

Detecção de Fraude em Transações

Relatório de Atividade

Seu Nome

Matrícula: 0000000

Disciplina XYZ

10 de maio de 2025

Sumário

1	Introdução	1
2	Metodologia (Parte Prática)	1
2.1	Análise Exploratório de Dados (EDA) e Pré-processamento	1
2.2	Abordagem 1: Modelagem da Classe Normal (Detecção de Anomalia) . . .	1
2.3	Abordagem 2: Classificação Supervisionada	1
2.4	Validação e Generalização	1
3	Resultados	1
4	Discussão e Comparação Final	2
5	Conclusão	2
6	Respostas Teórico-Conceituais (Parte 2)	2
6.1	Exercício 1: Fronteira de decisão em modelos generativos	2
6.2	Exercício 2: Verossimilhança, entropia cruzada e regressão logística	2
6.3	Exercício 3: Decomposição viés-variância	2
6.4	Exercício 4: Avaliação em dados desbalanceados	2
7	Referências	2

1 Introdução

- Contextualização do problema de detecção de fraude
- Objetivos do trabalho
- Descrição do dataset (fonte, número de amostras, desbalanceamento)

2 Metodologia (Parte Prática)

2.1 Análise Exploratório de Dados (EDA) e Pré-processamento

- Estatísticas descritivas e visualização das distribuições
- Tratamento de valores ausentes, codificações e escalonamento
- Estratégia para lidar com desbalanceamento

2.2 Abordagem 1: Modelagem da Classe Normal (Detecção de Anomalia)

- Escolha do modelo generativo probabilístico
- Estimação de parâmetros via MLE
- Cálculo do score de anomalia
- Visualização das distribuições de scores
- Seleção e justificativa do limiar

2.3 Abordagem 2: Classificação Supervisionada

- Treinamento de modelo probabilístico
- Geração de probabilidades $p(C_{\text{fraude}} | x)$
- Visualização das distribuições de probabilidade
- Escolha do limiar e justificativa

2.4 Validação e Generalização

- Estratégia de validação (cross-validation, hold-out)
- Seleção de modelo e otimização de hiperparâmetros

3 Resultados

- Métricas independentes de limiar: AUC-PR
- Métricas dependentes de limiar: Precisão, Recall, F1, matriz de confusão
- Gráficos de curva PR e distribuição de scores/probabilidades

4 Discussão e Comparação Final

- Comparação de performance entre as abordagens
- Vantagens, desvantagens e limitações
- Impacto prático da escolha de limiares
- Recomendações de uso

5 Conclusão

- Principais achados
- Sugestões de trabalhos futuros

6 Respostas Teórico-Conceituais (Parte 2)

6.1 Exercício 1: Fronteira de decisão em modelos generativos

6.2 Exercício 2: Verossimilhança, entropia cruzada e regressão logística

6.3 Exercício 3: Decomposição viés-variância

6.4 Exercício 4: Avaliação em dados desbalanceados

7 Referências

Referências

[1] Pedregosa et al., *Scikit-learn: Machine Learning in Python*, JMLR, 2011.