

Desafio para consultor de Data Science

Objetivo

O objetivo deste desafio é avaliar as habilidades de um consultor de Data Science para avaliar as capacidades de realizar análises técnicas e resolução de problemas.

Descrição do desafio

Seu cliente possui uma plataforma hub de restaurante. Essa plataforma possui uma função para avaliação dos restaurantes via comentários pelos clientes. Seu trabalho é auxiliar nas estratégias de melhoria de satisfação dos clientes a fim de melhoria do serviço dos estabelecimentos conveniados. Você deve criar um algoritmo de classificação por meio de análise de sentimentos que será alocado no cliente para reduzir custos de infraestrutura. O cliente te disponibilizou apenas uma base de comentários das avaliações dos clientes que frequentaram os restaurantes

Instruções

1. Utilize a linguagem de programação python para desenvolvimento da solução
2. Crie uma coluna chamada “target” na base de treino* contendo o sentimento dos comentários por meio de uma LLM gratuita. Para acesso à LLM, siga os seguintes passos:
 - a. Acesse <https://openrouter.ai/> e crie uma conta
 - b. No seu perfil, clique em ‘keys’ e crie sua chave da API
 - c. Para utilizar a api, recomenda-se ver código [API_link](#) no anexo como exemplo.
3. Faça análise com modelagem de tópicos dos comentários de acordo com a classificação criada na coluna “target” na base de treinamento*.
4. Crie um algoritmo de Machine Learning para classificação dos comentários por análise de sentimento com a base de treinamento*.
5. Utilize o algoritmo do item 4 para predição da base de validação**.

*Base de treinamento: [dataset_train.csv](#)

**Base de validação: [dataset_valid.csv](#)

Entrega **(Prazo: 7 dias corridos pós recebimento)**

1. Código com o desenvolvimento da análise dos textos e algoritmo utilizado de acordo com as instruções
2. Predição do sentimento dos comentários da base de validação** com uma coluna ‘target’ por meio do algoritmo de classificação desenvolvido
3. Apresentação dos resultados para os times de Ciência de Dados e Arquitetura da Atento

Aviso Importante: O time de Ciência de Dados e Arquitetura da Atento possui vasta experiência com modelos de IA Generativa. Incentivamos o uso de tais ferramentas apenas como apoio/pesquisa. Caso o time conclua que foram utilizadas ferramentas de IA Generativa para elaboração total da solução, ou seja, com pouco ou nenhuma intervenção do candidato, a desclassificação do processo seletivo será imediata.

Critérios de avaliação

1. Capacidade de entendimento do problema
2. Capacidade analítica
3. Conhecimento de frameworks
4. Conhecimento de metodologias de tratamento e análise de dados
5. Qualidade e organização do código
6. Resultado da classificação
7. Apresentação e Explicação da solução

Desafio para consultor de Data Science

Anexo

API_link.py

```
import os
```

```
import json
```

```
import requests
```

```
from dotenv import load_dotenv
```

```
# --- Carregar variáveis de ambiente do arquivo .env ---
```

```
load_dotenv()
```

```
api_key = os.getenv("OPENROUTER_API_KEY") # Certifique-se de definir esta variável de ambiente
```

```
url = "https://openrouter.ai/api/v1/chat/completions"
```

```
headers = {
```

```
    "Authorization": f"Bearer {api_key}",
```

```
    "Content-Type": "application/json",
```

```
    "HTTP-Referer": "http://localhost", # Opcional, para rankings
```

```
    "X-Title": "MeuApp" # Opcional, para identificação
```

```
}
```

```
data = {
```

```
    "model": "meta-llama/llama-4-maverick:free",
```

```
    "messages": [{"role": "user", "content": "Explique computação quântica em termos simples"}],
```

```
    "max_tokens": 150
```

```
}
```

```
response = requests.post(url, headers=headers, data=json.dumps(data))
```

```
print(response.json()[0][0]["message"][0]["content"])
```