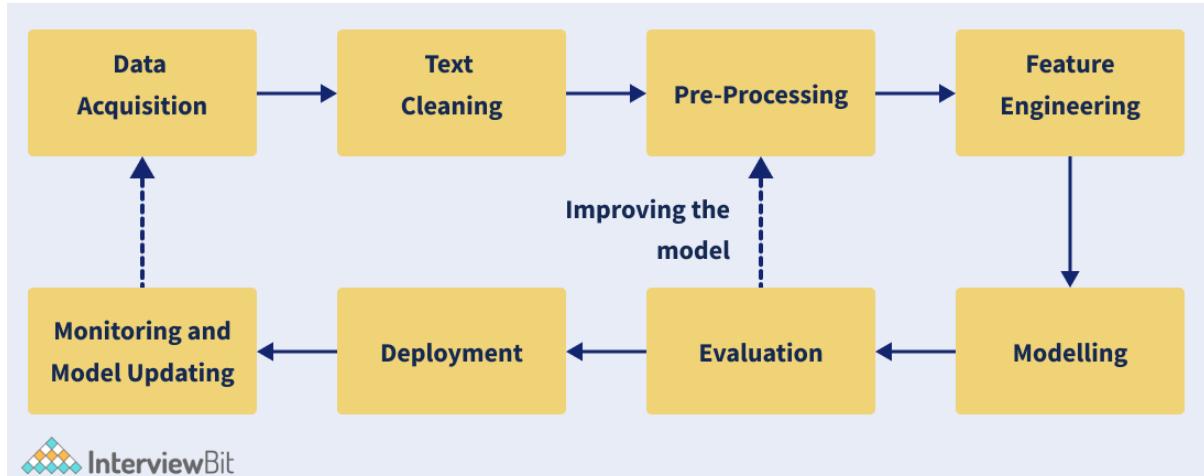


## 1. What are the stages in the lifecycle of a natural language processing (NLP) project?

Following are the stages in the lifecycle of a natural language processing (NLP) project:

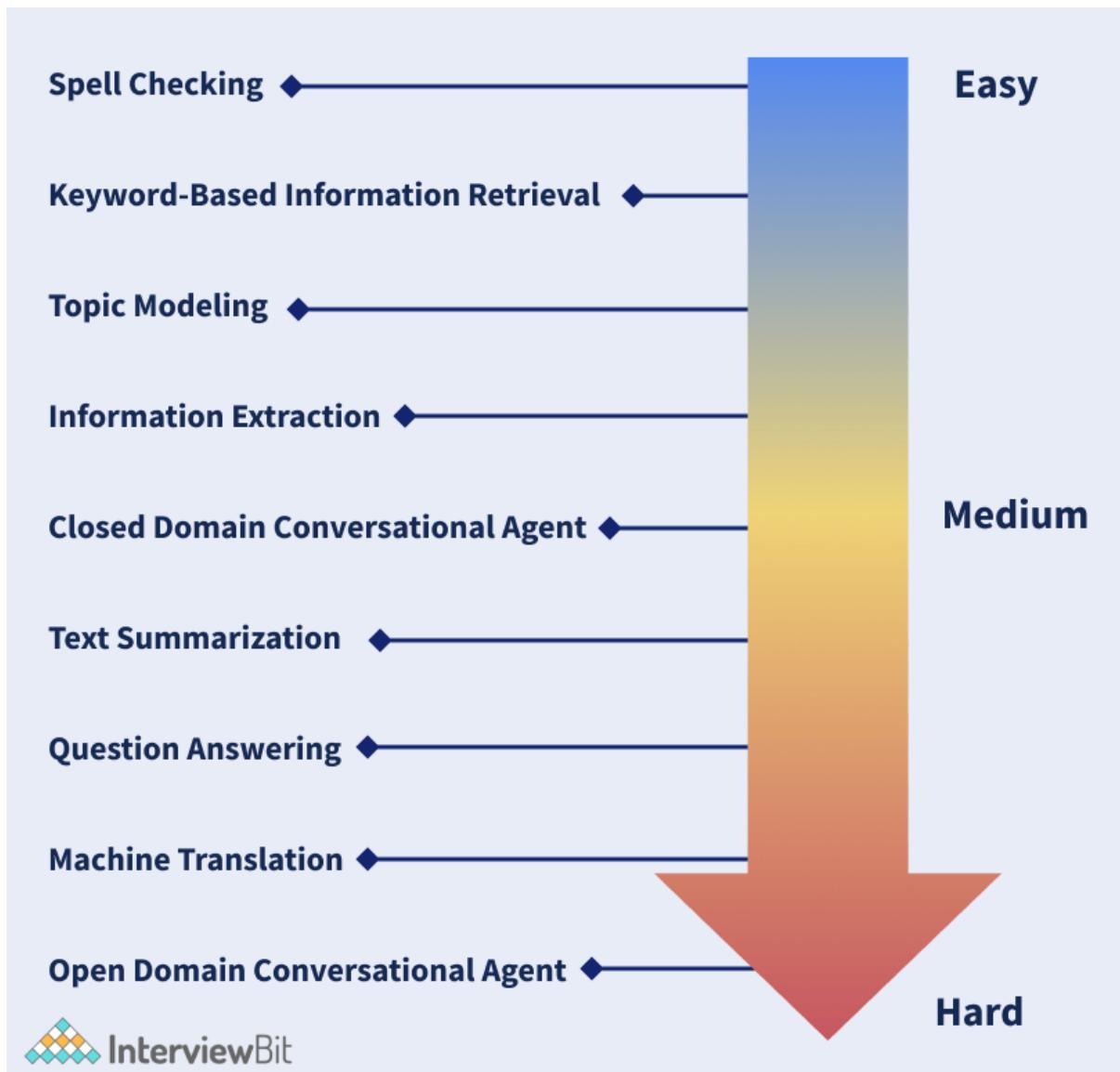


- **Data Collection:** The procedure of collecting, measuring, and evaluating correct insights for research using established approved procedures is referred to as data collection.
- **Data Cleaning:** The practice of correcting or deleting incorrect, corrupted, improperly formatted, duplicate, or incomplete data from a dataset is known as data cleaning.
- **Data Pre-Processing:** The process of converting raw data into a comprehensible format is known as data preparation.
- **Feature Engineering:** Feature engineering is the process of extracting features (characteristics, qualities, and attributes) from raw data using domain expertise.
- **Data Modeling:** The practice of examining data objects and their relationships with other things is known as data modelling. It's utilised to look into the data requirements for various business activities.
- **Model Evaluation:** Model evaluation is an important step in the creation of a model. It aids in the selection of the best model to represent our data and the prediction of how well the chosen model will perform in the future.
- **Model Deployment:** The technical task of exposing an ML model to real-world use is known as model deployment.
- **Monitoring and Updating:** The activity of measuring and analysing production model performance to ensure acceptable quality as defined by the use case is known as machine learning monitoring. It delivers alerts about performance difficulties and assists in diagnosing and resolving the core cause.

## 2. What are some of the common NLP tasks?

Some of the common tasks of NLP include:

- **Machine Translation:** This helps in translating a given piece of text from one language to another.
- **Text Summarization:** Based on a large corpus, this is used to give a short summary that gives an idea of the entire text in the document.
- **Language Modeling:** Based on the history of previous words, this helps uncover what the further sentence will look like. A good example of this is the auto-complete sentences feature in Gmail.
- **Topic Modelling:** This helps uncover the topical structure of a large collection of documents. This indicates what topic a piece of text is actually about.
- **Question Answering:** This helps prepare answers automatically based on a corpus of text, and on a question that is posed.
- **Conversational Agent:** These are basically voice assistants that we commonly see such as Alexa, Siri, Google Assistant, Cortana, etc.
- **Information Retrieval:** This helps in fetching relevant documents based on a user's search query.
- **Information Extraction:** This is the task of extracting relevant pieces of information from a given text, such as calendar events from emails.
- **Text Classification:** This is used to create a bucket of categories of a given text, based on its content. This is used in a wide variety of AI-based applications such as sentiment analysis and spam detection.



### 3. What are the different approaches used to solve NLP problems?

There are multiple approaches to solving NLP problems. These usually come in 3 categories:

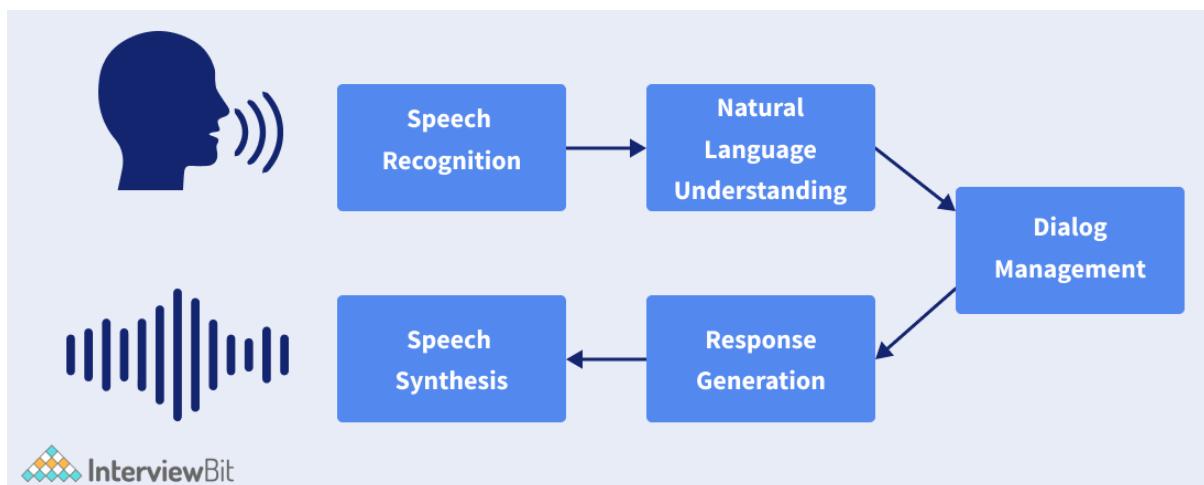
- Heuristics
- Machine learning
- Deep Learning

### 4. How do Conversational Agents work?

The following NLP components are used in Conversational Agents:

- **Speech Recognition and Synthesis:** In the first stage, speech recognition helps convert speech signals to their phonemes, and are then transcribed as words.

- **Natural Language Understanding (NLU):** Here, the transcribed text from stage one is further analysed through AI techniques within the natural language understanding system. Certain NLP tasks such as Named Entity Recognition, Text Classification, Language modelling, etc. come into play here.
- **Dialog Management:** Once the needed information from text is extracted, we move on to the stage of understanding the user's intent. The user's response can then be classified by using a text classification system as a pre-defined intent. This helps the conversational agent in figuring out what is actually being asked.
- **Generating Response:** Based on the above stages, the agent generates an appropriate response that is based on a semantic interpretation of the user's intent.



## 5. What is meant by data augmentation? What are some of the ways in which data augmentation can be done in NLP projects?

NLP has some methods through which we can take a small dataset and use that in order to create more data. This is called data augmentation. In this, we use language properties to create text that is syntactically similar to the source text data.

Some of the ways in which data augmentation can be done in NLP projects are as follows:

- Replacing entities
- TF-IDF-based word replacement
- Adding noise to data
- Back translation
- Synonym replacement
- Bigram flipping

## **6. How can data be obtained for NLP projects?**

There are multiple ways in which data can be obtained for NLP projects. Some of them are as follows:

- **Using publicly available datasets:** Datasets for NLP purposes are available on websites like Kaggle as well as Google Datasets.
- **By using data augmentation:** These are used to create additional datasets from existing datasets.
- **Scraping data from the web:** Using coding in Python or other languages once can scrape data from websites that are usually not readily available in a structured form.

## **7. What do you mean by Text Extraction and Cleanup?**

The process of extracting raw text from the input data by getting rid of all the other non-textual information, such as markup, metadata, etc., and converting the text to the required encoding format is called **text extraction and cleanup**. Usually, this depends on the format of available data for the required project.

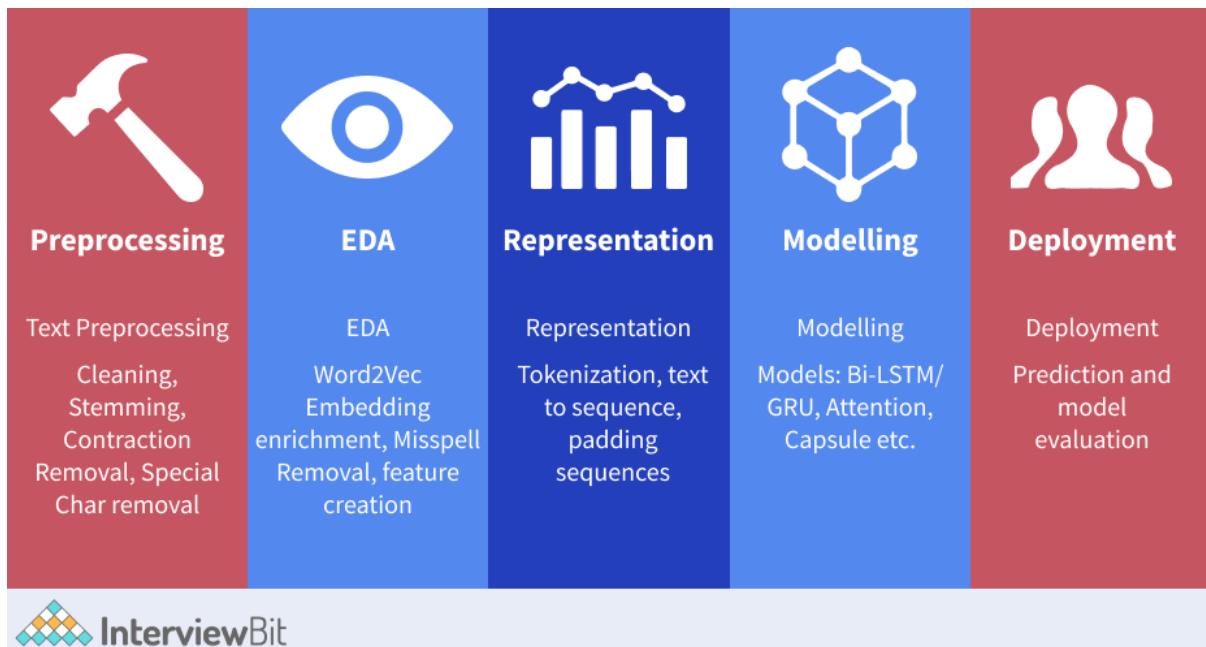
Following are the common ways used for Text Extraction in NLP:

- Named Entity Recognition
- Sentiment Analysis
- Text Summarization
- Aspect Mining
- Topic Modeling

## **8. What are the steps involved in preprocessing data for NLP?**

Here are some common pre-processing steps used in NLP software:

- **Preliminaries:** This includes word tokenization and sentence segmentation.
- **Common Steps:** Stop word removal, stemming and lemmatization, removing digits/punctuation, lowercasing, etc.
- **Processing Steps:** Code mixing, normalization, language detection, transliteration, etc.
- **Advanced Processing:** Parts of Speech (POS) tagging, coreference resolution, parsing, etc.



## 9. What do you mean by Stemming in NLP?

When we remove the suffixes from a word so that the word is reduced to its base form, this process is called stemming. When the word is reduced to its base form, all the different variants of that word can be represented by the same form (e.g., “bird” and “birds” are both reduced to “bird”).

We can do this by using a fixed set of rules. For instance: if a word ends in “-es,” we can remove the “-es”).

Even though these rules might not really make sense as a linguistically correct base form, stemming is usually carried out to match user queries in search engines to relevant documents. And in text classification, is done to reduce the feature space to train our machine learning (ML) models.

The code snippet given below depicts the way to use a well known NLP algorithm for stemming called Porter Stemmer using NLTK:

```
from nltk.stem.porter import PorterStemmer
stemmer = PorterStemmer()
word1, word2 = "bikes", "revolution"
print(stemmer.stem(word1), stemmer.stem(word2))
```

This gives “bike” as the stemmed version for “bikes,” but “revolut” as the stemmed form of “revolution,” even though the latter is not linguistically correct. Even if this might not affect the performance of the search engine, a derivation of the correct linguistic form becomes useful in some other cases.

This can be done by another process that is closer to stemming, known as lemmatization.

## **10. What do you mean by Lemmatization in NLP?**

The method of mapping all the various forms of a word to its base word (also called “lemma”) is known as Lemmatization. Although this may appear close to the definition of stemming, these are actually different. For instance, the word “better,” after stemming, remains the same. However, upon lemmatization, this should become “good.” Lemmatization needs greater linguistic knowledge. Modelling and developing efficient lemmatizers still remains an open problem in NLP research.

The application of a lemmatizer based on WordNet from NLTK is shown in the code snippet below:

```
from nltk.stem import WordNetLemmatizer  
lemmatizer = WordnetLemmatizer()  
print(lemmatizer.lemmatize("better", pos="a")) #a is for adjective
```

## **NLP Interview Questions for Experienced**

### **11. What is the meaning of Text Normalization in NLP?**

Consider a situation in which we’re operating with a set of social media posts to find information events. Social media textual content may be very exceptional from the language we’d see in, say, newspapers. A phrase may be spelt in multiple ways, such as in shortened forms, (for instance, with and without hyphens), names are usually in lowercase, and so on. When we’re developing NLP tools to work with such kinds of data, it’s beneficial to attain a canonical representation of textual content that captures these kinds of variations into one representation. This is referred to as text normalization.

Converting all text to lowercase or uppercase, converting digits to text (e.g., 7 to seven), expanding abbreviations, and so on are some frequent text normalisation stages.

### **12. Explain the concept of Feature Engineering.**

After a variety of pre-processing procedures and their applications, we need a way to input the pre-processed text into an NLP algorithm later when we employ ML methods to complete our modelling step. The set of strategies that will achieve this goal is referred to as feature engineering. Feature extraction is another name for it. The purpose of feature engineering is to convert the text’s

qualities into a numeric vector that NLP algorithms can understand. This stage is called "text representation".

### **13. What is an ensemble method in NLP?**

An ensemble approach is a methodology that derives an output or makes predictions by combining numerous independent similar or distinct models/weak learners. An ensemble can also be created by combining various models such as random forest, SVM, and logistic regression.

Bias, variance, and noise, as we all know, have a negative impact on the mistakes and predictions of any machine learning model. Ensemble approaches are employed to overcome these drawbacks.

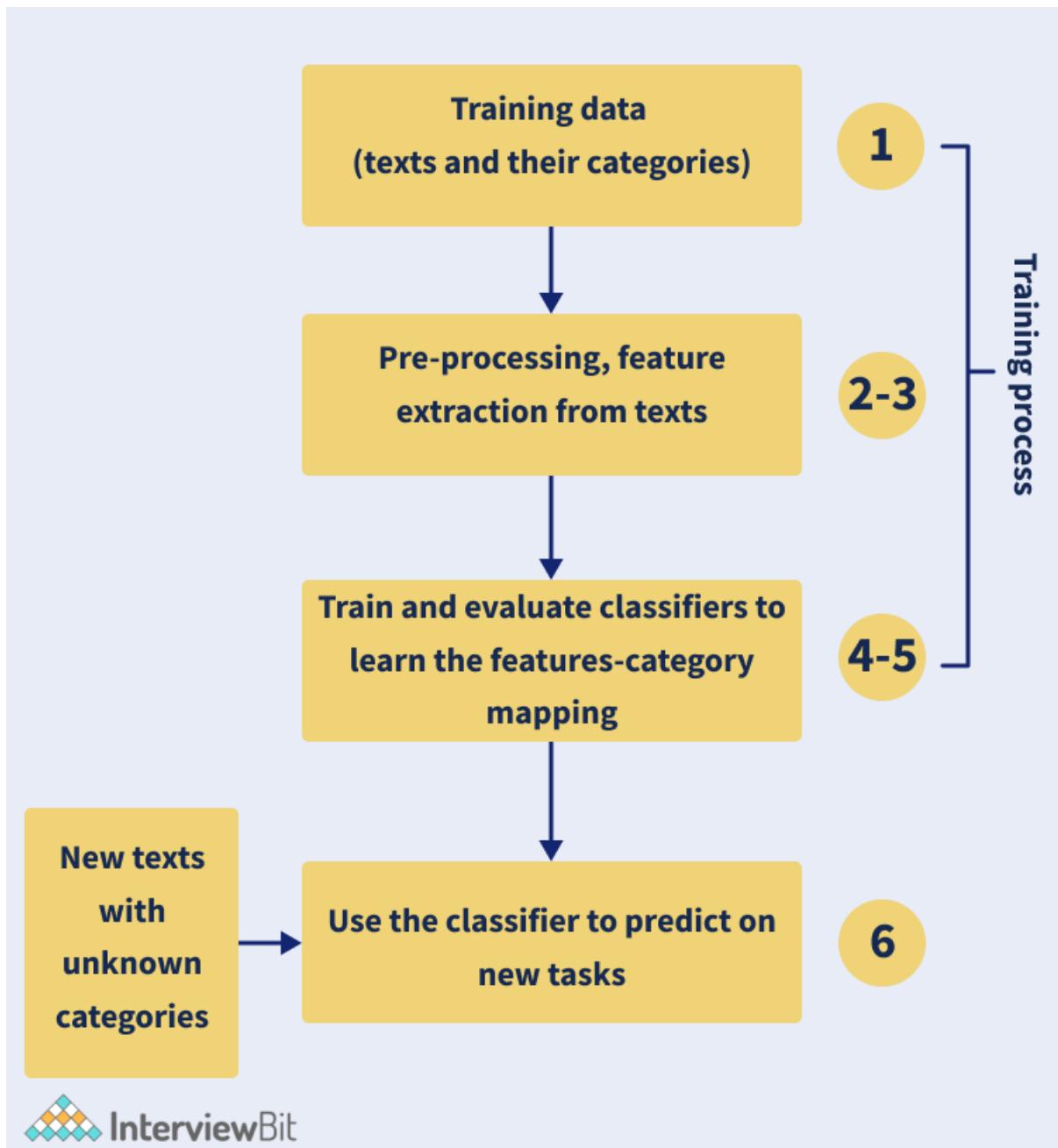
### **14. What do you mean by TF-IDF in Natural language Processing?**

TF-IDF also called **Term Frequency-Inverse Document Frequency** helps us get the importance of a particular word relative to other words in the corpus. It's a common scoring metric in information retrieval (IR) and summarization. TF-IDF converts words into vectors and adds semantic information, resulting in weighted unusual words that may be utilised in a variety of NLP applications.

