CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING AND DEEP LEARNING

Ankush Ramaswamy[1], Aman DA Mulimani[2], Suparna Pal[3], Abhijeet Kumar Singh[4], Mrs Monika Rani H G[5]

[1,2,3,4]*B.E. Student, Department of CSE, Sir M Visvesvaraya Institute of Technology, VTU, Bengaluru, India*
[5]*Assistant Professor, Department of CSE, Sir M Visvesvaraya Institute of Technology, VTU, Bengaluru, India*

---***---

Abstract - *The raison d'etre of analysis of data, is to expose inherent patterns and stats, to back the decisions that are informed made in various situations. Credit card frauds have been on a steady rise along with the unflinching progress made in the field of technology, showcasing the bitter side of the golden coin of Evolution. Credit card frauds make up a tiny portion of the millions of transactions that happen every day, making it a highly imbalanced set of publicly available data-sets. In the proposed paper, we exercise Supervised Machine Learning algorithms to identify fraudulent transactions by feeding real-world data-sets. Additionally, we engage our algorithms to utilise a classifier using machine learning methods.*

We button-down the most significant variables that could catapult the accuracy of detection of fraudulent credit card transactions. Deep-learning umbrellas a variety of topologies. Moreover, the various attributes used to develop the model (example:the count of neurons within the hidden layer of the neural network) also impacts the outcome.

In the following paper, we shall assess a subset of deep learning topologies from the traditional artificial neural network to topologies sporting inherent time and memory components such as a Long Short-term memory and other parameters concerning their efficacy in fraud exposure on a data-set of close to 8 Crore credit card transactions prelabelled for their credibility. We employ a high-performance distributed cloud computing environment to steer past the prevailing and familiar problems like class imbalance and scalability. Our investigation gives a far-reaching guide to sensitivity analysis of model parameters concerning performance in fraud detection.

We also showcase a framework to tune the parameters of deep-learning topologies for credit card fraud detection to facilitate financial institutions to curtail costs by preventing fraudulent activity.

***Key Words*: Fraud detection, Python3, Random Forest algorithm, Convolutional Neural Network (CNN), Deep Learning, Classification, Supervised learning, Naive-Bayes Classifier, Super asymmetry.**

1.INTRODUCTION

In the swift moving lives of the 21st century, looking back at the feats achieved by Mankind, We have come a long way in how we make our transactions. Long gone are the days of the Barter system. One can see the cash valued barter system also fading away with each new technological advancement. Every new invention or innovation has shown Mankind how inefficient the existing system of doing things has been. But this "EFFICIENCY" comes at a price of surety and security. Transactions today and made, involving two or more computers over a network. Our debit cards and credit cards are mere passkeys to enable these transactions.

These transactions are merely conversations and log-keeping that two computers indulge in (by sight or touch) handing over and reception of funds takes place.

This partial or sometimes complete un-involvement of the customer in this exchange makes an innuendo for fraudulent transactions. We have strived hard to develop this project and make the world a better place to live in!

1.1 Python

Python is a high-level programming language with an easy syntax. Its inbuilt data structures, along with dynamic typing and dynamic binding, can be used in the Rapid Application Development. It supports modules and packages, that improve the program modularity and code reuse. It not only offers readable code, but also is widely used in the fields of machine learning and artificial intelligence. Python's simple syntax permits the developers to write down reliable code whereas advanced algorithms stand behind machine learning and AI.

1.2 Random Forest Algorithm

The proposed model will be making use of decision trees as the basic data structure and a group of these, making a forest will allow us to employ the randomness of the universe into the model. Making our system truly natural. We will be providing each tree a non-similar set of attributes so as to resemble the inequality and beauty in the things around us. Since the Universe as we know it, is governed by super symmetry, we hope our approach to the problem statement using the Random Forest algorithm is the way to go about it. The two main reasons for us to be employing RFA is its capability to tackle overfitting and the ability to be used as both classification and a regression model.

