Que es minería de texto:

La minería de texto es un proceso de extracción de información útil y conocimiento significativo a partir de grandes cantidades de texto no estructurado.

Proceso de la minería de texto:

1. Data Collection: Gather text data from various sources.

2. Text Preprocessing: Normalize, tokenize, remove stopwords, correct spelling, and identify

parts of speech.

3. Feature Extraction: Represent text data using techniques like Bag of Words, TF-IDF, or

word embeddings.

4. Text Mining Algorithms: Apply clustering, classification, topic modeling, sentiment

analysis, or named entity recognition.

5. Model Evaluation and Validation: Assess the performance of algorithms using metrics like

accuracy and precision.

6. Visualization and Interpretation: Visualize results to gain insights and understand patterns

in the data.

7. Deployment and Integration: Implement text mining solutions for real-world applications.

Tokenizacion:

Tokenization is the process of breaking down text into smaller units, called tokens, to facilitate analysis and processing. This involves removing punctuation and special characters, and normalizing the text to ensure consistency.

letra por letra:

consiste en separar el texto en letra por letra como su nombre lo indica para sacar información detallada a un nivel muy alto

palabra por palabra:

la tokenizacion palabra por palabra se utiliza mas en un lenguaje natural ya que se entiende en el lenguaje básico humano

Método de frecuencia

contar la frecuencia de ocurrencia de palabras o términos específicos en un texto para sacar palabras clave

Método de sentimiento:

El método de sentimiento es en el cual se obtiene a través de la minería de texto si la palabra después de separarla es negativa, positiva, o neutra

JUEGO:

Oración: "El gato negro saltó sobre la cerca."

Instrucciones:

Lee la oración.

Separa la oración en palabra por palabra.

Escribe cada palabra en la secuencia correcta en la que aparecen en la oración original.

Ten en cuenta que la estructura debe ser la sig: [“hola”, “mundo”]