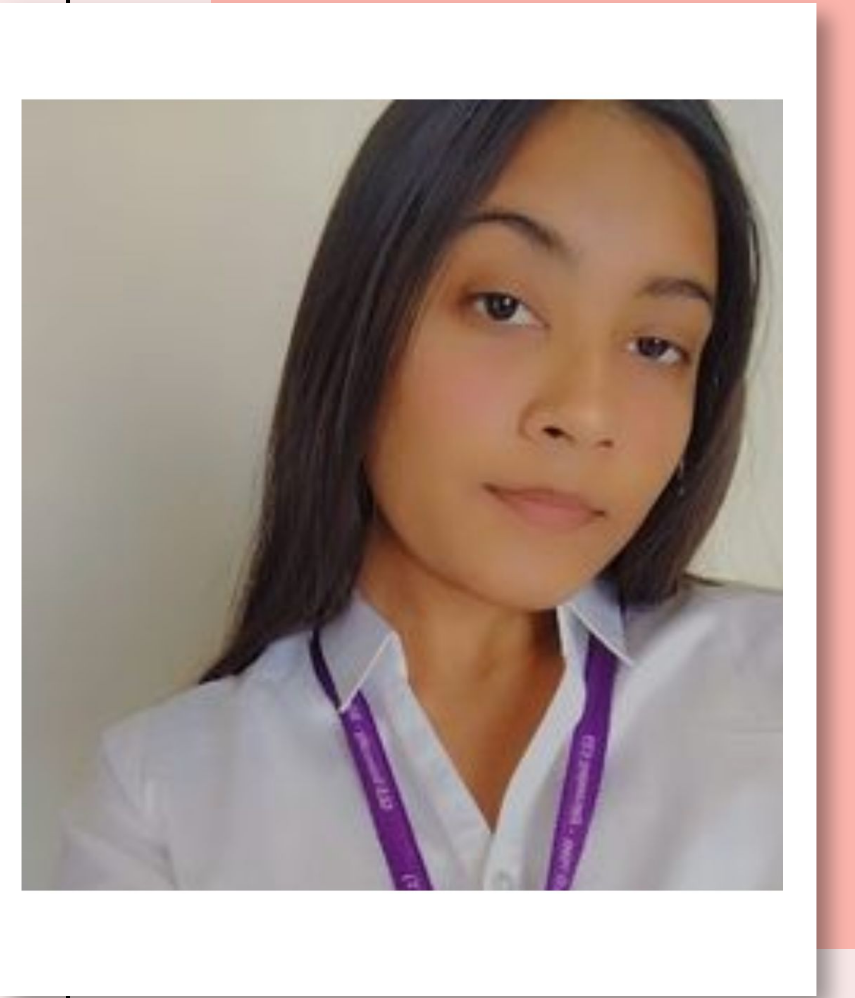
Online Payment Fraud Detection using Machine Learning

Hello! I’m Mauli Patel



A passionate data science enthusiast with a keen interest in uncovering insights from data to drive informed decisions and solve real-world problems. I'm constantly exploring the fascinating intersection of statistics, machine learning, and programming to extract meaningful patterns and trends from complex datasets. With a strong foundation in mathematics and programming languages like Python and R, I'm eager to dive deeper into the world of predictive analytics, data visualization, and artificial intelligence. Let's embark on this exciting data journey together!

Table of contents

01 INTRODUCTION

02 FEATURE ENGINEERING

03 DATA VISUALIZATION

04 MACHINE LEARNING MODEL

05 MODEL EVALUATION

06 Conclusion

01

Introduction

With the rise of online transactions, ensuring secure payment processing is critical for businesses. Online payment fraud poses a significant threat, leading to financial losses and damage to reputation. In this project, we aim to develop a machine learning model to detect fraudulent transactions in real-time, thus enhancing security measures for online payment systems using python

Did you know?

- The online payment method leads to fraud that can happen using any payment app.
- That is why Online Payment Fraud Detection is very important.

01

Feature Engineering

Feature engineering is the process of crafting informative input variables from raw data to enhance machine learning model performance. In our workflow, we transform categorical data into numerical representations using techniques like one-hot encoding. We then split the dataset, design features, and fine-tune models for optimal accuracy through techniques such as cross-validation and hyperparameter tuning. This iterative process ensures our models deliver robust and accurate predictions, driving value in real-world applications.

03

Data Visualization

We will utilize various data visualization techniques to gain insights into the dataset and identify patterns of fraudulent behavior. Potential visualizations include:

Histograms and density plots to visualize the distribution of transaction amounts.

Time series plots to analyze temporal patterns in transaction activity.

Box plots and scatter plots to identify outliers and correlations between features.

04

Machine Learning Model

We will develop a machine learning model to classify transactions as either fraudulent or legitimate. Potential algorithms to explore include:-

- Logistic Regression
- Random Forest
- Gradient Boosting
- Neural Networks

05

Model Evaluation

We will evaluate the performance of our machine learning model using metrics such as accuracy, precision, recall, and F1-score. Additionally, we will analyze the receiver operating characteristic (ROC) curve and area under the curve (AUC) to assess the model's ability to discriminate between fraudulent and legitimate transactions.

06

Conclusion

By implementing a robust machine learning model for online payment fraud detection, we can enhance security measures and mitigate financial risks associated with fraudulent transactions. This project demonstrates the importance of leveraging data-driven approaches to combat online payment fraud effectively.

