Cyber Fraud: Detection and Analysis of Crypto Ransomware

[1]Tolum Srihari, [2]S Srinivas
[1,2]Dept. of CSE, KIET, Kakinada, AP, India

Abstract
The internet users are increasing day by day causing internet traffic. Fraud detection is one of the main challenges in e-commerce transactions. As transactions are increasing, the quantity of fraud on online transaction also may increase. Huge amount transactions are often done on e-commerce websites. When large amounts of money are moved, there is a high risk that users will engage in fraudulent activities. Fraud prevention in e-commerce shall be developed using machine learning, this work to analyze the suitable machine learning algorithm such as Random Forest and Decision Tree algorithms. Machine learning is really great at detecting fraudulent activity. The goal of this challenge is to create a machine learning model that predicts the probability that a new user's first transaction will be fraudulent. Our proposed system can be applied on Kaggle dataset.

Keywords
Fraud Detection, Machine Learning, E-Commerce, Decision Tree, Random Forest, Fraud Prevention.

I. Introduction

Fraud detection research on e-commerce is only limited to the determination of features or attributes that will be used to determine the nature of the fraud or non-fraud transactions. Machine learning is really great at detecting fraudulent activity. Every website that you provide your credit card information to has a risk team that is responsible for preventing machine learning fraud. The goal of this project is to create a machine learning model that predicts the probability that a new user's first transaction will be fraudulent. With the evolution of big data, data mining and machine learning techniques, it is possible to perform analysis on the historic data and correlate it with seller behaviours to identify potential fraudulent moves.The number of credit card transactions is increasing as a result of technological developments and the rise of e-commerce. Around the world, the ratio of fraudulent transactions to legitimate transactions is around 0,006%. Although this rate may appear low, each fraudulent transaction damages a bank's reputation. As more than just a result, banks are expanding their investments in fraud detection. The number of fraudulent actions and approaches expands and changes every day. Detecting fraudulent activities alone by studying transactions is extremely difficult and costly. To retain client satisfaction and trust, it is vital to detect fraud swiftly and accurately. The frequency of credit card payments is increasing as a result of technological developments and the rise of e-commerce. Around the world, 375 billion card payments were processed in 2017 . In the same year, however, 16.7 million fraudulent transactions occurred . The proportion of fraudulent to valid transactions is roughly 0,006 percent all across the world. Although this rate may appear low, each fraudulent transaction damages a bank's reputation. As a result, banks are increasing their investments in fraud detection. Every day, the number of fraudulent activities and their methods grows and changes. Detecting fraudulent activities alone by studying transactions is extremely difficult and costly. It is critical to detect fraud quickly and accurately in order to maintain client pleasure and trust. In a recent study, machine learning methods were found to be more effective than the majority of rulebased systems. In terms of numbers, characteristics, and changes over time, the data sets utilized in the publications in which these results are published do not always correlate to the real banking environment. In this work, researchers conducted an unprecedented analysis on a real data set to uncover previously unknown aspects of fraud detection efforts. These findings were based on 245,000 fraudulent transactions and four billion non-fraudulent transactions gathered from various banks in 2017.

II. Related Work

Statistical classification methods in consumer credit scoring D. Hand, W. Henley In light of the rationale of logging, the model of the Chinese business bank was carried out in logging. When looking at test results, it is accepted that the example size, test size, and blunder worth can influence the quantity of mistakes. A comparison of neural networks and linear scoring models in the credit union environment V. Desai, J. Crook, G. A. Overstreet The motivation behind this paper is to assess conventional abilities, for example, multi-sensor and measured neural net sensors, segregation-based investigation, and consecutive FICO ratings in a credit agreeable.When the benchmarks clearly reveal bad credit, our findings suggest that adaptive businesses offer a lot of opportunities. Even if all other factors are equal, claiming one is still beyond the typical range. The standard model exhibition was slightly below average as the standard model, particularly on account of helpless credit positioning. Despite the fact that there is a major distinction between the three credit associations, our neural net organizations don't match the distinctions, recommending that a better approach for working might be expected to make an overall outline. Neural nets versus conventional techniques in credit scoring in Egyptian banking Hussein A. Abdou The quantity of non-performing advances has expanded as of late, expanding the worth of the model of credit in case of a downturn. This study provides FICO scores for the Neuro Fuzzy Inference System, which includes three types of information incorporation based on Neuro Fuzzy innovation. A given model's presentation is nearly identical to that of a typical and often used model. Instances of financial assessments are estimated utilizing a 10-time check technique and Visas given by the International Bank for Reconstruction and Development in Turkey. In terms of the appropriate norm of estimate and correlation of spurious attributes, the given model outperforms Discriminant Analysis, Logistic Regression Analysis, and the Artificial Neural Network (ANN). The model offered, like the ANN, can be pondered; but, unlike the ANN, the model does not remain in secret elements. Clarifications of the proposed change might give helpful data to financiers and purchasers, particularly with regards to why the advance application is inescapable. Credit Scoring Methods. Czech Journal of Economics and Finance Martin Vojtek, EvženKoèenda It is broadly used to identify rebelliousness with loaning models. To determine score precision, rules such as Gini esteem, Kolmogorov-Smirnov statistics (KS) insights, Lifting, Mahalanobis distance, and information can be used. This page sums up and shows how to utilize these aspects by and by. A

