

Mecanismo para Detecção de Transações de Cartão de Crédito Fraudulentas

Douglas Affonso Clementino
Rafael De Paulo Dias





Base de Dados

Variáveis

- Time
- Amount
- Class
- v1-v28



Base de Dados

Variáveis

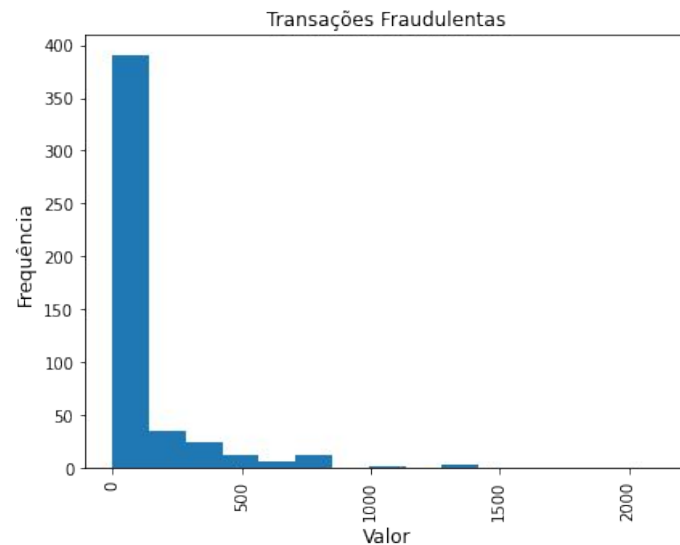
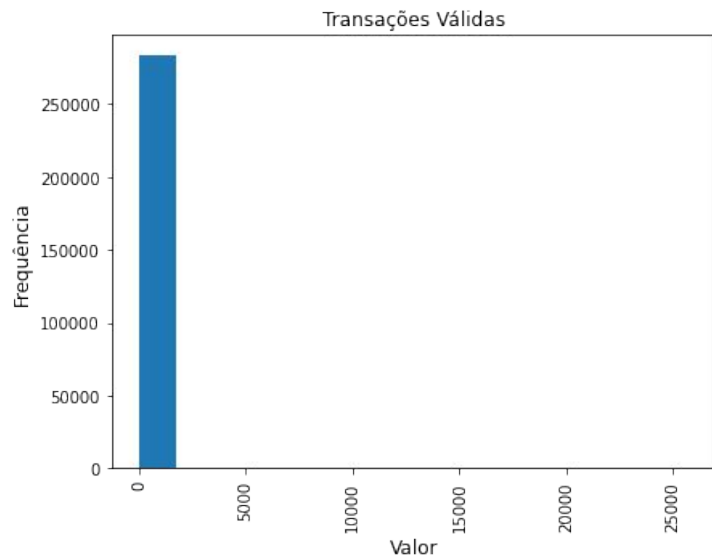
- Time
- Amount
- Class
- v1-v28

Distribuição

- Total de Amostras: 284.807
- Classe 0 (Válidas): 284.315 (99,82%)
- Classe 1 (Fraude): 492 (0,17%)

