Fraudit - SOI Jr AI Challenge 2025

Detecting Bank Account Fraud with Machine Learning

Organized by: HardlyHuman. AI-Club, IIT Dharwad

May 30, 2025

Overview

Bank account fraud is one of the most common threats in digital finance. In this project, your task is to develop a machine learning model that can classify whether a bank transaction is **fraudulent or not**.

You will be provided with a pre-cleaned dataset containing transaction records. The challenge is focused on learning to apply simple ML algorithms, visualize your insights, and communicate your results effectively.

Team Guidelines

- Each team must consist of **1 to 3 members**.
- You may work solo or collaborate with others.
- Mentors will be available throughout the project to assist with questions.

Dataset Details

- The dataset contains numerical and categorical features along with a binary label (1 = fraud, 0 = not fraud).
- It will be shared via a Google Drive link.

Task Requirements

Each team must:

1. **Explore** the dataset: Check feature distributions and basic patterns.
2. **Preprocess** the data: Handle missing values, encode categories, scale features if needed.

3. **Train at least two models** from this list (models are not limited to this list):
    - Logistic Regression
    - Decision Tree
    - Random Forest
    - K-Nearest Neighbors (KNN)
4. **Evaluate** your models using:
    - Accuracy, Precision, and Score
    - Confusion Matrix
5. **Visualize results**: Use plots such as bar charts, confusion matrices, or correlation heatmaps.
6. **Submit in a ZIP file**:
    - A brief report (PDF)
    - A short slide deck (max 5 slides)
    - Python code (.py or .ipynb)

Bonus Points

You can earn additional points by:

- Explaining which features were most important and why.

Evaluation Criteria

Submissions will be judged on:

- **Data Understanding** – Clarity in interpreting the dataset.
- **Model Implementation** – Accuracy and correctness of ML models.
- **Result Analysis** – Use of proper metrics and explanations.
- **Code Quality** – Organization, readability, and comments.
- **Presentation** – Simplicity, clarity, and completeness of report and slides.

Submission Instructions

- Submit your files through the official Google Form link (will be shared).
- Include:
  - `report.pdf`
  - `slides.pdf` or `slides.pptx`
  - `notebook.ipynb` or `code.py`
- Deadline: **July 15, 2025**, 11:59 PM
- File naming format: `TeamXX_FraudDetection.zip`

Resources (Optional)

- Scikit-learn Beginner Guide
- Confusion Matrix Explained – StatQuest
- Pandas Cheatsheet