CREDIT CARD FRAUD DETECTION

Micro Project for Practical Machine Learning

by

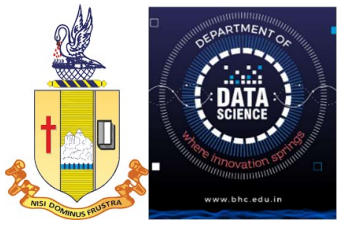
Rahini Devi S

225229129

Submitted To

Dr. K. RAJKUMAR

Course Instructor



DEPARTMENT OF DATA SCIENCE

BISHOP HEBER COLLEGE (AUTONOMOUS)

TIRUCHIRAPPALLI 620017

MARCH 2023

CERTIFICATE

I hereby acknowledge that this project is the original work done by me for the requirements for Micro Project in Practical Machine Learning Course. This micro project is not copied from internet or whatsoever.

Tiruchirappalli
20 March 2023 Your Name and signature

Table of Contents

Chapter 1. ABSTRACT

Credit card fraud is a growing concern in the financial industry. As more and more transactions are conducted online, the risk of fraud increases. To combat this issue, credit card companies and financial institutions have implemented various measures to detect and prevent fraud. One of the most important methods is through the use of fraud detection systems that analyze transaction data and identify suspicious patterns or behavior.

In recent years, machine learning algorithms have been used extensively to detect credit card fraud. These algorithms are capable of analyzing large amounts of data and identifying complex patterns that would be difficult for humans to detect. They can also learn from past data to improve their accuracy over time.

In this paper, we will discuss the different supervised machine learning techniques, approaches to credit card fraud detection. We will also examine the challenges and limitations of these methods and discuss some of the best practices for developing effective fraud detection systems.

Chapter 2. BACKGROUND AND MOTIVATION

Background :

To address this issue, credit card companies and financial institutions have implemented various security measures to detect and prevent fraud. These measures include authentication techniques, such as passwords and security codes, as well as fraud detection systems that analyze transaction data for suspicious patterns or behavior.

Traditional fraud detection systems used rule-based approaches that relied on predefined rules to identify fraudulent transactions. These rules were based on known patterns of fraud and could be updated as new types of fraud emerged. While rule-based systems were effective to a certain extent, they had limitations in detecting complex and evolving fraud patterns.

With the advent of machine learning, supervised and unsupervised (I use only supervised techniques) techniques have been employed to detect credit card fraud. These techniques are capable of analyzing large amounts of transaction data and identifying complex patterns that would be difficult for humans to detect. Machine learning algorithms can also learn from past data to improve their accuracy over time.

Motivation :

The motivation for this paper is to provide a comprehensive overview of machine learning techniques that have been used for Credit Card Fraud Detection. By reviewing the state of the art in this area, we aim to compare the different machine learning algorithms models and highlight which

machine learning algorithms models have more effective and efficient methods for detecting and preventing Credit Card Fraud .

Chapter3. PROBLEM STATEMENT AND DATASET DESCRIPTION

Problem Statement :

The dataset Credit Card Fraud Detection contains a large number of transactions, with a relatively small percentage of them being fraudulent. The objective is to develop a model that can accurately identify these fraudulent transactions, in order to prevent financial losses for both the credit card companies and their customers. The model should be able to take into account a variety of features associated with each transaction, Additionally, the model should be able to adapt and learn from new patterns of fraud as they emerge, in order to stay up-to-date and effective over time. The problem statement involves identifying the appropriate machine learning algorithm for the task, selecting relevant features, and preprocessing the data to ensure that it is suitable for analysis. The models should be evaluated to ensure that they are accurately detecting fraudulent claims while minimizing false positives.

Dataset Description :

- The Credit Card Fraud Detection dataset contains transactions made by credit cards in 2019 to 2020
- Two Dataset of Train to Test Dataset.
- In this dataset transactions have 7506 frauds out of 12,96675 transactions on train dataset.

- This dataset has both numerical and categorical variables.
- 14 categorical and 20 numerical columns.
- It has 1296675 rows × 34 columns in the Train set and 129667 rows × 34 columns in the Test set
- It has Zero Null values
- In the dataset, column named as 'is_fraud' have categorical values of 0 and 1
- I take it as a output feature.