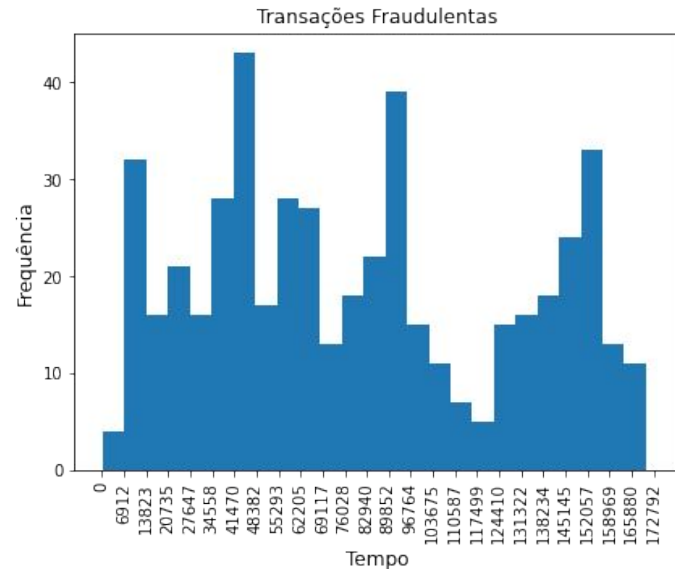
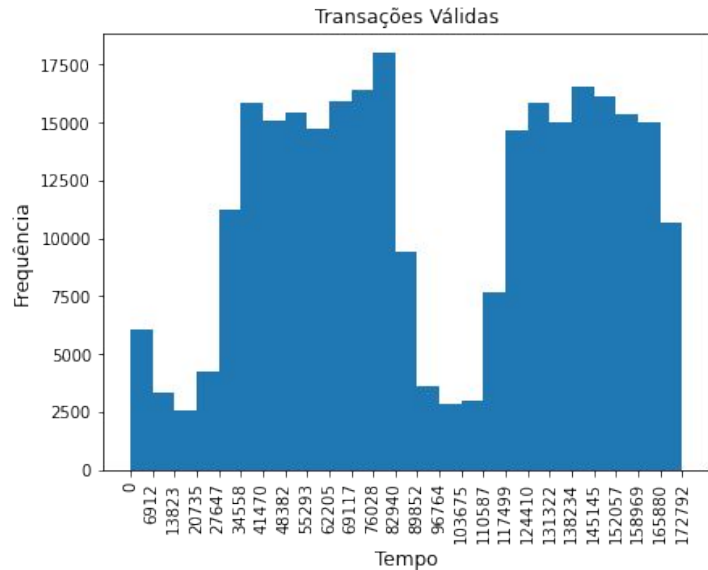


Análise de Dataset





Análise de Dataset





Experimentos

Escolha de Classificador

- KNeighborsClassifier (KNN)
- Multi-layer Perceptron (MLP)
- CatBoostClassifier (Random Forest)



Experimentos

Escolha de Classificador

- KNeighborsClassifier (KNN)
- Multi-layer Perceptron (MLP)
- CatBoostClassifier (Random Forest)

Tuning de Classificador (CatBoostClassifier)

- **learning_rate:** 0.0947
- **max_depth:** 7
- **n_estimators:** 700.0



Experimentos

Escolha de Classificador

- KNeighborsClassifier (KNN)
- Multi-layer Perceptron (MLP)
- CatBoostClassifier (Random Forest)

Tuning de Classificador (CatBoostClassifier)

- `learning_rate`: 0.0947
- `max_depth`: 7
- `n_estimators`: 700.0

Balanceamento de Bases de Treino

- Synthetic Minority Over-sampling (SMOTE)



Experimentos

Escolha de Classificador

- KNeighborsClassifier (KNN)
- Multi-layer Perceptron (MLP)
- CatBoostClassifier (Random Forest)

Tuning de Classificador (CatBoostClassifier)

- `learning_rate`: 0.0947
- `max_depth`: 7
- `n_estimators`: 700.0

Balanceamento de Bases de Treino

- Synthetic Minority Over-sampling (SMOTE)

