Universidade de Brasília
Departamento de Ciência da Computação
Projeto 2 , Introdução à Inteligência Artificial, Turma A, 2021/1
Prof. Díbio

Técnicas de Inteligência Artificial, especialmente Aprendizagem de Máquinas como Árvores de Decisão e Florestas Randômicas têm atingido excelentes resultados na predição/classificação diagnóstica de várias doenças. Um banco de dados foi coletado pelo NIDD em mulheres indígenas, com pelo menos 21 anos de idade, da etnia Pima. Oito (8) variáveis independentes e uma (1) alvo/classificação foram medidas em 768 pessoas. O arquivo com esses dados estão em:

https://www.kaggle.com/uciml/pima-indians-diabetes-database

O projeto visa aplicar algoritmos de Florestas Randômicas para predição de diabetes com base nesses dados. Sua solução deverá incluir:

1. Selecionar aleatoriamente 80% dos dados para treinamento e 20% para teste; (1,0)
2. Executar e mostrar os resultados para Floresta Randômica com números diferentes de estimadores n, sendo pelo menos n=10, n= 100, e n = número estimado pela sqrt(?) do estudo do algoritmo; (2,0)
3. Mostrar resultados de precisão, revocação, medida f1 de todos, assim como matrizes de confusão (2,0)
4. Plotar resultados com número de árvores variando de 10, 20, 50, 100, 200. 500, 1000 (2,0)
5. Comparar com os resultados do artigo Sisodia, D. & Sisodia, D. S. “Prediction of Diabetes using Classification Algorithms”, 10.1016/j.procs.2018.05.122 (1,0)
6. Anexar na entrega relatório sucinto com os resultados pedidos e conclusões. (2,0)

O código deve ser bem documentado, escrito em Python, por um (1) estudante individualmente do curso, e entregue somente via sistema http://aprender3.unb.br do curso, no prazo estipulado. **O estudante deve indicar no código se, e de onde, estão usando fontes públicas de outros, e realizar suas próprias alterações para entendimento. Códigos iguais, ou tendo indicativo de plágios, ou feitos por outros, poderão receber nota zero.**