

Lista #2

Curso: Ciência de Dados

Disciplina: Aprendizado de Máquina I

Prof^a. Cristiane Neri Nobre

Data de entrega: 15/09

Valor: 1,5 pontos

Questão 01

Considerando-se a base de dados sobre “**Esperar ou não pelo restaurante**” (verificar base de dados ‘**Restaurante.csv**’ disponibilizada no CANVAS), pede-se:

- 1) Calcular o ganho de informação de cada atributo. Que atributo é a raiz da árvore?
- 2) Que atributo estará no segundo nível da árvore. Faça os cálculos e apresente a árvore gerada até o segundo nível da árvore.

Questão 02

Considerando-se os códigos (**Lendo_e_tratando_arquivo_v2.ipynb** e **DecisionTree.ipynb**) disponibilizados no CANVAS, pede-se:

- 1) Com a codificação atual dos atributos de entrada, rode-os e plote a árvore de decisão obtida.
- 2) Altere a codificação do atributo **Tipo** para um atributo nominal ordinal e faça os experimentos acima novamente. O que você observa? Houve ganho?
- 3) Utilizando o código (**DecisionTree_com_gridsearch.ipynb**), ajuste os hiperparâmetros da árvore e veja que hiperparâmetros retornaram os melhores resultados. Investigue outros hiperparâmetros da árvore.

Questão 03

Leia o artigo “**A_comparative_study_of_decision_tree_ID3_and_C4.5.pdf**” que está no CANVAS e responda:

- 1) Quais as diferenças entre os algoritmos de árvore ID3 e C4.5?
- 2) Como o algoritmo C4.5 lida com os atributos de entrada que são numéricos?