Considerações

- Outliers
- CatBoostClassifier GPU

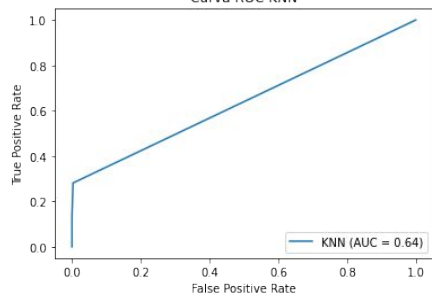


Escolha de Classificador

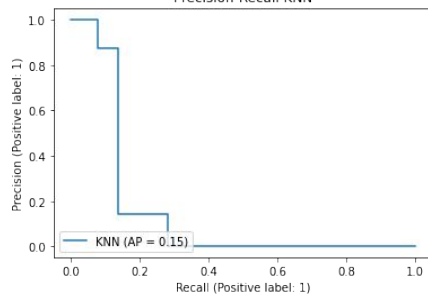


Escolha de Classificador

Curva ROC KNN



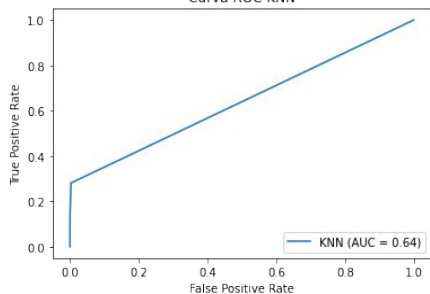
Precision-Recall KNN



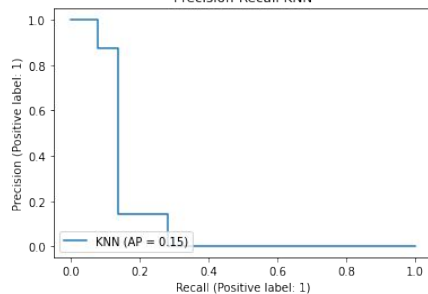


Escolha de Classificador

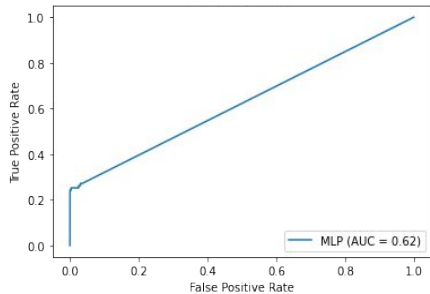
Curva ROC KNN



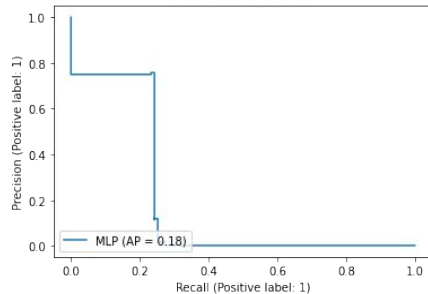
Precision-Recall KNN



Curva ROC MLP



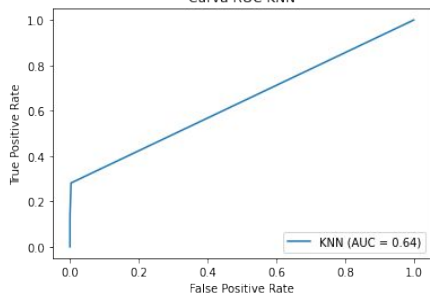
Precision-Recall MLP



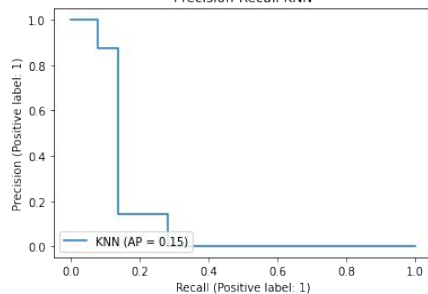


Escolha de Classificador

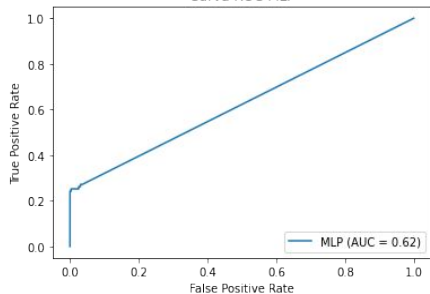
Curva ROC KNN



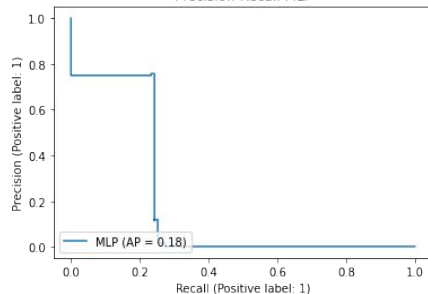
