Aprendizado de Máquina

Classificação Supervisionada

Tathiane Souza

Esta pesquisa utiliza dados sobre pagamentos de clientes de Taiwan fazendo uma análise sobre a inadimplência.

1.

Explorar os dados

Estudar os dados

2.

Preparar os dados

Identificar as colunas de características e a coluna alvo.

3.

Pré-processamento dos dados

Normalização
Binarização

4.

Treinar o modelo

Logistic Regression
Decision Tree
SVC
GaussianNB
Perceptron

5.

Redução de Dimensionalidade

Utilizando o PCA

6.

Refinar o modelo

Ajustar os parâmetros

7.

Avaliar o modelo

Acurácia
F1 score
Precision e recal
Curva roc
Matrix de convulsão

Amostra dos resultados obtidos

	Classificador	Tamanho	Tempo treino	Tempo predisão	F1 score - treino	F1 score - teste
0	DecisionTreeClassifier	5000	0.213	0.003	1.00000	0.7228
1	DecisionTreeClassifier	10000	0.403	0.001	0.99980	0.7076
2	DecisionTreeClassifier	20000	0.974	0.001	0.99940	0.7132
3	SVC	5000	5.197	2.114	0.99500	0.7700
4	SVC	10000	36.748	4.066	0.99600	0.7812
5	SVC	20000	262.170	7.895	0.99325	0.7656
6	GaussianNB	5000	0.007	0.004	0.38460	0.3800
7	GaussianNB	10000	0.014	0.000	0.36920	0.3712
8	GaussianNB	20000	0.022	0.002	0.36355	0.3704
9	LogisticRegression	5000	0.249	0.001	0.78100	0.7688
10	LogisticRegression	10000	0.516	0.004	0.77840	0.7800
11	LogisticRegression	20000	1.475	0.000	0.77935	0.7628
12	Perceptron	5000	0.014	0.002	0.78160	0.7688
13	Perceptron	10000	0.021	0.000	0.77600	0.7804
14	Perceptron	20000	0.040	0.004	0.77665	0.7604

Obrigada!

tathianers@alu.ufc.br

Referências

https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients