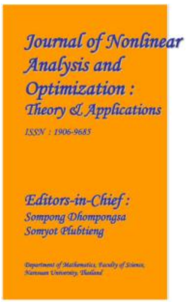
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 2, No.1 : 2024
ISSN : **1906-9685**


Journal of Nonlinear Analysis and Optimization : Theory & Applications
ISSN : 1906-9685
Editors-in-Chief :
Sompong Dhompongsa
Somyot Plubtieng
Department of Mathematics, Faculty of Science, Naresuan University, Thailand

PREVENT FRADULANT TRANSACTION AND DETECT CREDIT CARD FRAUDUSING MACHINE LEARNING ALGORITHM

GOWTHAM S Student, III Year (Digital Cyber Forensic Science) Rathinam College of Arts and Science, Coimbatore-21
Dr. T VELUMANI Assistant Professor Department of Information Technology Rathinam College of Arts and Science, Coimbatore–21

INTRODUCTION

Financial fraud is increases in modern communication world with in seconds. They stolen billion of dollars. This way company and financial institutes are losses there profit, mainly all the bank transactions are now converted in online. In online we have username andpassword .

So they provide credit card for purchasing and transaction purpose. Credit card is morerelevant for day by day for a person. One person behavior we understand followed by credit card usage. By detecting this fraud cases different software are developed but by chance it cannot existing much more years.

So we are going for next stage for detecting fraudulent cases in credit card by machinelearning approach. Machine learning approach is based on algorithm performance, so here weuse much accurate algorithm Random forest .this is the best algorithm for classification. These analysis has taken by choose different attributes of credit card.

MACHINE LEARNING

Machine learning the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on models and inference instead. It is seen as a subset of artificial intelligence.

Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmedto perform the task Machine learning algorithms are used in the applications of email filtering, detection of network intruders, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task.

Python is an open source programming language. Python was made to be easy-to-read and powerful. A Dutch programmer named Guido van Rossum made Python in 1991. He namedit after the television show Monty Python's Flying Circus. Many Python examples and tutorialsinclude jokes from the show. Python is an interpreted language. Interpreted languages do not need to be compiled to run. A program called an interpreter runs Python code on almost any kind of computer. This means that a programmer can change the code and quickly see the results. This also means Python is slower than a compiled language like C, because it is not running machine code directly.

Acknowledgment

This article / project is the outcome of research work carried out in the Department of **Computer Science under the DBT Star College Scheme. The** authors are grateful to the Department of Biotechnology (DBT), Ministry of Science and Technology, Govt. of India, New Delhi, and the Department of **Computer Science** for the support.

BIBLIOGRAPHY:

1. Smith, J. "Enhancing User Engagement in ATM Authentication: A Front-End PuzzleNumber Pad Approach." Proc. Int. Conf. on HCI, 2020.
2. Johnson, E. "Interactive Interfaces for ATM Authentication: Designing Engaging User Experiences." J. User Exp. Design, 5.3, 2019.
3. Patel, R. "Innovative Approaches to ATM Security: Leveraging Gamification in PINEntry." Int. J. of Info. Sec., 12.2, 2021.
4. Brown, S. "Improving User Interaction with ATM Systems: A Case Study of Puzzle- Based Authentication." Proc. ACM CHI, 2018.
5. Lee, D. "Designing Secure and User-Friendly ATM Interfaces: A Human-CenteredApproach." J. of HCI, 8.4, 2020.
6. Nielsen, M. "Usability Engineering for ATM Systems: Principles and Best Practices."Addison-Wesley, 2021.
7. Garcia, M. "Enhancing ATM Security with Biometric Authentication: A ComparativeStudy of User Acceptance." Int. J. of HCI, 15.1, 2019.
8. Thompson, J. "Designing Effective Error Feedback Mechanisms for ATM Interfaces."Proc. ACM CHI, 2017.
9. Kim, S. "Exploring the Role of Gamification in ATM Security Awareness: A User-Centered Design Approach." J. of Cybersec. Educ., 3.2, 2018.
10. Williams, M. "Usability Testing of ATM Interfaces: Methods, Challenges, and BestPractices." J. of Usability Stud., 7.3, 2020.

Version 3 -

Python 3.0 (also called "Python 3000" or "Py3K") was released on December 3, 2008 It was designed to rectify fundamental design flaws in the language—the changes requiredcould not be implemented while retaining full backwards compatibility with the 2.x series, which necessitated a new major version number. The guiding principle of Python 3 was: "reduce feature duplication by removing old ways of doing things".

