



Teste Técnico

Ciência de Dados – NLP/LLM

O teste técnico para a vaga de Cientista de Dados da A3Data consiste na exploração de uma base de dados com avaliações de livros, acesse os dados [aqui](#) e [aqui](#).

O objetivo deste teste é que você consiga explorar e demonstrar suas habilidades técnicas e de negócio, trazendo insights acionáveis para o negócio. **Para essa vaga, o conhecimento em NLP e LLM são essenciais. Nesta base temos colunas textuais e esperamos que você faça bom uso delas mostrando suas habilidades.**

Definição do problema:

A editora contratou a A3Data para explorar uma base de avaliações de livros. Atualmente o processo de exploração é feito manualmente pelos analistas e leva muito tempo. O objetivo é tanto de entender a performance de um certo autor, um gênero ou até buscar usuários com opiniões relevantes para uma entrevista. A editora gostaria de uma ferramenta que agilizasse e otimizasse este processo. Você pode desenvolver o teste em sua linguagem de programação de preferência.

Você deverá entregar os seguintes artefatos:

1. Apresentação dos resultados, um Power Point ou Google slides, simulando uma apresentação para o time de negócio. Ela deve conter:
 - a. Apresentação do desafio
 - b. Planejamento de entregáveis (inclusive futuros) – Roadmap
 - c. Explicação do processo utilizado
 - d. Hipóteses levantadas
 - e. Análise exploratória
 - f. Sumarização das informações textuais
 - g. Uso de bases de conhecimento
 - h. Modelo de Q&A
 - i. Estimativa de impacto das ações sugeridas
 - j. Desejável/Opcional: fine tuning de modelo open source para a aplicação
2. Código no GitHub

O que iremos avaliar:

1. Capacidade analítica
2. Qualidade do storytelling na apresentação do problema e no código
3. Qualidade e clareza do código
4. Criatividade

5. Planejamento e organização
6. Capacidade de se conectar com o negócio
7. Ser propositivo

Você irá apresentar o desenvolvimento deste desafio para a equipe da A3Data. Esta apresentação será realizada durante uma entrevista em um tempo máximo de 45 minutos (30 minutos apresentação + 15 para perguntas).