Fraud Risk Modelling (Credit Card) **Executive Summary**

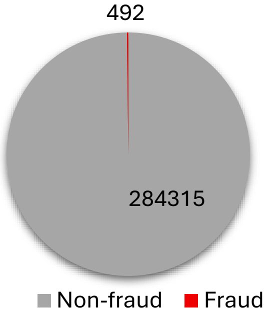
Author: Mohanad Alemam Date: 16 December 2025

Objective: To build production-ready Fraud Risk solution using Machine Learning (ML) best practises. Aiming for early detection of fraudulent activities to assist financial institutions mitigate Fraud Risk.

Exploratory Data Analysis (EDA)

Dataset is European Credit Card Transactions (covering two days provided by Kaggle). The dataset is severely imbalanced with fraud = **0.173%** of transactions (492 fraud vs 284,315 non-fraud) as illustrated in Figure 1 below. The total number of features is 30 columns V1 to V28, Time and Amount.

Figure 1: Class Imbalance



EDA highlighted that time feature has a multimodal distribution (peaks of activity), several features e.g. V13, V15, V18 and V19 have near-normal distribution. Some features have low variance (concentrated) while others show strong class separation. No missing values in the raw dataset. The EDA concluded that the problem is highly imbalanced and nonlinear. Indicating the need for targeted feature engineering and tree-based ensembles models to tackle the highlighted issues.

Feature Engineering (Targeted Approach)

Processes and decisions at this stage were guided by the results of the EDA. The feature engineering conducted is *targeted* (not generic). Transformations were decided based on each feature's statistical behaviour. For features with strong class separation, I applied non-monotonic transforms e.g. absolute and squaring to amplify the separation ability. Low variance features were scaled (MinMax). Time feature was converted into two features 'hour_of_day' and 'time_segment' to capture time/daily cycles.

The engineered features are found to be very impactful as *40% of the final production modes' most important predictors are engineered*, showing that targeted engineering improved patterns' detection.

Modelling And Evaluation (workflow summary)

The models selected and trained for this project are: Logistic Regression as a baseline, Random Forest, CatBoost, and LightGBM as advanced models. Models' tuning stage included hyperparameter search for 50 different hyperparameter combinations for each advanced model. Followed by Out of Fold (OOF) Evaluation of all models focused on the positive/fraud class performance metrics including PR AUC. Based on OOF metrics the best candidate is **LightGBM (tuned)** as it achieved the highest OOF PR AUC (0.863) and strong positive class F1 score of 0.883.

Calibration And Thresholding

Following OOF evaluation model probabilities' reliability calibration was conducted using Brier score achieving ~ **0.00037,** which is very low. This means the model produces realistic and reliable probabilities that can be interpreted as risk of fraud i.e. *actual likelihood of fraud*. After that three probability thresholds were selected and validated/adjusted to work as Fraud Risk cut-offs for production API. The aim was to balance interpretability and coverage for these thresholds based on the OOF probabilities *(High Risk ≥ 0.80; Medium Risk 0.30–0.80; Low Risk < 0.30*). Because the model's probabilities are well calibrated and its precision is strong, the model produces few mid-range scores, meaning the model produces a binary probability distribution.

Test (Holdout) Evaluation (Positive Class Focus)

All models were evaluated on unseen test data and ranked by test PR AUC. A small drop in performance metrics was observed compared to OOF metrics this was expected due to generalisation in unseen data *e.g. LightGBM OOF PR AUC = 0.863, test PR AUC = 0.857.*

Table 1 Final Production Model Test (Holdout) Metrics:

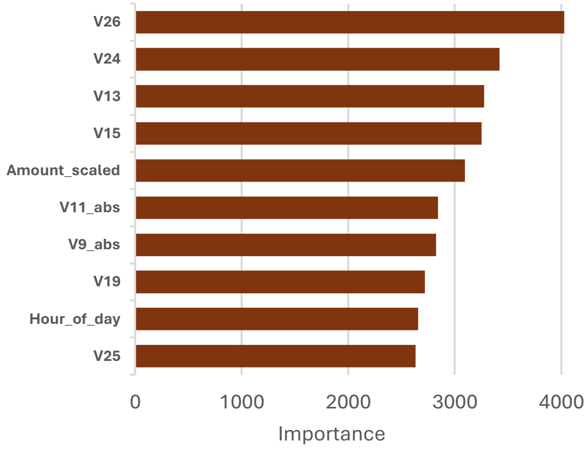
Class	Precision	Recall	F1-score	Support	Balanced Accuracy	PR AUC
Fraud (Class 1)	0.95	0.776	0.854	98	0.888	0.857
Non-fraud (Class 0)	1	1	1	56864	–	–
Macro Avg	0.975	0.888	0.927	56962	0.888	0.857

The **Final positive class test metrics** for the production model are **PR AUC 85.7%, Precision = 95%, Recall = 77.6%**. A normalized confusion matrix was computed and displayed with the following metric for the positive class/fraud class: 77.6% true positive, 22.4% false negative, it was further compared to the 2nd best performing model confusion model to demonstrate the former's superiority. For non-fraud class metrics show near-perfect true negative rate. The ranking of models based on PR AUC is consistent between OOF evaluation and Test data evaluation confirming the correctness of production model selection.

Feature Importance and Explainability

The production model's (LightGBM) ten most important predictors/features were computed based on gain i.e. contribution is loss reduction during training. The top 10 strongest predictors contain four engineered features 40%, as shown in Figure 2 including scaled amounts and hour_of_day engineered features. This demonstrate that the targeted engineering increased separability and interpretability.

Figure 2: Most Important Predictors



Client-facing API and Smoke Test

A user API i.e. fraud_risk_assessment() was created and tested. It was designed to return a client-ready table/DataFrame with fraud probability, risk level, risk indicator and an action code and text for each new transaction. While risk levels thresholds have been calibrated from a technical perspective *(High ≥ 0.80; Medium 0.30–0.80; Low < 0.30*) the API allows the client/ business to change these thresholds based on business knowledge and risk appetite. A smoke test was conducted sampling test rows across Risk tiers High/Medium/Low to assess output across different levels, an example of the output is shown in Table 2 below. The test verified schema, coverage and calculated fraud rates match expectations. The API is simple, interpretable and operation ready.

Table 2 Example of Fraud Risk Assessment API output (Fraud probability, risk level, indicator and recommended action)

Transaction ID	Fraud Probability	Risk Level	Risk Indicator	Action code	Action Text
26	1	High	Red	BLOCK	Block transaction; initiate fraud protocol.
93	0.42	Medium	Amber	REVIEW	Flag for manual review.
33	0.2	Low	Green	ALLOW	No action; process normally.

Conclusion (production readiness)

Based on the EDA, the target feature Engineering and the rigorous training, tuning and evaluation the recommended model is **LightGBM (tuned)**, This solution is wrapped in a simple three-tier user-friendly API for operational use i.e. fraud_risk_assessment. The solution is ready for production. This solution handles the severe imbalance (fraud prevalence = 0.173%), provides calibrated probabilities, strong performance in positive class (test PR AUC = 0.857, precision = 0.95, recall = 0.776). Explainability was addressed via feature importance. To find supporting materials see:

- /results/plots/ for figures (PR curves, importance, confusion matrices and etc).
- /results/tables/ for metric tables (OOF/test metrics and etc).