Python 3.0 was developed with the same philosophy as in prior versions. However, as Python had accumulated new and redundant ways to program the same task, Python 3.0 hadan emphasis on removing duplicative constructs and modules, in keeping with "There should be one— and preferably only one —obvious way to do it".

Nonetheless, Python 3.0 remained a multi-paradigm language. Coders still had options among object-orientation, structured programming, functional programming and otherparadigms, but within such broad choices, the details were intended to be more obvious in Python 3.0 than they were in Python 2.x.

An ordered dictionary type

New unittest features including test skipping, new assertmethods, and test discovery

A much faster io module

Automatic numbering of fields in the str.format() method

Float repr improvements backported from 3.x

DRAWBACKS OF EXISTING SYSTEM

In existing System, a research about a case study involving credit card fraud detection, where data normalization is applied before Cluster Analysis and with results obtained from the use of Cluster Analysis and Artificial Neural Networks on fraud detection has shown thatby clustering attributes neuronal inputs can be minimized. And promising results can be obtained by using normalized data and data should be MLP trained.

This research was based on unsupervised learning. Significance of this paper was find new methods for fraud detection and to increase the accuracy of results. The data set for this paper is based on real life transactional data by a large company and personal details in data is keptconfidential. Accuracy of an algorithm is around 50%. Significance of this paper was to find an algorithm and to reduce the cost measure. The result obtained was by 23% and the algorithm they find was Bayes minimum risk.

Disadvantage:

1. High false positive rates: Existing systems often generate a large number of false positives,which require manual review and can delay legitimate transactions.
2. Limited fraud detection capability: Existing systems may not be able to detect emerging or sophisticated fraud patterns, putting customers at risk of financial loss.
3. Slow detection times: Existing systems may take hours or even days to detect and respond to fraudulent activity, which can result in significant financial losses.
4. Inability to adapt to changing fraud patterns: Fraudsters are constantly evolving their tactics,and existing systems may not be able to keep up with these changes.
5. High costs: Existing systems can be expensive to implement and maintain, particularly forsmaller businesses.

ADVANTAGE OF PROPOSED SYSTEM

In proposed System, we are applying random forest algorithm for classify the creditcard dataset. Random Forest is an algorithm for classification and regression. Summarily, it isa collection of decision tree classifiers. Random forest has advantage over decision tree as it corrects the habit of over fitting to their training set.

A subset of the training set is sampled randomly so that to train each individual tree and then a decision tree is built; each node then splits on a feature selected from a randomsubset of the full feature set. Even for large data sets with many features and data instances training is extremely fast in random forest and because each tree is trained independently of theothers. The Random Forest algorithm has been found to provide a good estimate of the generalization error and to be resistant to over fitting.

Advantage:

- In this paper a new collative comparison measure that reasonably represents the gainsand losses due to fraud detection is proposed.
- A cost sensitive method which is based on Bayes minimum risk is presented using theproposed cost measure.
- Improved accuracy and efficiency: The proposed system could use advanced machinelearning and data analytics techniques to more accurately and efficiently detect fraudulent transactions, reducing the number of false positives and minimizing the need for manual reviews.
- Real-time fraud detection: The proposed system could detect fraudulent activity in real-time, allowing for prompt action to prevent financial losses.
- Adaptability to changing fraud patterns: The proposed system could be designed to adapt to changing fraud patterns and proactively identify potential risks before they become widespread.

NumPy is often used along with packages like SciPy (Scientific Python)and Mat−plotlib (plotting library). This combination is widely used as a replacement for MatLab, a popular platform for technical computing. However,Python alternative to MatLab is now seen as a more modern and complete programming language.

1. PANDAS

Pandas is an open-source, BSD-licensed Python library providing high- performance, easy-to-use data structures and data analysis tools for the Python programming language. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc. In this tutorial, we will learn the various features of Python Pandas and how to use them in practice.

2. MATPLOTLIB

It is a collection of command style functions that make matplotlib work like MATLAB.Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc.

In matplotlib.pyplot various states are preserved across function calls, so that it keeps trackof things like the current figure and plotting area, and the plotting functions are directed to the current axes (please note that &quot;axes&quot; here and in most places in the documentation refers to the axes part of a figure and not the strict mathematical term for more than one axis).

3. SKLEARN

Scikit-learn is a machine learning library for Python. It features several regression, classification and clustering algorithms including SVMs, gradient boosting, k- means, randomforests and DBSCAN. It is designed to work with
Python Numpy and SciPy .

