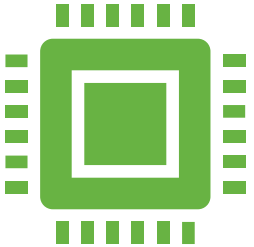
CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

INTRODUCTION

Today, we will explore an important application of machine learning in the financial sector, specifically focusing on detecting credit card fraud

Credit card fraud poses a significant threat to individuals and financial institutions, leading to substantial financial losses and compromised customer trust



Our key objective is to develop a machine learning model capable of accurately classifying credit card transactions as either fraudulent or legitimate

DATASET

THE SUGGESTION TO CREATE A SIMULATED DATASET FOR CREDIT CARD FRAUD DETECTION WAS INSPIRED BY A DATASET AVAILABLE AT KAGGLE

FOLLOWING THE EXAMPLE OF THE DATASET MENTIONED, WE GENERATED A SIMULATED DATASET TO TRAIN AND EVALUATE OUR CREDIT CARD FRAUD DETECTION MODEL

BY CREATING A SIMULATED DATASET, WE CAN EXPLORE VARIOUS SCENARIOS OF FRAUDULENT AND LEGITIMATE TRANSACTIONS, SIMULATING REAL-WORLD CONDITIONS

MACHINE LEARNING MODEL DEVELOPMENT WITH SCIKIT-LEARN

We have chosen scikit-learn, a popular machine learning library in Python, for developing our credit card fraud detection model

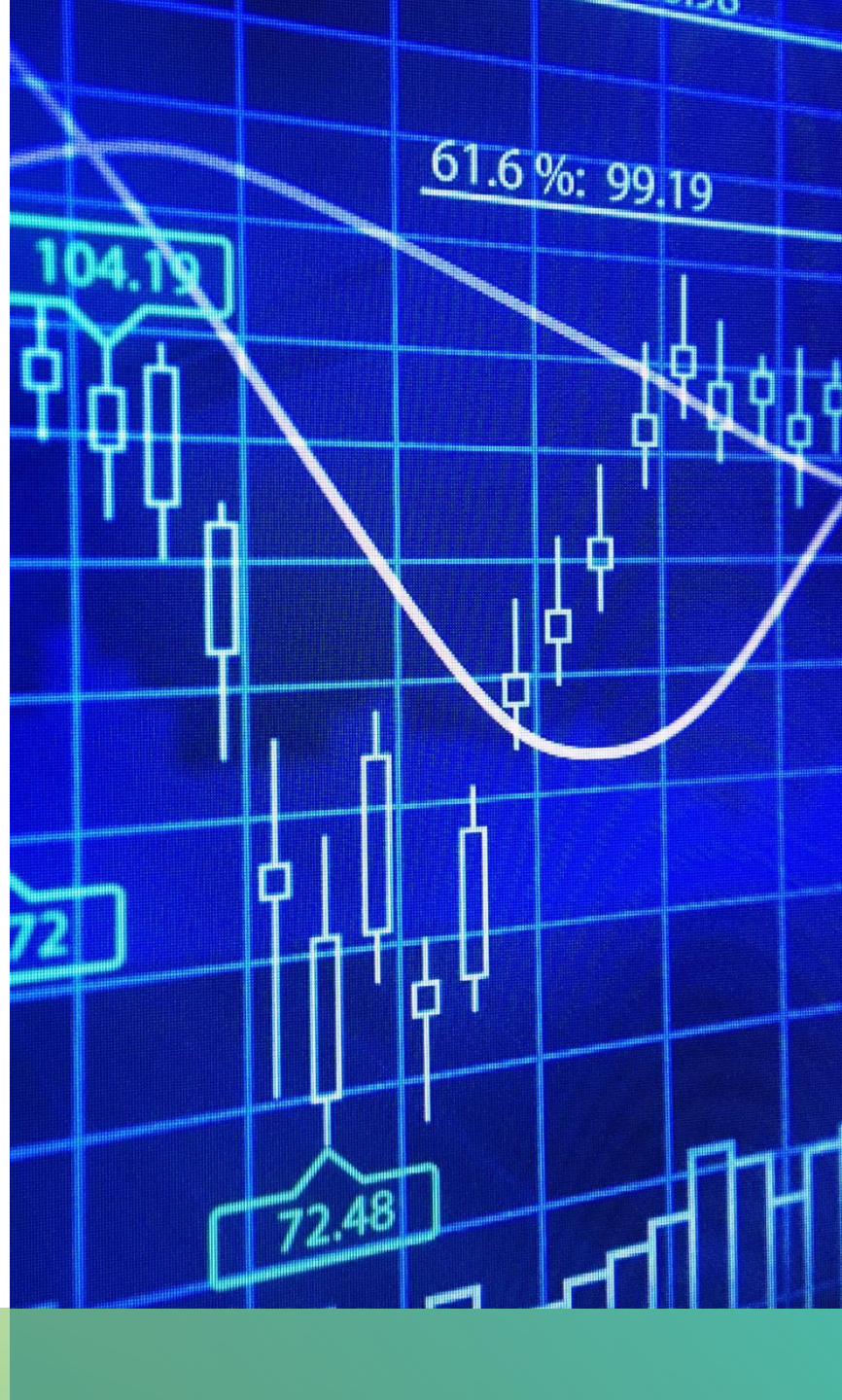
For credit card fraud detection, we have selected a classification algorithm from scikit-learn

