**Q.1.Explanation of Preprocessing Techniques**

1. **Text Cleaning:**
   * **What it does:**
     + Removes special characters and punctuation.
     + Converts text to lowercase for consistency.
   * **Impact:**
     + Standardizes the text, reducing variations
     + Makes data easier to analyze and improves model accuracy.
2. **Tokenization:**
   * **What it does:** Splits sentences into individual words or tokens.
   * **Impact:**
     + Enables granular analysis by breaking down text into smaller components.
     + Facilitates feature extraction for machine learning algorithms.
3. **Stopword Removal:**
   * **What it does:** Removes commonly used words like "is," "and," "the" that don't carry much context.
   * **Impact:**
     + Focuses on the meaningful parts of the text.
     + Reduces noise and improves the efficiency and accuracy of models by concentrating on relevant terms.

**Saving Processed Data**

* **Why save processed data?**
  + Preprocessing can be computationally expensive, especially for large datasets. Saving the processed data allows you to reuse it without repeating the steps.

**Impact of Preprocessing on Model Performance**

* **Improved Signal-to-Noise Ratio:** Cleaning removes irrelevant or noisy data.
* **Reduced Dimensionality:** Removing stopwords and focusing on meaningful tokens simplifies data representation.
* **Enhanced Generalization:** Standardization through cleaning and tokenization ensures that the model learns patterns rather than inconsistencies or artifacts.

This approach prepares the text data for further natural language processing or machine learning tasks, ensuring better results. Let me know if you'd like to proceed with additional steps, such as vectorization or applying a model!

Q3.

* Choose **Naive Bayes** for speed and simpler datasets.
* Choose **SVM** for more robust performance with complex, higher-dimensional datasets.