



EDITAL Nº 001/2024 – RESIDÊNCIA TI – DPE/RN
PROGRAMA DE RESIDÊNCIA EM TECNOLOGIA DA INFORMAÇÃO
DEFENSORIA PÚBLICA DO RIO GRANDE DO NORTE – DPE/RN

Prova de conhecimentos Específicos - BUSINESS INTELLIGENCE E ANALYTICS

1	C
2	C
3	A
4	D
5	C
6	ANULADA
7	A
8	A
9	B
10	D

11	C
12	A
13	D
14	D
15	C
16	C
17	D
18	A
19	A
20	D

21	C
22	B
23	A
24	B
25	A
26	D
27	D
28	A
29	B
30	A

Observação: Os candidatos receberam pontuação integral para as questões que foram anuladas.

Expectativa de Resposta Prova de conhecimentos Específicos (subjetiva)

1.



2.

a) Coleta e Preparação dos Dados

- Coleta de Dados: Utilizaria APIs das plataformas de mídia social (como Twitter, Facebook, Instagram) para coletar comentários mencionando palavras-chave relacionadas à Defensoria



Pública. Ferramentas como o Tweepy (para Twitter) podem ser usadas para automatizar essa coleta.

- Limpeza de Dados: Removeria ruídos dos dados coletados, como URLs, menções, hashtags e caracteres especiais, usando expressões regulares e bibliotecas de processamento de texto.
- Normalização: Converteria todos os textos para minúsculas para manter a consistência e usaria técnicas de tokenização e lematização para reduzir as palavras à sua forma base.
- Identificação de Linguagem: Utilizaria bibliotecas como a langdetect para filtrar comentários na língua desejada, considerando a diversidade linguística nas mídias sociais.

b) Técnicas de PLN para Análise de Dados

- Tokenização e Análise Morfológica: Dividiria os textos em palavras ou tokens, e aplicaria técnicas de análise morfológica para entender a função de cada palavra no contexto.
- Modelagem de Tópicos: Usaria técnicas como LDA (Latent Dirichlet Allocation) para identificar os principais tópicos discutidos nos comentários.
- Análise de Sentimentos: Implementaria modelos de PLN, como BERT ou LSTM, treinados em datasets específicos para identificar sentimentos, sarcasmo e linguagem figurada. Poderia utilizar bibliotecas como TextBlob ou VADER para uma análise inicial.
- Tratamento de Sarcasmo e Linguagem Figurada: Usaria modelos de PLN avançados que são especialmente treinados para identificar nuances como sarcasmo, utilizando datasets anotados para esse fim.

c) Lidando com Ambiguidade e Contexto Variável

- Contexto de Longo Alcance: Utilizaria modelos de linguagem que capturam contexto de longo alcance (como o Transformer) para entender melhor o contexto em que os comentários foram feitos.
- Análise de Co-ocorrência: Examinaria palavras que frequentemente aparecem juntas para entender melhor o contexto e a ambiguidade dos comentários.
- Feedback Humano: Incorporaria uma etapa de revisão humana para casos ambíguos, melhorando a precisão do sistema através de aprendizado ativo.

d) Métricas de Avaliação

- Precisão, Revocação e F1-Score: Avaliaria o sistema usando estas métricas padrão para entender a exatidão, a completude e o equilíbrio entre elas.
- Análise de Sentimentos Confusion Matrix: Usaria para visualizar o desempenho do modelo em classificar os sentimentos corretamente.
- Avaliação de Consistência Inter-Anotadores: Para garantir a confiabilidade das anotações manuais, compararia a consistência entre diferentes anotadores.
- Acompanhamento do Sentimento ao Longo do Tempo: Avaliaria como os sentimentos e as percepções mudam ao longo do tempo para ajustar e melhorar continuamente o sistema.

Natal-RN, 24 de fevereiro de 2024.

Profa. Dra. Iris Linhares Pimenta Gurgel

Coordenadora do Programa de Residência em Tecnologia da Informação do IMD-UFRN
Turma Defensoria Pública do RN