

Detecção de Fraudes em Cartões de Crédito

Projeto MachineCredit - panorama da avaliação do modelo final em dados desbalanceados.

Metricas Chave

- Acuracia: **0.94**
- Macro Avg F1-Score: **0.93**
- Precisao (fraude): **0.88**
- Recall (fraude): **0.91**

Matriz de Confusao

```
[243  11]  
[ 9   89]
```

- 9 fraudes perdidas (falsos negativos) e 11 alertas falsos.
- Recall acima de 0.90 prioriza a detecção de fraudes.

Tecnicas Aplicadas

- EDA: exploração do desbalanceamento e distribuições monetarias.
- Seleccionadores: SelectKBest, RFE (logistic) e importancia de Random Forest.
- Balanceamento: comparação entre SMOTE, Random Undersampling e SMOTEENN.
- Tuning: RandomizedSearchCV (3 folds) em Decision Tree, XGBoost e KNN.
- Ensembles: Voting e Stacking com modelos ajustados.

Impacto do Desbalanceamento

- Dataset com 1.267 transacoes legítimas e 492 fraudulentas.
- Sem reamostragem: recall médio ~0.84.
- Random undersampling elevou o recall base para 0.88.
- Configuração final (RFE + undersampling + XGBoost) entregou recall 0.91.
- Relatórios e gráficos em `outputs/reports` e `outputs/plots`.