

SYNOPSIS

Project Team No: 24SOCU1057

Register No:

1. 125003354

2. 125003153

3. 125003137

Name:

Sudharshanan S

S H Krishaant

Karthikeyan S

Project Title: Optimizing Detection of Credit Card Fraud Using Machine Learning Techniques

Name of the Guide: Ms. Helen W R, Assistant Professor – III CSE/SOC

Abstract:

Fraud involves criminal deception and false representations to gain an unfair advantage, particularly amplified by the growth in online transactions and technologies. The widespread use of online transaction systems and IoT devices has increased transaction volumes, heightening the risk of fraudulent activities. Given the prevalence of fraud, there is an urgent call for effective fraud detection systems. In general, fraud detection can be categorized into two types: misuse detection and anomaly detection. Misuse detection involves the use of machine learning-based classification models to distinguish between fraudulent and legitimate transactions. On the other hand, anomaly detection establishes a baseline from sequential records to define the characteristics of a typical transaction and creates a distinctive profile for it. We proposed a strategy for misuse detection that utilizes a combination of K-nearest neighbor (KNN), linear discriminant analysis (LDA), and linear regression (LR) models. Then we enhance the results with few modifications. The features extracted using this strategy demonstrated recall scores higher values across four tested fraud datasets. As a result, this approach surpasses other methods that rely on single machine learning models, particularly in terms of recall.

Specific Contribution:

- Data preprocessing (Data cleaning, Feature reduction), Model training for dataset 3. Changing model parameters, improving accuracy, comparing models, Report Writing and Presentation and visualization of the datasets and the models .

Specific Learning:

- Acquired Knowledge in various machine learning algorithms, namely K-Nearest Neighbors, Decision Tree, Linear Discriminant Analysis, Linear Regression, Naive Bayes, Quadratic Discriminant Analysis and Random Forest

Technical Limitations & Ethical Challenges faced:

- The dataset exhibited a significant disparity in the distribution of instances across the classes, leading to challenges in model training. This imbalance resulted in the model being biased towards the majority class, consequently yielding low recall for minority classes. This dataset contains 1,048,575 rows and 32 columns so our hardware proved inefficient while training and testing the models

Keywords: Recall scores, Fraud detection systems, K-nearest neighbor (KNN), Linear discriminant analysis (LDA), Linear regression (LR).

Name & Signature of the Student

Date: 15/04/2024

Signature of Guide