survey of credit and behavioral scoring: forecasting: financial risk of lending to customers L. Thomas Loaning and scoring practices are abilities that assist associations with choosing whether to loan to clients. This article examines factual exploration and the procedures utilized in research dependent on these choices. It likewise examines the significance of consolidating monetary designs into the scoring framework, looking at irregularities, and how the framework can change until the purchaser thinks about the advantages to the bank. This features the achievement of the not-for-profit area in the locale. Credit Decision-Making And Information Requirements Susan Cancino, Giovanni Cancino-Escalante The world economy has been hit hard by the downturn. The motivation behind the review was to comprehend the acquiring system and bank prerequisites. The review, directed in 22 Brazilian loaning establishments, found that the significance of ongoing utilization of resources, for example, Accounts are scrutinized, bank records are inspected, and credit experience is comparable, and security is the most important concern for a firm because of data inconsistencies.

II. Methodology

Algorithms or Methodology used in this project are Decision trees and Random Forest. A decision tree is in the form of a tree data structure. Decision tree is used for both classification and regression in machine learning. It is a supervised learning method. The output can be generated even for huge training sets as this method learns from the examples. The major disadvantage of using decision trees for huge data sets is it may be a time consuming process.

A. Random Forest

Random Forest is a combination of multiple decision trees. It can be used for regression as well as classification. It has more than one output. Accuracy and variable importance information can be provided with the results. A random forest is the classifier consisting of a collection of tree structured classifiers k, where the k is independently, identically distributed random trees and each random tree consist of the unit of vote for classification of input. The working of random forest is, a random seed is chosen which pulls out at random, a collection of samples from the training datasets while maintaining the class distribution.

B. Decision Tree

Decision tree induction is the learning of decision trees from class-labelled training tuples. A decision tree is a flowchart-like tree structure.

- Decision tree induction is a non-parametric approach for building classification models.
- Finding an optimal decision tree is an NP-complete problem
- Techniques developed for constructing decision trees are computationally inexpensive, making it possible to construct models even when the training set size is very large.
- Decision trees, especially smaller-sized trees, are relatively easy to interpret.
- Decision tree provide an expressive representation for learning discrete- valued functions.
- Decision tree algorithms are quite robust to the presence of noise, especially when methods for avoiding overfitting.

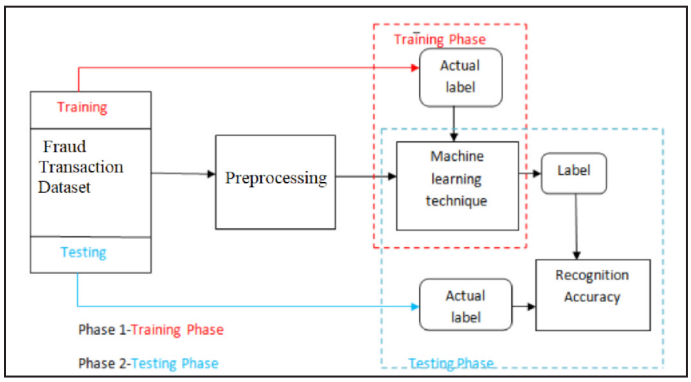

Figure 1: Processing

III. Modeling and Analysis

The fraud detection process mainly undergoes three processes or modules:

A. Dataset Collection:

Dataset is taken from Kaggle website which has features as user_id, signup_time, purchase_time, purchase_value, device_id, source browser, gender, age, ip_address and label as fraud or not. The dataset used here has a total of 151,112 records, the dataset classified as fraud is 14,151 records, the ratio of fraud data is 0.093 percent.

