Lista 2 - Mineração de texto

1. **Defina o conceito de mineração de texto (*text mining*). Quais são as etapas básicas de um processo de mineração de texto e quais tipos de informações podem ser extraídas a partir de grandes volumes de dados textuais?**
   R:

2. **No contexto de mineração de texto, explique a importância do pré-processamento, incluindo tokenização, remoção de stopwords e stemming. Como essas etapas impactam a qualidade dos modelos de aprendizado de máquina para análise de sentimentos ou classificação de textos?**
   R:

3. **No contexto de KDT, explique a importância da representação de texto (por exemplo, bag-of-words, TF-IDF ou embeddings). Como a escolha da representação afeta a análise e a extração de conhecimento?**
   R:

4. **Dado o seguinte texto: "O produto chegou antes do prazo, ótimo serviço!", escreva uma função que:**
   a. Converta o texto para minúsculas;
   b. Remova pontuações;
   c. Tokenize o texto em palavras;
   d. Remova stopwords;
   e. Aplique stemming usando RSLPStemmer.

5. **Faça uma análise de sentimentos das seguintes listas:**
   a. Lista_de_palavras;
   b. Posts;
   c. Lista_de_comentários.

   Obs.: A análise poderá ser feita em código Python ou em um software especializado como o Knime.

6. **Desenvolver ou utilizar um webcrawler para recuperar dados de uma rede social ou fórum online. Os dados coletados devem incluir:**

   Obs.: Use Reddit, GitHub ou outra plataforma gratuita.

- Posts completos dos usuários.
- Comentários feitos sobre esses posts.
- **Pré-processar os textos coletados, aplicando etapas como:**
    - Remoção de caracteres especiais e pontuação.
    - Conversão para minúsculas.
    - Tokenização.
    - Remoção de stopwords.
    - Stemming ou lematização.
- **Realizar a análise de sentimentos nos textos coletados, classificando cada post e comentário como positivo, negativo ou neutro.**
    - Discutir os resultados, considerando:
    - A diferença entre analisar palavras isoladas versus textos completos.
    - A presença de sarcasmo, ironia ou ambiguidade nos textos.
    - Limitações do método utilizado.
- **Opcional: Apresente uma visualização das proporções de sentimentos (por exemplo, gráfico de barras ou pizza) para os dados coletados.**

- **Opcional: Faça a análise de similaridade com o seguinte texto:**

  “O clima está bom e iremos à praia!”