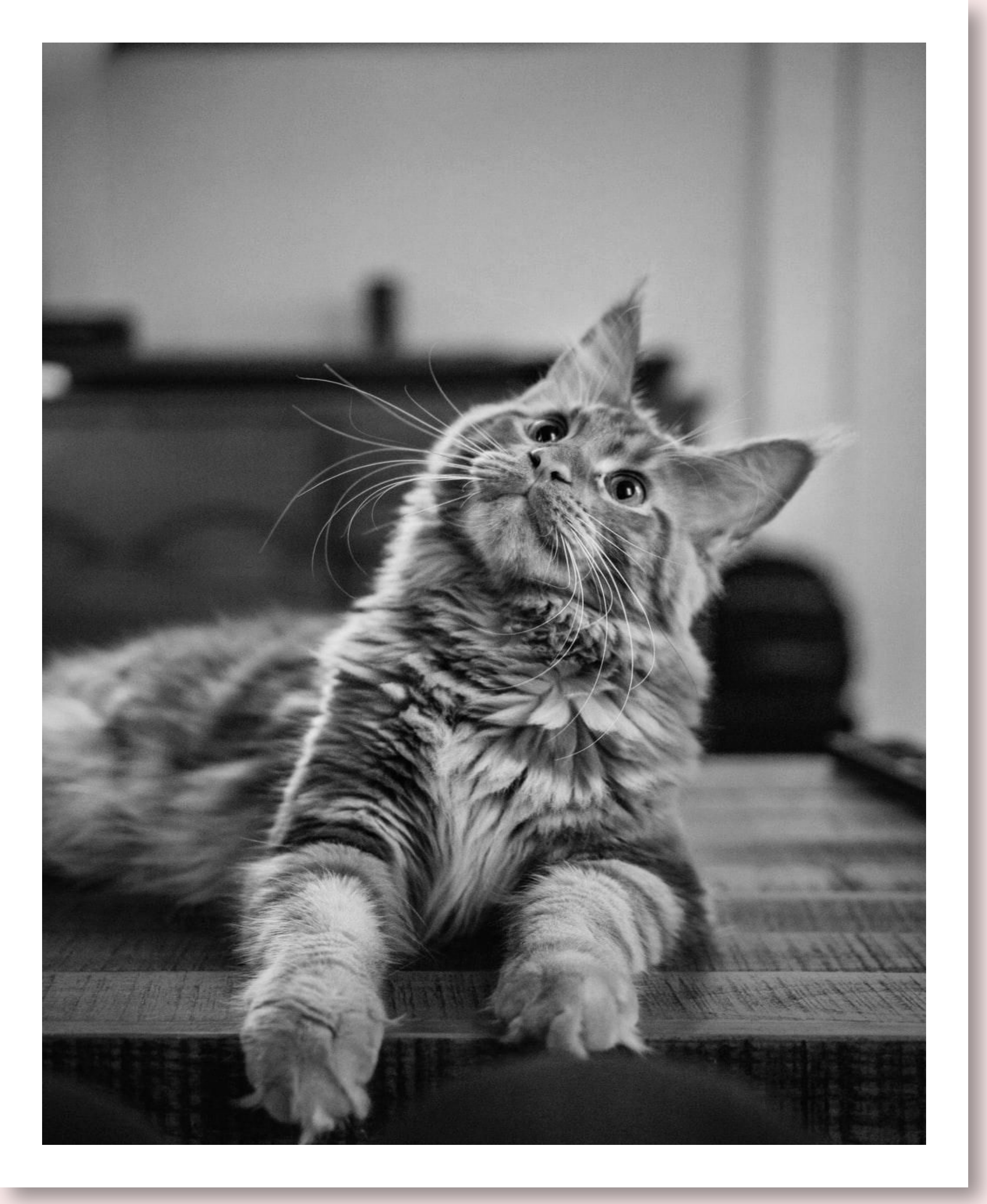
Importing Libraries and Datasets

The libraries used are : -

Pandas: This library helps to load the data frame in a 2D array format and has multiple functions to perform analysis tasks in one go.

Seaborn/Matplotlib: For data visualization.

Numpy: Numpy arrays are very fast and can perform large computations in a very short time.



Description with features

1

Step

This represents the step or unit of time at which the transaction occurred. It could be an arbitrary unit of time like an hour, day, or any other defined period.

2

Amount

Denotes the total amount involved in the transaction. It is essential for analyzing transaction values and identifying unusual or suspicious activities, such as unusually high-value transactions.

3

Type

Refers to the type of transaction performed, such as 'PAYMENT', 'TRANSFER', 'CASH_OUT', 'DEBIT', or 'CASH_IN'. Understanding the type of transaction is crucial for identifying patterns and potential fraud.

4

nameOrig

Represents the name or identifier of the account that initiated the transaction. It provides information about the sender of the transaction.

Description with features

5 OldbalanceOrg

Indicates the balance of the sender's account before the transaction took place. It serves as a reference point to track changes in the sender's balance due to the transaction.

7 NewbalanceOrg

Reflects the balance of the sender's account after the transaction is completed. It helps determine how the transaction affects the sender's account balance.

6 nameDest

Represents the name or identifier of the account that received the transaction. It provides information about the recipient of the transaction.

8 OldbalanceDest

Denotes the balance of the recipient's account before the transaction occurred. It serves as a reference point to track changes in the recipient's balance due to the transaction.

Description with features

9

NewbalanceDest

Reflects the balance of the recipient's account after the transaction is completed. It helps determine how the transaction affects the recipient's account balance.

10

isFraud

This binary variable indicates whether the transaction is fraudulent (1) or legitimate (0). It is the target variable for fraud detection models, and accurately predicting fraud is the main objective of the analysis.

The dataset includes the features like type of payment,
Old balance , amount paid, name of the destination, etc.

Thank you!

Do you have any questions?

maulipatel18112003@gmail.com