

Nome:

Data:

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

- 1) (5 pontos) Você trabalha na *Amazon* e foi convocado para apoiar os setores de vendas e marketing. A empresa possui um banco de dados detalhado contendo todas (ordem de dezenas de milhões de amostras) as transações de compra realizadas por cliente e suas avaliações (estrelas, comentários, data e hora) acerca dos produtos comprados. Cada cliente possui um perfil com características como: idade, endereço, profissão, idioma preferido, histórico de compras, tempo de navegação no site e categorias de produtos visualizadas. Sua tarefa é projetar um sistema de recomendação de produtos para ser integrado ao *Echo Dot* ("Alexa"), permitindo que os produtos sejam recomendados por voz e que os clientes possam adicionar itens ao carrinho e concluir a compra através de comandos de voz. Em uma análise inicial, percebeu-se que a idade é uma variável relevante para prever o comportamento de compra dos clientes, mas apresenta muito ruído nos dados, o que compromete o desempenho do modelo ao ser utilizada. Desenvolva uma estratégia para garantir que o sistema de recomendação se mantenha atualizado com as mudanças nos comportamentos de compra dos clientes e nas tendências de mercado e envolva o uso da variável idade pelo sistema. A resposta deve abranger: seleção das amostras, pré-processamento dos dados, explicação do uso do algoritmo ao problema.
- 2) (5 pontos) Você recebeu um conjunto de dados pré-processados contendo dados acerca do perfil de cada usuário e quais objetos ele comprou na plataforma eBay (ar-condicionado, cama, mesa, ventilador e geladeira). Neste conjunto de dados, além da atributos numéricos como idade, existem características nominais qualitativas como idioma preferencial, gênero (masculino, feminino, não-binário, outro), etc. Considerando essas informações, responda o que se pede:

ID_usuario	Característica 1	Característica 2	...	Característica N	Produto Comprado
000001					geladeira
000001					cama
000002					ar-condicionado
000003					ventilador
000003					ar-condicionado
000004					geladeira
...	...	...	...	...	...
120000					Mesa

- a) (2,5 pontos) Descreva **como você utilizaria o K-NN** para responder à pergunta: Qual produto um usuário com as características de 1 até N supostamente compraria nessa plataforma? A descrição deve abranger: pré-processamento dos dados, estratégia de criação do conjunto de dados de treinamento/validação (explicando como será realizado), estratégia de treinamento/validação do modelo e como o algoritmo será entregue para ser utilizado em produção (especificando sucintamente a classe a ser utilizada pela equipe de backend).
- b) (2,5 pontos) Utilizando o algoritmo *Case-based Reasoning*, projete um sistema para ser aplicado ao problema. O projeto deve abranger: criação do banco de dados inicial com base nos dados pré-processados e identificação clara dos 4Rs do sistema proposto.