

Сравнение различных моделей для детекции аномалий на выделенном **test**, с использованием timesplit и временной кросс валидацией.

Данные - [Credit Card Fraud Detection](#)

модель	Precision	Recall	F1	PR_AUC
IsolationForest (обучение только на нормальных данных)	0.03	0.004	0.03	0.05
LocalOutlierFactor (обучение только на нормальных данных)	0.43	0.69	0.53	0.43
OneClassSVM (обучение только на нормальных данных)	0.12	0.65	0.21	0.12
LocalOutlierFactor + rolling статистики (обучение только на нормальных данных)	0.38	0.40	0.39	0.2
Ансамбль LocalOutlierFactor (обучение только на нормальных данных)	0,55	0,68	0,60	0,51
Random Forest + custom GridSearchCV (supervised)	0,81	0,77	0,79	0,80
CatBoost +Optuna (supervised)	0.96	0.71	0.82	0.82
Простая нейросеть + Optuna + подбор порога (supervised)	0,90	0,71	0,8	0,81
TabNET + Optuna + подбор порога (supervised)	0.89	0.75	0.82	0.79
TabMixer + Optuna + подбор порога (supervised)	0.96	0.68	0.79	0.81
LSTM + Optuna + подбор порога (supervised)	0,88	0,74	0,8	0,7
Transformer + Optuna + подбор порога (supervised)	0,86	0,71	0,78	0,81

Краткий анализ:

Модели делятся на две группы:

Unsupervised (обучение только на нормальных данных) - не используют метки аномалий. Unsupervised модели плохо справляются с задачей, особенно если важно не пропустить аномалии и не перегрузить систему ложными срабатываниями.

Supervised (обучение с метками) - используют метки и подбирают параметры. Дают гораздо более надёжные и точные результаты. Особенно эффективны градиентный бустинг и нейросетис подбором порога.

Вывод:

Если есть метки аномалий - нужно выбирать supervised модели (CatBoost, Random Forest, нейросеть).

Если меток нет - ансамбль LOF может принести лучший результат.

Замечание:

В реале, наверное, лучше проводить оценку по денежному эквиваленту или **value-based evaluation**. Вместо того чтобы считать количество правильно найденных мошеннических транзакций, оценивают суммарную стоимость этих транзакций.