Seminar 7

Question 1:

Tokenize the following sentence: "Natural Language Processing is fascinating."

Answer 1:

Tokens: ["Natural", "Language", "Processing", "is", "fascinating", "."]

Question 2:

Remove stop words from the sentence: "The quick brown fox jumps over the lazy dog."

Answer 2:

Filtered sentence: "quick brown fox jumps lazy dog."

Question 3:

Lemmatize the words: "running," "better," "went," "driven."

Answer 3:

Lemmatized words: "run," "good," "go," "drive."

Question 4:

Assign parts of speech to the words in the sentence: "She walked to the store."

Answer 4:

- She (PRONOUN)

- walked (VERB)

- to (PREPOSITION)

- the (DETERMINER)

- store (NOUN)

Question 5:

In the sentence "Apple Inc. was founded by Steve Jobs in Cupertino," identify and label the named entities.

Answer 5:

- Apple Inc. (ORGANIZATION)

- Steve Jobs (PERSON)

- Cupertino (LOCATION)

Question 6:

Given a dataset of customer reviews, create a text classification model to classify reviews as positive or negative.

Review	Sentiment
"I love this product, it's fantastic!"	Positive
"The quality of this item is terrible."	Negative
"This is the best thing I've ever purchased!"	Positive
"I'm very disappointed with the service."	Negative
"The product arrived in perfect condition."	Positive
"This is a waste of money."	Negative

Answer 6:

In this manual example:

- We have a sample dataset of customer reviews, each labeled as "Positive" or "Negative" sentiment.
- We manually preprocess the text data, removing punctuation and converting it to lowercase.
- We manually extract features based on keyword matching. Positive reviews contain positive keywords, and negative reviews contain negative keywords.

positive_keywords = ["love", "fantastic", "best", "perfect"]

negative_keywords = ["terrible", "disappointed", "waste"]

- We evaluate the model's performance manually by comparing the predicted sentiments with the actual sentiments and calculating accuracy.

Question 7:

Create bigrams (2-grams) from the sentence: "Natural language processing is interesting."

Answer 7:

Bigrams: ["Natural language", "language processing", "processing is", "is interesting"]

Question 8:

Create a simplified three-dimensional representation of the words "king," "queen," and "man" in a vector space and explain their relationships in the vector space.

Instructions:

1. Represent the words "king," "queen," and "man" as vectors in a three-dimensional vector space.

2. Provide a brief explanation of the dimensions and how you've assigned values to each word.

3. Explain the relationships between these words in the vector space based on their meanings and similarities.

Answer:

Vector Representations:

In this simplified three-dimensional vector space, we'll represent the words "king," "queen," and "man" as follows:

- "king" represented as (4, 3, 6)

- "queen" represented as (4, 3, 5)

- "man" represented as (3, 1, 2)

Explanation:

- In this vector space, each dimension can be associated with certain semantic attributes. Here, we can consider three dimensions:

  - Dimension 1: Gender (higher values might represent more masculinity)

  - Dimension 2: Royalty (higher values might indicate higher royalty)

  - Dimension 3: Age (higher values might suggest older age)

Relationships in the Vector Space:

- "king" and "queen" are similar in the first two dimensions (gender and royalty) but differ in the third dimension (age). This is consistent with the fact that both "king" and "queen" have high values in dimensions related to royalty but differ in terms of gender and age.

- "king" and "man" are closer in dimensions 1 and 2, suggesting a similarity in terms of gender and royalty. However, they still have differences in the third dimension, representing differences in age.

- "queen" and "man" are closer in dimensions 1 and 3, indicating some similarity in terms of gender and age, but they differ in the second dimension (royalty).