2. LITERATURE SURVEY

Extortion go about as criminal misdirection planned to bring about monetary or individual advantage. This is a cognizant doing, that is illegal, rule, or strategy with a goal to accomplish unapproved monetary advantage. Various sorts of writing relating to peculiarity or extortion location in this space have been distributed as of now and are accessible for public utilization. A complete review directed by Clifton Phua and his partners has uncovered that the most well-known strategies and calculations utilized in this area incorporate information mining applications, robotized extortion identification, antagonistic location, KNN, CNN, SVM and profound neural organization.

Preparing information and testing information split:

When information are pre handled this framework needs to part information into division preparing information and testing information. Generally preparing ought to be huge for exact outcome.

Preparing measure:

In this strategy the preparation informational collection with name will be given to any of the AI procedures like arbitrary timberland, this module will remove the component from the mark information keep it prepared forecast measure.

Preparatory model:

In this strategy the test information without mark needs to give expectation module which produce utilizing preparing technique this forecast module acknowledge the test information and interaction. At long last it will create the precision module.

Information representation:

This strategy utilizes tangle plot lib python instrument for delivering chart from preparing information just as testing informational index.

3. PROPOSED SYSTEM

Our proposed framework recognizes the main factors that may prompt higher precision in Mastercard deceitful exchange discovery utilizing Profound Learning and Arbitrary woods calculation Display and assess the better model to take care of the given issues.

The approach is: implementing Random forest model and deep learning implementing an efficient and accurate platform that can help predict the possibilities of fraud.

Input: Twenty-Three different features that are essential to credit card fraud

Processing Pre-processing Methods: Standardization and Label Encoding

Classifier: Random Forest and CNN Algorithm is used as a classifier.

3.1 Methodology

Prediction:

Collection of data and pre-processing the same:

The data is uploaded into the system through the pandas data frame and the data is pre-processed to correct all the missing values.

Splitting the training and testing data:

Once the data is pre-processed, the system splits the data into training data and testing data in a given ratio which has to be passed as a parameter to the spit function. The training data should be more than the test data by at least 40% in order for the result to be accurate.

Training process:

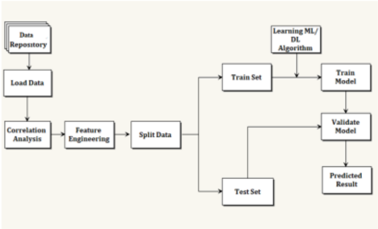
The system learns the data epoch number of times in this process. The labelled training dataset is fed into the supervised machine learning algorithm which is random forest in our case. This module takes all the features in the training data set and makes it ready for the next stage, that is, prediction process.

Prediction model:

The test data which is generated in the split stage, has to be fed into the prediction module excluding the label of the dataset. The prediction module is generated from the training process which is in the previous stage. This prediction module will take the data and the process, and produce the next module, that is, the accuracy module, from which the accuracy of the system is determined.

Data visualization:

This is an important stage in the machine learning process. Using the matplotlib library in python, we can create a graph of the training data and the testing data. Before using this module, we have to install this module in the terminal using the pip install matplotlib command.



4. CONCLUSION

In the present times, where everything is happening online, digital payments are emerging widely, online payments make use of only the credential information in the credit card to fulfil any desire, buy something and then deduction of the money happens. Because of all these reasons, it becomes very necessary to investigate on the number of frauds happening while any transaction happens online.

In this paper, we have scrutinized the accuracy of K-nearest neighbours, Naive-bayes and the deep learning models in the classification of not so balanced fraudulent data of the credit card system comparatively. The reasoning for probing the above mentioned three techniques is that, there is no much comparison done between these three models for the fraudulent credit card data in the past literatures.

REFERENCES

1. “Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy”,Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi, Gianluca Bontempi, IEEE on Neural Networks and Learning Systems,2018.

2. “Credit card Fraud Detection based on the transaction by using Data mining techniques”, B.Pushpalatha, C.Willson Joseph, Vol.5 , Issue 2.