Precision-Recall KNN



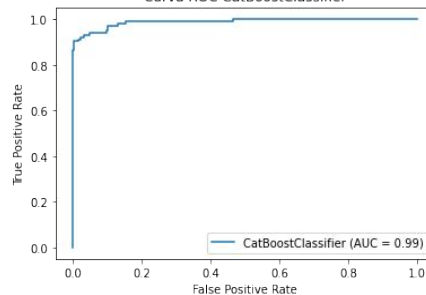
Curva ROC MLP



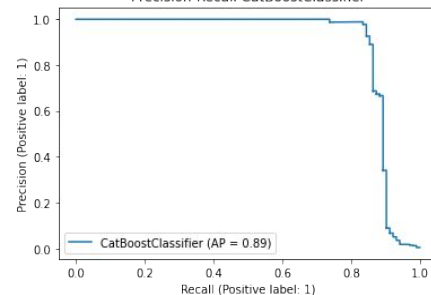
Precision-Recall MLP



Curva ROC CatBoostClassifier



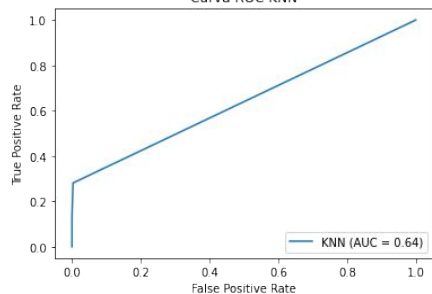
Precision-Recall CatBoostClassifier



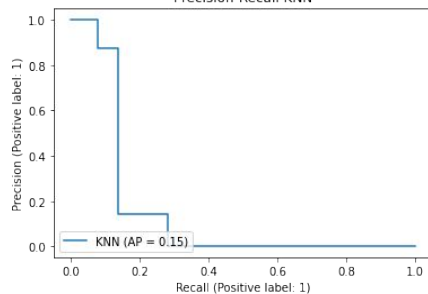


Escolha de Classificador

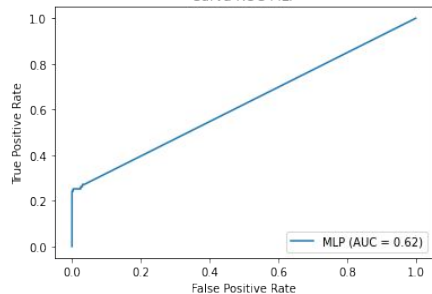
Curva ROC KNN



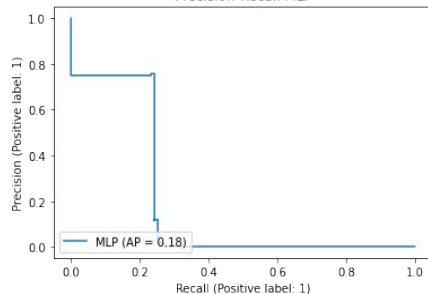
Precision-Recall KNN



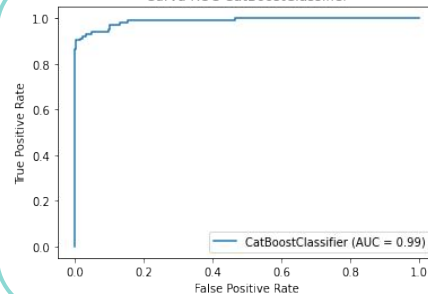
Curva ROC MLP



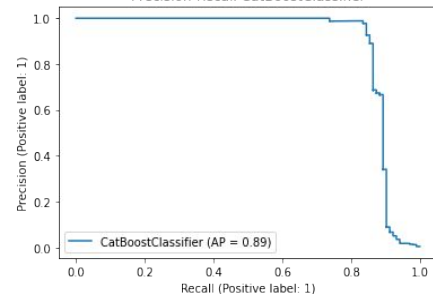
Precision-Recall MLP



Curva ROC CatBoostClassifier

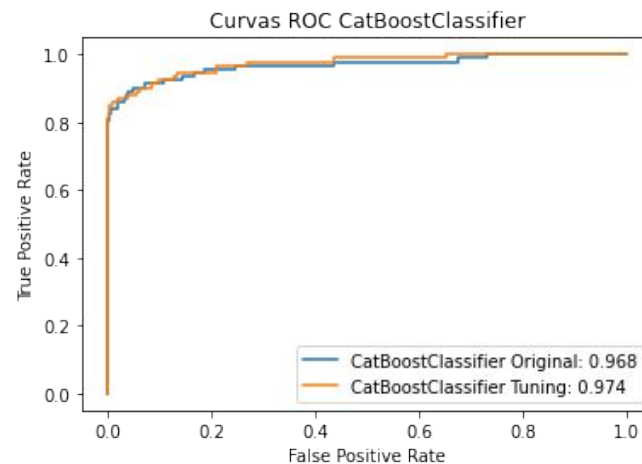
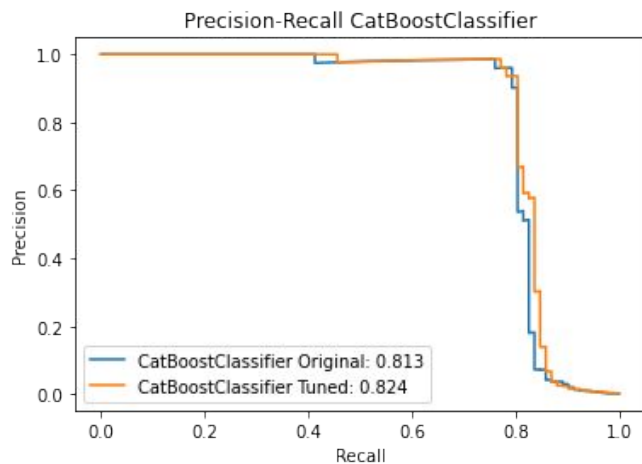


