1.Explain One-Hot Encoding

Ans **One-Hot Encoding** is a method used to convert categorical data (i.e., data that represents categories or labels) into a numerical format that can be used by machine learning algorithms.

2. Explain Bag of Words

Ans **Bag of Words (BoW)** is a simple and commonly used technique in natural language processing (NLP) and text mining for converting text into a numerical representation. In this model, a text (document, sentence, or set of words) is represented as a collection (or "bag") of its words, disregarding grammar and word order but keeping track of the frequency of each word.

3. Explain Bag of N-Grams

Ans **Bag of N-Grams** is an extension of the **Bag of Words (BoW)** model in natural language processing (NLP). It represents text data using **n-grams**, which are contiguous sequences of **n** items (words or characters) from the given text.

4. Explain TF-IDF

Ans **TF-IDF (Term Frequency-Inverse Document Frequency)** is a statistical measure used to evaluate the importance of a word in a document relative to a collection of documents (corpus)

5. What is OOV problem?

Ans The **Out-of-Vocabulary (OOV)** problem refers to the issue encountered in Natural Language Processing (NLP) when a model or system comes across words or terms that were not present in the training dataset or vocabulary.

6. What are word embeddings?

Ans **Word embeddings** are a type of word representation used in natural language processing (NLP) and machine learning

7. Explain Continuous bag of words (CBOW)

Ans **Continuous Bag of Words (CBOW)** is one of the two main architectures used in the **Word2Vec** algorithm, a technique for learning word embeddings from a corpus of text

8. Explain SkipGram

Ans **Skip-gram** is another model used in the **Word2Vec** algorithm, and it is the opposite of the **Continuous Bag of Words (CBOW)** model.

9. Explain Glove Embeddings.

Ans **GloVe (Global Vectors for Word Representation)** is an unsupervised machine learning algorithm for generating word embeddings, which represents words in a dense vector space based on their co-occurrence patterns in a given corpus of text