B. Preprocessing:

Preprocessing is used to extract, transform, normalize and scaling new features that will be used in the machine learning algorithm process to be used. Preprocessing is used to convert raw data into quality data. In given data set many unwanted features are used which are device_id, source browser, and user id. These are removed and time is converted to required format.

C. Split Dataset:

In this stage data is collected from dataset and divided to testing and training and given input to algorithm and fit to algorithm.

IV. Results and Discussion

From the given input we will find out whether it is a fraudulent or a genuine transaction.

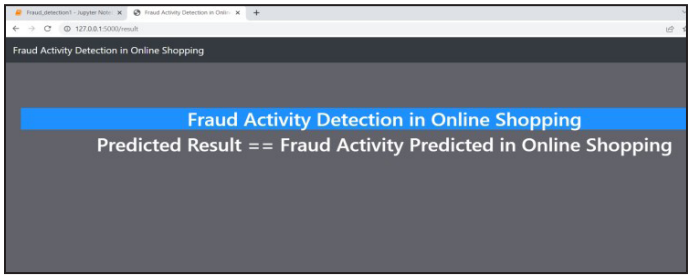

Figure 2: Fraudulent Transaction

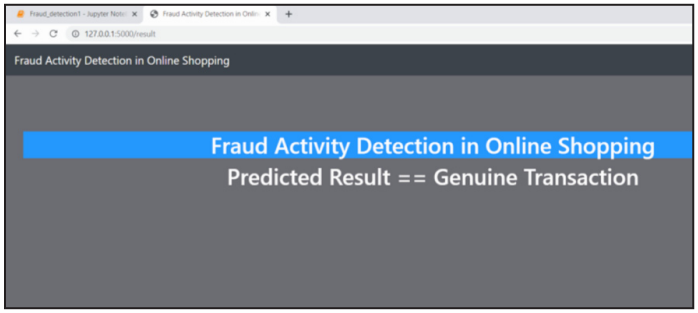

Figure 3: Genuine Transaction

V. Conclusion

For the past few years there is an increased focus on the researches on e-commerce domain, especially on marketplace scenarios mainly attributed to the increasing popularity and year on year growth being exhibited. Reputation is considered as a critical attribute for every online marketplace as consumers has a wide variety of options to choose from. The most important aspect of reputation for an online marketplace is how it protects its customers from fraudulent sellers. As this project finds out the whether the first transaction by a user is fraudulent or not, the results show that on using the Random Forest algorithm we can achieve higher accuracy in fraud detection and depending on the usage different machine learning algorithms can be implemented.

References

[1] AdiSaputra&Suharjito: "Fraud Detection using Machine Learning in e-commerce", International Journal of Advanced Computer Science and Applications, Vol 10, No.9, 2019.

[2] R. Jhangiani, D. Bein and A. Verma, "Machine Learning Pipeline for Fraud Detection and Prevention in E-Commerce Transactions," 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 0135-0140, 2019.

[3] H. Weng et al., "Online E-Commerce Fraud: A Large-Scale Detection and Analysis," 2018 IEEE 34th International Conference on Data Engineering (ICDE), pp. 1435-1440, 2018.

[4] E. Caldeira, G. Brandao and A. C. M. Pereira, "Fraud Analysis and Prevention in e-Commerce Transactions," 2014 9th Latin American Web Congress, pp. 42-49, 2014.

[5] H. Weng et al., "CATS: Cross-Platform E-Commerce Fraud Detection," 2019 IEEE 35th International Conference on Data Engineering (ICDE), pp. 1874-1885, 2019.

[6] J. Shaji and D. Panchal, "Improved fraud detection in e-commerce transactions," 2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA), pp. 121-126, 2017.

[7] I. M. Mary, M. Priyadharsini, K. K and M. S. F, "Online Transaction Fraud Detection System," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 14-16, 2021.

[8] A. K, A. K. Pani, M. M and P. Kumar, "An Approach for Detecting Frauds in E-Commerce Transactions using Machine Learning Techniques," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), pp. 826-831, 2021.