Data Mining and Machine Learning Project

Project Proposal

Credit Card Fraud Detection Among European Cardholders

TEMFACK Derick
Ali Asadnia

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

UNIVERSITÀ DI PISA

CROSSLAB
Innovation for industry 4.0

Problem description

- ***Problem:*** In today's digital age, credit card fraud is a major concern for financial institutions and customers. Our project is to experiment multiple machine learning techniques to detect fraudulent transactions.
- **Relevance of DMML techniques:** Fraudulent transactions follow complex patterns that are challenging to model with traditional programming. Therefore, it is essential to use data mining and machine learning techniques to learn from existing patterns and effectively identify unseen fraudulent transactions.
- **Our approach:** We are tackling a classification task. Our approach involves handling missing values and applying feature scaling as part of data preprocessing, followed by performing exploratory data analysis. We will address class imbalance using resampling techniques or class weight adjustments. Additionally, we will experiment with multiple machine learning models, including Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting Machines, to train our models and compare their results.

Dataset description

- **Description:** This dataset contains over 550,000 anonymized credit card transactions made by European cardholders in 2023. Each transaction is represented by 28 anonymized features (V1-V28), a transaction amount (Amount), and a binary label indicating whether the transaction is fraudulent (Class). The anonymized features are likely derived from techniques such as Principal Component Analysis (PCA) to ensure privacy while retaining the essential patterns for fraud detection.

- **Key Features:**
  - **id:** Unique identifier for each transaction
  - **V1-V28:** Anonymized features representing various transaction attributes
  - **Amount:** The transaction amount
  - **Class:** Binary label indicating whether the transaction is fraudulent (1) or not (0)

References

- https://www.kaggle.com/datasets/nelgiriyewithana/credit-card-fraud-detection-dataset-2023?resource=download