

Desafio técnico para processo seletivo

Vaga: Analista de dados Pleno com foco em Machine Learning

Descrição: Construção de um modelo de machine learning para recomendação personalizada de estabelecimentos em no app. O modelo será utilizado para ordenar os estabelecimentos na tela inicial do app com base na probabilidade de interesse dos usuários.

Introdução

Objetivo do desafio: Avaliar a capacidade do candidato em desenvolver um modelo de machine learning, considerando o ciclo de modelagem e implementação na nuvem.

Duração estimada: A tarefa está planejada para ser concluída em até 7 dias corridos.

Escopo da tarefa

Desenvolvimento do modelo

- **Descrição:** Construir um modelo de machine learning que estime ordene a lista de estabelecimentos exibidos no app para cada usuário. O modelo poderá considerar dados de comportamento histórico de navegação, transações passadas, preferências do usuário e características dos estabelecimentos;
- **Dados Disponíveis:** Forneceremos um conjunto de dados fictício para simular as informações do app, incluindo detalhes dos usuários, categorias de estabelecimentos, histórico de interações e atributos dos locais;
- **Modelo:** Fica a critério do candidato escolher a abordagem de modelagem (classificação binária, probabilidade, ranking, etc.) e os algoritmos mais adequados para maximizar a acurácia na predição de interesse dos usuários.

Preparação do Modelo para Produção

- **Descrição:** Crie um fluxograma do que seria necessário para disponibilizar o modelo proposto em um ambiente de nuvem, com processamento em lote realizado 1 vez ao dia. Ficando a critério do candidato escolher os serviços necessários.

Entregáveis:

- Código do modelo final com estrutura clara e documentação (.py ou .ipynb);
- Justificativa da escolha do modelo final (PDF);
- Fluxograma para deployment em nuvem (PDF).

Ferramentas sugeridas para o fluxograma:

- Miro: <https://miro.com/pt/>
 - Lucidchart: <https://www.lucidchart.com/pages/pt>
-

Critérios de avaliação

Desenvolvimento e Qualidade do Modelo

- Clareza do código, uso eficiente de bibliotecas, organização e práticas de codificação;
- Escolha e justificativa do modelo de machine learning, bem como técnicas de feature engineering e tuning de hiperparâmetros.

Desempenho do Modelo

- Precisão, recall e outras métricas de avaliação relevantes para o problema proposto;
- Análise e interpretação dos resultados obtidos pelo modelo.

Preparação para Produção e Deploy

- Estrutura do fluxo para colocar o modelo em produção;
- Uso de melhores práticas para escalabilidade e manutenção do modelo na nuvem.

Documentação e Comunicação

- Documentação clara do processo, incluindo descrições de decisões tomadas;
- Organização das informações para fácil leitura e entendimento.