Chapter 4. EXISTING METHODOLOGY

RandomForestClassifier where Used. Accuracy score of 0.99590

Source Code :

https://www.kaggle.com/code/rosicky1234/credit-card-fraud-detection

Chapter5.PROPOSED METHODOLOGY AND SOLUTION

Proposed Methodology

- DecisionTreeClassifier , RandomForestClassifier and Support Vector Machine (SVM) algorithms were Used.
- Find its Accuracy Score ,F1 score,Precision Score ,Classification Report.
- And find Score for X-train with Y-train and Score for X-test with Y-test.

Solution :

- DecisionTreeClassifier has the highest Accuracy Score comparatively.
- And it is faster then Support Vector Machine (SVM)
- Runtime is much higher in Support Vector Machine (SVM)

Chapter 6. MODEL IMPLEMENTATION

Hence I am going to import several classifiers such as,

Libraries :

- import numpy as np
- import pandas as pd

Model :

- from sklearn import svm
- from sklearn.ensemble import RandomForestClassifier
- from sklearn.tree import DecisionTreeClassifier

Standardizer :

- from sklearn.model_selection import train_test_split
- from sklearn.preprocessing import MinMaxScaler

Metrics :

- from sklearn.metrics import accuracy_score, mean_absolute_error ,mean_squared_error,confusion_matrix,median_absolute_error,class ification_report, f1_score,recall_score,precision_score

Plotting :

- from sklearn import tree

Chapter 7. Testing and Evaluation

I have evaluated and compared the results of DecisionTreeClassifier, RandomForestClassifier , and Support Vector Machine (SVM) models. Results have been brought up with differences in accuracy values by using several metrics additionally.

DecisionTreeClassifier

Score the X-train with Y-train is : 0.996322902809108

Score the X-test with Y-test is : 0.9962481038078598

Accuracy score : 0.9962481038078598

F1 score: 99.52 %

Precision: 0.990

RandomForestClassifier

Score the X-train with Y-train is : 0.9948229124491488

Score the X-test with Y-test is : 0.9947131553896843

Accuracy score : 0.9947131553896843

F1 score : 99.21 %

Precision : 0.000

Support Vector Machine (SVM)

Score the X-train with Y-train is : 0.9948229124491488

Score the X-test with Y-test is : 0.9947131553896843

Accuracy score : 0.9947131553896843

F1 score : 99.21%

Precision : 0.000

Chapter 8. MODEL ARCHIVAL IN GITHUB AND DEMO IN YOUTUBE

Github:

https://github.com/rahinidevi/Rahini-DevI-scredit-card-fraud detection

Chapter 9. CONCLUSION AND FUTURE WORK

Conclusion

Credit card fraud is a significant problem faced by the banking and financial industry. Machine learning techniques can be used to build fraud detection models that can help identify fraudulent transactions and prevent financial losses.

One of the most widely used techniques for credit card fraud detection is supervised learning, where a model is trained using historical data that includes both fraudulent and legitimate transactions. The model is then used to identify new transactions that are likely to be fraudulent.

However, it is important to note that fraudsters are constantly evolving their techniques, and fraud detection models need to be updated regularly to keep up with these changes. Continuous monitoring and evaluation of the model's performance is necessary to ensure its effectiveness.

Future Work :

Using more advanced feature engineering techniques to extract more meaningful features from the data. This could involve using unsupervised learning techniques, such as clustering, to identify patterns and anomalies in the data.

Exploring more advanced machine learning algorithms, such as deep learning, to improve the accuracy of fraud detection models. Deep learning models, such as convolutional neural networks and recurrent neural networks, have been shown to be effective in other domains, and may be able to identify more subtle patterns in the credit card transaction data.

Incorporating additional data sources, such as social media and web browsing history, to enhance fraud detection accuracy. By analyzing a wider range of data sources, machine learning models may be able to detect more sophisticated forms of fraud.

Improving the explainability of machine learning models to make it easier for analysts to understand how the models are making predictions. This could involve using techniques such as model visualization and feature importance analysis.

Developing real-time fraud detection systems that can identify fraudulent transactions in near-real-time. This would require designing models that can process large volumes of data quickly and efficiently, and integrating them into existing banking and financial systems.

Chapter 10. REFERANCE

https://www.kaggle.com/code/rosicky1234/credit-card-fraud-detection