The scikit-learn project kicked off as a Google Summer of Code (also known as GSoC) project by David Cournapeau as scikits.learn. It gets its name from “Scikit”, a separate third- party extension to SciPy.

4. FLASK:

What is Flask?

Flask is an API of Python that allows us to build up web-applications.

It was developed by Armin Ronacher. Flask’s framework is more explicit thanDjango’s framework and is also easier to learn because it has less base code to implement asimple web-Application.

A Web-Application Framework or Web Framework is the collection of modules and libraries that helps the developer to write applications without writing the low-level codes such as protocols, thread management, etc. Flask is based onWSGI(Web Server Gateway Interface) toolkit and Jinja2 template engine.

MODULE DESCRIPTION

- Data collections
- Data Pre-Processing
- Training Data and Test Data
- Model Creation
- Model Prediction
- Algorithm Implementation

➢ **DATA COLLECTIONS**

- Dataset source - Kaggle
- Atmpin,bankcode,ifsc code,pin no...etc(30features)
- Label 1 - Normal
- Label 0 -Fraud

➢ **Data Pre-Processing**

- Data preprocessing which mainly include data cleaning, integration, transformation andreduction, and obtains training sample data needed.
- It is a data mining technique that transforms raw data into an understandable formatSteps in Data Preprocessing

1. Import libraries
2. Read data
3. Checking for missing values

4. Checking for categorical data
5. Standardize the data
6. PCA transformation
7. Data splitting

➢ **Training Data and Test Data**

The training data set in Machine Learning is used to train the model for carryingout abundant actions. Detailed features are fetched from the training set to train the model. These structures are therefore combined into the prototype , if the training set istrained correctly, then the model will be able to acquire something from the comparisonimage. So for testing the model such type of data is used to check whether it is responding correctly or not.

➢ **Model Creation**

- Contextualise machine learning in your organisation.
- Explore the data and choose the type of algorithm.
- Prepare and clean the dataset.
- Split the prepared dataset and perform cross validation.
- Perform machine learning optimisation.
- Deploy the model.

➢ **Model Prediction**

Predictive modeling is a statistical technique using machine learning and data mining to predict and forecast likely future outcomes with the aid of historical and existing data. It works by analyzing current and historical data and projecting what it learns on a modelgenerated to forecast likely outcomes. In this Project, our final prediction is to predict whethera transaction should be be successfully done or any fraud activities are held .In future studies,the success ratio can be increased by strengthening the data set. Lung tomography can be usedin addition to chest radiographs. By developing different deep learning models, success ratio and performance can be increased.

➢ **Algorithm Implementation Random Forest Algorithm:-**

Random Forest is one of the most popular and commonly used algorithms by Data Scientists. Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takestheir majority vote for classification and average in case of regression.