Precision-Recall CatBoostClassifier



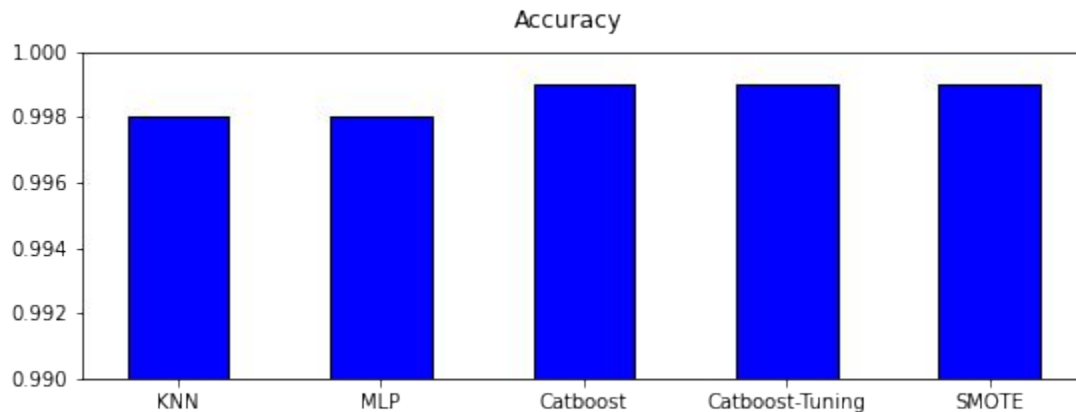


Tuning de Classificador



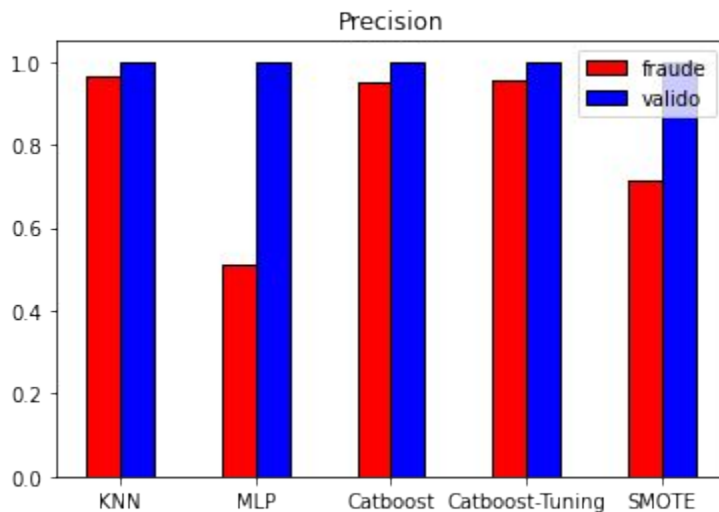


Médias de Métricas por Modelo



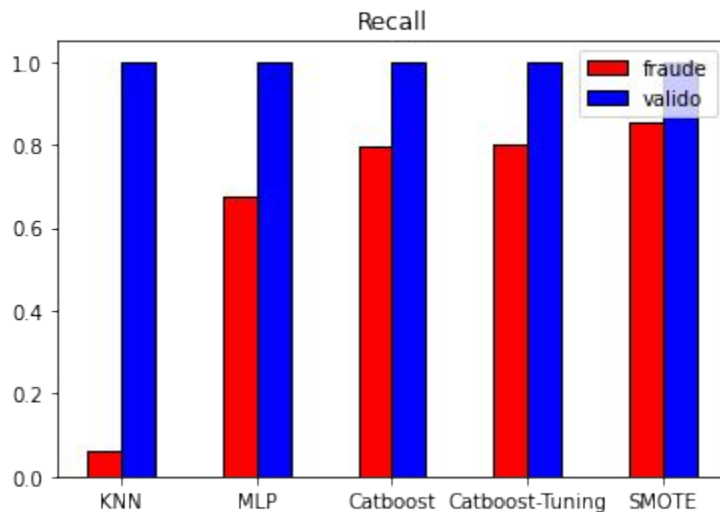


Médias de Métricas por Modelo



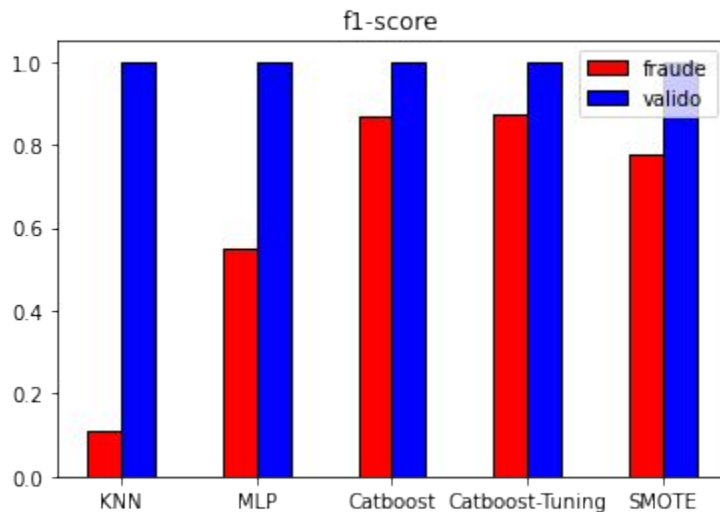


Médias de Métricas por Modelo





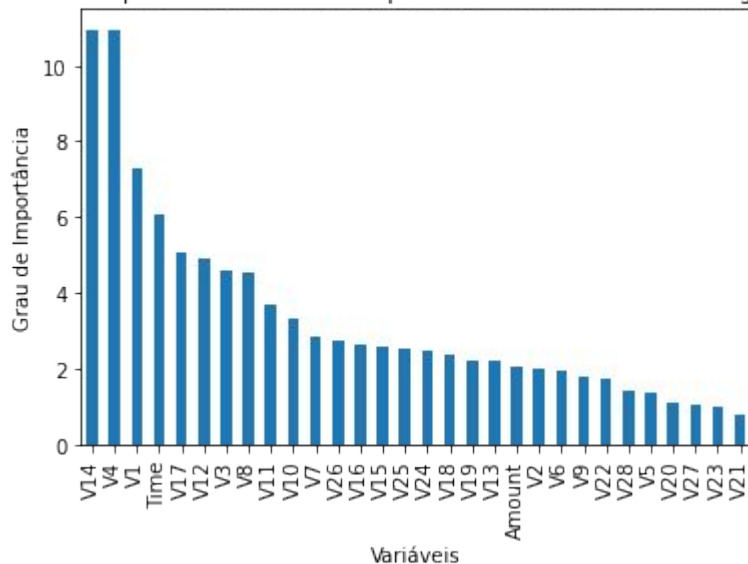
Médias de Métricas por Modelo





Importância de Variáveis (CatBoostClassifier-Tuning + SMOTE)

Análise de Importância de Variável para CatBoostClassifier-Tuning + SMOTE





Conclusão

- Classificador CatBoostClassifier foi o que atingiu as melhores métricas gerais e para a Classe 1 (Fraudes).
- Apesar de pouco, foi possível obter desempenho a partir do tuning do classificador CatBoostClassifier.
- Utilização da técnica de oversampling SMOTE contribuiu para o reconhecimento da transacções fraudulentas.
- De forma geral, as técnicas empregadas cooperaram com o objetivo principal do trabalho, aumentar a taxa de Verdadeiros positivos para Classe 1, Fraudes.



Obrigado

Link para repositório de trabalho: <https://github.com/Daffc/Deteccao-Fraudes-ML>