### **15. What are the steps to follow when building a text classification system?**

When creating a text classification system, the following steps are usually followed:

- Gather or develop a labelled dataset that is appropriate for the purpose.
- Decide on an evaluation metric after splitting the dataset into two (training and test) or three parts: training, validation (i.e., development), and test sets (s).
- Convert unprocessed text into feature vectors.
- Utilize the feature vectors and labels from the training set to train a classifier.
- Benchmark the model's performance on the test set using the evaluation metric(s) from Step 2.
- Deploy the model and track its performance to serve a real-world use case.



## 16. Explain how parsing is done in NLP.

Parsing is the process of identifying and understanding a text's syntactic structure. It is accomplished by examining the text's constituent pieces. The machine parses each word one by one, then two by two, three by three, and so on. It's a unigram when the system parses the text one word at a time. A bigram is a text that is parsed two words at a time. When the machine parses three words at a time, the set of words is called a **trigram**.

The following points will help us comprehend the importance of parsing in NLP:

1. Parsing helps in understanding the meaning of a sentence by identifying its grammatical structure.

2. Parsing is used in various NLP applications such as sentiment analysis, named entity recognition, and machine translation.

3. Parsing can be used to generate natural language text from a given semantic representation.

4. Parsing is a fundamental step in many NLP tasks, such as generating a parse tree for a sentence.

5. Parsing can be used to identify errors in a sentence, such as punctuation or grammar mistakes.

6. Parsing can be used to extract information from a sentence, such as the subject, verb, and object.

- Any syntax errors are reported by the parser.
- It aids in the recovery of often occurring errors so that the remainder of the programme can be processed.
- A parser is used to generate the parse tree.
- The parser is used to construct a symbol table, which is crucial in NLP.
- In addition, a Parser is utilised to generate intermediate representations (IR).

## 17. What do you mean by a Bag of Words (BOW)?

The **Bag of Words** model is a popular one that uses word frequency or occurrences to train a classifier. This methodology generates a matrix of occurrences for documents or phrases, regardless of their grammatical structure or word order.

A bag-of-words is a text representation that describes the frequency with which words appear in a document. It entails two steps:

- A list of terms that are well-known.
- A metric for determining the existence of well-known terms.

Because any information about the sequence or structure of words in the document is deleted, it is referred to as a "bag" of words. The model simply cares about whether or not recognised terms appear in the document, not where they appear.

## 18. What do you mean by Parts of Speech (POS) tagging in NLP?

A Part-Of-Speech Tagger (POS Tagger) reads the text in a language and assigns parts of speech to each word (and other tokens), such as noun, verb, adjective, and so on.

To label terms in text bodies, PoS taggers employ an algorithm. With tags like "noun-plural" or even more complicated labels, these taggers create more complex categories than those stated as basic PoS.

## 19. What is Latent Semantic Indexing (LSI) in NLP?

**Latent Semantic Indexing** (LSI), also known as Latent Semantic Analysis, is a mathematical method for improving the accuracy of information retrieval. It aids in the discovery of hidden(latent) relationships between words (semantics) by generating a set of various concepts associated with the terms of a phrase in order to increase information comprehension. Singular value decomposition is

the NLP technique utilised for this aim. It's best for working with small groups of static documents.

## 20. What is the difference between NLP and NLU?

Natural Language Processing (NLP)	Natural Language Understanding (NLU)
NLP is a system that manages end-to-end conversations between computers and people at the same time.	NLU aids in the solving of Artificial Intelligence's complex problems.
Humans and machines are both involved in NLP.	NLU allows machines to interpret unstructured inputs by transforming them into structured text.
NLP focuses on interpreting language in its most literal sense, such as what was said.	NLU, on the other hand, concentrates on extracting context and meaning, or what was meant.
NLP can parse text-based on grammar, structure, typography, and point of view.	It'll be NLU that helps the machine deduce the meaning behind the language content.

## 21. What are some metrics on which NLP models are evaluated?

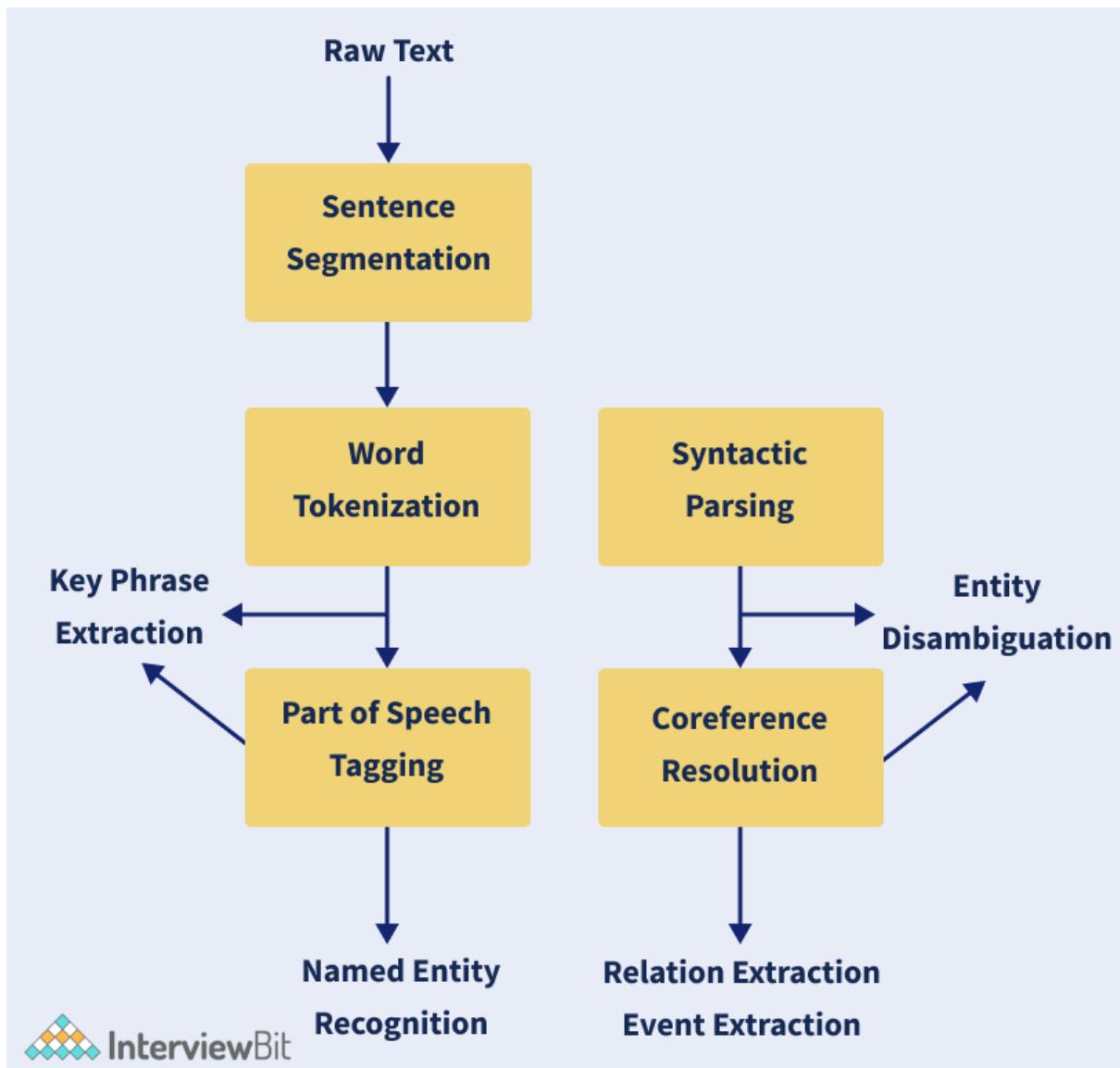
The following are some metrics on which NLP models are evaluated:

- **Accuracy:** When the output variable is categorical or discrete, accuracy is used. It is the percentage of correct predictions made by the model compared to the total number of predictions made.
- **Precision:** Indicates how precise or exact the model's predictions are, i.e., how many positive (the class we care about) examples can the model correctly identify given all of them?
- **Recall:** Precision and recall are complementary. It measures how effectively the model can recall the positive class, i.e., how many of the positive predictions it generates are correct.
- **F1 score:** This metric combines precision and recall into a single metric that also represents the trade-off between accuracy and recall, i.e., completeness and exactness.  
$$(2 \text{Precision} \text{Recall}) / (\text{Precision} + \text{Recall})$$
 is the formula for F1.
- **AUC:** As the prediction threshold is changed, the AUC captures the number of correct positive predictions versus the number of incorrect positive predictions.

## 22. Explain the pipeline for Information extraction (IE) in NLP.

In comparison to text classification, the typical pipeline for IE necessitates more fine-grained NLP processing. For example, we'd need to know the part-of-speech tags of words to identify named entities (people, organisations, etc.). We would require coreference resolution to connect various references to the same entity (e.g., Albert Einstein, Einstein, the scientist, he, etc.). It's worth noting that none of these stages are required for creating a text classification system. As a result, IE is a more NLP-intensive operation than text categorization. Not all steps in the pipeline are required for all IE jobs, as shown in the diagram, and the figure shows which IE tasks necessitate which degrees of analysis.

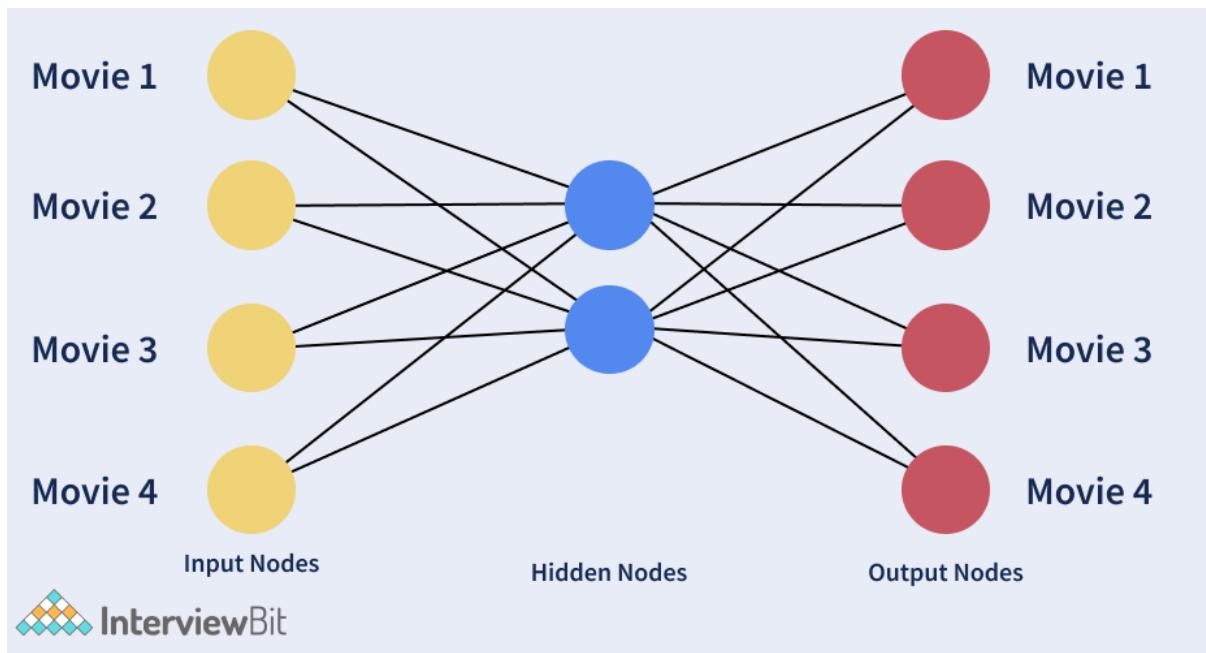
Other than named entity recognition, all other IE tasks require deeper NLP pre-processing followed by models developed for those specific tasks. Key phrase extraction is the task that requires the least amount of NLP processing (some algorithms also do POS tagging before extracting key phrases), whereas all other IE tasks require deeper NLP pre-processing followed by models developed for those specific tasks. Standard evaluation sets are often used to assess IE tasks in terms of precision, recall, and F1 scores. Because of the various levels of NLP pre-processing required, the accuracy of these processing steps has an impact on IE jobs. All of these factors should be considered when collecting relevant training data and, if necessary, training our own models for IE.



### 23. What do you mean by Autoencoders?

A network that is used for learning a vector representation of the input in a compressed form, is called an autoencoder. It is a type of unsupervised learning since labels aren't needed for the process. This is mainly used to learn the mapping function from the input. In order to make the mapping useful, the input is reconstructed from the vector representation. After training is complete, the vector representation that we get helps encode the input text as a dense vector. Autoencoders are generally used to make feature representations.

In the figure below, the hidden layer depicts a compressed representation of the source data that captures its essence. The input representation is reconstructed by the output layer called the decoder.



## 24. What do you mean by Masked language modelling?

Masked language modelling is an NLP technique for extracting the output from a contaminated input. Learners can use this approach to master deep representations in downstream tasks. Using this NLP technique, you may predict a word based on the other words in the sentence.

The following is the process for Masked language modelling:

- Our text is tokenized. We start with text tokenization, just as we would with transformers.
- Make a tensor of labels. We're using a labels tensor to calculate loss against — and optimise towards — as we train our model.
- Tokens in input ids are masked. We can mask a random selection of tokens now that we've produced a duplicate of input ids for labels.
- Make a loss calculation. We use our model to process the input ids and labels tensors and determine the loss between them.

## 25. What is the meaning of Pragmatic Analysis in NLP?

Pragmatic Analysis is concerned with outside word knowledge, which refers to information that is not contained in the documents and/or questions. The many parts of the language that require real-world knowledge are derived from a pragmatics analysis that focuses on what was described and reinterpreted by what it truly meant.

## 26. What is the meaning of N-gram in NLP?

Text N-grams are commonly used in text mining and natural language processing. They're essentially a collection of co-occurring words within a specific frame, and when computing the n-grams, you usually advance one word (although you can move X words forward in more advanced scenarios).

## 27. What do you mean by perplexity in NLP?

It's a statistic for evaluating the effectiveness of language models. It is described mathematically as a function of the likelihood that the language model describes a test sample. The perplexity of a test sample  $X = x_1, x_2, x_3, \dots, x_N$  is given by,

$$PP(X) = P(x_1, x_2, \dots, x_N)^{-1/N}$$

The total number of word tokens is  $N$ .

The more perplexing the situation, the less information the language model conveys.

1.

Which of the following are some of the use cases of NLP?

- Text Summarization
- Topic modeling
- Information extraction
- All of the above

2.

Which of the following are some of the components on NLP?

- Pragmatic analysis
- Entity extraction
- Syntactic analysis
- All of the above

3.

What is NER in Natural Language Processing?

- Named Entity Recognition
- Named Entity Response
- Name Explicit Response

- None of the above

4.

Which of the following techniques can be used to find the distance between 2-word vectors in NLP?

- N-grams
- Euclidean distance
- Lemmatization
- All of the above

5.

Which of the following are the keyword Normalization techniques in NLP?

- Part of Speech
- Named entity recognition
- Lemmatization
- None of the above

6.

Which of the following are some of the pre-processing techniques in NLP?

- Removal of stop words
- Converting to lowercase
- Stemming and Lemmatization
- All of the above

7.

Which of the following are NLP tools?

- Natural language Toolkit
- SpaCy
- CogcompNLP
- All of the above

8.

Information extraction in NLP includes which of the following?

- Network Graph Module

- Document Classification & Language Modeling Module
- Both (a) and (b)
- None of the above

9.

Which of the following are differences between NLTK and Spacey?

- NLTK can support word vectors while Spacey cannot
- NLTK supports a wider range of languages compared to Spacey (Spacey supports only 7 languages)
- While Spacey has a collection of programs to choose from, NLTK contains only the best-suited algorithm for a problem in its toolkit
- While NLTK has an object-oriented library, Spacey has a string processing library

10.

Which of the following are types of text summarization?

- Abstraction-based summarization
- Extraction-based summarization
- Both (a) and (b)
- None of the above.

## **1. What do you know about NLP?**

NLP stands for Natural Language Processing. It deals with making a machine understand the way human beings read and write in a language. This task is achieved by designing algorithms that can extract meaning from large datasets in audio or text format by applying machine learning algorithms.

## **2. Give examples of any two real-world applications of NLP.**

**1. Spelling/Grammar Checking Apps:** The mobile applications and websites that offer users correct grammar mistakes in the entered text rely on NLP algorithms. These days, they can also recommend the following few words that the user might type, which is also because of specific NLP models being used in the backend.

**2. ChatBots:** Many websites now offer customer support through these virtual bots that chat with the user and resolve their problems. It acts as a filter to the issues that do not require an interaction with the companies' customer executives.

### 3. What is tokenization in NLP?

Tokenization is the process of splitting running text into words and sentences.

#### New Projects

- Talend Real-Time Project for ETL Process Automation**[View Project](#)
- Build a Data Pipeline with Azure Synapse and Spark Pool**[View Project](#)
- Build CI/CD Pipeline for Machine Learning Projects using Jenkins**[View Project](#)
- Build Piecewise and Spline Regression Models in Python**[View Project](#)
- Build Streaming Data Pipeline using Azure Stream Analytics**[View Project](#)
- Build Serverless Pipeline using AWS CDK and Lambda in Python**[View Project](#)
- Build Piecewise and Spline Regression Models in Python**[View Project](#)
- Building Data Pipelines in Azure with Azure Synapse Analytics**[View Project](#)
- Getting Started with Pyspark on AWS EMR and Athena**[View Project](#)
- Getting Started with Pyspark on AWS EMR and Athena**[View Project](#)
- Talend Real-Time Project for ETL Process Automation**[View Project](#)
- Build a Data Pipeline with Azure Synapse and Spark Pool**[View Project](#)
- Build CI/CD Pipeline for Machine Learning Projects using Jenkins**[View Project](#)
- Build Piecewise and Spline Regression Models in Python**[View Project](#)
- Build Streaming Data Pipeline using Azure Stream Analytics**[View Project](#)
- Build Serverless Pipeline using AWS CDK and Lambda in Python**[View Project](#)
- Build Piecewise and Spline Regression Models in Python**[View Project](#)
- Building Data Pipelines in Azure with Azure Synapse Analytics**[View Project](#)
- Getting Started with Pyspark on AWS EMR and Athena**[View Project](#)
- Getting Started with Pyspark on AWS EMR and Athena**[View Project](#)

[View all New Projects](#)

#### **4. What is the difference between a formal language and a natural language?**

Formal Language	Natural Language
A formal language is a collection of strings, where each string contains symbols from a finite set called alphabets.	A natural language is a language that humans utilize to speak. It is usually a lot different from a formal language. These typically contain fragments of words and pause words like uh, um, etc.

