

## TESTE 1 –

### Exercício 1

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.

- a. 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.
- b. Como você avaliaria esse sistema de recomendação?
- c. Suponha que novos textos jurídicos sejam adicionados diariamente. Como você manteria o sistema de recomendação atualizado e garantiria que ele continue a fornecer recomendações relevantes?

### Exercício 2

Suponha que você tenha uma tarefa específica de processamento de linguagem natural (PLN) para a qual deseja criar um modelo de alto desempenho. Você tem acesso a um modelo de linguagem pré-treinado e deseja realizar o fine-tuning desse modelo para sua tarefa específica.

- a. Escolha uma tarefa de NLP específica, de sua preferência, para a qual você deseja adaptar o modelo de linguagem, como classificação de sentimentos, resumo de texto, geração de texto ou qualquer outra tarefa.
- b. Como você faria a coleta e processamento desses dados?
- c. Descreva o processo de fine-tuning do modelo de linguagem pré-treinado para a tarefa específica.
- d. Como você avaliaria o desempenho do modelo fine-tuned, especifique as métricas.
- e. Descreva como você implantaria o modelo fine-tuned em um ambiente de produção para realizar previsões em novos dados.
- f. EXTRA - você pode explorar o ajuste de hiperparâmetros do modelo fine-tuned, como o tamanho da rede, a taxa de aprendizado e o número de épocas de treinamento, para otimizar o desempenho.

## TESTE 2

1. O que é um desvio padrão e qual é o seu papel na medição da dispersão dos dados?
2. O que é aprendizado supervisionado e como ele difere do aprendizado não supervisionado?
3. O que é transfer learning e como ele é usado em deep learning?
4. O que são redes generativas adversárias (GANs) e quais são os possíveis usos dessas redes?
5. O que são modelos de linguagem? Qual a diferença entre LLMs e modelos de linguagem tradicionais?
6. Explique o processo de pré-treinamento e fine-tuning de LLMs em termos gerais. Por que esse processo é importante?
7. 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?

8. Explique o conceito de hipótese nula e hipótese alternativa em um teste ANOVA. Qual é o papel da estatística F no teste?
9. O que é Processamento de Linguagem Natural (NLP) e quais são algumas das principais aplicações dessa tecnologia?
10. Qual é a diferença entre a classificação de texto e o agrupamento (clustering) de texto em NLP? Em que situações cada um é mais apropriado?

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