- What is word Embeding?write the process how to do embeding.

Ans:-

➔ Word embeddings are a type of representation for text where words or phrases are mapped to vectors of real numbers.

➔ These vectors capture semantic meanings of words by placing similar words closer together in the vector space

Process of Word Embedding:

1. **Data Preparation:**
   - Collect and preprocess the text data.
   - Tokenize the text into words or phrases.
2. **Choose an Embedding Technique:**
   - **Pre-trained Embeddings:** Use embeddings from models like Word2Vec, GloVe, or FastText.
   - **Train Embeddings from Scratch:** Use a method like Word2Vec or GloVe to learn embeddings from your data.
3. **Training (if from scratch):**
   - Define the architecture (e.g., Skip-gram or CBOW for Word2Vec).
   - Feed the preprocessed text into the model and train it to adjust embeddings based on word contexts.
4. **Use Embeddings:**
   - Convert words into their corresponding vectors.
   - Use these vectors as input features for machine learning models or for other natural language processing (NLP) tasks.

- Explin about Tokanization

Ans:-

What is Tokenization? Tokenization is the process of splitting text into smaller units called tokens, which can be words, phrases, or subwords. It is a fundamental step in text preprocessing for NLP tasks.

Types of Tokenization:

1. **Word Tokenization:** Splits text into individual words (e.g., "ChatGPT is amazing!" → ["ChatGPT", "is", "amazing"]).
2. **Subword Tokenization:** Splits text into subword units or characters (e.g., "unhappiness" → ["un", "happiness"]).

3. **Sentence Tokenization:** Splits text into sentences (e.g., "Hello! How are you?" → ["Hello!", "How are you?"]).

Process of Tokenization:

1. **Select Tokenization Strategy:** Based on the NLP task, choose whether to tokenize by word, subword, or sentence.
2. **Apply Tokenizer:** Use a tokenizer tool or library (e.g., NLTK, spaCy, or Hugging Face's tokenizers) to convert text into tokens.
3. **Post-processing:** Clean and adjust tokens as needed (e.g., removing punctuation or lowercasing).

- For what purpose we use tranformer libray.

Ans:-

Transformer Library

Purpose of Transformer Library:

➔ Transformer libraries are used for building and training models based on the Transformer architecture.

➔ These libraries provide tools to handle various NLP tasks such as text classification, translation, and generation.

Common Uses:

1. **Model Implementation:** Create and train Transformer-based models like BERT, GPT, or T5.
2. **Pre-trained Models:** Access and fine-tune pre-trained models for specific tasks.
3. **Text Processing:** Use built-in tools for tokenization, encoding, and decoding text data.
4. **Transfer Learning:** Leverage pre-trained models to improve performance on NLP tasks with limited data.

Popular Libraries:

1. **Hugging Face Transformers:** A comprehensive library for Transformer models and tools.
2. **TensorFlow and PyTorch:** Frameworks that offer support for implementing Transformer models.