#### **5. What is the difference between stemming and lemmatization?**

Both stemming and lemmatization are keyword normalization techniques aiming to minimize the morphological variation in the words they encounter in a sentence. But, they are different from each other in the following way.

Stemming	Lemmatization
This technique involves removing the affixes added to a word and leaving us with the rest of the word.	Lemmatization is the process of converting a word into its lemma from its inflected form.
Example: 'Caring' → 'Car'	Example: 'Caring' → 'Care'

#### **6. What is NLU?**

NLU stands for Natural Language Understanding. It is a subdomain of NLP that concerns making a machine learn the skills of reading comprehension. A few applications of NLU include Machine translation (MT), Newsgathering, and Text categorization. It often goes by the name Natural Language Interpretation (NLI) as well.

**Get Closer To Your Dream of Becoming a Data Scientist with 70+ Solved End-to-End ML Projects**

#### **7. List the differences between NLP and NLU.**

Natural Language Processing	Natural Language Understanding
NLP is a branch of AI that deals with designing programs for machines that will allow them to process the language that humans use. The idea is to make machines imitate the way humans utilize language for communication.	In NLU, the aim is to improve a computer's ability to understand and analyze human language. This aim is achieved by transforming unstructured data into a machine-readable format.

## **8. What do you know about Latent Semantic Indexing (LSI)?**

LSI is a technique that analyzes a set of documents to find the statistical coexistence of words that appear together. It gives an insight into the topics of those documents.

LSI is also known as Latent Semantic Analysis.

## **9. List a few methods for extracting features from a corpus for NLP.**

1. Bag-of-Words

2. Word Embedding

## **10. What are stop words?**

Stop words are the words in a document that are considered redundant by NLP engineers and are thus removed from the document before processing it. Few examples are ‘is’, ‘the’, ‘are’, ‘am’.

## **11. What do you know about Dependency Parsing?**

Dependency parsing is a technique that highlights the dependencies among the words of a sentence to understand its grammatical structure. It examines how the words of a sentence are linguistically linked to each other. These links are called dependencies.

## **12. What is Text Summarization? Name its two types.**

Text Summarization is a method of converting a long-form text into a summary. The summary thus generated is expected to have critical ideas of the lengthy text. Two main types of Text Summarization are:

1. Extraction-based Summarization
2. Abstraction-based Summarization

**Get FREE Access to Machine Learning Example Codes for Data Cleaning, Data Munging, and Data Visualization**

## **13. What are false positives and false negatives?**

If a machine learning algorithm falsely predicts a negative outcome as positive, then the result is labeled as a false negative.

And, if a machine learning algorithm falsely predicts a positive outcome as negative, then the result is labeled as a false positive.

#### **14. List a few methods for part-of-speech tagging.**

Rule-based tagging, HMM-tagging, transformation-based tagging, and memory-based tagging.

#### **15. What is a corpus?**

‘Corpus’ is a Latin word that means ‘body.’ Thus, a body of the written or spoken text is called a corpus.

#### **Recommended Reading: 10 NLP Techniques Every Data Scientist Should Know**

### **NLP Algorithm Interview Questions with Answers**

Most recruiters usually try to understand how well you know the models that are used widely in NLP. Take a look at these interview questions in NLP with answers that will help you upgrade your NLP algorithm skills.

#### **1. List a few real-world applications of the n-gram model.**

1. Augmentive Communication
2. Part-of-speech Tagging
3. Natural language generation
4. Word Similarity
5. Authorship Identification
6. Sentiment Extraction
7. Predictive Text Input

#### **2. What does TF\*IDF stand for? Explain its significance.**

TF\*IDF stands for Term-Frequency/Inverse-Document Frequency. It is an information-retrieval measure that encapsulates the semantic significance of a word in a particular document N, by degrading words that tend to appear in a

variety of different documents in some huge background corpus with D documents.

Let  $n_w$  denote the frequency of a word  $w$  in the document  $N$ ,  $m$  represents the total number of documents in the corpus that contain  $w$ . Then, TF\*IDF is defined as

$$\text{TF-IDF}(w) = \text{TF}(w) \times \text{IDF}(w)$$

### 3. What is perplexity in NLP?

It is a metric that is used to test the performance of language models. Mathematically, it is defined as a function of the probability that the language model represents a test sample. For a test sample  $X = x_1, x_2, x_3, \dots, x_n$ , the perplexity is given by,

$$\text{Perplexity} = \exp\left(-\frac{1}{N} \sum_{i=1}^N \log P(x_i | \theta)\right)$$

where  $N$  is the total number of word tokens.

Higher the perplexity, lesser is the information conveyed by the language model.

### 4. Which algorithm in NLP supports bidirectional context?

BERT

### 5. What is the Naive Bayes algorithm?

Naive Bayes is a classification machine learning algorithm that utilizes Baye's Theorem for labeling a class to the input set of features. A vital element of this algorithm is that it assumes that all the feature values are independent.

---

## Explore Categories

Data Science Projects in R		Deep Learning Projects		Neural Network
Projects	Tensorflow Projects	H2O R Projects	IoT	
Projects	Keras Deep Learning Projects			NLP
Projects	Pytorch	Data Science Projects in Banking and		

<b>Finance</b>	<b>Data Science Projects in Retail &amp; Ecommerce</b>	<b>Data Science</b>
<b>Projects in Entertainment &amp; Media</b>		<b>Data Science Projects in</b>
<b>Telecommunications</b>		

## **6. What is Part-of-Speech tagging?**

Part-of-speech tagging is the task of assigning a part-of-speech label to each word in a sentence. A variety of part-of-speech algorithms are available that contain tagsets having several tags between 40 and 200.

## **7. What is the bigram model in NLP?**

A bigram model is a model used in NLP for predicting the probability of a word in a sentence using the conditional probability of the previous word. For calculating the conditional probability of the previous word, it is crucial that all the previous words are known.

## **8. What is the significance of the Naive Bayes algorithm in NLP?**

The Naive Bayes algorithm is widely used in NLP for various applications. For example: to determine the sense of a word, to predict the tag of a given text, etc.

## **9. What do you know about the Masked Language Model?**

The Masked Language Model is a model that takes a sentence with a few hidden (masked) words as input and tries to complete the sentence by correctly guessing those hidden words.

## **10. What is the Bag-of-words model in NLP?**

Bag-of-words refers to an unorganized set of words. The Bag-of-words model in NLP is a model that assigns a vector to a sentence in a corpus. It first creates a dictionary of words and then produces a vector by assigning a binary variable to each word of the sentence depending on whether it exists in the bag of words or not.

### **Recommended Reading:**

- [\*\*20 Linear Regression Interview Questions and Answers\*\*](#)
- [\*\*Top 20 Logistic Regression Interview Questions and Answers\*\*](#)
- [\*\*Best NLP Books- What Data Scientists Must Read?\*\*](#)
- [\*\*50 Artificial Intelligence Interview Questions and Answers\*\*](#)

## **11. Briefly describe the N-gram model in NLP.**

N-gram model is a model in NLP that predicts the probability of a word in a given sentence using the conditional probability of n-1 previous words in the sentence. The basic intuition behind this algorithm is that instead of using all the previous words to predict the next word, we use only a few previous words.

## **12. What is the Markov assumption for the bigram model?**

The Markov assumption assumes for the bigram model that the probability of a word in a sentence depends only on the previous word in that sentence and not on all the previous words.

## **13. What do you understand by word embedding?**

In NLP, word embedding is the process of representing textual data through a real-numbered vector. This method allows words having similar meanings to have a similar representation.

## **14. What is an embedding matrix?**

A word embedding matrix is a matrix that contains embedding vectors of all the words in a given text.

## **15. List a few popular methods used for word embedding.**

Following are a few methods of word embedding.

1. Embedding Layer
2. Word2Vec
3. Glove

## **1. How will you use Python's concordance command in NLTK for a text that does not belong to the package?**

The concordance() function can easily be accessed for a text that belongs to the NLTK package using the following code:

```
>>>from nltk.book import *
>>>text1.concordance("monstrous")
```

However, for a text that does not belong to the NLTK package, one has to use the following code to access that function.

```
>>>import nltk.corpus  
>>>from nltk.text import Text  
>>>NLTKtext = Text(nltk.corpus.gutenberg.words('Your_file_name_here.txt'))  
>>>NLTKtext.concordance('word')
```

Here, we have created a Text object to access the concordance() function. The function displays the occurrence of the chosen word and the context around it.

## **2. Write the code to count the number of distinct tokens in a text?**

len(set(text))

## **3. What are the first few steps that you will take before applying an NLP machine-learning algorithm to a given corpus?**

- Ans:
1. Removing white spaces
  2. Removing Punctuations
  3. Converting Uppercase to Lowercase
  4. Tokenization
  5. Removing Stopwords
  6. Lemmatization

## **4. For correcting spelling errors in a corpus, which one is a better choice: a giant dictionary or a smaller dictionary, and why?**

Initially, a smaller dictionary is a better choice because most NLP researchers feared that a giant dictionary would contain rare words that may be similar to misspelled words. However, later it was found (Damerau and Mays (1989)) that in practice, a more extensive dictionary is better at marking rare words as errors.

## **5. Do you always recommend removing punctuation marks from the corpus you're dealing with? Why/Why not?**

No, it is not always a good idea to remove punctuation marks from the corpus as they are necessary for certain NLP applications that require the marks to be counted along with words.

For example: Part-of-speech tagging, parsing, speech synthesis.

## **6. List a few libraries that you use for NLP in Python.**

NLTK, Scikit-learn, GenSim, SpaCy, CoreNLP, TextBlob.

## **7. Suggest a few machine learning/deep learning models that are used in NLP.**

Support Vector Machines, Neural Networks, Decision Tree, Bayesian Networks.

## **8. Which library contains the Word2Vec model in Python?**

GenSim

### **1. What are homographs, homophones, and homonyms?**

Homographs	Homophones	Homonyms
“Home”=same “graph”=write	“Home”=same “phone”=sound	“Homo”=same, “onym” = name
These are the words that have the same spelling but may or may not have the same pronunciations.	These are the words that sound similar but have different spelling and different meanings.	These are the words that have the same spelling and pronunciation but different meanings.
To <u>live</u> a life, airing a show <u>live</u>	<u>Eye</u> , <u>I</u>	River <u>Bank</u> , <u>Bank</u> Account

## **2. Is converting all text in uppercase to lowercase always a good idea?**

**Explain with the help of an example.**

No, for words like The, the, THE, it is a good idea as they all will have the same meaning. However, for a word like brown which can be used as a surname for someone by the name Robert Brown, it won't be a good idea as the word ‘brown’ has different meanings for both the cases. We, therefore, would want to treat them differently. Hence, it is better to change uppercase letters at the beginning of a sentence to lowercase, convert headings and titles to which are all in capitals to lowercase, and leave the remaining text unchanged.

## **3. What is a hapax/hapax legomenon?**

The rare words that only occur once in a sample text or corpus are called hapaxes. Each one of them is called an hapax or hapax legomenon (greek for ‘read-only once’). It is also called a singleton.

**4. Is tokenizing a sentence based on white-space ‘ ‘ character sufficient? If not, give an example where it may not work.**

Tokenizing a sentence using the white space character is not always sufficient.

Consider the example,

“ One of our users said, ‘I love Dezyre’s content’. ”

Tokenizing purely based on white space would result in the following words:

‘I      said,    content’.

**5. What is a collocation?**

A collocation is a group of two or more words that possess a relationship and provide a classic alternative of saying something. For example, ‘strong breeze’, ‘the rich and powerful’, ‘weapons of mass destruction’.

**6. List a few types of linguistic ambiguities.**

**1. Lexical Ambiguity:** This type of ambiguity is observed because of homonyms and polysemy in a sentence.

**2. Syntactic Ambiguity:** A syntactic ambiguity is observed when based on the sentence’s syntax, more than one meaning is possible.

**3. Semantic Ambiguity:** This ambiguity occurs when a sentence contains ambiguous words or phrases that have ambiguous meanings.

**7. What is a Turing Test? Explain with respect to NLP-based systems.**

Alan Turing developed a test, called Turing Test, that could differentiate between humans and machines. A computer machine is considered intelligent if it can pass this test through its use of language. Alan believed that if a machine could use language the way humans do, it was sufficient for the machine to prove its intelligence.

**8. What do you understand by regular expressions in NLP?**

Regular expressions in natural language processing are algebraic notations representing a set of strings. They are mainly used to find or replace strings in a text and can also be used to define a language in a formal way.

**9. Differentiate between orthographic rules and morphological rules with respect to singular and plural forms of English words.**

Orthographic Rules	Morphological Rules
These are the rules that contain information for extracting the plural form of English words that end in 'y'. Such words are transformed into their plural form by converting 'y' into 'i' and adding the letters 'es' as suffixes.	These rules contain information for words like fish; there are null plural forms. And words like goose have their plural generated by a change of the vowel.

**10. Define the term parsing concerning NLP.**

Parsing refers to the task of generating a linguistic structure for a given input. For example, parsing the word 'helping' will result in **verb-pass + gerund-ing**.

**11. Use the minimum distance algorithm to show how many editing steps it will take for the word 'intention' to transform into 'execution'.**

or

**Calculate the Levenshtein distance between two sequences 'intention' and 'execution'.**

The image above can be used to understand the number of editing steps it will take for the word intention to transform into execution.

1. The first step is deletion (d) of 'I.'
2. The next step is to substitute (s) the letter 'N' with 'E.'
3. Replace the letter 'T' with 'X.'
4. The letter E remains unchanged, and the letter 'C' is inserted (i).
5. Substitute 'U' for the letter 'N.'

Thus, it will take five editing steps for transformations, and the Levenshtein distance is five.

**12. What are the full listing hypothesis and minimum redundancy hypothesis?**

These are the two hypotheses relating to the way humans store words of a language in their memory.

**Full Listing Hypothesis:** This hypothesis suggests that all humans perceive all the words in their memory without any internal morphological structure. So, words like tire, tiring, tired are all stored separately in the mental lexicon.

**Minimum Redundancy Hypothesis:** This hypothesis proposes that only the raw form of the words (morphemes) form the part of the mental lexicon. When humans process a word like tired, they recall both the morphemes (tire-d).

## **1. What is Naive Bayes algorithm, When we can use this algorithm in NLP?**

Naive Bayes algorithm is a collection of classifiers which works on the principles of the Bayes' theorem. This series of NLP model forms a family of algorithms that can be used for a wide range of classification tasks including sentiment prediction, filtering of spam, classifying documents and more.

Naive Bayes algorithm converges faster and requires less training data. Compared to other discriminative models like logistic regression, Naive Bayes model it takes lesser time to train. This algorithm is perfect for use while working with multiple classes and text classification where the data is dynamic and changes frequently.

## **2. Explain Dependency Parsing in NLP?**

Dependency Parsing, also known as Syntactic parsing in NLP is a process of assigning syntactic structure to a sentence and identifying its dependency parses. This process is crucial to understand the correlations between the “head” words in the syntactic structure.

The process of dependency parsing can be a little complex considering how any sentence can have more than one dependency parses. Multiple parse trees are known as ambiguities. Dependency parsing needs to resolve these ambiguities in order to effectively assign a syntactic structure to a sentence.

Dependency parsing can be used in the semantic analysis of a sentence apart from the syntactic structuring.

## **3. What is text Summarization?**

Text summarization is the process of shortening a long piece of text with its meaning and effect intact. Text summarization intends to create a summary of any given piece of text and outlines the main points of the document. This technique has improved in recent times and is capable of summarizing volumes of text successfully.

Text summarization has proved to be a blessing since machines can summarise large volumes of text in no time which would otherwise be really time-consuming. There are two types of text summarization:

- Extraction-based summarization
- Abstraction-based summarization

#### **4. What is NLTK? How is it different from Spacy?**

NLTK or Natural Language Toolkit is a series of libraries and programs that are used for symbolic and statistical natural language processing. This toolkit contains some of the most powerful libraries that can work on different ML techniques to break down and understand human language. NLTK is used for Lemmatization, Punctuation, Character count, Tokenization, and Stemming. The difference between NLTK and Spacey are as follows:

- While NLTK has a collection of programs to choose from, Spacey contains only the best-suited algorithm for a problem in its toolkit
- NLTK supports a wider range of languages compared to Spacey (Spacey supports only 7 languages)
- While Spacey has an object-oriented library, NLTK has a string processing library
- Spacey can support word vectors while NLTK cannot

#### **5. What is information extraction?**

Information extraction in the context of Natural Language Processing refers to the technique of extracting structured information automatically from unstructured sources to ascribe meaning to it. This can include extracting information regarding attributes of entities, relationship between different entities and more. The various models of information extraction includes:

- Tagger Module
- Relation Extraction Module
- Fact Extraction Module
- Entity Extraction Module
- Sentiment Analysis Module
- Network Graph Module
- Document Classification & Language Modeling Module

#### **6. What is Bag of Words?**

Bag of Words is a commonly used model that depends on word frequencies or occurrences to train a classifier. This model creates an occurrence matrix for documents or sentences irrespective of its grammatical structure or word order.

#### **7. What is Pragmatic Ambiguity in NLP?**

Pragmatic ambiguity refers to those words which have more than one meaning and their use in any sentence can depend entirely on the context. Pragmatic

ambiguity can result in multiple interpretations of the same sentence. More often than not, we come across sentences which have words with multiple meanings, making the sentence open to interpretation. This multiple interpretation causes ambiguity and is known as Pragmatic ambiguity in NLP.

## 8. What is Masked Language Model?

Masked language models help learners to understand deep representations in downstream tasks by taking an output from the corrupt input. This model is often used to predict the words to be used in a sentence.

## 9. What is the difference between NLP and CI(Conversational Interface)?

The difference between NLP and CI is as follows:

Natural Language Processing (NLP)	Conversational Interface (CI)
NLP attempts to help machines understand and learn how language concepts work.	CI focuses only on providing users with an interface to interact with.
NLP uses AI technology to identify, understand, and interpret the requests of users through language.	CI uses voice, chat, videos, images, and more such conversational aid to create the user interface.

## 10. What are the best NLP Tools?

Some of the best NLP tools from open sources are:

- SpaCy
- TextBlob
- Textacy
- Natural language Toolkit (NLTK)
- Retext
- NLP.js
- Stanford NLP
- CogcompNLP

## 11. What is POS tagging?

Parts of speech tagging better known as POS tagging refer to the process of identifying specific words in a document and grouping them as part of speech,

based on its context. POS tagging is also known as grammatical tagging since it involves understanding grammatical structures and identifying the respective component.

POS tagging is a complicated process since the same word can be different parts of speech depending on the context. The same general process used for word mapping is quite ineffective for POS tagging because of the same reason.

## **12. What is NES?**

Name entity recognition is more commonly known as NER is the process of identifying specific entities in a text document that are more informative and have a unique context. These often denote places, people, organizations, and more. Even though it seems like these entities are proper nouns, the NER process is far from identifying just the nouns. In fact, NER involves entity chunking or extraction wherein entities are segmented to categorize them under different predefined classes. This step further helps in extracting information.

## **13. Which of the following techniques can be used for keyword normalization in NLP, the process of converting a keyword into its base form?**

- a. Lemmatization
- b. Soundex
- c. Cosine Similarity
- d. N-grams

**Answer:** a)

Lemmatization helps to get to the base form of a word, e.g. are playing -> play, eating -> eat, etc. Other options are meant for different purposes.

## **14. Which of the following techniques can be used to compute the distance between two-word vectors in NLP?**

- a. Lemmatization
- b. Euclidean distance
- c. Cosine Similarity
- d. N-grams

**Answer:** b) and c)

Distance between two-word vectors can be computed using Cosine similarity and Euclidean Distance. Cosine Similarity establishes a cosine angle between the vector of two words. A cosine angle close to each other between two-word vectors indicates the words are similar and vice versa.

E.g. cosine angle between two words “Football” and “Cricket” will be closer to 1 as compared to the angle between the words “Football” and “New Delhi”.

