

Lista 9 - IA

Lista 9 - Pre-processamento e Agrupamento (base creditcard.csv)

Dados e limpeza - Linhas originais: 284807; duplicatas removidas: 9144; apos limpeza: 275663. - Outliers removidos (zscore>3 apenas classe 0): 36686; linhas finais: 238977. - Colunas removidas: Time (irrelevante), V2 (corr alta com Amount); total de colunas usadas: ['V1', 'V3', 'V4', 'V5', 'V6', 'V7', 'V8', 'V9', 'V10', 'V11', 'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20', 'V21', 'V22', 'V23', 'V24', 'V25', 'V26', 'V27', 'V28', 'Amount']. - Valores ausentes detectados: 0. - Distribuicao de classe apos limpeza: {0: 238504, 1: 473}. - Pares com correlacao > 0.5: V21-V22:0.575, V27-V28:0.522.

Normalizacao, split e balanceamento - Escala MinMax (0-1) ajustada no treino; split estratificado 80/20. - Conjuntos de treino: {'Sem balanceamento': {0: 190803, 1: 378}, 'TomekLinks': {0: 190799, 1: 378}, 'SMOTE': {0: 190803, 1: 190803}, 'Tomek+SMOTE': {0: 190799, 1: 190799}}.

Modelo MLP (relu, hidden_layer=(n_features+1)/2, lr adaptive, lr_init=0.175, max_iter=7000) - Teste avaliado em todos os cenarios acima (mesmo X_test). - Resultados: - Sem balanceamento: acc=0.9996, prec=1.000, recall=0.779, f1=0.876, matriz=[[47701, 0], [21, 74]] - TomekLinks: acc=0.9996, prec=1.000, recall=0.821, f1=0.902, matriz=[[47701, 0], [17, 78]] - SMOTE: acc=0.9093, prec=0.021, recall=0.958, f1=0.040, matriz=[[43371, 4330], [4, 91]] - Tomek+SMOTE: acc=0.9550, prec=0.040, recall=0.937, f1=0.076, matriz=[[45554, 2147], [6, 89]]

Agrupamento (dados processados sem coluna Class; amostra <=40k, padronizada) - KMeans: clusters=2, silhouette=0.06631427355989863, DB=3.9235810551640027, CH=2214.8514845798895, ruido=None, params={'n_clusters': 2} - DBSCAN: clusters=152, silhouette=0.1461169438772722, DB=1.239497992109749, CH=221.20398660425212, ruido=0.6136, params={'eps': 1.6, 'min_samples': 10} - SOM: clusters=2, silhouette=0.03900037133139597, DB=5.075692629892714, CH=1328.1208830165742, ruido=None, params={'grid': '1x2', 'iterations': 800}

Arquivos - creditcard_processado.csv (dados limpos da Q1, sem Time e sem V2). - outputs/results_summary.json, outputs/relatorio_lista9.md, outputs/relatorio_lista9.pdf.