1. Explain One-Hot Encoding

One-hot encoding in machine learning is the conversion of categorical information into a format that may be fed into machine learning algorithms to improve prediction accuracy. One-hot encoding is a common method for dealing with categorical data in machine learning.

2. Explain Bag of Words

The Bag-of-words model is an orderless document representation — only the counts of words matter. For instance, in the above example "John likes to watch movies. Mary likes movies too", the bag-of-words representation will not reveal that the verb "likes" always follows a person's name in this text.

3. Explain Bag of N-Grams

A bag-of-n-grams model records the number of times that each n-gram appears in each document of a collection. An n-gram is a collection of n successive words. bagOfNgrams does not split text into words. To create an array of tokenized documents, see tokenizedDocument .

4. Explain TF-IDF

TF-IDF (Term Frequency - Inverse Document Frequency) is a handy algorithm that uses the frequency of words to determine how relevant those words are to a given document. It's a relatively simple but intuitive approach to weighting words, allowing it to act as a great jumping off point for a variety of tasks.

5. What is OOV problem?

Out-of-vocabulary (OOV) are terms that are not part of the normal lexicon found in a natural language processing environment. In speech recognition, it's the audio signal that contains these terms. Word vectors are the mathematical equivalent of word meaning.

6. What are word embeddings?

Word Embeddings in NLP is a technique where individual words are represented as real-valued vectors in a lower-dimensional space and captures inter-word semantics. Each word is represented by a real-valued vector with tens or hundreds of dimensions.

7. Explain Continuous bag of words (CBOW)

The continuous bag-of-words (CBOW) model is a neural network for natural language processing tasks such as language translation and text classification. It is based on predicting a target word given the context of the surrounding words

8. Explain SkipGram

Skip-gram is one of the unsupervised learning techniques used to find the most related words for a given word. Skip-gram is used to predict the context word for a given target word. It's reverse of CBOW algorithm. Here, target word is input while context words are output.

9. Explain Glove Embeddings.

The basic idea behind the GloVe word embedding is to derive the relationship between the words from statistics. Unlike the occurrence matrix, the co-occurrence matrix tells you how often a particular word pair occurs together. Each value in the co-occurrence matrix represents a pair of words occurring together