Python code to implement CosineSimilarity function would look like this:

```
def cosine_similarity(x,y):  
    return np.dot(x,y)/( np.sqrt(np.dot(x,x)) * np.sqrt(np.dot(y,y)) )  
  
q1 = wikipedia.page('Strawberry')  
  
q2 = wikipedia.page('Pineapple')  
  
q3 = wikipedia.page('Google')  
  
q4 = wikipedia.page('Microsoft')  
  
cv = CountVectorizer()  
  
X = np.array(cv.fit_transform([q1.content, q2.content, q3.content,  
q4.content]).todense())  
  
print ("Strawberry Pineapple Cosine Distance", cosine_similarity(X[0],X[1]))  
  
print ("Strawberry Google Cosine Distance", cosine_similarity(X[0],X[2]))  
  
print ("Pineapple Google Cosine Distance", cosine_similarity(X[1],X[2]))  
  
print ("Google Microsoft Cosine Distance", cosine_similarity(X[2],X[3]))  
  
print ("Pineapple Microsoft Cosine Distance", cosine_similarity(X[1],X[3]))  
  
Strawberry Pineapple Cosine Distance 0.8899200413701714  
  
Strawberry Google Cosine Distance 0.7730935582847817
```

Pineapple Google Cosine Distance 0.789610214147025

Google Microsoft Cosine Distance 0.8110888282851575

Usually Document similarity is measured by how close semantically the content (or words) in the document are to each other. When they are close, the similarity index is close to 1, otherwise near 0.

The **Euclidean distance** between two points is the length of the shortest path connecting them. Usually computed using Pythagoras theorem for a triangle.

### **15. What are the possible features of a text corpus in NLP?**

- a. Count of the word in a document
- b. Vector notation of the word
- c. Part of Speech Tag
- d. Basic Dependency Grammar
- e. All of the above

**Answer:** e)

All of the above can be used as features of the text corpus.

### **16. You created a document term matrix on the input data of 20K documents for a Machine learning model. Which of the following can be used to reduce the dimensions of data?**

- 1. Keyword Normalization
  - 2. Latent Semantic Indexing
  - 3. Latent Dirichlet Allocation
- a. only 1
  - b. 2, 3
  - c. 1, 3
  - d. 1, 2, 3

**Answer:** d)

### **17. Which of the text parsing techniques can be used for noun phrase detection, verb phrase detection, subject detection, and object detection in NLP.**

- a. Part of speech tagging
- b. Skip Gram and N-Gram extraction
- c. Continuous Bag of Words
- d. Dependency Parsing and Constituency Parsing

**Answer:** d)

**18. Dissimilarity between words expressed using cosine similarity will have values significantly higher than 0.5**

- a. True
- b. False

**Answer:** a)

**19. Which one of the following is keyword Normalization techniques in NLP**

- a. Stemming
- b. Part of Speech
- c. Named entity recognition
- d. Lemmatization

Answer: a) and d)

Part of Speech (POS) and Named Entity Recognition(NER) is not keyword Normalization techniques. Named Entity helps you extract Organization, Time, Date, City, etc., type of entities from the given sentence, whereas Part of Speech helps you extract Noun, Verb, Pronoun, adjective, etc., from the given sentence tokens.

**20. Which of the below are NLP use cases?**

- a. Detecting objects from an image
- b. Facial Recognition
- c. Speech Biometric
- d. Text Summarization

Ans: d)

- a) And b) are Computer Vision use cases, and c) is the Speech use case.  
Only d) Text Summarization is an NLP use case.

**21. In a corpus of N documents, one randomly chosen document contains a total of T terms and the term “hello” appears K times.**

What is the correct value for the product of TF (term frequency) and IDF (inverse-document-frequency), if the term “hello” appears in approximately one-third of the total documents?

- a.  $KT * \log(3)$
- b.  $T * \log(3) / K$
- c.  $K * \log(3) / T$
- d.  $\log(3) / KT$

**Answer:** (c)

formula for TF is  $K/T$

formula for IDF is  $\log(\text{total docs} / \text{no of docs containing “data”})$

$$= \log(1 / (\frac{1}{3}))$$

$$= \log(3)$$

Hence, the correct choice is  $K \log(3) / T$

**22. In NLP, The algorithm decreases the weight for commonly used words and increases the weight for words that are not used very much in a collection of documents**

- a. Term Frequency (TF)
- b. Inverse Document Frequency (IDF)
- c. Word2Vec
- d. Latent Dirichlet Allocation (LDA)

**Answer:** b)

**23. In NLP, The process of removing words like “and”, “is”, “a”, “an”, “the” from a sentence is called as**

- a. Stemming
- b. Lemmatization
- c. Stop word
- d. All of the above

**Ans: c)**

In Lemmatization, all the stop words such as a, an, the, etc.. are removed. One can also define custom stop words for removal.

**24. In NLP, The process of converting a sentence or paragraph into tokens is referred to as Stemming**

- a. True
- b. False

**Answer: b)**

The statement describes the process of tokenization and not stemming, hence it is False.

**25. In NLP, Tokens are converted into numbers before giving to any Neural Network**

- a. True
- b. False

**Answer: a)**

In NLP, all words are converted into a number before feeding to a Neural Network.

**26. Identify the odd one out**

- a. nltk
- b. scikit learn
- c. SpaCy
- d. BERT

**Answer: d)**

All the ones mentioned are NLP libraries except BERT, which is a word embedding.

**27. TF-IDF helps you to establish?**

- a. most frequently occurring word in document
- b. the most important word in the document

**Answer:** b)

TF-IDF helps to establish how important a particular word is in the context of the document corpus. TF-IDF takes into account the number of times the word appears in the document and is offset by the number of documents that appear in the corpus.

- TF is the frequency of terms divided by the total number of terms in the document.
- IDF is obtained by dividing the total number of documents by the number of documents containing the term and then taking the logarithm of that quotient.
- Tf.idf is then the multiplication of two values TF and IDF.

Suppose that we have term count tables of a corpus consisting of only two documents, as listed here:

Term	Document 1 Frequency	Document 2 Frequency
This	1	1
is	1	1
a	2	
Sample	1	
another		2
example		3

The calculation of tf–idf for the term “this” is performed as follows:

for "this"

-----

$$tf("this", d1) = 1/5 = 0.2$$

$$tf("this", d2) = 1/7 = 0.14$$

$$idf("this", D) = \log (2/2) = 0$$

hence tf-idf

$$tfidf("this", d1, D) = 0.2 * 0 = 0$$

$$tfidf("this", d2, D) = 0.14 * 0 = 0$$

for "example"

-----

$$tf("example", d1) = 0/5 = 0$$

$$tf("example", d2) = 3/7 = 0.43$$

$$idf("example", D) = \log(2/1) = 0.301$$

$$tfidf("example", d1, D) = tf("example", d1) * idf("example", D) = 0 * 0.301 = 0$$

$$tfidf("example", d2, D) = tf("example", d2) * idf("example", D) = 0.43 * 0.301 = 0.129$$

In its raw frequency form, TF is just the frequency of the “this” for each document. In each document, the word “this” appears once; but as document 2 has more words, its relative frequency is smaller.

An IDF is constant per corpus, and accounts for the ratio of documents that include the word “this”. In this case, we have a corpus of two documents and all of them include the word “this”. So TF-IDF is zero for the word “this”, which implies that the word is not very informative as it appears in all documents.

The word “example” is more interesting – it occurs three times, but only in the second document. To understand more about NLP, check out these NLP projects.

**28. In NLP, The process of identifying people, an organization from a given sentence, paragraph is called**

- a. Stemming
- b. Lemmatization
- c. Stop word removal
- d. Named entity recognition

**Answer:** d)

**29. Which one of the following is not a pre-processing technique in NLP**

- a. Stemming and Lemmatization
- b. converting to lowercase
- c. removing punctuations
- d. removal of stop words
- e. Sentiment analysis

**Answer:** e)

Sentiment Analysis is not a pre-processing technique. It is done after pre-processing and is an NLP use case. All other listed ones are used as part of statement pre-processing.

**30. In text mining, converting text into tokens and then converting them into an integer or floating-point vectors can be done using**

- a. CountVectorizer
- b. TF-IDF
- c. Bag of Words
- d. NERs

**Answer:** a)

CountVectorizer helps do the above, while others are not applicable.

```
text =["Rahul is an avid writer, he enjoys studying understanding and presenting. He loves to play"]
```

```
vectorizer = CountVectorizer()  
  
vectorizer.fit(text)  
  
vector = vectorizer.transform(text)  
  
print(vector.toarray())
```

## Output

```
[[1 1 1 1 2 1 1 1 1 1 1 1 1 1]]
```

The second section of the interview questions covers advanced NLP techniques such as Word2Vec, GloVe word embeddings, and advanced models such as GPT, Elmo, BERT, XLNET-based *questions, and explanations*.

### **31. In NLP, Words represented as vectors are called Neural Word Embeddings**

- a. True
- b. False

**Answer:** a)

Word2Vec, GloVe based models build word embedding vectors that are multidimensional.

### **32. In NLP, Context modeling is supported with which one of the following word embeddings**

1. a. Word2Vec
2. b) GloVe
3. c) BERT
4. d) All of the above

**Answer:** c)

Only BERT (Bidirectional Encoder Representations from Transformer) supports context modelling where the previous and next sentence context is

taken into consideration. In Word2Vec, GloVe only word embeddings are considered and previous and next sentence context is not considered.

**33. In NLP, Bidirectional context is supported by which of the following embedding**

- a. Word2Vec
- b. BERT
- c. GloVe
- d. All the above

**Answer:** b)

Only BERT provides a bidirectional context. The BERT model uses the previous and the next sentence to arrive at the context. Word2Vec and GloVe are word embeddings, they do not provide any context.

**34. Which one of the following Word embeddings can be custom trained for a specific subject in NLP**

- a. Word2Vec
- b. BERT
- c. GloVe
- d. All the above

**Answer:** b)

BERT allows Transform Learning on the existing pre-trained models and hence can be custom trained for the given specific subject, unlike Word2Vec and GloVe where existing word embeddings can be used, no transfer learning on text is possible.

**35. Word embeddings capture multiple dimensions of data and are represented as vectors**

- a. True
- b. False

**Answer:** a)

**36. In NLP, Word embedding vectors help establish distance between two tokens**

- a. True
- b. False

**Answer: a)**

**One can use Cosine similarity to establish the distance between two vectors represented through Word Embeddings**

**37. Language Biases are introduced due to historical data used during training of word embeddings, which one amongst the below is not an example of bias**

- a. New Delhi is to India, Beijing is to China
- b. Man is to Computer, Woman is to Homemaker

**Answer: a)**

Statement b) is a bias as it buckets Woman into Homemaker, whereas statement a) is not a biased statement.

**38. Which of the following will be a better choice to address NLP use cases such as semantic similarity, reading comprehension, and common sense reasoning**

- a. ELMo
- b. Open AI's GPT
- c. ULMFit

**Answer: b)**

Open AI's GPT is able to learn complex patterns in data by using the Transformer models Attention mechanism and hence is more suited for complex use cases such as semantic similarity, reading comprehensions, and common sense reasoning.

**39. Transformer architecture was first introduced with?**

- a. GloVe
- b. BERT
- c. Open AI's GPT
- d. ULMFit

**Answer:** c)

ULMFit has an LSTM based Language modeling architecture. This got replaced into Transformer architecture with Open AI's GPT.

**40. Which of the following architecture can be trained faster and needs less amount of training data**

- a. LSTM-based Language Modelling
- b. Transformer architecture

**Answer:** b)

Transformer architectures were supported from GPT onwards and were faster to train and needed less amount of data for training too.

**41. Same word can have multiple word embeddings possible with \_\_\_\_\_?**

- a. GloVe
- b. Word2Vec
- c. ELMo
- d. nltk

**Answer:** c)

ELMo word embeddings support the same word with multiple embeddings, this helps in using the same word in a different context and thus captures the context than just the meaning of the word unlike in GloVe and Word2Vec. Nltk is not a word embedding.

**42. For a given token, its input representation is the sum of embedding from the token, segment and position**

**embedding**

- a. ELMo
- b. GPT
- c. BERT
- d. ULMFit

**Answer:** c)

BERT uses token, segment and position embedding.

**43. Trains two independent LSTM language model left to right and right to left and shallowly concatenates them.**

- a. GPT
- b. BERT
- c. ULMFit
- d. ELMo

**Answer:** d)

ELMo tries to train two independent LSTM language models (left to right and right to left) and concatenates the results to produce word embedding.

**44. Uses unidirectional language model for producing word embedding.**

- a. BERT
- b. GPT
- c. ELMo
- d. Word2Vec

**Answer:** b)

GPT is a bidirectional model and word embedding is produced by training on information flow from left to right. ELMo is bidirectional but shallow. Word2Vec provides simple word embedding.

**45. In this architecture, the relationship between all words in a sentence is modelled irrespective of their position. Which architecture is this?**

- a. OpenAI GPT
- b. ELMo

- c. BERT
- d. ULMFit

**Ans:** c)

BERT Transformer architecture models the relationship between each word and all other words in the sentence to generate attention scores. These attention scores are later used as weights for a weighted average of all words' representations which is fed into a fully-connected network to generate a new representation.

**46. List 10 use cases to be solved using NLP techniques?**

- Sentiment Analysis
- Language Translation (English to German, Chinese to English, etc..)
- Document Summarization
- Question Answering
- Sentence Completion
- Attribute extraction (Key information extraction from the documents)
- Chatbot interactions
- Topic classification
- Intent extraction
- Grammar or Sentence correction
- Image captioning
- Document Ranking
- Natural Language inference

**47. Transformer model pays attention to the most important word in Sentence.**

- a. True
- b. False

**Ans:** a) Attention mechanisms in the Transformer model are used to model the relationship between all words and also provide weights to the most important word.

**48. Which NLP model gives the best accuracy amongst the following?**

- a. BERT
- b. XLNET

- c. GPT-2
- d. ELMo

**Ans:** b) XLNET

XLNET has given best accuracy amongst all the models. It has outperformed BERT on 20 tasks and achieves state of art results on 18 tasks including sentiment analysis, question answering, natural language inference, etc.

#### **49. Permutation Language models is a feature of**

- a. BERT
- b. EMMo
- c. GPT
- d. XLNET

**Ans:** d)

XLNET provides permutation-based language modelling and is a key difference from BERT. In permutation language modeling, tokens are predicted in a random manner and not sequential. The order of prediction is not necessarily left to right and can be right to left. The original order of words is not changed but a prediction can be random. The conceptual difference between BERT and XLNET can be seen from the following diagram.

#### **50. Transformer XL uses relative positional embedding**

- a. True
- b. False

**Ans:** a)

Instead of embedding having to represent the absolute position of a word, Transformer XL uses an embedding to encode the relative distance between the words. This embedding is used to compute the attention score between any 2 words that could be separated by n words before or after.

1. **What is Pragmatic Analysis, exactly?**
2. **What is POS tagging?**

3. **In NLP, what are stop words?**
4. **What exactly is NES?**
5. **What is the definition of information extraction?**
6. **Give two instances of real-world NLP uses.**
7. **List a few ways for tagging parts of speech.**
8. **Explain the N-gram model in NLP in a few words.**
9. **What exactly do you mean by word embedding?**
10. **List a few popular word embedding techniques.**

## **1. What are the natural language processing (NLP) project's lifecycle stages?**

The stages of a natural language processing (NLP) project's lifespan are as follows:

- **Data collection:** Refers to the process of gathering, measuring, and assessing accurate insights for study following defined recognized techniques.
- **Data Cleaning:** The process of removing or repairing inaccurate, corrupted, incorrectly formatted, duplicate, or incomplete data from a dataset is called data cleaning.
- **Data Pre-Processing:** Data preparation is transforming raw data into a usable format. Feature Engineering extracts the characteristics, traits, and attributes from raw data.

- **Data modeling:** Studying data items and their connections with other objects. It is used to investigate data requirements for a variety of commercial operations.
- **Model Evaluation:** A crucial phase in building a model is model evaluation. It aids in selecting the optimal model to describe our data and predict how well that model will function in the future.
- **Model Deployment:** Process of making an ML model available for real-world application.
- **Monitoring and Updating:** Evaluate and analyze production model performance to achieve acceptable quality defined by the use case. It sends out notifications when there are problems with performance and aids in detecting and treating the root cause.

## 2. What are some examples of typical NLP tasks?

NLP is used to do various tasks, including

- **Machine Translation:** This assists in translating a text from one language to another.
- **Text Summarization:** This is used to generate a concise summary of the complete text in the document based on a significant corpus.
- **Language modeling:** This determines how the following phrase will seem based on the history of preceding words. The auto-complete sentences tool in Gmail is an excellent illustration of this.

- **Topic modeling:** A technique for determining the subject organization of many documents. It identifies the actual topic of a piece of writing.
- **Query Answering:** This aids in automatically preparing replies based on a corpus of text and an asked question.
- **Conversational Agent:** These are voice assistants like Alexa, Siri, Google Assistant, Cortana, and others that we see all the time.
- **Information Retrieval:** This aids in retrieving relevant documents in response to a user's search query.
- **Information Extraction:** The extraction of useful information from a text, such as an email calendar event, is known as information extraction.
- **Text Classification:** This is used to categorize a given text into a set of categories based on its content. It is also utilized in many AI-based applications, including sentiment analysis and spam identification.

### **3. How do Conversational Agents work?**

Conversational Agents use the following NLP components:

- **Voice Recognition and Synthesis:** Speech recognition aids in converting speech impulses to phonemes, which are subsequently transcribed as words.
- **Natural Language Understanding (NLU):** The transcribed text from stage one is analyzed using AI algorithms within the natural language

understanding system. Named Entity Recognition, Text Classification, Language Modeling, and other NLP tasks are relevant here.

- **Management of Conversations:** After extracting the necessary data from text, we proceed to the stage of determining the user's purpose. The user's response can then be categorized as a pre-defined intent using a text classification system. This assists the conversational agent in determining what is being asked.
- **Generating Answer:** The agent creates an appropriate response based on the initial phases' semantic understanding of the user's purpose.

#### **4. What does data augmentation imply? What are some examples of data augmentation techniques used in NLP projects?**

NLP offers various ways for taking a small dataset and combining it with other data to build larger datasets. Data augmentation is the term for this. Language attributes are used to generate text that is syntactically comparable to the original text data.

The following are some examples of how data augmentation may be used in NLP projects:

- Entities are being replaced.
- Word substitution based on the TF-IDF
- Back data translation with noise
- Synonym replacement
- Bigram flipping.

#### **5. What procedures should you take while creating a text categorization system?**

The following stages are commonly performed while establishing a text categorization system:

1. Gather or create a labeled dataset suitable for the task.
2. After separating the dataset into two (training and test) or three sections (training, validation (i.e., development), and test sets), choose an assessment measure (s).
3. Unprocessed text is converted into feature vectors.
4. To train a classifier, use the feature vectors and labels from the training set.
5. Using the evaluation metric(s) from Step 2, compare the model's performance on the test set.
6. To service a real-world use case, deploy the model and track its performance to service a real-world use case.

## **6. Describe how parsing is done in NLP?**

The process of recognizing and comprehending a text's syntactic structure is parsing. It's done by dissecting the text's fundamental parts. Each word is parsed one at a time, then two at a time, three at a time. When the machine parses the text one word at a time, it's called a unigram. A bigram is a text in which two words are analyzed at a time. A trigram is a group of three words that the computer parses simultaneously.

The following points will help us understand why parsing is vital in NLP:

- The parser will report any syntax mistakes.
- It assists in the recovery of frequently recurring mistakes, allowing the rest of the program to be processed.
- The parse tree is created using a parser.

- The parser creates a symbol table, vital in natural language processing.
- In addition, to construct intermediate representations, a Parser is used (IR).

## **7. What exactly is a "Bag of Words" (BOW)?**

The Bag of Words model, which employs word frequency or occurrences to train a classifier, is a popular one. This approach creates a matrix of occurrences for texts or phrases, regardless of grammatical structure or word order.

A bag-of-words is a text representation that indicates how often words appear in a document. It consists of two steps:

A glossary of well-known terms—a metric for determining if they exist.

The document is a "bag" of words since all information about the order or organization of words is removed. The model is just concerned with whether or not recognized terms occur in the document, not with the place of those phrases.

## **8. What is Regular Grammar?**

Regular grammar represents a common language.

$A \rightarrow a$ ,  $A \rightarrow aB$ , and many additional rules exist in regular grammar. The rules allow for the identification and analysis of strings to be automated.

In regular grammar, there are four tuples:

- ‘N’ represents the non-terminal set.
- ‘ $\Sigma$ ’ represents the set of terminals.
- ‘P’ stands for the set of productions.
- ‘ $\epsilon N$ ’ denotes the start of non-terminal.

## **9. What is Latent Semantic Indexing (LSI) in Natural recovering using?**

The mathematical approach of Latent Semantic Indexing (LSI), also known as Latent Semantic Analysis, is used to improve the accuracy of information

retrieval. It facilitates the finding of hidden (latent) associations between words (semantics) by developing a collection of various ideas connected with a phrase's terms to improve information understanding. The NLP approach used for this is singular value decomposition. It works well with small sets of static content.