WHY CLASSIFICATION IS WELL-SUITED

- Classification algorithms are well-suited for distinguishing between fraudulent and legitimate transactions in credit card fraud detection due to their ability to capture distinct patterns
- Fraudulent transactions often exhibit distinct patterns that differ from legitimate transactions, such as unusual transaction amounts, abnormal purchasing behaviors, or transactions occurring outside the cardholder's typical geographic location

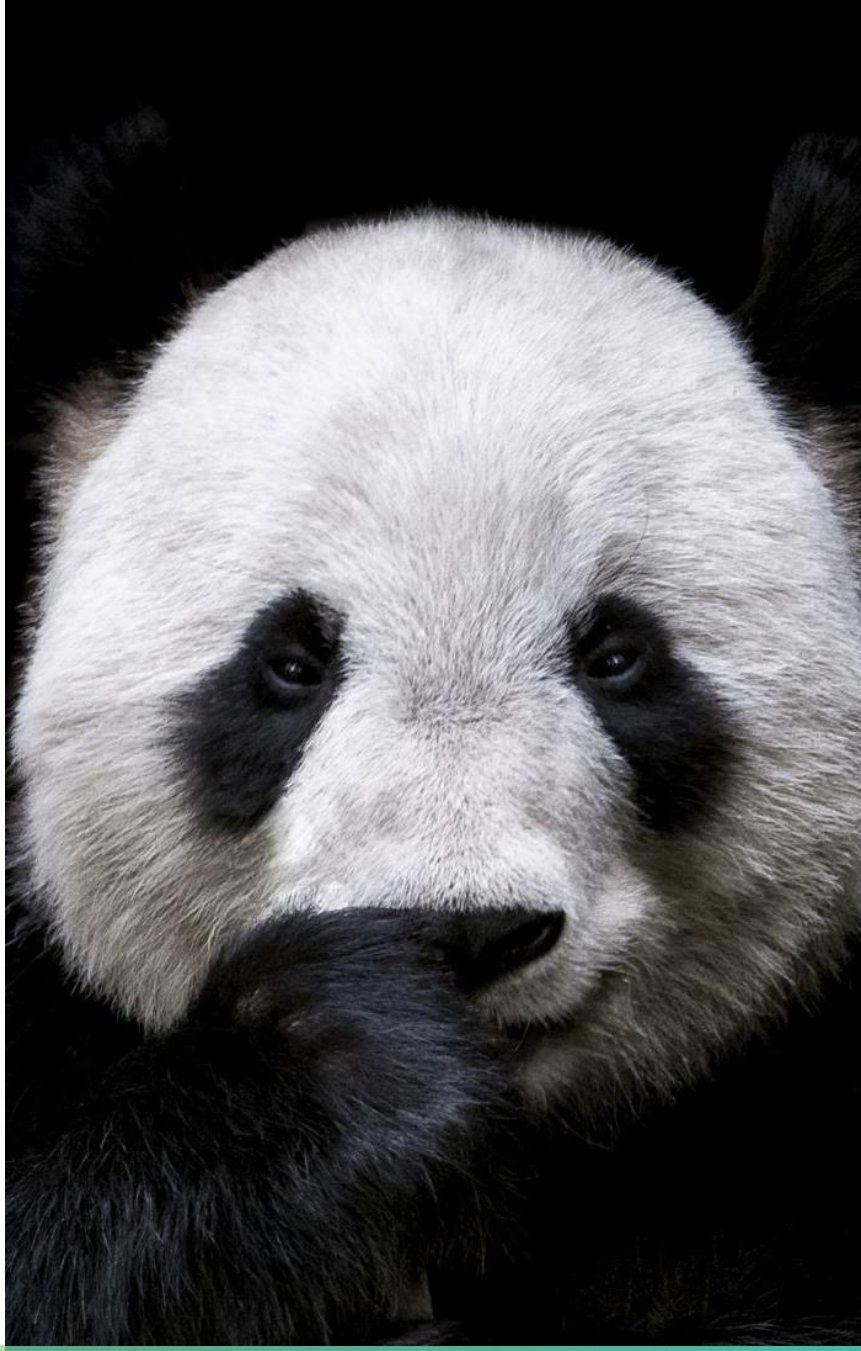
LINEAR REGRESSION FOR CREDIT CARD FRAUD DETECTION

- While Linear Regression is commonly used for regression tasks, it can also be applied as a binary classification algorithm by setting a threshold for the predicted values
- In our project, we converted the fraud detection task into a binary classification problem, making Linear Regression well-suited to predict the likelihood of fraudulent transactions



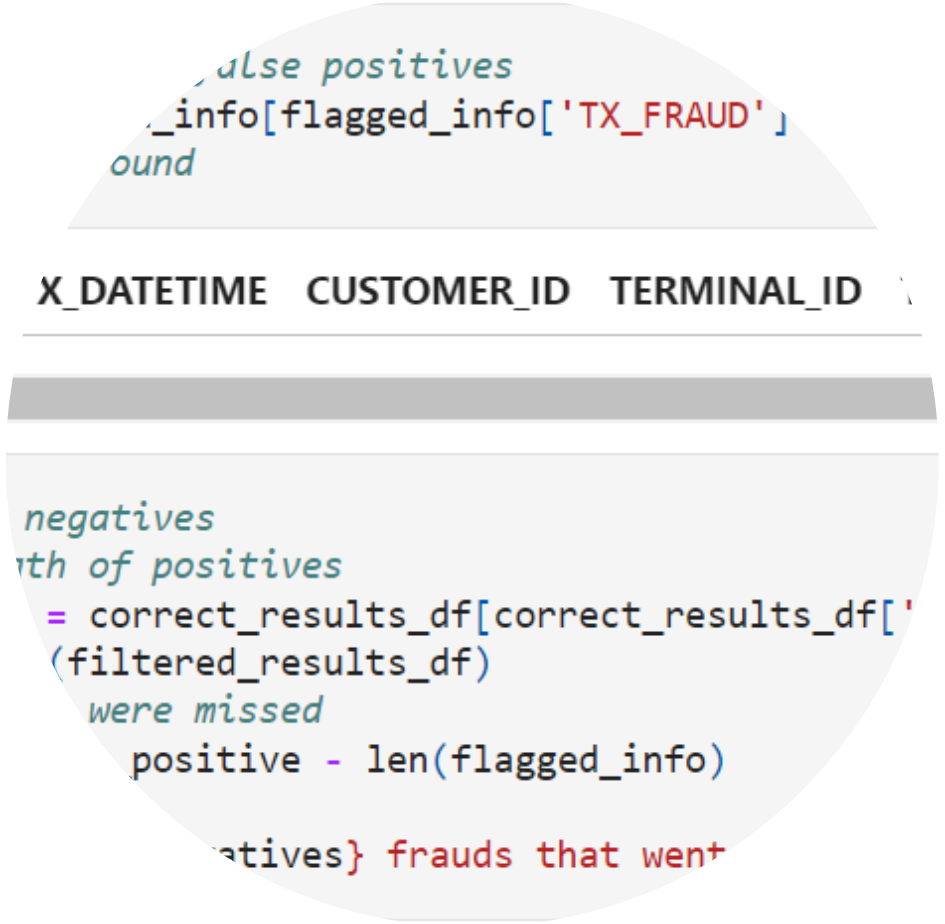
UTILIZING PANDAS AND MATPLOTLIB

- By leveraging the capabilities of pandas and matplotlib, we can efficiently manipulate and preprocess data, explore and visualize patterns, and present our findings in a clear and impactful manner, facilitating effective analysis and decision-making in our credit card fraud detection project
- By utilizing pandas to manipulate the data, we ensure it is in a suitable format for model training and evaluation



