Cairo University-Faculty of Engineering
Computer Department
Machine Learning

Project Proposal

Team 6

Full Name in ARABIC	**SEC**	**BN**
اسامة صالح فرج السيد	1	12
عبدالرحمن محمد عبدالفتاح محمود	1	23
عبدالرحمن محمد حفني	1	22
عمرو صلاح الدين فؤاد	1	32

Submitted to:
Eng. Mohamed Shawky

2025

Idea 1 - Customer Transaction Prediction

Problem Definition

Develop a machine learning model to predict whether a customer will make a specific transaction in the future irrespective of the amount of money transacted, based on anonymized numerical features.

Problem Motivation

Accurately predicting customer transactions enables financial institutions to enhance customer engagement strategies, personalize marketing efforts, and optimize product offerings, resulting in improving customer satisfaction and increasing revenue.

Evaluation Metrics

- **Accuracy**: Measures the overall correctness of the model's predictions.
- **Precision & Recall**
- **F1-score**: Balances precision and recall, especially if the dataset is imbalanced.
- **Area Under the ROC** (AUC-ROC): Also useful for imbalanced dataset.

Dataset Link and information

https://www.kaggle.com/c/santander-customer-transaction-prediction/data?select=train.csv

Dataset has 200K rows and 200+ anonymized features, and imbalanced classes with '0' class spanning 90% of the dataset.

Idea 2 - Large-scale Energy Anomaly Detection

Problem Definition

Create a machine learning model to identify anomalies in hourly smart electricity meter readings, distinguishing between normal consumption patterns and unusual energy usage.

Problem Motivation

Detecting energy consumption anomalies is crucial for energy providers to prevent fraud, identify faulty meters, and promote energy efficiency. Early detection can lead to significant cost savings and more reliable energy distribution.

Evaluation Metrics

- **Accuracy**
- **Precision & Recall**
- **F1-score**
- **Area Under the ROC** (AUC-ROC)

Dataset Link and information

https://www.kaggle.com/competitions/energy-anomaly-detection/data?select=train_features.csv

Dataset has 1.75M rows and 57 features, missing values, and imbalanced classes with ‘0’ class spanning 97% of the dataset.

Idea 3 - Credit Card Fraud Detection

Problem Definition

Develop a machine learning model to detect fraudulent credit card transactions using anonymized transaction data, aiming to differentiate between legitimate and fraudulent activities.

Problem Motivation

Credit card fraud results in significant financial losses annually. An effective fraud detection system helps financial institutions minimize losses, protect customers, and maintain trust. Given the dynamic nature of fraudulent activities, accurate detection methods are essential.

Evaluation Metrics

- **Precision & Recall**
- **F1-score**
- **Area Under the ROC** (AUC-ROC)

Dataset Link and information

https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud/data

Dataset has 285K rows and 31 features, and imbalanced classes with ‘0’ class spanning 99% of the dataset