## **10. What are some of the measures used to assess NLP models?**

The following are some of the measures used to evaluate NLP models:

- Accuracy is utilized when the output variable is categorical or discrete. It is the model's proportion of right predictions relative to the total number of forecasts.
- The precision parameter indicates how accurate or exact the model's predictions are, i.e., how many positive (the class we're interested in) instances can the model properly identify.
- F1 score: This measure combines precision and recalls into a single metric that indicates the accuracy-recall trade-off, i.e., completeness and exactness.
- The formula for F1 is  $(2 \text{ Precision Recall}) / (\text{Precision} + \text{Recall})$ .
- The AUC measures the number of correct positive predictions vs the number of wrong positive predictions as the prediction threshold is modified.

## **11. What is Pragmatic Analysis, exactly?**

In NLP, pragmatic analysis is a crucial job for understanding knowledge that exists outside of a given document. Using pragmatic analysis aims to concentrate on a specific component of a document or text in a language. This necessitates a thorough understanding of the real world. The pragmatic analysis

helps software programs to know the true meaning of phrases and words through critical interpretation of real-world data.

## **12. How can data for NLP projects be obtained?**

There are several methods for obtaining data for NLP projects. The following are a few:

**Using publicly accessible datasets:** Datasets for NLP may be found on sites such as Kaggle and Google Datasets.

**Using data augmentation:** This technique produces new datasets from current ones.

**Scraping data from the web:** Using Python or other programming languages, one may scrape data from websites that aren't generally available in an organized format.

## **13. What do Text Extraction and Cleanup imply?**

Text extraction and cleaning is the process of extracting raw text from input data while removing all non-textual information such as markup, metadata, and other non-textual information and converting the text to the needed encoding type. This is usually determined by the format of the available data for the project.

The following are some of the most frequent methods for text extraction in NLP.

- Sentiment Analysis for Named Entity Recognition
- Summarization of Text
- Topic Modeling using Aspect Mining

## **14. What actions are required in resolving an NLP issue?**

The steps for addressing an NLP problem are as follows:

- Obtain the text via scraping the web or using the provided dataset.
- For text cleaning, use stemming and lemmatization.

- Feature engineering strategies should be used.
- Use word2vec to embed
- Neural networks or other Machine Learning techniques can train the created model.
- Examine the model's results.
- Make the necessary adjustments to the model.
- Set up the model.

## **15. What are Regular Expressions?**

To match and tag words, a regular expression is employed. It is made up of a set of characters that are used to match strings.

If A and B are regular expressions, then they must satisfy the following conditions:

- It is a regular language, then for it is a regular expression.
- $A + B$  is a regular expression within the language A, B if A and B are regular expressions.
- The concatenation of A and B ( $A \cdot B$ ) is a regular expression if A and B are regular expressions.
- $A^*$  (A occurring multiple times) is a regular expression if A is a regular expression.

## **16. What is the difference between Natural Language Processing (NLP) and Natural Language Understanding (NLU)?**

**Natural Language Processing (NLP)**

- NLP is a system that handles simultaneous end-to-end talks between computers and people.
- In NLP, both humans and robots are engaged.
- NLP is concerned with understanding language in its purest form, as stated.
- Grammar, structure, typography, and point of view may all be used to parse text using NLP.

### **Natural Language Understanding(NLU)**

- NLU assists in resolving Artificial Intelligence's most complex challenges.
- NLU transforms unstructured inputs into structured text, allowing machines to comprehend them
- NLU, on the other hand, focuses on obtaining context and meaning or determining what was intended.
- NLU will assist the machine in deducing the meaning of the linguistic material.

### **17. What is a Masked Language Model, and how does it work?**

By generating an output from the defective input, masked language models assist learners in comprehending deep representations in downstream tasks. This approach is frequently used to anticipate the words in a phrase.

### **18. What is POS tagging?**

POS tagging, or parts of speech tagging, is the basis for identifying individual words in a document and classifying them as part of speech based on their

context. Because it entails analyzing grammatical structures and selecting the appropriate component, POS tagging is also known as grammatical tagging.

Because the same word might be several parts of speech depending on the context, POS tagging is a complicated procedure. Because of the same reason, the same general approach used for word mapping is unsuccessful for POS tagging.

## **19. What exactly is NES?**

The practice of recognizing certain entities in a text document that are more informative and have a distinct context is known as named entity recognition (NER). These are frequently referred to as places, individuals, organizations, and others. Even though these things appear to be proper nouns, the NER approach does not recognize them. In reality, NER entails entity chunking or extraction, which includes segmenting entities into many specified classes. This stage also aids in the extraction of data.

## **20. What exactly is NLTK? What distinguishes it from Spacy?**

Natural Language Toolkit (NLTK) is a set of libraries and applications for processing symbolic and statistical natural language. This toolkit includes some of the most sophisticated libraries for breaking down and understanding human language using machine learning approaches. Lemmatization, Punctuation, Character Count, Tokenization, and Stemming are all done with NLTK. The following are the differences between NLTK and Spacey:

- While NLTK provides various programs to pick from, Spacey's toolkit only contains the best-suited algorithm for a given scenario.
- In comparison to Spacey, NLTK supports many languages (Spacey supports only seven languages)
- NLTK provides a string processing library, but Spacey has an object-oriented library. Spacey can handle word vectors, whereas NLTK cannot.

## **21. What is the definition of information extraction?**

In the context of Natural Language Processing, information extraction refers to the process of mechanically extracting structured information from unstructured

sources to assign meaning to it. This might involve retrieving entity properties, relationships between entities, and more. The following are some examples of information extraction models:

- Module for Taggers
- Module for Extracting Relationships
- Module for Fact Extraction
- Module for Extracting Entities
- Module for Sentiment Analysis
- Module for Network Graphs
- Module for Document Classification and Language Modeling.

## **22. What are the most effective NLP tools?**

Some of the most excellent open-source NLP tools are

- SpaCy
- TextBlob
- Textacy
- Natural language Toolkit (NLTK)
- Retext
- NLP.js
- Stanford NLP
- CogcompNLP.

## **23. List 10 use cases to be solved using NLP techniques?**

- Sentiment Analysis

- Language Translation (English to German, Chinese to English, etc..)
- Document Summarization
- Question Answering
- Sentence Completion
- Attribute extraction (Key information extraction from the documents)
- Chatbot interactions
- Topic classification
- Intent extraction
- Grammar or Sentence correction
- Image captioning
- Document Ranking
- Natural Language Inference.

#### **24. In NLP, what are stop words?**

Stop words are common words that appear in sentences and provide weight to the phrase. These stop words serve as a link between phrases, ensuring grammatically accurate. Stop words are taken out before natural language data is processed, and they are a frequent pre-processing strategy.

#### **25. In NLP, what is stemming?**

Stemming is the process of extracting the root word from a given term. With efficient and well-generalized principles, all tokens may be broken down to retrieve the root word or stem. It's a rule-based system that's well-known for its ease of use.

#### **26. Give two instances of real-world NLP uses?**

**1. Spelling/Grammatical Checking Apps:** NLP algorithms are used in mobile applications and websites that help users fix grammar problems in the submitted

text. These days, they may even suggest the next few words that the user might input, thanks to the employment of particular NLP models on the backend.

**2. ChatBots:** Many websites now provide customer service via virtual bots that talk with users and help them solve problems. It functions as a filter for concerns that do not require engagement with the customer service representatives of the firms.

## **27. Define Dependency Parsing?**

Dependency parsing is a technique for understanding grammatical structure by highlighting the relationships between its components. It investigates how the words of a phrase are related linguistically. Dependencies are the names given to these connections.

## **28. What is the difference between false positives and false negatives?**

A false negative occurs when a machine learning system incorrectly forecasts a negative outcome as positive.

A false positive is defined as a machine learning system that incorrectly forecasts a positive outcome as a negative.

## **29. List a few ways for tagging parts of speech?**

Rule-based tagging, HMM tagging, transformation-based tagging, and memory-based tagging are all examples of tagging techniques.

## **30. List a few examples of how the n-gram model is used in the real world?**

1. Communication Enhancement
2. Tagging of parts of speech
3. Generation of natural language
4. The similarity of Words
5. Identification of Authorship
6. Sentiment Analysis
7. Text Input Predictive.

## **31. In NLP, what is the bigram model?**

A bigram model is an NLP model that uses the conditional probability of the preceding word to predict the likelihood of a word in a phrase. It is critical to know all of the last words to calculate the conditional probability of the previous word.

### **32. What are your impressions of the Masked Language Model?**

The Masked Language Model is a model that takes a phrase as input and attempts to finish it by accurately predicting a few concealed (masked) words.

### **33. List the types of sorts available in linguistic ambiguity?**

**1. Lexical Ambiguity:** This sort of ambiguity occurs when a phrase has homonyms and polysemy.

**2. Syntactic Ambiguity:** Syntactic ambiguity occurs when the grammar of a statement allows for many interpretations.

**3. Semantic Ambiguity:** When a statement comprises ambiguous words or phrases with unclear meanings, this ambiguity occurs.

### **34. Explain the N-gram model in NLP in a few words?**

The N-gram model is an NLP model that predicts the likelihood of a word in a phrase based on the conditional probability of n-1 preceding terms. The essential idea behind this method is that rather than utilising all of the previous words to predict the future word, we simply utilise a handful of them.

### **35. What is the bigram model's Markov assumption?**

For the bigram model, the Markov assumption assumes that the probability of a word in a phrase depends solely on the preceding word in that sentence rather than all last words.

### **36. What exactly do you mean by word embedding?**

Word embedding is the method of expressing textual data using a real-number vector in natural language processing. This technique allows words with similar meanings to be represented simultaneously.

### **37. What is an embedding matrix, and how does it work?**

A word embedding matrix is a matrix that contains all of the words in a text's embedding vectors.

### **38. List a few popular word embedding techniques?**

A few word embedding approaches are listed below.

- Word2Vec Glove
- Embedding Layer.

### **39. What are the first few steps you'll take before applying a natural language processing (NLP) machine-learning algorithm on a corpus?**

1. Eliminating white spaces
2. Eliminating Punctuation
3. Lowercase to Uppercase Conversion
4. Tokenisation
5. Getting Rid of Stopwords
6. Lemmatization.

### **40. What is the difference between an hapax and an hapax legomenon?**

Hapaxes are unusual words that only appear once in a sample text or corpus. Each one is referred to as an hapax or hapax legomenon ('read-only once' in Greek). It's also known as a singleton.

#### **1. Who is the person who created Python?**

- Dennis Ritchie
- James Gosling
- Guido Van Rossum
- Graham Bell

#### **2. What are some of the Python Software Foundation's (PSF) well-known sponsors?**

- Google, Bloomberg, Meta
- AWS, NVIDIA
- Microsoft, Salesforce, CapitalOne, Corning

All the above

**3. After finding inspiration in the language, Guido Van Rossum began creating the Python programming language.**

- C
- C
- C++
- Java
- ABC

**4. Python is a programming language at level .**

- Low level
- Medium Level
- High Level
- None

**5. The version of Python that is most frequently used is:**

- 1.5
- 2.0
- 2.7
- 3.1

1) What is the full form of NLP? / What is Natural Language Processing?

NLP stands for "**Natural Language Processing**". NLP is a field of computer science that deals with communication between computer systems and humans. This technique uses Artificial Intelligence and Machine Learning to create automated software that helps understand the human spoken languages and extract useful information from the data gathered from the audio.

The techniques used in NLP allow computer systems to process and interpret data in the form of natural languages. It designs algorithms that can extract meaning from large datasets in audio or text format by applying machine learning algorithms. In other words, we can say that NLP is software that uses artificial

intelligence and machine learning algorithms to understand natural languages or the way human beings read and write in a language and extracts required information from such data.

---

2) What are some real-life applications / real-world examples of Natural Language Processing (NLP)?

Some real-life applications of NLP or Natural Language Processing are as follows:

**Spelling/Grammar Checking Apps:** Spelling and grammar checking applications are real-life examples of Natural Language Processing. These apps are mainly used in mobile applications and websites that facilitate users to correct grammar mistakes in the entered text rely on NLP algorithms. They also recommend the best possible substitutes that the users might type. This is possible because of specific NLP models being used in the backend.

**Google Translate:** Google Translate is the most famous application of Natural Language Processing. Using this, you can convert your written or spoken sentences into any language. You can also get the correct pronunciation and meaning of a word by using Google Translate. The Google Translate application uses some advanced techniques of Natural Language Processing to provide translation of sentences into various languages.

**Chatbots apps:** Chatbots applications provide a better customer support service. Many websites and companies use this to offer customer support through these virtual bots that chat with the user and resolve their problems. Many companies use chatbots for 24/7 service to resolve the basic queries of customers. Generally, it filters the basic issues that do not require an interaction with the companies' customer executives. It makes the customers feel that the customer support team quickly attends them. If a chatbot cannot resolve any user's query, it forwards it to the support team while still engaging the customer. Chatbots also make companies capable of building cordial relations with customers. These all are only possible because of Natural Language Processing.

---

3) What are the most used NLP (Natural Language Processing) Terminologies?

Following is the list of most used NLP (Natural Language Processing) Terminologies:

- **Preprocessing:** This is a method used to remove unwanted text or noise from the given text and make it "clean." It is the first step of any NLP task.
  - **Documents:** Documents are the body of text and are collectively used to form a corpus.
  - **Corpus, or Corpora (Plural):** It is a collection of text of similar type, for example, movie reviews, social media posts, etc.
  - **Vocabulary:** It is a group of terms used in a text or speech.
  - **Out of Vocabulary:** It specifies the terms not included in the vocabulary. We put the terms created during the model's training in this category.
  - **Tokenization:** It is used in NLP to break down large sets of text into small parts for easy readability and understanding. Here, the small parts are referred to as 'text' and provide a piece of meaningful information.
  - **N-grams:** It specifies the continuous sequence (similar to the power set in number theory) of n-tokens of a given text.
  - **Parts of Speech (POS):** It specifies the word's functions, such as a noun, verb, etc.
  - **Parts of Speech Tagging:** It is the process of tagging words in the sentences into different parts of speech.
- 

#### 4) What are the most used NLP (Natural Language Processing) Terminologies?

Following is the list of most used NLP (Natural Language Processing) Terminologies:

- **Embeddings (Word):** This process is used to embed each token as a vector and then pass it into a machine learning model. We can apply embeddings also on phrases and characters, apart from words.
- **Stop Words:** These are used to remove the unwanted text from further text processing, for example, a, to, can, etc.
- **Transformers:** Transformers are deep learning architectures that can parallelize computations. They are used to learn long-term dependencies.
- **Normalization:** This is a process of mapping similar terms to a canonical form, i.e., a single entity.

- **Lemmatization:** Lemmatization is a type of normalization used to group similar terms to their base form according to their parts of speech. For example, talking and talked can be mapped to a single term, talk.
  - **Stemming:** Stemming is also a type of normalization similar to lemmatization. But, it is different in the term that it segregates the words without the parts of speech tags. It is faster than lemmatization and also be more precise in some cases.
- 

5) What are some of the major components of Natural Language Processing?

Following is a list of some of the major components of Natural Language Processing:

**Entity extraction:** It is used for segmenting a sentence to identify and extract entities, such as a person (real or fictional), organization, geographies, events, etc.  
85

**Pragmatic analysis:** Pragmatic analysis extracts information from the input text. It is part of the process of data extraction.

**Syntactic analysis:** Syntactic analysis is used for the proper ordering of words.

---

6) What do you understand by Dependency Parsing in NLP or Natural Language Processing?

In Natural Language Processing, Dependency Parsing is a process of assigning syntactic structure to a sentence and identifying its dependency parses. This is an important process to understand the correlations between the "head" words in the syntactic structure. That's why it is also known as syntactic parsing.

The process of dependency parsing becomes a little complex if there are more sentences that have more than one dependency parses. Multiple parse trees are known as ambiguities. The main task of dependency parsing is to resolve these ambiguities to assign a syntactic structure to a sentence effectively. It is also used in semantic analysis apart from syntactic structuring.

---

7) What are some most common areas of usage of Natural Language Processing?

Following is a list of some most common areas of usage of Natural Language Processing:

- Semantic Analysis
- Text classification
- Automatic summarization
- Questioning Answering

Some real-life examples of Natural Language Processing are chatbots, IOS Siri, Google Assistant, Amazon echo, Spelling, grammar checking apps, and Google translate.

---

8) What do you understand by NLTK in Natural Language Processing?

In Natural Language Processing, NLTK stands for Natural Language Toolkit. It is a Python library used to process data in human spoken languages. NLTK facilitates developers to apply parsing, tokenization, lemmatization, stemming techniques, and more to understand natural languages. It is also used for categorizing text, parsing linguistic structure, analyzing documents, etc.

Following is the list of some libraries of the NLTK package that are often used in NLP:

- DefaultTagger
  - UnigramTagger
  - RegexpTagger
  - backoff\_tagger
  - SequentialBackoffTagger
  - UnigramTagger
  - BigramTagger
  - TrigramTagger
  - treebank
  - wordnet
  - FreqDist
  - Patterns etc.
-

9) What is the use of TF-IDF? Why is it used in Natural language Processing?

In Natural language Processing, tf-idf, TF-IDF, or TFIDF stands for Term Frequency-Inverse Document Frequency. It is a numerical statistic used to specify how important a word is to a document in a collection or the collection of a set.

---

10) What is the difference between formal and natural languages?

The main difference between a formal language and a natural language is that a formal language is a collection of strings. Each string contains symbols from a finite set called alphabets. On the other hand, a natural language is a language that humans use to speak. This is completely different from a formal language as it contains fragments of words and pause words like uh, um, etc.

---

11) What are the tools used for training NLP models?

The most common tools used for training NLP models are NLTK, spaCY, PyTorch-NLP, openNLP, etc.

---

12) What do you understand by information extraction? What are the various models of information extraction?

In Natural Language Processing, information extraction is a technique of automatically extracting structured information from unstructured sources to get useful information. It extracts information such as attributes of entities, the relationship between different entities, and more.

Following is a list of various models of information extraction in Natural Language Processing:

- Fact Extraction Module
- Entity Extraction Module
- Sentiment Analysis Module
- Tagger Module
- Relation Extraction Module

- Network Graph Module
  - Document Classification and Language Modeling Module
- 

### 13) What are the stop words in Natural Language Processing?

In Natural Language Processing, stop words are regarded as useless data for a search engine. It includes the words like articles, prepositions, was, were, is, am, the, a, an, how, why, and many more. The algorithm used in Natural Language Processing eliminates the stop words to understand and analyze the meaning of the sentences. Eliminating the stop words is one of the most important tasks for search engines to process data.

Software developers design the algorithms of search engines so that they ignore the use of stop words and only show the relevant search result for a query.

---

### 14) What is Bag of Words in Natural Language Processing?

Bag of Words is a commonly used model in Natural Language Processing that depends on word frequencies or occurrences to train a classifier. This model creates an occurrence matrix for documents or sentences without depending on their grammatical structure or word order.

---

### 15) What do you understand by semantic analysis? What are the techniques used for semantic analysis?

Semantic analysis is a process that makes a machine understand the meaning of a text. It uses several algorithms to interpret the words in sentences. It is also used to understand the structure of a sentence.

Following are the techniques used for semantic analysis:

**Named entity recognition:** This technique is used to specify the process of information retrieval that helps identify the entities like the name of a person, organization, place, time, emotion, etc.

**Natural language generation:** This technique specifies a process used by the software to convert the structured data into human spoken languages. By using

natural language generation, organizations can automate content for custom reports.

**Word sense disambiguation:** It technique is used to identify the sense of a word used in different sentences.

---

16) What is pragmatic ambiguity in NLP?

Pragmatic ambiguity is used to specify words with more than one meaning, and they can be used in any sentence depending on the context. In pragmatic ambiguity, words have multiple interpretations.

Pragmatic ambiguity occurs when the meaning of the words is not specific. For example, if a word gives different meanings. Due to pragmatic ambiguity, a sentence can have multiple interpretations. Sometimes, we come across sentences that have words with multiple meanings, making the sentence open to interpretation.

---

17) What is Latent Semantic Indexing (LSI)? What is the use of this technique?

LSI or Latent Semantic Indexing is a mathematical technique used in Natural Language Processing. This technique is used to improve the accuracy of the information retrieval process. The LSI algorithm is designed to allow machines to detect the latent correlation between semantics.

The machines generate various concepts to enhance information understanding. The technique used for information understanding is called singular value decomposition. It is mainly used to handle static and unstructured data. This is one of the best-suited models to identify components and group them according to their types.

Latent Semantic Indexing or LSI is based on a principle that specifies that words carry a similar meaning when used in a similar context. The computational LSI models are slow compared to other models, but they can improve a text or document's analysis and understanding.

