

Lista #3

Curso: Ciência da Computação

Disciplina: Inteligência Artificial

Prof^a. Cristiane Neri Nobre

Data de entrega: 31/08

Valor: 1 ponto

Observação:

Para todas as listas que envolvem a implementação das funções em Python, solicita-se que:

1. Todas as discussões das questões devem estar contidas na lista. Ou seja, todas as decisões e explicações necessárias precisam estar na lista, que só pode ser entregue em PDF. **Listas em qualquer outro formato serão zeradas.**
2. Os links para os códigos desenvolvidos devem estar inseridos na lista, com as devidas permissão de acesso. **Listas sem permissão de acesso serão zeradas.**

Questão 01

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

- 1) Gerar a árvore para a base de dados **Restaurante**, alterando a codificação do atributo **cliente** para conter a seguinte codificação:
 Cliente_Nenhum = 0
 Cliente_Algun = 1
 Cliente_Cheio = 2

Questão 02

Baseado nestes códigos acima, encontrar o padrão de pessoas que sobreviveram ao desastre do TITANIC, que matou mais de 1.500 pessoas em 1912. A base de dados do TITANIC está no CANVAS.

1. Visualize a base de dados primeiro, veja como estão os atributos e suas distribuições.
2. Investigue a melhor forma de codificar cada atributo da base de dados.
3. Forneça as regras que mostre o padrão de mortalidade.

Questão 03

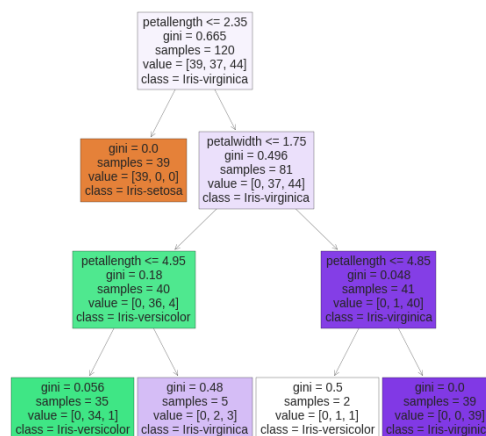
Leia o artigo “**A_comparative_study_of_decision_tree_ID3_and_C4.5.pdf**” que está no CANVAS e responda:

- 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?

Questão 04

A figura abaixo mostra uma árvore de decisão construída por um algoritmo de aprendizado indutivo a partir de um conjunto de dados em que as instâncias são descritas por quatro atributos: **Tamanho** da Pétala, **largura** da Pétala, **Tamanho** da Sépala e **Largura** da Sépala.

Dado um objeto de classe desconhecida, essa árvore classifica o objeto nas classes: **Iris_Setosa**, **Iris_Virgínica** e **Iris_Versicolor**. Esta árvore foi gerada com os hiperparâmetros (DecisionTreeClassifier(criterion='gini', max_depth=3)), usando a linguagem Python.



Com base nestas informações, qual as saídas da árvore para os seguintes **registros de teste**, respectivamente?

Registros de teste	Tamanho da Pétala	Largura da Pétala	Tamanho da Sépala	Largura da Sépala
Instância 1	3.46	0.87	2.45	1.78
Instância 2	1.67	1.89	0.78	1.32
Instância 3	2.56	2.34	2.45	1.78
Instância 4	6.67	2.34	2.45	1.78

- a) Iris_Virgínica, íris_Setosa, Iris_Versicolor, Iris_Virgínica
- b) Iris_Setosa, íris_Setosa, Iris_Virgínica , Iris_Versicolor
- c) Iris_Versicolor, íris_Setosa, Iris_Versicolor, Iris_Virgínica
- d) Íris_Setosa, Iris_Virgínica, Iris_Virgínica , Iris_Versicolor
- e) Iris_Versicolor, Íris_Setosa, Iris_Versicolor, Íris_Setosa

Questão 05

Considerando a árvore da questão anterior, e as seguintes afirmações:

- I. Esta árvore possui 5 regras de classificação
- II. Das regras geradas, há apenas uma com cobertura por classe de 100%
- III. A menor cobertura por classe é de 6.8% e corresponde à classe Iris_Virgínica

É **correto** o que se afirma em:

- a) I, apenas.
- b) III, apenas.
- c) I e II, apenas.
- d) I e III, apenas.
- e) I, II e III.

Questão 06

Considere a seguinte matriz de confusão obtida por meio do classificador, **Árvore de decisão**, para um problema de quatro classes:

		Foi classificado como			
		A	B	C	D
Era da classe	A	10	4	2	1
	B	1	15	2	0
	C	2	3	20	5
	D	4	1	2	50

Quais os valores para as métricas abaixo para cada uma das classes A, B, C e D?

	Precisão	Recall	F1Score	TVP	TFN	TFP	TVN
A							
B							
C							
D							