FRAUDULENT CLAIM DETECTOR FOR SAFEDRIVE INSURANCE

Training Bootcamp: AI and GenAI with Python

Project Overview

- Objective: Detect fraudulent insurance claims using machine learning.
- Business Context: Reduce financial losses from fraud.
- Approach: Build a binary classification model using pandas and scikit-learn.

Data Generation & Preprocessing

- Generated synthetic dataset with realistic insurance fields.
- Categorical features encoded using one-hot encoding.
- Missing values filled and features scaled.
- Applied SMOTE to handle class imbalance.

Model Training & Evaluation

- Model: Random Forest Classifier
- Metrics: Precision, Recall, F1 Score, ROC-AUC
- Used Stratified Train-Test Split and Cross Validation
- Feature importance visualized using bar chart

Fraud Flagging & Export

- Predicted fraud probability for each test record.
- Flagged claims with probability > 0.4
- Top 10 suspicious claims printed and analyzed
- Exported results to CSV: 'flagged_claims_improved.csv'

Conclusion & Next Steps

- ✔ Successfully built fraud detection pipeline
- ✔ Evaluated using real-world metrics
- 🔍 Future Work: Model explainability (SHAP, LIME)
- 💾 Save model, deploy to API or dashboard