---

18) What do you understand by MLM in Natural Language Processing?

In Natural Language Processing, MLM is a term that stands for Masked Language Model. It helps learners understand deep representations in downstream tasks by taking the output from the corrupt input.

This model is mainly used to predict the words used in a sentence.

---

19) What are the most commonly used models to reduce data dimensionality in NLP?

The most commonly used models to reduce the dimensionality of data in NLP are TF-IDF, Word2vec/Glove, LSI, Topic Modelling, Elmo Embeddings, etc.

---

20) What is Lemmatization in Natural Language Processing?

Lemmatization is a process of doing things properly using a vocabulary and morphological analysis of words. It is mainly used to remove the inflectional endings only and return the base or dictionary form of a word, known as the lemma. It is just like cutting down your beard or shaving to get the original shape of your face.

**For example:** girl's = girl, bikes= bike, leaders= leader etc.

So, the main task of Lemmatization is to identify and return the root or original words of the sentence to explore various additional information.

---

21) What is Stemming in Natural Language Processing?

Stemming is a process of extracting the base form of a word by removing the affixes from them. It is just like cutting down the branches of a tree to its stems.

**For example:** After stemming, the words go, goes, and going would be 'go'.

Search engines use stemming for indexing the words. It facilitates them to store only the stems rather than storing all forms of a word. By using stemming, the search engines reduce the size of the index and increase the retrieval accuracy.

---

22) What is the difference between Stemming and Lemmatization in NLP?

Stemming and Lemmatization are both the text normalization techniques used in Natural language Processing. Both are used to prepare text, words, and documents for further processing. They seem very similar techniques, but there are quite differences between them. Let's see the main differences between them:

Stemming	Lemmatization
Stemming is the process of extracting the base form of a word by removing the affixes from them. It produces the morphological variants of a root/base word. Stemming programs are commonly known as stemming algorithms or stemmers.	Lemmatization is a more advanced process and looks beyond word reduction, just like stemming. It considers a full vocabulary of a language and applies a morphological analysis to words. For example, the lemma of 'went' is 'go', and the lemma of 'mice' is 'mouse'.
Stemming is not as much informative as Lemmatization. It is a somewhat crude method for cataloging related words. It essentially cuts letters from the end until the stem is reached.	Lemmatization is much more informative than simple Stemming; that is why Spacy has opted to only have Lemmatization available instead of Stemming.
Stemming is not as efficient as Lemmatization. This method works fairly well in most cases, but unfortunately, English has many exceptions requiring a more sophisticated process.	Lemmatization is more efficient than Stemming as it works well in exceptional words.
Following are some examples of Stemming: run: runner: running: ran: runs: easily: fairly: fair etc.	Following are some examples of Lemmatization: run: runner: running: ran: runs: goes: go: went: saw: mice: mouse

---

23) Which NLP techniques use a lexical knowledge base to obtain the correct base form of the words?

The NLP techniques that use a lexical knowledge base to obtain the correct base form of the words are Lemmatization and stemming.

---

#### 24) What is tokenization in Natural Language Processing?

In Natural Language Processing, tokenization is a method of dividing the text into various tokens. These tokens are the form of the words, just like a word forms into a sentence. In NLP, the program computers process large amounts of natural language data. These large amounts of natural language data have to be cut into shorter forms. So, tokenization is an important step in NLP that cuts the text into minimal units for further processing.

---

#### 25) What are some open-source libraries used in NLP?

Some popular open-source libraries used in NLP are NLTK (Natural Language ToolKit), SciKit Learn, Textblob, CoreNLP, spaCY, Gensim, etc.

---

#### 26) What are the key differences between NLP and NLU?

Following is the list of key differences between NLP and NLU:

NLP	NLU
NLP is a short form of Natural Language Processing.	NLU is a short form of Natural Language Understanding.
NLP or Language Processing is used to create a system that can make and establish communication between humans and computers.	NLU or Natural Language Understanding provides techniques that can solve complicated problems related to machine understanding.
It includes all the techniques required for the interaction between computers and humans.	It converts the uncategorized input data into a structured format and allows the computers to understand the data.
It includes the techniques focused on analyzing "what is said?"	It includes the techniques to understand "what is meant?"

---

27) What are the key differences between NLP (Natural Language Processing) and CI (Conversational Interface)?

Following is the list of key differences between NLP (Natural Language Processing) and CI (Conversational Interface):

Natural Language Processing (NLP)	Conversational Interface (CI)
The full form of NLP is Natural Language Processing.	The full form of CI is Conversational Interface.
The main focus of NLP is to make computers understand and learn how the normal human being languages' concepts work.	The main and only focus of CI is to provide users with an interface to interact with.
Natural Language Processing uses AI technology to identify, understand, and interpret users' requests through languages.	CI or Conversational Interface uses voice, chat, videos, images, and other conversational aid to create the user interface for communication.

---

28) What do you understand by Pragmatic Analysis?

Pragmatic analysis is an important task used in Natural Language Processing for interpreting knowledge lying outside a given document. It is mainly implemented to focus on exploring a different aspect of the document or text in a language. It requires a comprehensive knowledge of the real world to make software applications capable of critical interpretation of the real-world data to know the actual meaning of sentences and words.

**For example, see the following sentence:**

'Do you know what time it is?'

This sentence can be used to ask for knowing the time or for yelling at someone to make them note the time. It completely depends on the context in which this sentence is used.

---

29) What are the best open sources of NLP Tools available in the market?

Some of the best open sources NLP tools available in the market are:

- SpaCy
  - TextBlob
  - Textacy
  - Natural language Toolkit (NLTK)
  - Retext
  - NLP.js
  - Stanford NLP
  - CogcompNLP etc.
- 

30) How can you differentiate Artificial Intelligence, Machine Learning, and Natural Language Processing?

Following are the key differences between Artificial Intelligence, Machine Learning, and, Natural Language Processing:

<b>Artificial Intelligence</b>	<b>Machine Learning</b>	<b>Natural Language Processing</b>
Artificial Intelligence is a technique used to create smarter machines and computers.	Machine Learning is a term used for systems that learn from experience.	Natural Language Processing or NLP is the set of systems that can understand the languages used by humans and process these languages to make them understood by computers.
Artificial Intelligence requires human intervention. Without human intervention, it is not possible to create intelligent machines.	Machine Learning doesn't require human intervention. It purely involves the working of computers and machines.	Natural Language Processing uses both computer and human languages to work properly.
Artificial Intelligence is a broader concept than Machine Learning. It includes a lot of working fields.	Machine Learning is a narrow concept and is a subset of Artificial Intelligence.	Natural Language Processing uses the concept of both Artificial Intelligence and Machine Learning to make the tools that can process human language and make it understandable by machines.

---

31) What do you understand by POS tagging?

The full form of POS tagging is Parts of speech tagging. It is most commonly known as POS tagging. According to their context, it specifies a process of identifying specific words in a document and groups them as part of speech.

POS tagging is also known as grammatical tagging because it involves understanding grammatical structures and identifying the respective component. It is a very complicated process because the same word can be different parts of speech depending on the situation and the structure of the sentence.

---

32) What is NES in Natural Language Processing? Why is it used?

NES is an acronym that stands for Name Entity Recognition. It is used in Natural Language Processing and is most commonly known as NER. It is the process of identifying specific entities in a text document that is more informative and have a unique context. It includes places, people, organizations, and more. After identification, it extracts these entities and categorizes them under different predefined classes. Later, this step helps in extracting information.

---

33) What is parsing in Natural Language Processing? What are the different types of parsing used in NLP?

Parsing is a technique or a method of analyzing the sentences automatically according to their syntactic structure.

Following is a list of different types of parsing used in Natural Language Processing:

**Dependency parsing / Syntactic parsing:** Dependency parsing is also known as syntactic parsing. It recognizes a dependency parse of a sentence and assigns a syntax structure to the sentence. It mainly focuses on the relationship between different words.

**Semantic parsing:** Semantic parsing is a method of converting the natural language into machine language that a computer can understand and process.

**Constituency parsing:** Constituency parsing is a specific parsing method where a division of sentences is divided into sub-parts or constituencies. It is mainly

used to extract a constituency-based parse tree from the constituencies of the sentences.

**Shallow parsing / Light parsing:** Shallow parsing is also known as light parsing and chunking. It identifies constituents of sentences and then links them to different groups of grammatical meanings.

---

34) What is language modeling in NLP?

In Natural Language Processing, language modeling creates a probability distribution of a sequence of words. It provides probability to all the words present in that sequence.

---

35) What is topic modeling in NLP?

In NLP, topic modeling is finding abstract topics in a document or set of documents to find hidden semantic structures.

---

36) What is the key difference between dependency parsing and shallow parsing?

The key difference between dependency parsing and shallow parsing is that dependency parsing is the process of finding a relation between all the different words. On the other hand, shallow parsing is the parsing of a selected limited part of the information.

---

37) What do you understand by Pragmatic Ambiguity in NLP?

In Natural Language Processing, pragmatic ambiguity specifies multiple descriptions of a word or a sentence. It occurs when the words of the sentence may have different meanings, and the correct meaning of the sentence is not clear. In this case, it becomes very difficult for a machine to understand a sentence's meaning, which causes pragmatic ambiguity.

**For example, see the following sentence:**

"Are you feeling hungry?"

The above sentence could be either a generally asked question or a formal way of offering food.

---

### 38) What are the steps used to solve an NLP problem?

Following is a list of steps used to solve an NLP problem:

- In the first step, get the text from the available dataset.
  - Now, apply stemming and lemmatization to clean the text.
  - Now, apply feature engineering techniques to the received text.
  - Embed using word2vec.
  - Now, train the built model using neural networks or other Machine Learning techniques.
  - Now it turns to evaluate the model's performance.
  - Make the appropriate changes in the model.
  - Now, the model is complete. Deploy the model.
- 

### 39) What is noise removal in NLP? Why is it used?

Noise removal is one of the NLP techniques. As the name specifies, it is used to remove unnecessary pieces of text from the sentences.

#### 1. What Is Natural Language Processing?

While this may sound like a softball NLP interview question, the way you answer it will clue the interviewer into how well you grasp NLP as a whole.

Natural language processing (NLP) is an automated way to understand or analyze the nuances and overall meaning of natural language, extracting key information from typed or spoken language by applying machine learning algorithms. Since meaning is largely derived from its context, NLP seeks to understand language beyond the literal and allow machines to learn through experience.

#### 2. What is an NLP pipeline, and what does it consist of?

Generally, NLP problems can be solved by navigating the following steps (referred as a pipeline):

- Gathering text, whether it's from web scraping or the use of available datasets
- Cleaning text (through the processes of stemming and lemmatization)
- Representation of the text (bag-of-words method)
- Word embedding and sentence representation (Word2Vec, SkipGram model)
- Training the model (via neural nets or regression techniques)
- Evaluating the model
- Adjusting the model, as needed
- Deploying the model

### 3. What does “parsing” mean in the world of NLP?

To “parse” a document, in the context of NLP, is to make sense of its grammatical structure. For example, an NLP application parses text by determining the relationship of words and phrases within the text (e.g., which words are the subject, or object, of a given verb?). Parsing will differ from one set of text to another, since its goal is to understand the grammar and what the writer or speaker is trying to convey.

### 4. What is “named entity recognition”?

This will likely be one of the NLP interview questions you will get. Named entity recognition (NER) is an NLP process that separates out the components of a sentence to summarize it into its main components, similar to sentence diagramming in grade school. For example, the sentence “Bob moved to New York City in 1997” may be categorized as:

- Bob = name
- New York City = city/location
- 1997 = time

NER helps machines understand the context of the document by identifying data related to “who, what, when, and where.” It’s very useful for scanning documents and responding to chatbots in a customer service environment.

## 5. What is a “stop” word?

Articles such as “the” or “an,” and other filler words that bind sentences together (e.g., “how,” “why,” and “is”) but don’t offer much additional meaning are often referred to as “stop” words. In order to get to the root of a search and deliver the most relevant results, search engines routinely filter out stop words.

## 6. What is “feature extraction” and how is it accomplished using NLP?

The process of feature extraction involves the identification of certain key words or phrases that put it into a particular category, often based on the author’s purported sentiment. For example, a product review by a customer that uses the word “great” or the phrase “good quality” could be summarized as a positive review. The feature extraction process in NLP could enable a given phrase or use of certain words to be “tokenized” into the positive review category.

## 7. How do you test an NLP model? What metrics are used?

NLP models should be tested for accuracy, but also must consider the likelihood of false positives and false negatives due to the complexity and nuances of language. Therefore, while accuracy is important, you also want to test an NLP model using the following metrics:

- Recall. This is expressed by the following equation:

$$\text{True Positive} / (\text{True Positive} + \text{False Negative}) = \text{True Positive} / \text{Total Actual Positive}$$

- Precision. This is expressed by the following equation:

$$\text{True Positive} / (\text{True Positive} + \text{False Positive}) = \text{True Positive} / \text{Total Predicted Positive}$$

- F1. This is a combination of recall and precision and is expressed by the following equation:

$$F1 = 2 \times \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$$

## 8. What are two applications of NLP used today?

There are several real-world NLP applications in use today, including:

- Chatbots. Chatbots (powered by NLP) are often the starting point for customer service interactions, designed to resolve basic customer queries and funnel them to the right personnel if the chatbot is unable to provide resolution. These provide efficiency and cost savings for companies.
- Online translation. Services such as Google Translate use NLP to convert both written and spoken language into other languages and can also help with pronunciation.

## 9. What is “term frequency-inverse document frequency?”

Term frequency-inverse document frequency (**TF-IDF**) is an indicator of how important a given word is in a document, which helps identify key words and assist with the process of feature extraction for categorization purposes. While “TF” identifies how frequently a given word or phrase (“W”) is used, “IDF” measures its importance within the document. The formulas to answer this NLP interview question are as follows:

- $TF(W) = \text{Frequency of } W \text{ in a document} / \text{Total number of terms in the document}$
- $IDF(W) = \log_e (\text{Total number of documents} / \text{Number of documents having the term } W)$

Using these formulas, you can determine just how important a given word or phrase is within a document. If the TF-IDF is high, then the frequency of that term is lower; if the TF-IDF is low, then its frequency is higher. Search engines use this to help them rank sites.

## 10. What is “latent semantic indexing?”

Undoubtedly, this will be one of the NLP interview questions that you will be asked: What is latent semantic indexing? Latent semantic indexing (LSI) is used to extract useful information from unstructured data by identifying different words and phrases that have the same or similar meanings within a given

context. It's a mathematical method for determining context and obtaining a deeper understanding of the language, widely used by search engines.

1. What is the NLG (Natural Language Generation)?
2. What is the order of steps in natural language understanding?
3. What is signal processing in NLP?
4. What is pragmatic analysis in NLP?
5. What is syntactic analysis in NLP?
6. What is semantic analysis in NLP?
7. What is sentiment analysis in NLP?
8. What is discourse analysis in NLP?
9. What is pragmatic ambiguity in NLP?
10. What are the major applications of NLP?
11. List any real-world application of NLP?
12. What are the common NLP techniques?
13. What are the components of NLP?
14. What are the tools used for training NLP models?
15. Which NLP technique uses a lexical knowledge base to obtain the correct base form of the words?
16. List the models to reduce the dimensionality of data in NLP.
17. List some open-source libraries for NLP.
18. Explain the masked language model.
19. What is the bagofwords (bag of words) model?
20. What is CBOW in NLP?
21. What is TF-IDF, and what are its uses?
22. What are POS and tagging?
23. What is n-gram in NLP?
24. What is skip-gram?
25. What is the corpus in NLP?
26. What are the features of the text corpus in NLP?
27. What is normalization in NLP?
28. What is keyword normalization?
29. What is lemmatization in NLP?
30. What is stemming in NLP?
31. What is ambiguity in NLP?
32. What is tokenization in NLP?
33. What are stop words in NLP?
34. How to find word similarity in NLP?
35. How to find sentence similarity in NLP?
36. How to find document similarity in NLP?
37. What are transformers?
38. What are punctuations in NLP, and how can we remove them?
39. What is latent semantic indexing (LSI)?

- 40.What is named entity recognition (NER)?
- 41.What is NLTK in NLP?
- 42.What is spaCY?
- 43.What is openNLP?
- 44.What is the difference between NLTK and openNLP?
- 45.What is parsing?
- 46.What is dependency parsing?
- 47.What is semantic parsing?
- 48.What is constituency parsing?
- 49.What is shallow parsing?
- 50.What are the differences between dependency parsing and shallow parsing?
- 51.What is language modeling?
- 52.What is topic modeling?
- 53.What is text summarization in NLP?
- 54.What is the difference between a regular expression (RegEx) and regular grammar?
- 55.What is perplexity in NLP?
- 56.What is the Naive Bayes algorithm, and where is it used in NLP?
- 57.What is the PageRank algorithm?
- 58.What is noise removal?
- 59.What is word embedding?
- 60.What are the word embedding libraries?
- 61.What is word2vec?
- 62.What is doc2vec?
- 63.What is a document-term matrix?
- 64.What is wordnet?
- 65.What is GloVe in NLP?
- 66.What is a flexible string matching?
- 67.What is cosine similarity?
- 68.What is information extraction?
- 69.What is object standardization, and when is it used?
- 70.What is text generation, and when is it done?
- 71.How can we estimate the entropy of the English language?
- 72.What is Latent Dirichlet Allocation?
- 73.What are the conditional random fields?
- 74.What are the hidden Markov random fields?
- 75.What is a coreference resolution?
- 76.What is PAC learning?
- 77.What is sequence learning?
- 78.What is an ensemble method?

## **1. What is the NLG (Natural Language Generation)?**

Natural Language Generation is a part of AI and generates natural language texts from structured data to produce an output. It can be seen as NLP's reverse process, where NLP is used to understand and interpret the natural language to form data, and NLU is used to generate outputs in natural language from structured data.

## **2. What is the order of steps in natural language understanding?**

The order of steps that are to be followed in Natural Language Understanding is as follows:

1. Signal Processing
2. Syntactic Analysis
3. Semantic Analysis
4. Pragmatics

## **3. What is signal processing in NLP?**

Signal processing is a method that enables software to analyze, modify, and synthesize signals. In NLP, these can be sound or text signals.

## **4. What is pragmatic analysis in NLP?**

The pragmatic analysis is the process of information extraction from the given text. It is a set of linguistic and logical tools that enable us to churn out the meaning of the given structure of a text.

## **5. What is syntactic analysis in NLP?**

The syntactic analysis, also referred to as parsing and syntax analysis, is the phase in which we try to process the given text's structure. This process tries to draw meaning from the text by comparing it to formal grammar rules or syntax.

## **6. What is semantic analysis in NLP?**

The semantic analysis is the process of understanding the meaning of the text in the way humans perceive and communicate. It focuses on larger parts of data for processing, as compared to other analysis techniques.

## **7. What is sentiment analysis in NLP?**

The sentiment analysis, also known as opinion mining and emotion AI, is a process of detecting the polarity of the opinion in the text or can be a part of it. It is majorly used to identify, extract, and quantify user or customer reviews' polarity, survey responses, or social media opinions.

## **8. What is discourse analysis in NLP?**

Discourse is a structured group of the sentence. Discourse analysis can be termed as an approach to analyzing the discourse, i.e., text or language. It involves text interpretations and interactions.

## **9. What is pragmatic ambiguity in NLP?**

Pragmatic ambiguity can be referred to as a condition where words have multiple interpretations. This condition arises when the meaning of words is not specific; i.e., it can give different meanings.

## **10. What are the major applications of NLP?**

The major applications of NLP are:

1. Machine Translation
2. Speech Recognition
3. Sentiment Analysis
4. Text Classification
5. How is long term dependency maintained while building a language model?

Language models can be built using the following popular methods –  
Using n-gram language model n-gram language models make assumption for the value of n. Larger the value of n, longer the ...

6. You have come up with a Spam classifier. How do you measure accuracy ?

Spam filtering is a classification problem. In a classification problem, the following are the common metrics used to measure efficacy : True positives : Those data points where the outcome ...

7. What are the advantages and disadvantages of using naive bayes for spam detection?

Disadvantages: Naive bayes is based on the conditional independence of features assumption – an assumption that is not valid in many real world scenarios. Hence it sometimes oversimplifies the problem ...

8. What is the difference between stemming and lemmatisation?

Stemming is about replacing each word with its origin stem word in order to remove the suffixes like “es”, “ies”, “s”. For ex., “cats” => “cat”, “computers” => “computer” etc. ...

9. What are some common tools available for NER ? Named Entity Recognition ?

Notable NER platforms include: GATE supports NER across many languages and domains out of the box, usable via a graphical interface and a Java API. OpenNLP includes rule-based and statistical named-entity recognition. SpaCy ...

