TESTE 1 –

Exercício 1

Suponha que você possui uma base de dados rotulada com 10 classes não balanceadas, essa base é formada por 40 features de metadados e mais 3 de dados textuais abertos.

Para todos os itens: Informe as bibliotecas usadas, se necessário, o motivo de cada decisão, explore as possibilidades.

1. Descreva como faria a modelagem dessas classes.
2. Ao finalizar essa modelagem, como iria apresentar essa modelagem para a área contratante?
3. Como faria a validação desse modelo?
4. Supondo que esses dados são recebidos diariamente, como iria trabalhar com esse desafio?
5. Como levaria esse projeto para um ambiente produtivo?
6. EXTRA - Existe mais algo que gostaria de relatar sobre esse caso?

Exercício 3

Suponha que você tenha uma base de dados contendo textos jurídicos, como decisões judiciais, petições e documentos legais. A base de dados inclui informações sobre o conteúdo do texto, data, jurisdição e outras informações relevantes. Seu objetivo é criar um sistema de recomendação que sugira textos jurídicos semelhantes a um texto de referência.

Para todos os itens: Informe as bibliotecas usadas, se necessário, o motivo de cada decisão, explore as possibilidades.

1. Descreva como você desenvolveria o sistema de recomendação que recebe um texto de referência e sugere os textos mais semelhantes a ele na base de dados.
2. Como você avaliaria esse sistema de recomendação?
3. Suponha que novos textos jurídicos sejam adicionados diariamente. Como você manteria o sistema de recomendação atualizado e garantiria que ele continue a fornece recomendações relevantes?

TESTE 2 -

1. O que é um desvio padrão e qual é o seu papel na medição da dispersão dos dados?
2. Como funciona o teste de hipóteses e qual é a sua finalidade na análise estatística?
3. O que é aprendizado supervisionado e como ele difere do aprendizado não supervisionado?
4. O que é transfer learning e como ele é usado em deep learning?
5. Você está conduzindo um experimento A/B em um site de comércio eletrônico para determinar a eficácia de uma nova página de destino na conversão de visitantes em clientes. Como você projetaria o experimento, escolheria as métricas apropriadas para avaliação e realizaria a análise estatística para tirar conclusões significativas?
6. Após realizar um teste ANOVA e obter um valor de F significativo, como você determinaria quais grupos são estatisticamente diferentes entre si?
7. Suponha que você tenha um conjunto de dados com três ou mais grupos para comparar e deseja determinar se há diferenças significativas entre eles. Descreva como você escolheria entre o teste ou outras técnicas estatísticas
8. Qual é a importância do pré-processamento de texto em tarefas de NLP? Quais são as etapas comuns no pré-processamento de texto?
9. Descreva o processo de vetorização de texto e como modelos de linguagem como o Word2Vec ou o TF-IDF podem ser usados para representar palavras e documentos.
10. Como você lidaria com problemas de desequilíbrio de classe em tarefas de classificação de texto em NLP? Quais estratégias seriam eficazes?

**OBS: Se for necessário, crie um repositório público no GitHub e compartilhe com a gente para a avaliação.**