods to the latest one. Proposed a novel solution in the form of fission computing. The proposed approach relies on the edge-crowd integration for maintenance of trust and preservation of privacy rules in social IoT environment. They performed analysis through numerical simulations by using a safe network and presented a case study on the detection of fake news sources in social IoT environment [4]. Also, a pervasive trust management framework is presented for Pervasive Online Social Networks (POSNs), which is capable of generating high trust value between the users with a lower cost of monitoring [5]. As a solution to identify anomalies in IoT environment, a model was proposed on the basis of cognitive tokens, which provide an Intelligent Sensing Model for Anomalies (ISMA) detection by deliberately inducing faulty data to attract the anomalous users [6]. Van Wyk Hartman suggested automatic network topology detection and fraud detection. If fraud is detected in the distribution network, the system schedules the follow-up and field investigation to investigate and fix the fraud [7]. Also, systems and methods for online fraud detection have been proposed. The front end device generates a first dynamic device identification based on dynamic device characteristics and the back end device generates a second dynamic device identification based on the dynamic device characteristics of the front end device for an authentication [8].

III. METHODOLOGY

Fraud detection is a technique which is used to find the fraud. Fraud detection techniques helps organizations simpler and with efficiency by facultative them to run deep

analyses on their complete information . Fraud detection techniques help in investigation departments departments. The detection scheme gives added value because insight in trends and developments, enabling the organization to react quickly and effectively. This makes tracking down criminals more efficient and more focused .

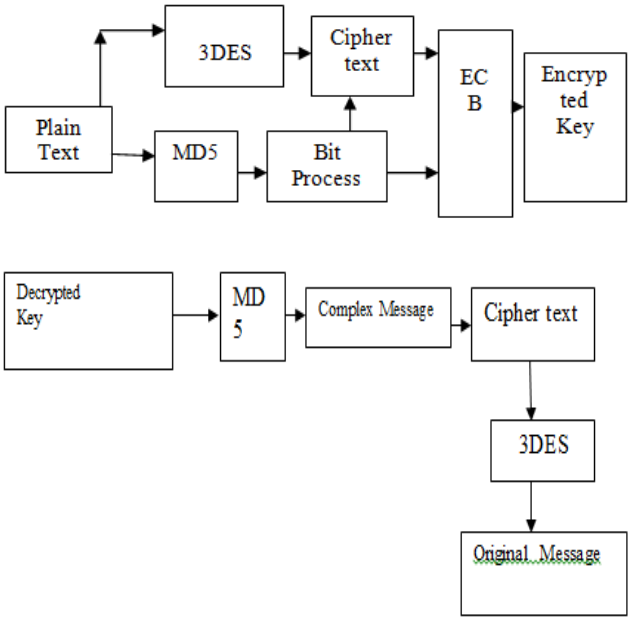

Fig.1. Dataflow diagram

The typical setting involves two parties: one that seeks information from the other that is either motivated, or compelled, to share the requested information. Consequently, in numerous occasions, there is a tension between information sharing and privacy. On the one hand, sensitive data needs to be kept confidential; on the other hand, data owners may be willing, or forced, to share information.

IV. RESULTS AND DISCUSSION

We extensively evaluate the accuracy of our solution with several types of datasets under a multitude of real-world data leak scenarios. This module allows the user to register their identity into the system with proper input parameters. The key generation centers play a vital role in it, which generates public/ secret parameters . The key authorities consist of a central authority and multiple local authorities. Assume that there are secure and reliable communication channels between a central authority and each local authority during the initial key setup and generation phase. Each local authority manages different attributes and issues corresponding attribute keys to users. They grant differential access rights to individual users based on the users' attributes. The key authorities are assumed to be honest but curious. That is, they will honestly execute the assigned tasks in the system however they would like to learn information of encrypted contents as much as possible.

V. PROPOSED SYSTEM

The purpose of this proposed work is to provide the approach functions as a privacy shield to protect parties from disclosing more than the required minimum of their respective sensitive information. PPSSI deployment prompts several challenges, which are addressed in this project. Extensive experimental results attest to the practicality of attained privacy features and show that our approach incurs quite low overhead. For better attack detection, big data incorporates attack graph analytical procedures into the intrusion detection processes. We must note that the design of does not intend to improve any of the existing intrusion detection algorithms; indeed, employs a reconfigurable virtual networking approach to detect and counter the attempts to compromise VMs, thus preventing zombie VMs.

The proposed method has several advantages.
1. To avoid the attacker.
2. Secrecy of the data should be maintained.
3. The scheme is robust to withstand brute force attacks.

Understanding privacy in data mining requires understanding how privacy can be violated and the possible means for preventing privacy violation. In general, one major factor contributes to privacy violation in data mining: the misuse of data. Users' privacy can be violated in different ways and with different intentions. the absence of adequate safeguards, violate informational privacy. Privacy can be violated if personal data are used for other purposes subsequent to the original transaction between an individual and an organization when the information was collected.

One of the sources of privacy violation is called data magnets (Rezgui et al., 2003). Data magnets are techniques and tools used to collect personal data. Examples of data magnets include explicitly collecting information through on-line registration, identifying users through IP addresses, software downloads that require registration, and indirectly collecting information for secondary usage. In many cases, users may or may not be aware that information is being collected or do not know how that information is collected. In particular, collected personal data can be used for secondary usage largely beyond the users' control and privacy laws. This scenario has led to an uncontrollable privacy violation not because of data mining itself, but fundamentally because of the misuse of data.