10. What is negative sampling when training the skip-gram model ?

Recap: Skip-Gram model is a popular algorithm to train word embeddings such as word2vec. It tries to represent each word in a large text as a lower dimensional vector in ...

11. What is the significance of n-grams in a language model ?

n-grams is a term used for a sequence of n consecutive words/tokens/grams. In general, n-grams can either preserve the ordering or indicate what level of dependency is required in order to ...

12. What is a language model ? How do you create one ? Why do you need one ?

A language model is a probability distribution over sequences of words  $P(w_1, \dots, w_m)$ . It enables us to measure the relative likelihood of different phrases. Measuring the likelihood of a sequence ...

13. What are knowledge graphs? When would you need a knowledge graph over say a database to store information?

A knowledge graph organizes real world knowledge as entities and relationships between entities. Creating a knowledge graph often involves scraping / ingesting unstructured data and creating structure out of it ...

14. What are the different independence assumptions in hMM & Naive Bayes ?

Both the hMM and Naive Bayes have conditional independence assumption. hMM can be expressed by the equation below : Second equation implies a conditional ...

15. What is the difference between word2Vec and Glove ?

Both word2vec and glove enable us to represent a word in the form of a vector (often called embedding). They are the two most popular algorithms for word embeddings that bring ...

16. How do you deal with dataset imbalance in a problem like spam filtering ?

Class imbalance is a very common problem when applying ML algorithms. Spam filtering is one such application where class imbalance is apparent. There are many more non-spam emails in a ...

17. You want to find food related topics in twitter – how do you go about it ?

One can use any of the topic models above to get topics. However, to direct the topics to contain food related information, specialized topic modeling algorithms are available. However, one ...

18. Why is smoothing applied in language model ?

Because there might be some n-grams in the test set but may not be present in the training set. For ex., If the training corpus is and ...

19. You are given some documents and asked to find prevalent topics in the documents – how do you go about it ?

This is typically called topic modeling. Topic modeling is a type of statistical modeling for discovering the abstract “topics” that occur in a collection of documents. For instance, two statements ...

20. What are popular ways of dimensionality reduction in NLP tasks ? Do you think this is even important ?

Common representation is bag of words that is very high dimensional given high vocab size. Commonly used ways for dimensionality reduction in NLP : TF-IDF : Term frequency, inverse document ...

## 21. How do you find the most probable sequence of POS tags from a sequence of text?

This problem can be solved with a HMM. Using a HMM involves finding the transition probabilities (what is the probability of going from one POS tag to another) and emission/output ...

## 22. How do you generate text using a Hidden Markov Model (HMM) ?

The HMM is a latent variable model where the observed sequence of variables are assumed to be generated from a set of temporally connected latent variables . The joint distribution ...

## 23. Given the following two sentences, how do you determine if Teddy is a person or not? “Teddy bears are on sale!” and “Teddy Roosevelt was a great President!”

This is an example of Named Entity Recognition(NER) problem. One can build a sequence model such as an LSTM to perform this task. However, as shown in both the sentences above, ...

## 24. What is speaker segmentation in speech recognition ? How do you use it ?

Speaker diarization or speaker segmentation is the process of automatically assigning a speaker identity to each segment of the audio file. Segmenting by speaker is very useful in several applications ...

## 25. Which is better to use while extracting features character n-grams or word n-grams? Why?

Both have their uses. Character n-grams are great where character level information is important : Example: spelling correction, language identification, writer identification (i.e. fingerprinting), anomaly detection. While word n-grams are ...

## 26. What is PMI ?

PMI : Pointwise Mutual Information, is a measure of correlation between two events x and y. As you can see from above ...

## 27. How many parameters are there for an hMM model?

Let us calculate the number of parameters for bi-gram hMM given as Let be the total number of states and be the vocabulary size ...

## 28. Suppose you build word vectors (embeddings) with each word vector having dimensions as the vocabulary size(V) and feature values as pPMI between corresponding words: What are the problems with this approach and how can you resolve them ?

Problems As the vocabulary size (V) is large, these vectors will be large in size. They will be sparse as a word may not have co-occurred with all possible words. Resolution Dimensionality Reduction using ...

## 29. What can you say about the most frequent and most rare words ? Why are they important or not important ?

Most frequent words are usually stop words like Rare words could be because of spelling mistakes or due to the word being sparsely used in the data set. Usually ...

## 30. Can you find the antonyms of a word given a large enough corpus? For ex. Black => white or rich => poor etc. If yes then how, otherwise justify your answer.

Pre existing Databases: There are several curated antonym databases such as wordpress and so on from which you can directly check if you can get antonyms of a given word. Hearst ...

## 31. How to measure the performance of the language model ?

While building language model, we try to estimate the probability of the sentence or a document. Given sequences(sentences or documents) like Language model(bigram language model) will be ...

## 32. What will happen if you do not convert all characters to a single case (either lower or upper) during the pre-processing step of an NLP algorithm?

When all words are not converted to a single case, the vocabulary size will increase drastically as words like Up/up or Fast/fast or This/this will be treated differently which isn't ...

## 33. Explain latent dirichlet allocation – where is it typically used ?

Latent Dirichlet Allocation is a probabilistic model that models a document as a multinomial mixture of topics and the topics as a multinomial mixture of words. Each of these multinomials ...

#### 34. How do you train a hMM model in practice ?

The joint probability distribution for the HMM model is given by the following equation where are the observed data points and the corresponding latent states: ...

#### 35. What are some knowledge graphs you know. What is different between these ?

DBPedia : Entities and relationships are automatically extracted from wikipedia. Wordnet: Lexical database of english language. Groups english words as synsets and provides various relationships between words in a synset. ...

#### 36. Given a bigram language model, in what scenarios do we encounter zero probabilities? How Should we handle these situations ?

Recall the Bi-gram model can be expressed as : Scenario 1 – Out of vocabulary(OOV) words – such words may not be present during training and hence ...

#### 37. Why are bigrams or any n-grams important in NLP(task like sentiment classification or spam detection) or important enough to find them explicitly?

There are mainly 2 reasons Some pair of words always occur together more often than they occur individually. Hence it is important to treat such co-occurring words as a single entity ...

#### 38. You are trying to cluster documents using a Bag of Words method. Typically words like if, of, is and so on are not great features. How do you make sure you are leveraging the more informative words better during the feature Engineering?

Words like if, of, ... are called stop words. Typical pre-processing in standard NLP pipeline involves identifying and removing stop-words (except in some cases where context/ word adjacency information ...

#### 39. How can you increase the recall of a search query (on search engine or e-commerce site) result without changing the algorithm ?

Since we are not allowed to change the algorithm, we can only play with modifying or augmenting the search query. (Note, we either change the algorithm/model or the data, here ...)

#### 40. What is shallow parsing

Typically we have a generative grammar that tells us how a sentence is generated from a set of rules. Parsing is the process of finding a parse tree that is ...

#### 41. What order of Markov assumption does n-grams model make ?

An n-grams model makes order  $n-1$  Markov assumption. This assumption implies: given the previous  $n-1$  words, probability of word is independent of words prior to words. Suppose we have  $k$  words ...

#### 42. Say you've generated a language model using Bag of Words (BoW) with 1-hot encoding , and your training set has lot of sentences with the word "good" but none with the word "great". Suppose I see sentence "Have a great day" $p(\text{great})=0.0$ using this language model. How can you solve this problem leveraging the fact that good and great are similar words?

BoW with 1-hot encoding doesn't capture the meaning of sentences, it only captures co-occurrence statistics. We need to build the language model using features which are representative of the meaning ...

#### 43. What are common tools for speech recognition ? What are the advantages and disadvantages of each?

There are several ready tools for speech recognition, that one can use to train custom models given the appropriate dataset. CMU Sphinx : Used more in an academic setting, one ...

#### 44. Suppose you are modeling text with a HMM, What is the complexity of finding most probable sequence of tags or states from a sequence of text using brute force algorithm?

Assume there are total states and let be the length of the largest sequence. Think how we generate text using an hMM. We first have a state sequence and ...

#### 45. How will you build an auto suggestion feature for a messaging app or google search?

Auto Suggestion feature involves recommending the next word in a sentence or a phrase. For this, we need to build a language model on large enough corpus of “relevant” data. ...

46. You are building a natural language search box for a website. How do you accommodate spelling errors?

If you have a dictionary of words, edit distance is the simplest way of incorporating this. However, sometimes corrections based on context make sense. For instance, suppose I type “bed ...

47. What would you care more about – precision or recall for spam filtering problem?

False positive means it was not a spam and we called it spam, false negative means it was a spam and we didn't label it spam Precision =  $(TP / TP + FN)$  ...

48. How will you build the automatic/smart reply feature on an app like gmail or linkedIn?

Generating replies on the fly: Smart Reply can be built using sequence to sequence modeling. An incoming mail acts as the input to the model and the reply will be ...

49. If you don't have a stop-word dictionary or are working on a new language, what approach would you take to remove stop words?

TF-IDF (term frequency Inverse document frequency) is a popular approach that can be leveraged to eliminate stop words. This technique is language independent. The intuition here is that commonly occurring ...

50. What is the difference between paraphrasing and textual entailment ?

Textual entailment is the process of determining if a source T implies the hypothesis text H. Example :It is a unidirectional relationship : text: If you help the needy, God ...

51. Where would you not want to remove stop words ?

Stop words can be removed in most applications when you are doing bag of words features. Some exceptions can involve sentiment analysis where ‘not’ cannot be removed because it is ...

## 52.What are the advantages and disadvantages of using Rule based approaches in NLP?

Cold start: Many a times when we have the cold start problem (No data to begin with) in Machine Learning, rule based approaches make sense. For example, you want to recommend ...

## 53.How do you design a system that reads a natural language question and retrieves the closest FAQ answer?

There are multiple approaches for FAQ based question answering  
Keyword based search (Information retrieval approach): Tag each question with keywords. Extract keywords from query and retrieve all relevant questions answers. Easy ...

## 54.How do you deal with out of vocabulary words during run time when you build a language model ?

Out of vocabulary words are words that are not in the training set, but appear in the test set, real data. The main problem is that the model assigns a ...

## 55.BLUE Score

<https://youtu.be/UV2ymKoMcyw> This brief video describes the BLEU score, a popular evaluation metric used for several tasks such as machine translation, text summarization and so on. What is BLEU Score? BLEU stands for ...

## 56.What is the state of the art technique for Machine Translation ?

Rule based machine translation (Older techniques) : Uses dictionary between words of the two languages along with syntactic, semantic morphological analysis of the source sentence to define context.  
Linguistic Rules ...

## 57.What are the different ways of representing documents ?

Bag of words: Commonly called BOW involves creating a vocabulary of words and representing the document as a count vector, dimension equivalent to the vocabulary size – each dimension representing ...

## 58.What is the difference between translation and transliteration

Transliteration is the process of converting a word written in one language into another language, phoneme by phoneme. Enabling transliteration for your search engine allows your site visitors to type ...

### 1)How can machines make meaning out of language?

Popular NLP procedure is to use stemming and lemmatization methods along with the parts of speech tagging. The way humans use language varies with context and everything can't be taken too literally.

Stemming approximates a word to its root i.e identifying the original word by removing the plurals or the verb forms. For example, 'rides' and 'riding' both denote 'ride'. So, if a sentence contains more than one form of ride, then all those will be marked to be identified as the same word. Google used stemming back in 2003 for its search engine queries.

Whereas, lemmatization is performed to correctly identify the context in which a particular word is used. To do this, the sentences adjacent to the one under consideration are scanned too. In the above example, riding is the lemma of the word ride.

Removing stop words like a, an, the from a sentence can also enable the machine to get to the ground truth faster.

### 2)What does a NLP pipeline consist of?

Any typical NLP problem can be proceeded as follows:

1. Text gathering(web scraping or available datasets)
2. Text cleaning(stemming, lemmatization)
3. Feature generation(Bag of words)
4. Embedding and sentence representation(word2vec)
5. Training the model by leveraging neural nets or regression techniques
6. Model evaluation
7. Making adjustments to the model
8. Deployment of the model.

### 3)What is Parsing in the context of NLP?

Parsing a document means to work out the grammatical structure of sentences, for instance, which groups of words go together (as “phrases”) and which words are the subject or object of a verb. Probabilistic parsers use knowledge of language gained from hand-parsed sentences to try to produce the most likely analysis of new sentences.

#### 4) What is Named Entity Recognition(NER)?

Named entity recognition is a method to divide a sentence into categories.

Neil Armstrong of the US had landed on the moon in 1969 will be categorized as

Neil Armstrong- name; The US – country; 1969 – time(temporal token).

The idea behind NER is to enable the machine to pull out entities like people, places, things, locations, monetary figures, and more.

#### 5) Where can NER be used?

Scanning documents for classification, customer support(chatbots, understanding feedback) and entity identification in molecular biology(names of genes etc.,)

#### 6) How is feature extraction done in NLP

The features of a sentence can be used to conduct sentiment analysis or document classification. For example if a product review on Amazon or a movie review on IMDB consists of certain words like ‘good’, ‘great’ more, it could then be concluded/classified that a particular review is positive.

Bag of words is a popular model which is used for feature generation. A sentence can be tokenized and then a group or category can be formed out of these individual words, which further explored or exploited for certain characteristics(number of times a certain word appears etc).

#### 7) Name some popular models other than Bag of words?

Latent semantic indexing, word2vec.

8) Explain briefly about word2vec

Word2Vec embeds words in a lower-dimensional vector space using a shallow neural network. The result is a set of word-vectors where vectors close together in vector space have similar meanings based on context, and word-vectors distant to each other have differing meanings. For example, apple and orange would be close together and apple and gravity would be relatively far. There are two versions of this model based on skip-grams (SG) and continuous-bag-of-words (CBOW).

9) What is Latent Semantic Indexing?

Latent semantic indexing is a mathematical technique to extract information from unstructured data. It is based on the principle that words used in the same context carry the same meaning.

In order to identify relevant (concept) components, or in other words, aims to group words into classes that represent concepts or semantic fields, this method applies Singular Value Decomposition to the Term-Document matrix. As the name suggests this matrix consists of words as rows and document as columns.

LSI is computation heavy when compared to other models. But it equips an NLP model with better contextual awareness, which is relatively closer to NLU

10) What are the metrics used to test an NLP model?

Accuracy, Precision, Recall and F1. Accuracy is the usual ratio of the prediction to the desired output. But going just be accuracy is naive considering the complexities involved. Whereas, precision and recall consider false positive and false negative making them more reliable metrics.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$= \frac{\text{True Positive}}{\text{Total Actual Positive}}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$= \frac{\text{True Positive}}{\text{Total Predicted Positive}}$$

And, F1 is the sweet spot between precision and recall.

$$F1 = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

11) What are some popular Python libraries used for NLP

Stanford's CoreNLP, SpaCy , NLTK and TextBlob.

- What is part of speech (POS) tagging? What is the simplest approach to building a POS tagger that you can imagine?
- How would you build a POS tagger from scratch given a corpus of annotated sentences? How would you deal with unknown words?
- How would you train a model that identifies whether the word “Apple” in a sentence belongs to the fruit or the company?
- How would you find all the occurrences of quoted text in a news article?
- How would you build a system that auto corrects text that has been generated by a speech recognition system?
- What is latent semantic indexing and where can it be applied?
- How would you build a system to translate English text to Greek and vice-versa?
- How would you build a system that automatically groups news articles by subject?

- What are stop words? Describe an application in which stop words should be removed.
- How would you design a model to predict whether a movie review was positive or negative?
- What is entropy? How would you estimate the entropy of the English language?
- What is a regular grammar? Does this differ in power to a regular expression and if so, in what way?
- What is the TF-IDF score of a word and in what context is this useful?
- How does the PageRank algorithm work?
- What is dependency parsing?

- What are the difficulties in building and using an annotated corpus of text such as the Brown Corpus and what can be done to mitigate them?
- **1. Discuss real-life apps based on Natural Language Processing (NLP).**
- **Chatbot:** Businesses and companies have realized the importance of chatbots, as they assist in maintaining good communication with customers, any queries that a chatbot fails to resolve gets forwarded. They help keep the business moving as they are used 24/7. This feature makes use of natural language processing.
- **Google Translate:** Spoken words or written text can be converted into another language, proper pronunciation is also available of words, Google Translate makes use of advanced NLP which makes all of this possible.
- **2. What is meant by NLTK?**
- Natural language toolkit is a Python library that processes human languages, different techniques including tokenization, stemming, parsing, lemmatization are used for grasping the languages. Also used for classification of text, and assessing documents. Some libraries include DefaultTagger, wordnet, patterns, treebank, etc.
- **3. Explain parts of speech tagging (POS tagging).**
- POS is also known as parts of speech tagging is Implemented for assigning tags onto words like verbs, nouns, or adjectives. It allows the software to understand the text, then recognize word differences using algorithms. The purpose is to make the machine comprehend the sentences correctly.
- **Example:-**
- ```

import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize, sent_tokenize
stop_words = set(stopwords.words('english'))
txt = "A, B, C are longtime classmates."

```

- ## Tokenized via sent\_tokenize  
tokenized\_text = sent\_tokenize (txt)
- ## Using word\_tokenizer to identify a string's words and punctuation then removing stop words  
for n in tokenized\_text:  
wordsList = nltk.word\_tokenize(i)  
wordsList = [w for w in wordsList if not w in stop\_words]
- ## Using POS tagger  
tagged\_words = nltk.pos\_tag(wordsList)  
print (tagged\_words)
- **Output:-**  
[('A', 'NNP'), ('B', 'NNP'), ('C', 'NNP'), ('longtime', 'JJ'), ('classmates', 'NNS')]
- **4. Define pragmatic analysis**
- In a given data of human language, different meaning exists, in order to understand more, pragmatic analysis is used for discovering different facets of the data or document. Actual meaning of words or sentences can be understood by the systems, and for this purpose pragmatic analysis is deployed.
- **5. Elaborate on Natural language processing components**
- These are the major NLP components:-
- **1. Lexical/morphological analysis:** word structure is made comprehensible via analysis through parsing.
- **2. Syntactic analysis:** specific text meaning is assessed
- **3. Entity extraction:** information like the place, institution, individual gets retrieved via sentence dissection. Entities present in a sentence get identified.
- **4. Pragmatic analysis:** helps in finding real meaning and relevancy behind the sentences.
- **6. List the steps in NLP problem-solving**
- The steps in NLP problem-solving include:-
- 1. Web scraping or collecting the texts from the dataset.
- 2. For text cleaning, making use of lemmatization and stemming.
- 3. Use feature engineering
- 4. Use word2vec for embedding
- 5. Using machine learning techniques or with neural networks, start training the models.
- 6. Assess the performance.
- 7. Do the required model modifications and deploy.
- **7. Elaborate stemming with examples**
- When a root word is gained by detaining the prefix or suffix involved, then that process is known as stemming. For instance, the word 'playing' can be minimized to 'play' by removing the rest.

- Different algorithms are deployed for implementation of stemming, for example, PorterStemmer which can be imported from NLTK as follows:-
- from nltk.stem import PorterStemmer  
pst = PorterStemmer()  
pst.stem("running"), pst.stem("cookies"), pst.stem("flying")

- **Output:-**

- ('run', 'cooki', 'fly')

- **8. Define and implement named entity recognition**

- For retrieving information and identifying entities present in the data for instance location, time, figures, things, objects, individuals, etc. NER (named entity recognition) is used in AI, NLP, machine learning, implemented for making the software understand what the text means. Chatbots are a real-life example that makes use of NER.

- **Implementing NER with spacy package:-**

- import spacy  
nlp = spacy.load('en\_core\_web\_sm')  
Text = "The head office of Tesla is in California"  
document = nlp(text)  
for ent in document.ents:  
print(ent.text, ent.start\_char, ent.end\_char, ent.label\_)

- **Output:-**

- Office 9 15 Place  
Tesla 19 25 ORG  
California 32 41 GPE

- **9. Explain checking word similarity with spacy package**

- Spacy library allows the implementation of word similarity techniques for detecting similar words. The evaluation is done with a number between 0 & 1 (where 0 tends towards less similar and 1 tends toward highly similar).

- import spacy  
nlp = spacy.load('en\_core\_web\_md')  
print ("Enter the words:")  
input\_words = input()  
tokens = nlp(input\_words)  
for i in tokens:  
print(i.text, i.has\_vector, i.vector\_norm, i.is\_oov)  
token\_1, token\_2 = tokens[0], tokens[1]  
print("Similarity between words:", token\_1.similarity(token\_2))

- **Output:-**

- hot True 5.6898586 False  
cold True 6.5396233 False  
Similarity: 0.597265

- This implies that the similarity between the two words cold and hot is 59%.

