

ANA RAQUEL

POSTECH

DATA ANALYTICS

DATATHON

FASE 5

Datathon

Olá estudante!

Chegou o momento de colocar em prática todo o aprendizado adquirido ao longo do curso, desenvolvendo as habilidades trabalhadas até aqui para solucionar uma dor real de uma empresa. O tema central deste Datathon será a **aplicação de Inteligência Artificial em processos de recrutamento e seleção**, utilizando como base o estudo de caso da empresa Decision, que atua no setor de bodyshop, focada em alocar talentos ideais para os clientes de forma eficiente.

Contexto sobre a empresa:

A Decision é especializada em serviços de bodyshop e recrutamento, com foco em conectar talentos qualificados às necessidades específicas dos clientes. A Decision atua principalmente no setor de TI, onde a agilidade e a precisão no “match” entre candidatos e vagas são diferenciais essenciais. O objetivo da empresa é entregar profissionais que não apenas atendam aos requisitos técnicos, mas também se alinhem à cultura e aos valores das empresas contratantes.

A Decision desempenha um papel fundamental no processo de recrutamento ao combinar tecnologia e expertise humana para identificar os melhores talentos. O foco da empresa é garantir que cada decisão de contratação seja baseada em dados concretos e análises profundas.

Como funciona o processo atual da Decision?

O time de hunters busca candidatos utilizando a base interna da empresa e plataformas como o LinkedIn, InfoJobs, Catho, grupos de WhatsApp, entre outros. Um grande desafio da empresa é encontrar o candidato ideal para cada vaga em tempo hábil. Hoje a Decision possui algumas dores como:

- Falta de padronização em entrevistas, o que pode gerar perda de informações valiosas.

- Dificuldade em identificar o real engajamento dos candidatos.

Embora as entrevistas demandam tempo, elas são cruciais para entender o alinhamento do candidato com a vaga e o nível de comprometimento. Muitas vezes, hunters pulam essa etapa para agilizar o processo, o que pode comprometer a qualidade da seleção.

Fatores que tornam a entrevista parte essencial do processo de match:

- **Análise Técnica:** verificamos se o candidato possui as habilidades e conhecimentos exigidos pela vaga (ex.: linguagens de programação como Java, Python, etc.).
- **Fit Cultural:** avaliamos se o candidato se encaixa nos valores e cultura da empresa contratante.
- **Engajamento e Motivação:** identificamos se o candidato está realmente interessado na vaga e motivado para assumir o desafio.

Você como cientista de dados tem o desafio de propor uma inteligência artificial para solucionar as principais dores da Decision, criando um MVP da solução. O objetivo é desenvolver soluções criativas e viáveis usando IA para melhorar o processo de recrutamento. Você pode pensar em alguns cases, como por exemplo:

- Melhoria de CV com IA;
- Vaga otimizada com IA;
- Bot de entrevista (escrita/falada) com IA;
- Entrevista no app por vídeo com IA;
- Análise pós-entrevista com IA;
- Agendamento de entrevista com IA;
- Provas de vaga integrada no aplicativo com IA.

Essas são algumas opções propostas pela Decision, mas você pode surpreender na solução e utilizar a criatividade para resolver esse problema. Utilize os conhecimentos obtidos ao longo do curso, tal como Machine Learning e Deep Learning para fazer este desafio!

Base de dados: [Base de dados Decision](#)

Como os arquivos se relacionam?

- Cada Vaga em vagas.json tem um conjunto de candidatos inscritos em prospects.json.
- Cada candidato listado em prospects.json tem informações detalhadas que estão armazenadas em applicants.json.

Exemplo:

Para a vaga XPTO em vagas.json, temos:

- Primeiro a pessoa se cadastra no sistema por meio do applicants.json;
- Depois os candidatos que passaram (ou selecionados/avaliados) para alguma vaga, ficam em prospects.json, onde se relaciona com vagas.json;

PS: Você pode ver mais detalhes [aqui](#)

Requisitos da entrega

1 - Etapas do Pipeline de Machine Learning completa:

Pré-processamento dos Dados: limpeza de dados, tratamento de valores ausentes, encoding de variáveis categóricas, normalização etc.

Engenharia de Features: criação de novas variáveis relevantes, binning, agregações, interações, etc.

Treinamento e Validação: separe os dados para treino e validação do modelo e justifique a generalização do modelo na base de validação.

Seleção de Modelo: justificativa da escolha com base em métricas e interpretabilidade

2 - Streamlit

Código-Fonte Organizado: organize o código fonte no repositório do GitHub. Você pode seguir o seguinte exemplo:

app/

- app.py: Script principal com a interface Streamlit.
- pages/: Scripts adicionais para páginas múltiplas (se aplicável).

src/

- preprocessing.py: Funções de pré-processamento.
- feature_engineering.py: Transformações de variáveis.
- model_utils.py: Funções de carregamento e inferência do modelo.
- train.py: Treinamento e validação do modelo.
- evaluate.py: Avaliação de métricas.
- utils.py: Funções auxiliares reutilizáveis.

models/

- modelo.pkl ou modelo.joblib: Modelo serializado.
- Outros artefatos, se necessário.

notebooks/

- Notebooks de EDA, testes de modelo, validação etc.

tests/

- Testes unitários das funções (Pytest recomendado, mas não obrigatório).

Documentação técnica: adicione uma documentação técnica no README.md do projeto no GitHub. Incluindo:

- Descrição do projeto;
- Stack utilizada;
- Como rodar o app localmente;
- Instruções de instalação;
- Como treinar o modelo novamente.

Versão das bibliotecas: adicione um arquivo requirements.txt para adicionar as bibliotecas e versões utilizadas.

Serialização do modelo: adicione um arquivo .pkl ou .joblib com o modelo treinado.

Entrega:

1. Link do repositório do GitHub com o modelo e a aplicação do Streamlit.
2. Link da aplicação no streamlit já deployado.
3. Vídeo de até 5 (cinco) minutos no formato gerencial com a solução proposta do seu projeto.

Boa sorte!

The background is a dark blue gradient with a complex pattern of wavy, overlapping lines in shades of teal, yellow, and red. Scattered throughout are small, light blue dots. Several geometric shapes are visible: a thin vertical line on the left, a circle containing the number '7' in the upper center, a circle on the left, an 'X' and a '0' near the bottom left, and a hexagon in the bottom right.

POSTECH