Machine Learning Project Proposal

Team Number: 10

Team Members

Name	Section	Bench Number
Ahmed Osama Helmy	1	5
Abdallah Ahmed	1	25
Aliaa Abdelazize Gheis	1	27
Omar Mahmoud	1	29

1. Problem Definition and Motivation

Problem: IEEE-CIS Fraud Detection

Task: Predict the probability of whether an online transaction is fraudulent (`isFraud`) based on transactional and identity-related features.

Motivation

- Fraud is a billion-dollar issue growing annually, with procurement fraud ranking among the top three most disruptive economic crimes globally (PwC Global Economic Crime Survey 2024).
- The dataset, provided by **Vesta Corporation**, includes real-world e-commerce transactions with features spanning device information, payment details, and engineered features.
- Solving this problem can help businesses mitigate financial losses and enhance trust in digital transactions.

2. Evaluation Metric

- **Primary Metric: AUC (Area Under the ROC Curve)** – Measures the model's ability to distinguish between fraudulent and non-fraudulent transactions.

3. Dataset and References

Dataset

- **Source:** IEEE-CIS Fraud Detection on Kaggle
- **Files:**
  - `transaction.csv`: Transactional data (e.g., amount, timestamps, product codes).
  - `identity.csv`: Identity-related features (e.g., device type, device info).
- **Key Features:**

- **Transaction Data:** `TransactionDT` (timestamp), `TransactionAmt`, `ProductCD`, card/address details, and Vesta-engineered features (`Vxxx`).
- **Identity Data:** `DeviceType`, `DeviceInfo`, and anonymized features (`id_01`–`id_38`).

References

- Kaggle Competition: IEEE-CIS Fraud Detection
- Vesta's Real-World Data: Vesta Corporation
