

Desafio Oncase

Data Scientist

Quem somos?

Somos uma empresa de tecnologia movida por pessoas apaixonadas e cheias de propósito. Acreditamos no poder transformacional do uso de dados e Inteligência Artificial para um mundo melhor. Por isso, apoiamos pessoas e organizações a solucionar desafios ousados para a transformação digital com uma cultura orientada por dados. Nos empenhamos para que a Oncase seja um ambiente colaborativo, focado e criativo que incentiva a evolução e valoriza as habilidades de toda a equipe. Também trabalhamos pelo equilíbrio entre vida profissional e pessoal.

Etapa do processo seletivo:

Estamos muito animados em ter você participando do processo seletivo para se juntar ao time de Data Scientist da Oncase!

Para você entender um pouquinho da sua trajetória com a gente e quais são as próximas etapas, fizemos esse esqueminha abaixo.

No dia previamente agendado, você irá apresentar seu desafio para a equipe.



Inscrição
inicial



Entrevista com
o time de Pessoas



Desafio
Oncase

Você está aqui!



Entrevista com
gestor e/ou equipe



Carta
proposta

Desafio - Data Scientist

Ficamos felizes em saber que você chegou a esta fase e pelo interesse na Oncase. Neste passo, gostaríamos de conhecer um pouco mais sobre você através das suas habilidades em Ciência de Dados. Para tal, avaliamos os pontos a seguir, mas salientamos que os quesitos não estão, necessariamente, em ordem de prioridade:

- Sua capacidade na compreensão e modelagem dos dados e problemas;
- Qualidade e proficiência codificado na linguagem de sua escolha (R, Python);
- Qualidade das soluções encontradas para responder às perguntas;
- Comunicação na hora de questionar itens que não estejam claros;
- Comunicação visual eficaz (escrita e gráficos) no embasamento as respostas;
- Utilização de métodos adequados para cada problema;
- Análise de estudo e performance quando a resposta envolver a criação de modelos.

A avaliação será composta por duas macro tarefas, por favor não se limite a estes pontos, mas sinta-se livre para explorar e pôr em prática toda sua criatividade para solucionar os problemas.

1. Análise descritiva dos dados (EDA)

- Construa uma análise descritiva extraindo conhecimento das variáveis e apresentando quais insights podem ser obtidos a partir delas;
- Construa graficamente um storytelling a partir das variáveis apresentadas no problema;
- Descreva o caminho escolhido para sua EDA;
- Descreva quais outras técnicas poderiam ser aplicadas e porquê você não as escolheu;
- Utilize os dados: eda_data.zip

2. Teste técnico

Nessa parte, será necessário implementar um algoritmo de acordo com o paradigma do problema. Cada problema tem um conjunto de métricas que são requeridas. A variável alvo sempre será a coluna de nome "target", exceto para os problemas não-supervisionados.

a. Construa um classificador e identifique quais variáveis exercem maior impacto sobre o “target” e informe o porquê interpretando os resultados obtidos.

i. métricas: precision, recall e F2-score;

ii.dados: classification_data.zip;

b. Sugira o número de clusters ótimo a partir da análise dos dados.

i. métricas: elbow method, silhouette score, calinski-harabasz e davies-bouldin;

ii. dados: clusterization_data.zip;

c. Construa um regressor e identifique quais variáveis exercem maior impacto sobre o “target” e informe o porquê interpretando os resultados obtidos.

i. métricas: RMSE, R2 e $\text{cor}(\text{target_observado}, \text{target_predito})$;

ii. dados : regression_data.zip.

Disponibilização

Submeta seu projeto compartilhando um repositório (github, bitbucket, gitlab);
Envie o link de acesso por email.

Boa sorte e, qualquer dúvida ou sugestão, conta com a gente!

E-mail: giovanna.ribeiro@oncase.com.br

Whatsapp: (13) 98125-1428