

PjBL

Alunos: _____

O PjBL deverá ser desenvolvido em grupos de até 5 pessoas é composto de 4 atividades, cada uma equivale a 25% da nota final do TDE. Os alunos deverão selecionar um dataset para os problemas de classificação e outro para os problemas de regressão. Lembrando que em certos cenários, a excelente performance dos modelos está ligada diretamente a qualidade do dataset.

OBS: Os datasets não podem ser iguais ao do TDE e nem dos utilizados em sala de aula, pois trata-se de uma atividade diferente e importante para o entendimento da disciplina.

Sendo assim, após a escolha das bases de dados, realize a aplicação dos modelos para complementar as 4 tabelas abaixo:

A) (25%) Classificação com protocolo Hold-out (65% para treinamento e 35% para teste).

Indutor	Taxa de Acerto (%)	F1 (%)	Precisão (%)	Sensibilidade (%)	Especificidade (%)	Nota
KNN						5%
Decision Tree						5%
Naive Bayes						5%
MLP						5%
SVM						5%
Ensemble (somatória)						5%
Random Forest						10%
Bagging						10%
Boosting						10%
Stacking						15%
Blending						15%
Adicional						10%

B) (25%) Classificação com protocolo experimental validação cruzada com 5 folds.

Indutor	Taxa de Acerto (%)	F1 (%)	Precisão (%)	Sensibilidade (%)	Especificidade (%)	Nota
KNN						5%
Decision Tree						5%
Naive Bayes						5%
MLP						5%
SVM						5%
Ensemble (somatória)						5%
Random Forest						10%
Bagging						10%
Boosting						10%
Stacking						15%
Blending						15%
Adicional						10%

C) (25%) Regressão com protocolo Hold-out (65% para treinamento e 35% para teste)

Indutor	Coefficiente de Determinação (R2)	MSE	RMSE	MAE	Nota
Regressão Linear					5%
KNN					5%
Árvore de Decisão					5%
MLP					5%
SVM					5%
Ensemble (Média)					5%
Random Forest					10%
Bagging					10%
Boosting					10%
Stacking					15%
Blending					15%
Adicional					10%

D) Regressão com protocolo experimental validação cruzada com 5 folds.

Indutor	Coefficiente de Determinação (R2)	MSE	RMSE	MAE	Nota
Regressão Linear					5%
KNN					5%
Árvore de Decisão					5%
MLP					5%
SVM					5%
Ensemble (Média)					5%
Random Forest					10%
Bagging					10%
Boosting					10%
Stacking					15%
Blending					15%
Adicional					10%

Observações:

- 1. Os alunos deverão apresentar os códigos do programa e realizar a defesa dos códigos.**