Credit Card Fraud Detection Project

Problem Definition and Design Thinking

Problem Definition

Objective: Develop a machine learning-based system for real-time credit card fraud detection, with a focus on accurately identifying fraudulent transactions while minimizing false positives.

Project Overview: The project involves several key components, including data preprocessing, feature engineering, model selection, model training, and evaluation. The ultimate goal is to create a robust fraud detection system.

Design Thinking

Data Source:

- Utilize a dataset containing transaction data, which should include features such as transaction amount, timestamp, merchant information, and card details.
- Ensure the dataset is representative of real-world credit card transactions, encompassing both legitimate and fraudulent cases.

Data Preprocessing:

- Perform data cleaning to address any inconsistencies or errors in the dataset.
- Handle missing values appropriately, either through imputation or data removal.
- Normalize features to ensure consistency and comparability across different attributes.

Feature Engineering:

- Create additional features that may enhance fraud detection accuracy. Examples include:
- Transaction frequency: Analyze the historical transaction frequency for each cardholder.
- Transaction amount deviations: Identify significant deviations from a cardholder's typical transaction amounts.
- Time-based features: Extract time-related patterns, such as day of the week or time of day when fraud is more likely to occur.

Model Selection:

- Evaluate and choose suitable machine learning algorithms for fraud detection. Options may include:

Logistic Regression

Random Forest

Gradient Boosting

Neural Networks (if applicable and data-rich)

Model Training:

- Utilize the preprocessed data to train the selected machine learning model.
- Consider techniques like cross-validation to optimize hyperparameters and avoid overfitting.
- Implement techniques for handling class imbalance if necessary (e.g., oversampling of minority class).

Evaluation:

- Assess the model's performance using a range of relevant metrics, including:

Accuracy

Precision

Recall

F1-score

* Receiver Operating Characteristic Area Under the Curve (ROC-AUC)

* Continuously monitor and refine the model as new data becomes available to maintain its effectiveness in detecting fraud.

Done by:

Jeripiti Venkata Tarun