- ***Individual privacy preservation*:**

The primary goal of data privacy is the protection of personally identifiable information. In general, information is considered personally identifiable if it can be linked, directly or indirectly, to an individual person. Thus, when personal data are subjected to mining, the attribute values associated with individuals are private and must be protected from disclosure. Miners are then able to learn from global models rather than from the characteristics of a particular individual.

- ***Collective privacy preservation*:**

Protecting personal data may not be enough. Sometimes, we may need to protect against learning sensitive knowledge representing the activities of a group. We refer to the protection of sensitive knowledge as collective privacy preservation. The goal here is quite similar to that one for statistical databases, in which security control mechanisms provide aggregate information about groups (population) and, at the same time, prevent disclosure of confidential information about individuals. However, unlike as is the case for statistical databases, another objective of collective privacy preservation is to protect sensitive knowledge that can provide competitive advantage in the business world.

In the case of collective privacy preservation, organizations have to cope with some interesting conflicts. For instance, when personal information undergoes analysis processes that produce new facts about users' shopping patterns, hobbies, or preferences, these facts could be used in recommender systems to predict or affect their future shopping patterns. In general, this scenario is beneficial to both users and organizations. However, when organizations share data in a collaborative project, the goal is not only to protect personally identifiable information but also sensitive knowledge represented by some strategic patterns.

Result

Significant as countries join together to securely share information at the global level. Information sharing in a secure fashion is a daunting challenge, since we must deal with information content that ranges from the simple to the complex (e.g., intelligence reports, financial information, travel records, citizenship records, military positions and logistical data, map data, etc.) in an interoperable environment that is constantly changing. Recently, numerous mandates have emerged to address information sharing.

Table 1. Dataflow

Characteristics	Exiting Scheme	Developed Scheme
Platform	.Net framework	.Net framework
Keys Used	Same Key Is Used For Encryption And Decryption Purpose.	Same Key Is Used For Encryption And Decryption But Additional Authentication Key Is Used.
Scalability	It Is Scalable Algorithm Due To Varing The Key Size.	It Is Scalable Algorithm Due To Varing The Key Size And Used Of Different Keys For Authentication.
Security Applied To	Only From Providers Side.	Both Providers And Client Side.

VI. CONCLUSION AND FUTURE SCOPE

The Previous Technique contents, the low Encryption Method, Single layer Still it required more time for the encryption of data. Since, our technique consists of hybridization of two Method still, it required less time as compare to the previous method. The y axis give the data packet size and the x axis gives time require for encryption. the previous method only protect data from insider attacks but it does not protect the data from outsider attacks so it only has the data security upto 70% but in your method of hybrid we protect the data from insider as well as outsider so your method give 90 % of secured data system.

We further explore the technique of bilinear aggregate signature to extend our main result into a multi-user setting, where can perform multiple auditing tasks simultaneously. Extensive security and performance analysis shows the proposed schemes are provably secure and highly efficient.

REFERENCES

[1] Bolton, R. & Hand, D.. *Statistical Fraud Detection: A Review* (With Discussion). Statistical Science 17(3): 235–255.

[2] G.K. Palshikar, *The Hidden Truth – Frauds and Their Control*: A Critical Application for Business Intelligence, Intelligent Enterprise, **vol. 5, no. 9, 28, pp. 46–51, May 2002.**

[3] Mark. "*Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations*". Hoboken, NJ: John Wiley & Sons Inc. ISBN 978-0-47089046-2.

[4] Pang-Ning Tan, Micheal Steinbach, Vipin Kumar, "*Introduction to Data Mining*", Pearson Education, ISBN: 81-317-1472-1, **2006**.

[5] Heyer, L.J., Kruglyak, S. and Yooseph, S., "*Exploring Expression Data: Identification and Analysis of Coexpressed Genes*", Genome Research 9: **pp:1106-1115, 1999.**

[6] Ayse Yasemin SEYDIM "*Intelligent Agents: A Data Mining Perspective*" Southern Methodist University, Dallas, **1999**

[7] T.Dean, J.Allen, Y.Aloimonos, "*Artificial Intelligence: heory and Practice*", The Benjamin/Cummings Publishing Co. Inc., **1995**.

[8] Eleni Mangina, "*Intelligent Agent-Based Monitoring Platform for Applications in Engineerings*", International Journal of Computer Science & applications **Vol.2, No.1, pp. 38-48, 2005.**

[9] L. Yang, R. Karim, V. Ganapathy, and R. Smith, "*Improving NFA-based signature matching using ordered binary decision diagrams*," in *Proc.13th Int. Symp. Recent Adv. Intrusion Detect.*, **pp. 58–78, Sep. 2010.**

[10]A. Z. Broder, "*Identifying and filtering near-duplicate documents*," in *Proc. 11th Annu. Symp. Combinat. Pattern Matching*, **pp. 1–10, 2000.**

AUTHORS PROFILE

Saleha Farheen is a Graduate student within the Computer Science program at the Anjuman College of Engineering & Technology. She will graduate with an MS in Computer Science. She has a certification in CISCO Network (CCNA). She has also Worked in a Company Universal Enginering, Nagpur Maharashtra.

Monika Raghuwanshi Working as an Assistant Professor in Bhabha Engineering Research Institute Bhopal (April 2018 to till date).she has completed her masters From TIT Bhopal.

