#### **Introdução ao Desafio**

Estamos em busca de um Cientista de Dados Pleno apaixonado por tecnologia de ponta e pela criação de soluções de IA que impactem diretamente o negócio da empresa. Este desafio técnico é projetado para avaliar suas habilidades em planejamento de solução de problemas, análise de dados, modelagem preditiva e interpretação de resultados.

Considere o desafio da competição "LMSys Chatbot Arena" no Kaggle (acessivel em <https://www.kaggle.com/competitions/lmsys-chatbot-arena/overview> ), que contém prompts e respostas geradas por dois modelos distintos. A proposta desse desafio do Kaggle é criar uma solução que possa inferir qual dos dois modelos gerou a resposta de melhor qualidade, considerando que o empate é uma opção válida.

Seu objetivo será **propor e defender uma solução** **viável** de ser implementada para solucionar o desafio. **Não será obrigatório a implementação da mesma**, mas lembre-se que você deverá explicar seu funcionamento e ter justificativas para cada decisão técnica. Durante a entrevista, iremos conversar sobre a solução proposta.

#### **Detalhes**

* **Base de Dados**: Ao se inscrever no desafio, o Kaggle irá liberar acesso a um conjunto de dados de treino e de teste com prompts e as respostas de dois modelos diferentes para cada prompt.
* **Tarefa do desafio**: Propor, documentar e defender uma solução que possa determinar qual resposta é melhor, ou se há um empate, com base em critérios definidos por você. A forma de documentação e apresentação é livre, e a defesa ocorrerá durante a entrevista.
* **Critérios de Avaliação**:
  + Justificativa da escolha do modelo e das técnicas utilizadas.
  + Clareza e eficácia da explicação dos resultados e do processo.
  + Complexidade da solução.
  + Discussão de prós e contras da solução e alternativas.

#### **Dicas**

* **Geral:** Você não terá muito tempo para ficar pensando nas coisas, então saiba priorizar.
* **Sobre Kaggle**: Se você não conhece o Kaggle, esse pode ser um bom lugar para se familiarizar com as competições (<https://www.kaggle.com/docs/competitions>), e se tiver interesse, aproveite os códigos públicos para se inspirar.
* **Feature Engineering**: Conheça os dados e considere criar novas features a partir dos dados brutos que possam ajudar o modelo a fazer previsões mais precisas.
* **Modelagem**: Fique a vontade para considerar diferentes algoritmos e técnicas.
* **Material externo**: Sinta se à vontade para utilizar e aproveitar de materiais e ferramentas externas como ChatGPT, Copilot e afins.

#### **Considerações Finais**

Durante a entrevista, pediremos que você apresente a sua solução para o problema. Ao final da nossa conversa, pediremos que você submeta o material que tiver construído para a Blip, por email.

1. Uma breve apresentação no seu formato favorito (Slides, Diagrama, etc) apresentando um desenho de uma possível solução.
   1. Não precisa ser algo formal, mas queremos entender sua linha de raciocínio.
2. [**Opcional**] - O link com o código completo submetido publicamente de forma similar ao exemplo <https://www.kaggle.com/code/awsaf49/lmsys-kerasnlp-starter> .
   1. Atenção! Compartilhamento de códigos de competições ativas no Kaggle devem ser públicas

Este desafio não é apenas uma oportunidade de mostrar suas habilidades técnicas, mas também sua capacidade de comunicar resultados complexos de forma clara e eficiente. Boa sorte!

Se tiver dúvidas durante o desafio, não hesite em entrar em contato. Estamos ansiosos para ver sua abordagem e solução para este problema!