1. Explain One-Hot Encoding

**Ans : One hot encoding is one method of converting data to prepare it for an algorithm and get a better prediction. With one-hot, we convert each categorical value into a new categorical column and assign a binary value of 1 or 0 to those columns. Each integer value is represented as a binary vector.**

1. Explain Bag of Words

**Ans : A bag-of-words model, or BoW for short, is a way of extracting features from text for use in modeling, such as with machine learning algorithms. A bag-of-words is a representation of text that describes the occurrence of words within a document. It involves two things: A vocabulary of known words.**

1. Explain Bag of N-Grams

**Ans : A bag-of-n -grams model is a way to represent a document, similar to a [bag-of-words][/terms/bag-of-words/] model. A bag-of-n -grams model represents a text document as an unordered collection of its n -grams.**

1. Explain TF-IDF

**Ans : TF-IDF is a popular approach used to weigh terms for NLP tasks because it assigns a value to a term according to its importance in a document scaled by its importance across all documents in your corpus, which mathematically eliminates naturally occurring words in the English language, and selects words that are more**

1. What is OOV problem?

**Ans : These words that are unknown by the models, known as out-of-vocabulary (OOV) words, need to be properly handled to not degrade the quality of the natural language processing (NLP) applications, which depend on the appropriate vector representation of the texts.**

1. What are word embeddings?

**Ans : A word embedding is a learned representation for text where words that have the same meaning have a similar representation. It is this approach to representing words and documents that may be considered one of the key breakthroughs of deep learning on challenging natural language processing problems.**

1. Explain Continuous bag of words (CBOW)

**Ans : In the CBOW model, the distributed representations of context (or surrounding words) are combined to predict the word in the middle .**

1. Explain SkipGram

**Ans : SkipGram is an algorithm that is used to create word embeddings i.e. high-dimensional vector representation of words. These embeddings are meant to encode the semantic meaning of words such that words that are semantically similar will lie close to each other in that vector's space.**

1. Explain Glove Embeddings.

**Ans : GloVe stands for global vectors for word representation. It is an unsupervised learning algorithm developed by Stanford for generating word embeddings by aggregating global word-word co-occurrence matrix from a corpus. The resulting embeddings show interesting linear substructures of the word in vector space.**