Natural language processing

Text processing:

Text Processing pertains to the analysis of text data using a programming language such as Python. Text Processing is an essential task in NLP as it helps to clean and transform raw data into a suitable format used for analysis or modeling.Unstructured text data can be automatically analyzed and sorted by text processing to obtain useful information.

Text processing methods:

1.Word frequency:

This statistical method accurately determines the most frequently used words or expressions in a particular section of text. With this specific insight, you can address problematic situations, identify areas of success, and more.

2.Collocation:

This method helps identify co-occurring words – meaning they commonly occur together. The most frequent kinds of collocations in text are bigrams (two adjacent words) and trigrams (three adjacent words). For example, keeping in touch or launching a product are standard connections.

3. Concordance:

By examining how particular words are employed in various settings, concordances effectively help to decipher the ambiguity of human language. The term "problem," for instance, can refer to a number of situations, including an issue, a situation, a topic, or the process of supplying something.

4. TF-IDF:

TF-IDF stands for Inverse Document Frequency. This metric measures how important a word is to a document but is offset by the number of documents that contain the word.

IDF is calculated as follows where ***t*** is the term (word) we are looking to measure the commonness of and ***N*** is the number of documents (d) in the corpus (D).. The denominator is simply the number of documents in which the term, ***t***, appears in.so tf-idf(t, d) = tf(t, d) * idf(t).

5. Text Summarization

Text summarization is the practice of applying natural language processing to reduce complex technical, scientific, or other jargon to its most straightforward components.

6. Text classification:

Again, text classification organizes large amounts of unstructured text (meaning the raw text data you receive from your customers). Text classification includes several subdivisions, including topic modeling, sentiment analysis, and keyword extraction (which we'll discuss next). Text classification takes your text dataset and then structures it for further analysis. It is often used to extract valuable data from customer reviews and customer service logs.

7. Keyword extraction:

The last key to the text analysis puzzle, keyword extraction, is a broader form of the techniques we've already discussed. The most pertinent information from text is automatically extracted using machine learning and artificial intelligence (AI) techniques.

8. Lemmatization and stemming:

Stemming is the process of getting the root form of a word. Stem or root is the part to which inflectional affixes (-ed, -ize, -de, -s, etc.) are added. The stem of a word is created by removing the prefix or suffix of a word. So, stemming a word may not result in actual words.Example:looked ---> look,denied ---> deni.

Like stemming, lemmatization also converts a word to its root form. The only difference is that lemmatization ensures that the root word belongs to the language. We will get valid words if we use lemmatization. In NLTK, we use the WordNetLemmatizer to get the lemmas of words. We also need to provide a context for the lemmatization. So, we add the part-of-speech as a parameter.Example:Looked->look,denied->deny.

Deep learning

Deep learning is a method in artificial intelligence (AI) that teaches computers to process data in a way that is inspired by the human brain. Deep learning models can recognize complex patterns in pictures, text, sounds, and other data to produce accurate insights and predictions. You can use deep learning methods to automate tasks that typically require human intelligence, such as describing images or transcribing a sound file into text.

Uses of deep learning:

- Self-driving cars use deep learning models to automatically detect road signs and pedestrians.

- Defense systems use deep learning to automatically flag areas of interest in satellite images.

- You can group these various use cases of deep learning into four broad categories—computer vision, speech recognition, natural language processing (NLP), and recommendation engines.

How does deep learning works:

Deep learning algorithms are neural networks that are modeled after the human brain. For example, a human brain contains millions of interconnected neurons that work together to learn and process information. Similarly, deep learning neural networks, or artificial neural networks, are made of many layers of artificial neurons that work together inside the computer.

Components of deep learning:

Input layer-An artificial neural network has several nodes that input data into it. These nodes make up the input layer of the system

Hidden layer-The input layer processes and passes the data to layers further in the neural network. These hidden layers process information at different levels, adapting their behavior as they receive new information. Deep learning networks have hundreds of hidden layers that they can use to analyze a problem from several different angles.For example, if you were given an image of an unknown animal that you had to classify, you would compare it with animals you already know. For example, you would look at the shape of its eyes and ears, its size, the number of legs, and its fur pattern. You would try to identify patterns, such as the following:

- The animal has hooves, so it could be a cow or deer.
- The animal has cat eyes, so it could be some type of wild cat.

Output layer.-The output layer consists of the nodes that output the data. Deep learning models that output "yes" or "no" answers have only two nodes in the output layer. On the other hand, those that output a wider range of answers have more nodes.

Benefits of deep learning:

Efficient processing of unstructured data :

Machine learning methods find unstructured data, such as text documents, challenging to process because the training dataset can have infinite variations. On the other hand, deep learning models can comprehend unstructured data and make general observations without manual feature extraction. For instance, a neural network can recognize that these two different input sentences have the same meaning.

Hidden relationships and pattern discovery

A deep learning application can analyze large amounts of data more deeply and reveal new insights for which it might not have been trained. For example, consider a deep learning model that is trained to analyze consumer purchases. The model has data only for the items you have already purchased. However, the artificial neural network can suggest new items that you haven't bought by comparing your buying patterns to those of other similar customers.

Unsupervised learning

Deep learning models can learn and improve over time based on user behavior. They do not require large variations of labeled datasets. For example, consider a neural network that automatically corrects or suggests words by analyzing your typing behavior.