Sinhgad Institutes

Synopsis for project

On

Credit Card Fraud Detection And Representation By Using Data Science And Machine Learning

Submitted by:

Mr. Omesh Anil Satpute

Mr. Vishal Thange

Guide:

Prof. S. D. Bhopale

Academics year 2023-2024

DEPARTMENT OF INFORMATION TECHNOLOGY SINHGAD INSTITUTE OF TECHNOLOGY, LONAVALA

Gat No. 309/310, Kusgaon (Bk)Off Mumbai-Pune Expressway, Lonavala, Tal. Maval, Dist. Pune- 410401

Tel.: 02114304355 Ext. 401

SIT, Dept. of Information Technology

ABSTRACT

The project aims to revolutionize modern education by developing an Android-based Virtual Classroom application, leveraging the powerful capabilities of Firebase. This innovative application will provide a seamless and interactive learning experience, transcending geographical boundaries and time constraints. Through real-time communication, collaborative document sharing, and synchronized multimedia content, our Virtual Classroom will foster engagement and active participation among students and educators alike. Firebase's robust cloud infrastructure will ensure secure data storage, authentication, and efficient data synchronization, while its analytics tools will offer valuable insights into user behavior. By amalgamating cutting-edge technology with the pursuit of knowledge, this project aspires to create an inclusive and accessible platform that transforms traditional education methods and facilitates continuous learning in the digital age.

In the contemporary landscape of education, the integration of technology has become pivotal in enhancing the learning experience. This project embarks on the development of an Android-based Virtual Classroom application, leveraging the versatile and powerful Firebase platform by Google. The overarching goal of this endeavor is to redefine the boundaries of traditional education, making learning accessible, engaging, and interactive regardless of geographical constraints. By harnessing Firebase's capabilities, we aim to create a robust, user-friendly, and feature-rich application that not only facilitates real-time communication but also ensures data security, seamless synchronization, and the efficient management of educational resources. This abstract delves into the project's key objectives, technical aspects, potential impact, and the broader vision of transforming education through technology.

Keyword:- Credit Card Fraud Detection, Fraud Detection, Fraudulent Transactions, Logistic Regression, KNearest Neighbors, Support Vector Machine, NaïveBayes.

INTRODUCTION

Credit Card Fraud can be defined as a case where a person uses someone else's credit card for personal reasons while the owner and the card issuing authorities are unaware of the fact that the card is being used. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behaviour, which consist of fraud, intrusion, and defaulting.

Due to rise and acceleration of E- Commerce, there has been a tremendous use of credit cards for online shopping which led to High amount of frauds related to credit cards. In the era of digitalization, the need to identify credit card frauds is necessary. Fraud detection involves monitoring and analyzing the behavior of various users in order to estimate detect or avoid undesirable behavior. In order to identify credit card fraud detection effectively, we need to understand the various technologies, algorithms and types involved in detecting credit card frauds. Algorithm can differentiate transactions which are fraudulent or not. Find fraud, they need to passed dataset and knowledge of fraudulent transaction. They analyze the dataset and classify all transactions.

These are not the only challenges in the implementation of a real-world fraud detection system, however. In real world examples, the massive stream of payment requests is quickly scanned by automatic tools that determine which transactions to authorize.

Machine learning algorithms are employed to analyses all the authorized transactions and report the suspicious ones. These reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent.

Objectives:-

The main objective of this thesis is to perform predictive analysis on credit card transaction dataset using machine learning techniques and detect the fraudulent transactions from the given dataset. The focus is to identify if a transaction comes under normal class or fraudulent class using predictive models. Different sampling techniques will be implemented to tackle the class imbalance problem and series of machine learning algorithms like logistic regression, random forest and xgboost will be implemented on the dataset, and the results will be reported.

LITERATURE SURVEY

With growing advancement in the electronic commerce field, fraud is spreading all over the world, causing major financial losses. In the current scenario, Major cause of financial losses is credit card fraud; it not only affects tradesperson but also individual clients. Decision tree, Genetic algorithm, Metalearning strategy, neural network, HMM are the presented methods used to detect credit card frauds. In contemplating system for fraudulent detection, artificial intelligence concept of Support Vector Machine (SVM) & decision tree is being used to solve the problem. Thus by the implementation of this hybrid approach, financial losses can be reduced to greater extent.

I have proposed a model based on a decision tree and a combination of Luhn's and Hunt's algorithms. Luhn's algorithm is used to determine whether an incoming transaction is fraudulent or not. It validates credit card numbers via the input, which is the credit card number. Address Mismatch and Degree of Outlierness are used to assess the deviation of each incoming transaction from the cardholder's normal profile. e. In the final step, the general belief is strengthened or weakened using Bayes Theorem, followed by recombination of the calculated probability with the initial belief of fraud using an advanced combination heuristic.

To detect counterfeit transactions, three machine-learning algorithms were presented and implemented. There are many measures used to evaluate the performance of classifiers or predictors, such as the Vector Machine, Random Forest, and Decision Tree. These metrics are either prevalence-dependent or prevalence-independent. Furthermore, these techniques are used in credit card fraud detection mechanisms, and the results of these algorithms have been compared.

supervised algorithms were presented Deep learning, Logistic Regression, Nave Bayesian, Support Vector Machine (SVM), Neural Network, Artificial Immune System, K Nearest Neighbour, Data Mining, Decision Tree, Fuzzy logic based System, and Genetic Algorithm are some of the techniques used. Credit card fraud detection algorithms identify transactions that have a high probability of being fraudulent. We compared machine-learning algorithms to prediction, clustering, and outlier detection.

For training the behavioral characteristics of credit card transactions, the Random Forest classifier was used. The following types are used to train the normal and fraudulent behavior features Random forest-based on random trees and random forest based on CART. To assess the model's effectiveness, performance measures are computed.

For fraud detection, different algorithms like Anomaly Detection Algorithm, K-Nearest Neighbor, Random Forest, K-Means and Decision Tree were used. Based on a given scenario, presented several techniques and predicted the best algorithm to detect deceitful transactions. To predict the fraud result, the system used various rules and algorithms to generate the Fraud score for that certain transaction.

I have proposed a deep network algorithm for fraud detection A deep neural network algorithm for detecting credit card fraud was described in the paper. It has described the neural network algorithm approach as well as deep neural network applications. The preprocessing methods and focal loss; for resolving data skew issues in the dataset.

RELEVANCE

The development of an Android-based Virtual Classroom application using Firebase forlearning is exceptionally relevant in the contemporary educational landscape. Its flexible learning schedules empower students to balance their education with various life commitments. By fostering real-time communication and interactive features, it promotes student engagement,a cornerstone of effective learning. The use of Firebase ensures efficient data management and analytics, allowing for data-driven improvements and personalized learning experiences. Data security measures inspire trust among users, crucial in an era where privacy is paramount. In conclusion, the development of this Android-based Virtual Classroom application not only addresses current educational needs but also positions education on the path of adaptability and technological advancement, enhancing its relevance in our ever-evolving world.

Methodology :-

Density-Based Anomaly Detection : Density-based anomaly detection is based on the k-nearest neighbors' algorithm.

The nearest set of data points are evaluated using a score, which could be Euclidian distance or a similar measure dependent on the type of the data (categorical or numerical). They could be broadly classified into two algorithms:

K-nearest neighbor: k-NN is a simple, non-parametric lazy learning technique used to classify data based on similarities in distance metrics such as Euclidian, Manhattan, Minkowski, or Hamming distance.

Relative density of data: This is better known as local outlier factor (LOF). This concept is based on a distance metric called reachability distance.

Clustering-Based Anomaly Detection: Clustering is one of the most popular concepts in the domain of unsupervised learning.

K-means is a widely used clustering algorithm. It creates 'k' similar clusters of data points. Data instances that fall outside of these groups could potentially be marked as anomalies.

Support Vector Machine-Based Anomaly Detection: A support vector machine is another effective technique for detecting anomalies.

A SVM is typically associated with supervised learning, but there are extensions (OneClassCVM, for instance) that can be used to identify anomalies as an unsupervised problem (in which training data are not labeled).

The algorithm learns a soft boundary in order to cluster the normal data instances using the training set, and then, using the testing instance.

