

Universidade Federal do Ceará.

Campus de Quixadá

Avaliação Parcial 2

Disciplina: Mineração de Dados

Professora: Livia Almada Cruz

Aluno: _____

Data: ____/____/____

Matrícula: _____

Importante:

Acesse o Moodle da disciplina e baixe o diretório mineracao_ap2 e a documentação para consulta.

Desconecte o cabo de rede do computador.

Somente é permitida consulta à documentação disponibilizada.

Não é permitido o uso de rede, pendrive ou qualquer outro meio de armazenamento externo de dados.

Ao concluir a avaliação, compacte somente a sua pasta mineracao_ap2 com dataset e resolução, e altere o nome do arquivo compactado para conter a matrícula e o nome do aluno (<matr>-<nome>-

mineracao_ap2). Somente depois disso, chame a professora para entregar sua resolução, que deverá ser possível, use uma semente (seed) ou random_state com o valor 42.

Crie um Jupyter Notebook para responder as questões a seguir.

***** Para todas as questões e etapas, documente ao máximo o seu notebook explicando o quê (e porquê) está sendo feito para facilitar a correção.**

Questão 1)

Problema: Diagnóstico de câncer de mama

Tipo: Classificação

Dataset: breast cancer wisconsin dataset

Avalie os métodos Ensemble (BaggingClassifier com DecisionTreeClassifier, RandomForestClassifier e GradientBoostingClassifier).

Para isso:

- 1 – Separe o conjunto de dados em 80% para treino e 20% para teste;
- 2 – Faça a padronização dos atributos e trate atributos faltante;
- 3 – Use o K-Fold Cross Validation **sobre os dados de treino** com $k = 3$ para resolver os problemas considerando os algoritmos especificados e valide em cada etapa usando a parte dos dados que ficou para validação;

Avalie as seguintes combinações de hiper-parâmetros:

- learning_rate: 0.1, 0.05, 0.01 (somente para o Gradient Boosting)
- n_estimators: 50, 100, 200
- max_depth: 3, 5, 7

- 4 – Compare os resultados segundo Accuracy, Precision, Recall.

- 5 – Escolha a melhor técnica para o modelo final, treine com todo o conjunto de treino, teste e exiba o resultado obtido.

Questão 2)

- 1 – Crie uma rede Multi Layer Perceptron usando o Keras para resolver o problema da questão 1.

- 2 – Explique a arquitetura da rede criada.

- 3 – Considere 70% de dados para treino, 10% para validação e 20% para teste. Apresente a acurácia do seu modelo para dados de teste. Apresente a acurácia nos dados de validação durante a evolução das épocas.

- 4 – Qual a época que obteve melhor desempenho?

- 5 – Re-treine o seu modelo com os dados de treino+validação, considerando a melhor época. Avalie seu modelo para os dados de teste.