

RECUPERAÇÃO DA AVALIAÇÃO AP1 N2 – 2024.2  
DISCIPLINA DE INTELIGÊNCIA ARTIFICIAL  
Naive Bayes e Árvore de Decisão

Nome:

Data:

**Diretriz:** responda essas questões utilizando um editor de texto no computador e, ao término, envie para valter.costa@ifce.edu.br.

- 1) Uma empresa de recrutamento on-line deseja avaliar o potencial de seus assinantes que estão em busca de emprego. Para isso, reuniu um conjunto de dados contendo informações sobre candidatos que aplicaram para uma vaga, juntamente com o resultado final da análise da empresa empregadora: "**Aceito**" ou "**Rejeitado**".

Cada candidato possui três atributos:

**Experiência** (Alta/Baixa)

**Graduação** (Sim/Não)

**Habilidades Técnicas** (Boa/Ruim)

A empresa deseja construir uma árvore de decisão para prever se um candidato será aceito ou rejeitado com base nesses atributos.

- a) (2,5 pontos) Explique como a **entropia** e o **ganho de informação** são utilizados para decidir qual atributo deve ser escolhido como raiz da árvore de decisão.
- b) (2,5 pontos) Suponha que a entropia inicial do conjunto seja **0.94**. Após dividir os dados com base no atributo **Experiência**, obtemos: Candidatos com **Experiência: 42 Aceitos e 7 Rejeitados**; Candidatos sem **Experiência: 12 Aceitos e 78 Rejeitados**. Calcule o **ganho de informação** desse atributo e interprete o resultado.
- 2) (5 pontos) Uma empresa de e-commerce deseja prever se um cliente comprará ou não um produto após visualizar a página do item. Para isso, foi analisado um conjunto de 200 interações de clientes e coletados os seguintes atributos:

**Tempo na Página** (Curto ou Longo)

**Dispositivo** (Mobile ou Desktop)

**Origem do Tráfego** (Orgânico ou Pago)

A tabela a seguir resume os dados coletados para treinamento de um algoritmo de classificação:

Característica	Comprou (Sim)	Não Comprou (Não)
Tempo na Página = Longo	60	30
Tempo na Página = Curto	20	90
Dispositivo = Desktop	50	50
Dispositivo = Mobile	30	70
Origem do Tráfego = Orgânico	40	40
Origem do Tráfego = Pago	40	60

Sabemos que **80 clientes compraram o produto e 120 não compraram**.

Suponha que um novo usuário acessa a página do produto com as seguintes características:

Tempo na Página = **Longo**; Dispositivo = **Desktop**; Origem do Tráfego = **Orgânico**

Com base nos dados de treinamento, como o algoritmo Naive Bayes classificaria este cliente?

Considere:

$$P(A|B, \dots, Z) = \frac{P(B|A)P(\dots|A)P(Z|A)P(A)}{P(B)P(\dots)P(Z)}$$