

# Trabalho Prático Final: Algoritmos Supervised Machine Learning para aprendizagem de modelos de classificação de dados

---

Valor: 6 valores (a teórica terá um valor total de 14 valores) | Mínimos: 30%

Data de entrega: 08-01-2021 | Data da defesa: dia da frequência

Grupos: 3 elementos | Defesa: Presencial com a participação de todo o grupo

Descrição: Desenvolva em Python, utilizando as bibliotecas numpy, pandas e sklearn, um conjunto de programas que permita determinar a exatidão (accuracy) e a matriz de confusão (confusion matrix) de cada um dos modelos machine learning criados utilizando os seguintes algoritmos:

1. Naive Bayes;
2. Decision Tree;
3. Random Forest;
4. kNN

Estes algoritmos de aprendizagem devem ser aplicados a dois datasets, disponíveis no site da [UCI](#), na construção de cada um dos modelos machine learning para classificação de dados:

1. Census Income Data Set
2. Congressional Voting Records Data Set

O programa deve garantir a preparação dos dados de forma a poderem ser utilizados na construção dos diferentes modelos pelos diferentes algoritmos. Verifique a evolução dos resultados quando utiliza diferentes percentagens do set de testes:

1. 15%
2. 30%
3. 50%

Baseie-se na tabela em baixo para apresentar um resumo dos resultados obtidos:

Precision	Naive Bayes	Decision Tree	Random Forest	kNN
labelencoder				
labelEncoder+standardScaler				
labelEncoder+oneHotEncoder				
labelEncoder+oneHotEncoder+standardScaler				

Deve ser entregue o código desenvolvido bem como um relatório explicativo do trabalho efetuado. Espera-se uma análise crítica dos resultados obtidos.