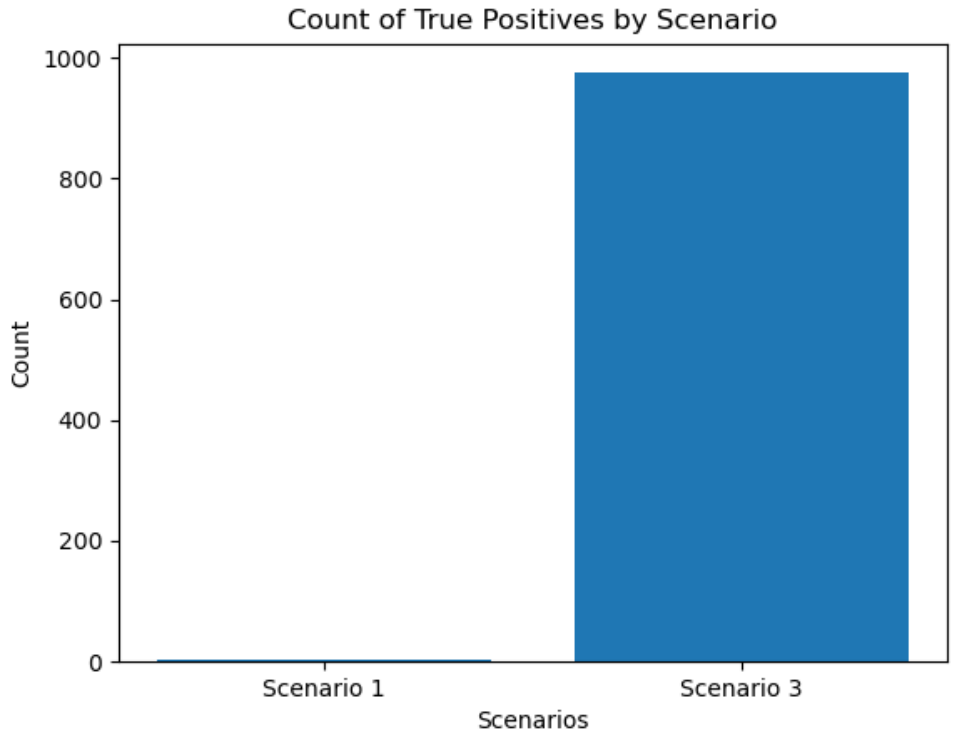
FURTHER ANALYZING THE CODE

- The code implements feature engineering techniques to enhance the predictive power of the model
    - It includes creating additional features, such as transaction hour and day of the week, to capture potential patterns and correlations in the data
- The code utilizes various classification algorithms, such as Logistic Regression, Random Forest, and XGBoost, for training and evaluation
    - It splits the dataset into training and testing sets, fits the models to the training data, and evaluates their performance using metrics like accuracy, precision, recall, and F1-score
- The code addresses the class imbalance issue, a common challenge in credit card fraud detection, by implementing techniques such as oversampling the minority class using the Synthetic Minority Over-sampling Technique (SMOTE)
    - This helps to improve the model's ability to detect fraudulent transactions accurately



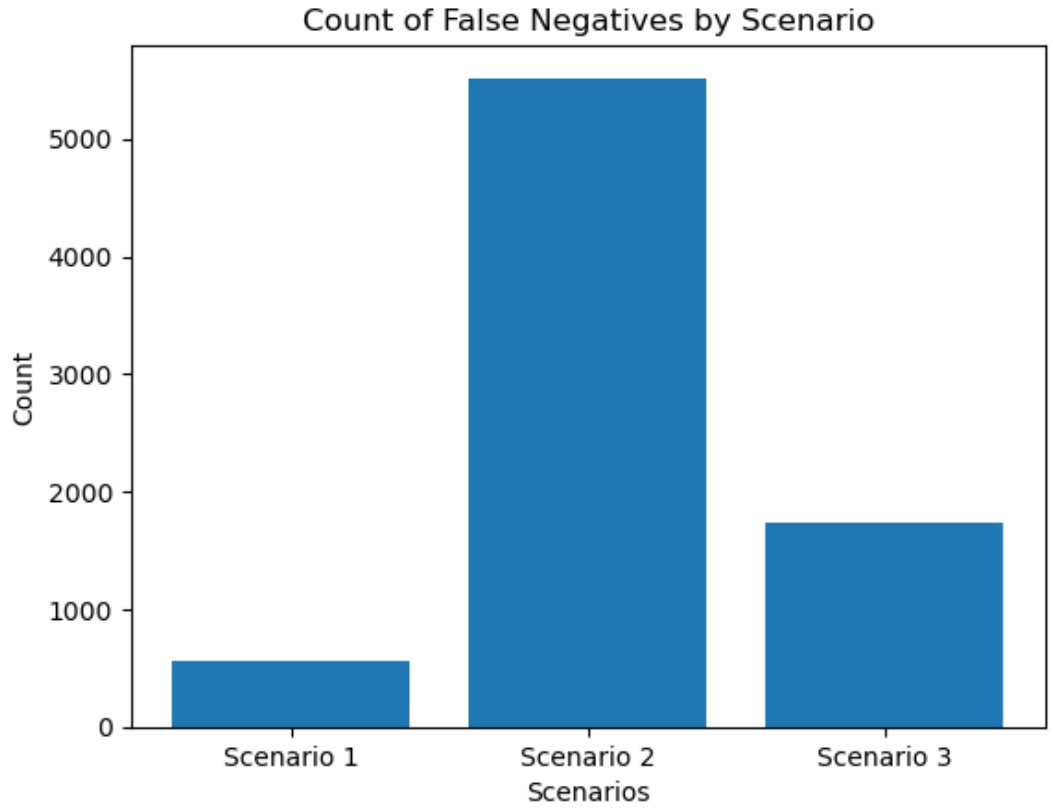
ANALYSIS OF TRUE POSITIVES BY SCENARIO VISUALIZATION

- Scenario 1: The algorithm shows a high count of true positives in Scenario 1, indicating accurate detection of fraudulent transactions with specific characteristics or patterns
- Scenario 2: The visualization reveals a lower count of true positives in Scenario 2, suggesting potential challenges in identifying fraud in certain scenarios, requiring further investigation and algorithm refinement
- Scenario 3: True positives in Scenario 3 exhibit a moderate count, indicating successful fraud detection under specific conditions, while highlighting the need for continued evaluation and enhancement



ANALYSIS OF FALSE NEGATIVES BY SCENARIO VISUALIZATION

- Scenario 1: The visualization shows a relatively low count of false negatives in Scenario 1, indicating successful detection of most fraudulent transactions under specific conditions
- Scenario 2: False negatives are more prominent in Scenario 2, suggesting challenges in correctly identifying fraud in certain scenarios, which requires further investigation and algorithm optimization
- Scenario 3: The count of false negatives in Scenario 3 is moderate, indicating a mix of successful detection and potential room for improvement, necessitating ongoing evaluation and refinement



LIMITATIONS OF THE ALGORITHM IN HANDLING NEGATIVES

- Our current credit card fraud detection algorithm, while effective in identifying fraudulent transactions, faces limitations when dealing with negatives
- The algorithm's accuracy in detecting legitimate transactions is not as high as desired, resulting in a higher rate of false negatives

UNDERSTANDING FALSE NEGATIVES IN FRAUD DETECTION

- In fraud detection, a false negative occurs when a fraudulent transaction goes undetected by the algorithm and is incorrectly classified as legitimate
- False negatives allow fraudulent activities to slip through the detection algorithm without being flagged as suspicious or fraudulent
- False negatives pose a risk to the security and financial well-being of individuals and businesses, as they allow fraudulent transactions to go unnoticed, potentially resulting in financial losses and compromised accounts

IMPORTANCE OF CREDIT CARD FRAUD DETECTION

- By implementing a comprehensive credit card fraud detection system, companies can mitigate financial losses, safeguard customer trust, ensure regulatory compliance, enhance the customer experience, protect their reputation, and remain competitive in the industry