1. 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. Bag-of-Words (BoW) is not a word embedding technique; it's a simple feature extraction method that counts the occurrence of words within a document. It uses a representation of text that is based on an unordered collection (a "bag") of words. It is used in natural language processing and information retrieval (IR).
3. In natural language processing (NLP), the Bag of N-Grams Model is a method for representing text input in an organized way that machine learning algorithms may exploit. An N-gram consists of a continuous series of 'N' elements from a specific voice or text sample.
4. TF-IDF stands for Term Frequency Inverse Document Frequency of records. It can be defined as the calculation of how relevant a word in a series or corpus is to a text. The meaning increases proportionally to the number of times in the text a word appears but is compensated by the word frequency
5. Out-of-vocabulary (OOV) are terms that are not part of the normal lexicon found in a natural language processing environment.
6. Word embeddings are used to convert words or phrases into numerical representations in NLP by representing words as dense vectors in a high-dimensional space, capturing semantic and syntactic relationships between words.
7. CBOW is a neural network-based algorithm that predicts a target word given its surrounding context words. It is a type of “unsupervised” learning.
8. Skip-gram is a natural language processing (NLP) algorithm that creates word embeddings by predicting the words around a specific target word.
9. GloVe Embeddings are a type of word embedding that encode the co-occurrence probability ratio between two words as vector differences.