

Sistemas de Informação Multimedia (SIM) – 2019/2020

Trabalho prático - Parte II

Trabalho a ser realizado (segundo momento):

1. Usar partição 75% *training* e 25% *testing* do dataset obtido no ponto (10) da primeira parte deste trabalho;
2. Calcular, usando sklearn, *precision* e *recall*, interpretar o resultado de forma sucinta numa célula markup no seu caderno Jupyter;
3. Calcular, usando sklearn, *f1-score* e *mcc*, interpretar o resultado de forma sucinta numa célula markup no seu caderno Jupyter;
4. Escrever uma avaliação final sobre o modelo obtido via regressão logística. Recomendaria usar este modelo, ou recomendaria investigar outras abordagens?

Bonus (pergunta de investigação, para extra 25% na nota prática)

Investigar como aplicar *k-fold cross validation* neste *pipeline*, e aplicar usando *k=5* folds usando como scoring function o *mcc*. Uma vez obtidos os resultados, adicionar uma interpretação sucinta numa célula *markup*.

Dicas:

https://chrisalbon.com/machine_learning/model_evaluation/cross-validation/
https://scikit-learn.org/stable/modules/model_evaluation.html