HARDWARE REQUIREMENTS

Hardware requirements :

Processor :	Intel CORE i3
RAM :	8 GB(min)
Hard Disk :	512 GB
Key Board :	Standard Windows Keyboard
Mouse :	Two or Three Button Mouse
Monitor :	LCD/LED

Software requirements :

IDE : Anaconda , VS

Code

Language :	Python
Operating System :	Windows 10
Database type:	NoSQL

SYSTEM ARCHITECTURE

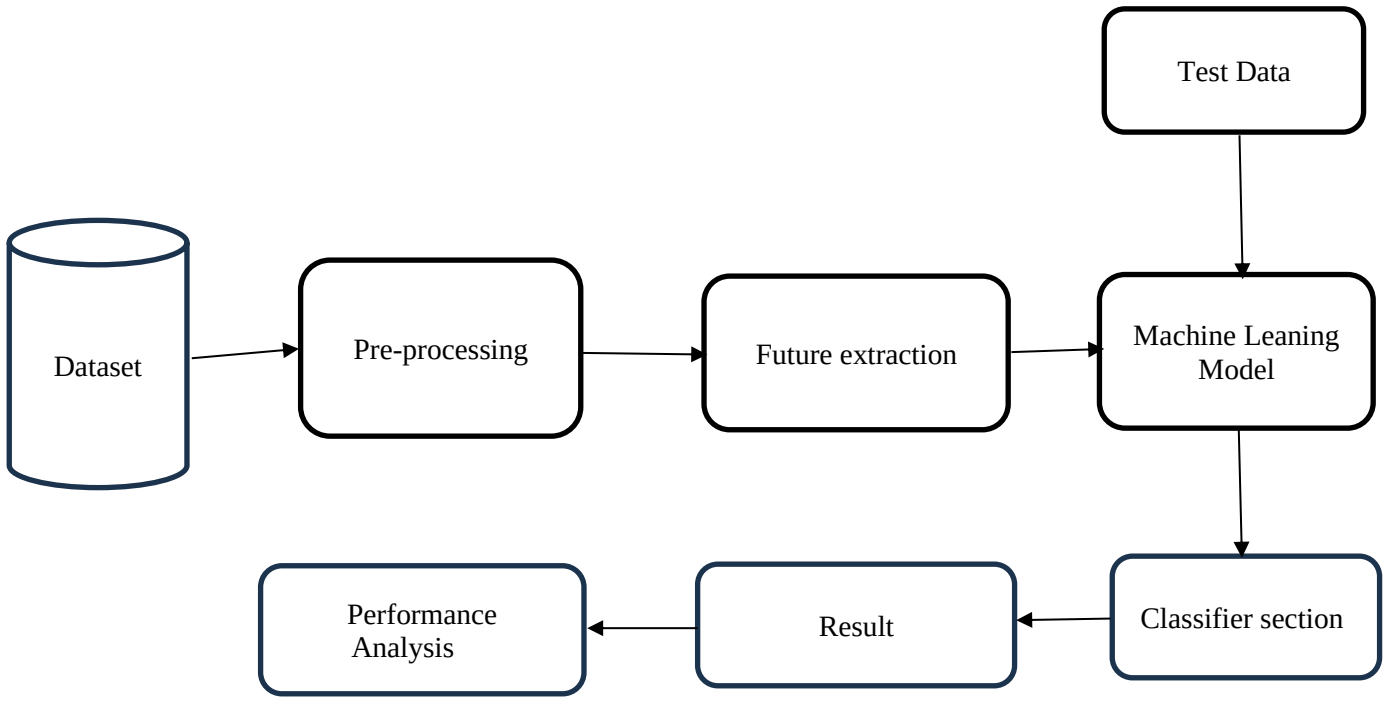

Fig. 1 System Architecture

CONCLUSION

Fraud detection is a complex issue that requires a substantial amount of planning before throwing machine learning algorithms at it. Nonetheless, it is also an application of data science and machine learning for the good, which makes sure that the customer's money is safe and not easily tampered with.

Future work will include a comprehensive tuning of the Random Forest algorithm I talked about earlier. Having a data set with non-anonymized features would make this particularly interesting as outputting the feature importance would enable one to see what specific factors are most important for detecting fraudulent transactions.

REFERENCES

[1] Credit Card Fraud Detection Based on Transaction Behavior -by John Richard D. Kho, Larry A. Vea published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

[2] L.J.P. van der Maaten and G.E. Hinton, Visualizing High-Dimensional Data Using t-SNE (2014), Journal of Machine Learning Research

[3] Machine Learning Group — ULB, Credit Card Fraud Detection (2018), Kaggle

[4] Nathalie Japkowicz, Learning from Imbalanced Data Sets: A Comparison of Various Strategies (2000), AAAI Technical Report WS-00–05

Name of student:	Signature of Student:
1……..	1……..
2. ……	2. ……
3. ……	3. ……
4. ……	4. ……

(Prof. S. D. Bhopale)	……
Name of Guide	Signature of Guide