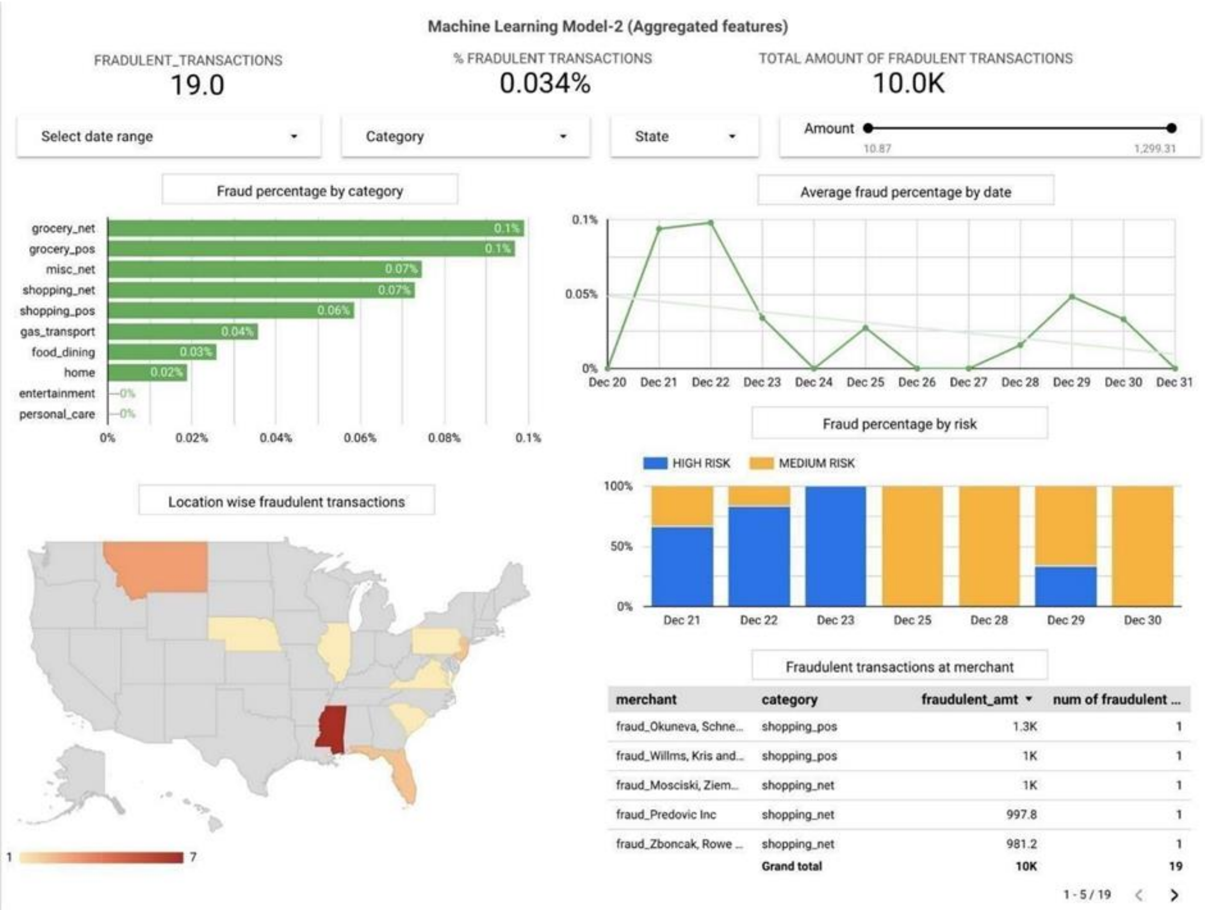
One of the most important features of the Random Forest Algorithm is that it can handlethe data set containing continuous variables, as in the case of regression, and categorical variables, as in the case of classification. It performs better for classification and regression tasks. In this tutorial, we will understand the working of random forest and implement randomforest on a classification task.

Working of Random Forest Algorithm

Before understanding the working of the random forest algorithm in machine learning, we must look into the ensemble learning technique. Ensemble simply means combining multiple models. Thus a collection of models is used to make predictions rather than an individual model.

Ensemble uses two types of methods:

1. Bagging– It creates a different training subset from sample training data withreplacement & the final output is based on majority voting. For example, Random Forest.
2. Boosting– It combines weak learners into strong learners by creating sequential modelssuch that the final model has the highest accuracy. For example, ADA BOOST, XG BOOST.



Machine Learning Model-2 (Aggregated features)

FRADULENT_TRANSACTIONS 19.0

% FRADULENT TRANSACTIONS 0.034%

TOTAL AMOUNT OF FRADULENT TRANSACTIONS 10.0K

Fraudulent transactions at merchant

merchant	category	fraudulent_amt ▾	num of fraudulent ...
fraud_Okuneva, Schne...	shopping_pos	1.3K	1
fraud_Willms, Kris and...	shopping_pos	1K	1
fraud_Mosciski, Ziem...	shopping_net	1K	1
fraud_Predovic Inc	shopping_net	997.8	1
fraud_Zboncak, Rowe ...	shopping_net	981.2	1
	Grand total	10K	19

1 - 5 / 19

OBJECTIVE OF THE PROJECT

The main aim of this project is the detection of credit card fraudulent transactions, as its important to figure out the fraudulent transactions so that customers don’t get charged for the purchase of products that they didn’t buy.

The detection of the credit card fraudulent transactions will be performed with multipleML techniques then a comparison will be made between the outcomes and results of each techniqueto find the best and most suited model in the detection of the credit card transaction that are fraudulent, graphs and numbers will be provided as well. In addition, exploring previous literatures and different techniques used to distinguish the fraud with in the dataset.

Predicting whether it is the cardholders or the fraudsters using the credit cards through credit card profiling. Using outlier detection methods to identify considerably different transactions(or 'outliers') from regular credit cards transactions to detect credit card fraud.

Classifying whether credit card transactions are authentic or fraudulent using algorithmssuch as logistic regression, random forests, support vector machines (SVMs), deep neural networksalong with autoencoders, long short-term memory (LSTM) networks, and convolutional neural networks (CNNs)

Increasing the efficiency and accuracy of the credit card fraud detection process tominimize false positives and reduce the number of manual reviews required. Enhancing the security of credit card transactions and preventing unauthorized access to sensitive data.