- **10. Describe recall and precision. Also, explain TF-IDF.**

- **Precision and recall**
- Precision, F1 and Recall, accuracy are NLP model testing metrics. The ratio of predictions with required output provides for a model's accuracy.
- **Precision:** The ratio of positive instances and total predicted instances.
- **Recall:** The ratio between real positive instances and total (real + unreal) positive instances.
- **TF-IDF**
- Term frequency-inverse document frequency is used for retrieval of information via numerical Statistics. It helps in identifying keywords present in any document. The real usage of it revolves around getting information from important documents using Statistical data. It's also useful in filtering out the stop words and text summarizing plus classification in the documents. With TF one can calculate the ratio of term frequency in a document wrt total terms, whereas IDF implies the significance of the term.
- **TF IDF calculation formula:**
- **TF** = frequency of term 'W' in a document / total terms in document
- **IDF** =  $\log(\frac{\text{total documents}}{\text{total documents with the term 'W'}})$
- If **TF\*IDF** appears higher then term frequency is likely less.
- Google implements TF-IDF for deciding search results index, which helps in optimization or ranking the relevant quality content higher.

## 1) What is NLP?

Natural Language Processing or **NLP** is an automated way to understand or analyze the natural languages and extract required information from such data by applying machine learning Algorithms.

Powered By

Play

Unmute

Loaded: 1.02%

Fullscreen

## **2) List some Components of NLP?**

Below are the few major components of NLP.

- **Entity extraction:** It involves segmenting a sentence to identify and extract entities, such as a person (real or fictional), organization, geographies, events, etc.
- **Syntactic analysis:** It refers to the proper ordering of words.
- **Pragmatic analysis:** Pragmatic Analysis is part of the process of extracting information from text.

## **3) List some areas of NLP?**

Natural Language Processing can be used for

- Semantic Analysis
- Automatic summarization
- Text classification
- Question Answering

Some real-life example of NLP is **IOS Siri**, the **Google assistant**, **Amazon echo**.

## **4) Define the NLP Terminology?**

NLP Terminology is based on the following factors:

- **Weights and Vectors:** TF-IDF, length(TF-IDF, doc), Word Vectors, Google Word Vectors
- **Text Structure:** Part-Of-Speech Tagging, Head of sentence, Named entities
- **Sentiment Analysis:** Sentiment Dictionary, Sentiment Entities, Sentiment Features
- **Text Classification:** Supervised Learning, Train Set, Dev(=Validation) Set, Test Set, Text Features, LDA.
- **Machine Reading:** Entity Extraction, Entity Linking, dbpedia, FRED (lib) / Pikes

## **5) What is the significance of TF-IDF?**

**tf-idf** or **TFIDF** stands for **term** frequency–inverse **document frequency**. In information retrieval TFIDF is a numerical statistic that is intended to reflect how important a word is to a document in a collection or in the collection of a set.

**Download Free : NLP Interview Questions PDF**

## **6) What is part of speech (POS) tagging?**

According to The Stanford Natural Language Processing Group :

A **Part-Of-Speech Tagger** (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc.

**PoS taggers** use an algorithm to label terms in text bodies. These taggers make more complex categories than those defined as basic PoS, with tags such as “noun-plural” or even more complex labels. Part-of-speech categorization is taught to school-age children in English grammar, where children perform basic PoS tagging as part of their education.

## **7) What is Lemmatization in NLP?**

**Lemmatization** generally means to do the things properly with the use of vocabulary and morphological analysis of words. In this process, the endings of the words are removed to return the base word, which is also known as Lemma.

**Example:** boy's = boy, cars= car, colors= color.

So, the main attempt of Lemmatization as well as of stemming is to identify and return the root words of the sentence to explore various additional information.

## **8) What is stemming in NLP?**

## **9) What is tokenization in NLP?**

**Natural Language Processing** aims to program computers to process large amounts of natural language data. **Tokenization in NLP** means the method of dividing the text into various tokens. You can think of a token in the form of the word. Just like a word forms into a sentence. It is an important step in NLP to split the text into minimal units.

## **10) What is latent semantic indexing and where can it be applied?**

**Latent Semantic Indexing (LSI)** also called Latent semantic analysis is a mathematical method that was developed so that the accuracy of retrieving information can be improved. It helps in finding out the hidden(latent) relationship between the words(semantics) by producing a set of various concepts related to the terms of a sentence to improve the information understanding. The technique used for the purpose is called Singular value decomposition. It is generally useful for working on small sets of static documents.

**Take Free:** [Nlp MCQ & Quiz](#)

## **11) What is dependency parsing?**

## **12) Differentiate regular grammar and regular expression.**

**Regular expression** is a sequence of characters that define a search pattern, mainly for use in pattern matching with strings, or string matching. It includes the following elements:

Example: A and B are regular expressions then

- The regular expression is A. B (concatenation)
- The regular expression (alternation) is A | B
- The regular expression (Kleene Star) is A\*

## **Regular Grammars**

There are 4 tuples in Regular Grammars ( $N, \Sigma, P, S \in N$ ). In this formula,  $N$  stands for the non-terminals' sets,  $\Sigma$  means the set of terminals,  $P$  is the set of productions to change the start symbol,  $P$  has its productions from one of the types and lastly  $S$  is the start non-terminal.

**13) List some tools for training NLP models?**

**14) Describe dependency parsing?**

**15) Explain Named entity recognition (NER)?**

**Named-entity recognition (NER)** is the method of extracting information. It arranges and classifies named entity in the unstructured text in different categories like locations, time expressions, organizations, percentages, and monetary values. NER allows the users to properly understand the subject of the text.

**16) What is NLTK?**

**17) List some OpenSource Libraries for NLP?**

## **18) What is the difference between NLP and NLU?**

**Difference between NLP and NLU are**

| <b>Natural Language Processing</b>                                                                           | <b>Natural Language Understanding</b>                                                                      |
|--------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| NLP is the system that works simultaneously to manage end-to-end conversations between computers and humans. | NLU helps to solve the complicated challenges of Artificial Intelligence.                                  |
| NLP is related to both humans and machines.                                                                  | NLU allows converting the unstructured inputs into structured text for easy understanding by the machines. |

## **19) What is the difference between NLP and CI(Conversational Interfaces)?**

The **difference between NLP and CI(Conversational Interfaces)** are as follows -

| <b>Natural Language Processing</b>                                                                                                                        | <b>Conversational Interfaces</b>                                                                           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| NLP is a kind of artificial intelligence technology that allows identifying, understanding and interpreting the request of users in the form of language. | CI is a user interface that mixes voice, chat and another natural language with images, videos or buttons. |
| NLP aims to make users understand a particular concept.                                                                                                   | A conversational Interface provides only what the users need and not more than that.                       |

## **20) List few differences between AI, Machine Learning, and NLP?**

## Differences between AI, Machine Learning, and NLP

| Artificial Intelligence                                                                     | Machine Learning                                                                     | Natural Language Processing                                                |
|---------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| It is the technique to create smarter machines used for systems that learn from experience. | Machine Learning is the term used for systems that learn from experience.            | This is the set of system that has the ability to understand the language. |
| AI includes human intervention                                                              | Machine Learning purely involves the working of computers and no human intervention. | NLP links both computer and human languages.                               |
| Artificial intelligence is a broader concept than Machine Learning                          | ML is a narrow concept and is a subset of AI.                                        |                                                                            |

### 21) Explain the Masked Language Model?

**Masked language modelling** is the process in which the output is taken from the corrupted input. This model helps the learners to master the deep representations in downstream tasks. You can predict a word from the other words of the sentence using this model.

### 22) What is latent semantic indexing? Where it is applied.

### 23) What is pragmatic analysis in NLP?

**Pragmatic Analysis:** It deals with outside word knowledge, which means knowledge that is external to the documents and/or queries. Pragmatics analysis that focuses on what was described is reinterpreted by what it actually meant, deriving the various aspects of language that require real-world knowledge.

## **24) Explain dependency parsing in NLP?**

**Dependency Parsing** is also known as Syntactic Parsing. It is the task of recognizing a sentence and assigning a syntactic structure to it. The most widely used syntactic structure is the parse tree which can be generated using some parsing algorithms. These parse trees are useful in various applications like grammar checking or more importantly it plays a critical role in the semantic analysis stage.

## **25) What is pragmatic ambiguity in NLP?**

**Pragmatic Ambiguity** can be defined as words that have multiple interpretations. Pragmatic Ambiguity arises when the meaning of words in a sentence is not specific; it concludes with different meanings. There are various sentences in which the proper sense is not understood due to the grammar formation of the sentence; this multi-interpretation of the sentence gives rise to ambiguity.

**For Example-** "do you want a cup of coffee", the given word is either an informative question or a formal offer to make a cup of coffee.

## **26) What is perplexity in NLP?**

The word "perplexed" means "puzzled" or "confused", thus Perplexity in general means the inability to tackle something complicated and a problem that is not specified. Therefore, Perplexity in NLP is a way to determine the extent of uncertainty in predicting some text.

In NLP, perplexity is a way of evaluating language models. Perplexity can be high and low; Low perplexity is ethical because the inability to deal with any complicated problem is less while high perplexity is terrible because the failure to deal with a complicated is high.

## **27) What is ngram in NLP?**

**N-gram in NLP** is simply a sequence of n words, and we also conclude the sentences which appeared more frequently, for example, let us consider the progression of these three words:

- New York (2 gram)
- The Golden Compass (3 gram)
- She was there in the hotel (4 gram)

Now from the above sequence, we can easily conclude that sentence (a) appeared more frequently than the other two sentences, and the last sentence(c) is not seen that often. Now if we assign probability in the occurrence of an n-gram, then it will be advantageous. It would help in making next-word predictions and in spelling error corrections.

## 28) What is Meta Model in NLP?

The **meta-model** in neuro-linguistic programming is a set of questions designed to specify information, challenge and expand the limits to a person's model of the world. It responds to the distortions, generalizations, and deletions in the speaker's language.

## 29) Please explain Milton Model?

**NLP Milton Model** is a set of language patterns used to help people to make desirable changes and solve difficult problems. It is also useful for inducing trance or an altered state of consciousness to access our all powerful unconscious resources.

### 1. What is NLP?

NLP stands for Natural Language Processing. It is a branch of AI that refers to the ability of a computer device to understand the human language, as it is spoken.

The auto-completion of words or queries in Google search, personal assistants Alexa, Siri, Google home are some of the examples of NLP being put to practical use in the real world.

NLP is an evolving field. It is a good idea to go prepared with some concepts that direct this field.

## 2. What do you know about Syntactic and Semantic Analysis in NLP?

NLP uses two important techniques called as Syntactic and Semantic Analysis.

**Syntactic analysis** studies the arrangement of words in a sentence to derive meaning from them based on the grammar rules of a language.

**Some of the techniques used for Syntactic analysis are:**

- i.) Parsing
- ii.) Word Segmentation
- iii.) Sentence breaking
- iv.) Morphological segmentation
- v.) Stemming
- vi.) Lemmatization

**Semantics** refer to the meaning that a text conveys. Semantic analysis makes use of computer algorithms to understand the meaning and interpretation of words and see how sentences are structured. Some of the techniques used for Semantic analysis are:

- i.) **Named Entity Recognition** - Here you identify and categorize the words into preset categories. E.g. Names of people, places, animals etc.
- ii.) **Word sense disambiguation** - Gives meaning to a word based on the context it is used in.
- iii.) **Natural language generation** - Deriving semantic intentions from the database and converting them into human language.

## 3. Explain Stemming.

Stemming and Lemmatization are both data pre-processing steps.

However, they both work to establish that the two given words are the different forms of the same word, the approach the two follow is absolutely different.

**Stemming** - It follows a heuristic approach and works by cutting out the prefix or the suffix to find the stem.

For example - If we take the words Asking and Asked, stemming would cut out the tails "ing" and "ed" and get the stem "Ask" in both the cases.

But this is not an optimal method because it can accidentally cut out the wrong letters and the stem won't really make any sense in that case.

So, we say that Stemming may fall a victim to "overstemming" or "understemming".

**The three algorithms commonly used for stemming are:**

- i.) Porter Stemmer
- ii.) **Snowball Stemmer** - also called as Porter2 Stemmer algorithm.
- iii.) **Lancaster Stemmer** - This is the most aggressive algorithm out of the three and can sometimes render absolutely stems that don't make sense.

#### 4. What is Lemmatization?

Lemmatization is a calculated approach to reach the base or the root of any word. Rather than just snipping the words at the head or the tail, this approach resolves the word to its dictionary meaning. This needs more efforts to prepare the system.

Here, the system maps the word to its origin. For e.g. the words "come", "came", "coming" are all mapped to the lemma "come".

Now, if it was a stemming algorithm, it would simply snip off the tail of the word and then won't know that all the three words share the same origin.

#### 5. What are "Stop words"?

The words like "is, am, are, the, of, for, was, were, how, why" that we use to make a sentence are categorized as stop words because we don't want them to make the search engine or the application focus on them rather than the words that really matter.

#### 6. What do you know about Zipf's law in NLP?

Zipf's law was named after the American linguist George Kingsley Zipf.

It states that in a corpus of given natural language utterances, the frequency of a word is inversely proportional to its rank in the frequency table.

This implies that the first most frequent word will appear approximately twice as often as the word that appears at 2nd position.

7. Can you name some commonly used Python NLP libraries?

**Some of the commonly used Python NLP Libraries are:**

- i.) **NLTK** - This is a well known and complete NLP Library. It has a lot of third-party extensions and approaches for various tasks. Supports the largest number of languages.
- ii.) Spacy
- iii.) Scikit-learn
- iv.) Gensim
- v.) Pattern
- vi.) Polyglot

## Table of Contents



- [How do you define natural language processing?](#)
- [What are the different Machine Translation Models?](#)
- [How can you differentiate between NLG and NLU?](#)
- [Give an example of NLG, NLU and also give an example that used both?](#)
- [What is the different type of Approaches used for Hate detection?](#)
- [How are the processes of spam classification and machine translation different from each other?](#)
- [What are some example problems of NLP/Text Mining?](#)
- [What are the typical steps of a text mining system involving machine learning?](#)
- [What is the difference between tokenizing the space-delimited language and unsegmented language?](#)
- [What are the various Pre Processing techniques in Text Mining?](#)
- [What is tokenization?](#)
- [What are Stemming and Lemmatization?](#)
- [Suppose you are developing machine translation, which preprocessing steps you will do OR will not do?](#)
- [For which task is an \(MLP|RNN|CNN\) useful?](#)

- What is the difference b/w Neural Networks and RNN
- What is the vanishing gradient problem and how it is tackled?
- How does LSTM work? Why is it useful and when?
- Explain the following in a language, and give examples: Morphemes, lexemes, syntax, and context.
- What is a POS tag? Make a sentence, and POS tag it.
- Describe the multi-class, multi-label, and multi-task classification problems.
- Name two similarity measurement metrics for vectorized semantic representation, and tell which one is preferred? Why?
- What are the different types of documents, what is document classification, provide some examples?
- What do you understand by Gender Bias?
- What is semantics and how we can represent it? Specifically, discuss formal and vectorized semantic representation and give examples of each.
- What is a virtual assistant and explain briefly how does it work?
- Justify why we use regularization methods, list the three major categories of regularization methods we discussed, and explain one of them in more detail.
- What is the difference between Syntactic and Semantic analysis, how parser can be applied as a method of syntactic analysis?
- In human language we have polysemy and synonymy, what is the difference between them? Based on the above, what is synset in NLP?
- Can CNN apply for NLP or just for computers, justify your answer?
- Explain the main idea behind FastText word representation?
- List some of the practical issues that NN might face and come up with some suggestions for the solution?
- What is the idea behind using N-Grams in NLP?
- What is the difference between tokenizing the space-delimited language and unsegmented language?
- What is Word Embedding (Also referred to as Feature Extraction techniques in NLP?)
- What is popular Word Embedding Techniques
- Describe a language model, and explain two methods one can use to evaluate it (one extrinsic, and one intrinsic)
- What are context vectors, encoders, and decoders in recurrent neural networks?
- Describe the model architecture of the encoder-decoder network for translation, in both phases of training and inference.
- What is word2vec
- What is Word2Vec – Skip Gram?
- What is Word2Vec- CBOW (Continuous Bag of Words)
- What is the difference b/w Skip-gram and CBOW

- What is Bag of Words: Word embedded Technique?
- What is the demerit of Bag of Words Technique?
- What is a Good NLP model?

## **How do you define natural language processing?**

Natural Language Processing is the technology used to aid computers to understand the human's natural language. In simple terms, converting text into structured data. The ultimate objective of NLP is to read, decipher, understand, and make sense of human languages in a manner that is valuable to machines/computers.

## **What are the different Machine Translation Models?**

**Statistical:** Statistical method find a word-to-word correspondence in order to translate the language.

**Rule-based:** Rule-based methods use the grammar rules for the translation.

**Hybrid:** Hybrid takes advantage of both Statistical and Rule-based methods.

**Neural:** The recent advancement in MT is neural networks.

## **How can you differentiate between NLG and NLU?**

**NLU (Natural language understanding)** is the process of reading and interpreting language. Converting Natural language to structure data. Ex- sentiment analysis, Profanity filters, Topic/Document classification, Entity detection.

**NLG (Natural Language Generation)** is the process of writing or generating language. Converting structure data to Natural language. Ex- Automated Journalism.

## **Give an example of NLG, NLU and also give an example that used both?**

**NLU:** Sentiment analysis, Profanity filters, Topic/Document classification, Entity detection

**NLG:** Automated Journalism

**NLU and NLG:** Virtual Assistant, Chatbot

## **What is the different type of Approaches used for Hate detection?**

- **Keyword approach**
- **Source Meta Data**

**Keyword approach:** For identifying hate speech, by using dictionary text that contains potentially hateful keywords are identified. Keyword-based approaches are fast and straightforward to understand. However, there is a severe limitation. For example: including terms that could or not always hateful. For example, trash, swine. Further keywords-based approaches could not identify hate speech. For example, hate speech that does not have any hateful keywords. Build that wire literally means constructing a wire. However, with the political context, some interpreted these as a condemnation of some immigrants in the United States.

**Source MetaData:** Additional information from social media can potentially lead to better identification approach. Information such as demographics of the posting user, location, timestamp, or even social engagement on the platform can all give further understanding. However, this information is not legally available to the external researcher because sensitive user information/data raises privacy concerns.

## **How are the processes of spam classification and machine translation different from each other?**

**Spam classification is NLU and Machine translation is both NLU and NLG,** also spam uses binary classification, while machine translation multi-class classification

Spam classification is a binary classification problem and not necessarily require understanding the context and semantics of sentences. We can

implement a simple dictionary-based spam classifier. On the other hand, machine translation is a sequence-to-sequence problem. The MT model must understand the semantics, relationships and context of the words. The model needs to store the state. RNN architectures are necessary such as Encoder-Decoder architecture, transformer architecture, and so on.

### **What are some example problems of NLP/Text Mining?**

- - - Sentiment Analysis
    - Topic Classification
    - Profane filters
    - Spam Filtering
    - Fraud detection
    - Risk Management
    - Knowledge Management
    - Cybercrime Prevention
    - Customer Care Service
    - Fraud Detection Through Claims Investigation
    - Contextual Advertising
    - Business Intelligence
    - Content Enrichment
    - Spam Filtering
    - Social Media Data Analysis
    - Contextual Advertisement

### **What are the typical steps of a text mining system involving machine learning?**

- - - Tokenization

- Part of speech Tagging
- Text Feature Extraction

**Tokenization:** Tokenization is one of the first steps in NLP, and it's the task of splitting a sequence of text into units. We have word-level tokenization, character-based tokenization, Subword level tokenization.

**Part of Speech Tagging:** The act of establishing the part of speech of each token in a text and then marking it as such is known as Part of Speech tagging (or PoS tagging). We employ PoS tagging to determine whether a token identified as a noun, verb, adjective, adverb, etc. It is not a simple task: Book a hotel (here book is a verb).

**Text feature extraction:** Requires machine learning

## What is the difference between tokenizing the space-delimited language and unsegmented language?

It consists only of splitting a sentence by the whitespace and punctuation marks. Many logographic (character-based) languages, such as Chinese, have no space breaks between words. Tokenizing these languages requires the use of machine learning.

## What are the various Pre Processing techniques in Text Mining?

Tokenization | Stemming, Lemmatization, Contractions, and Typos

Filtering: Stop Words | Accented Words | URLs | Special characters | Numbers | Punctuations

**Tokenization** is one of the first steps in NLP, and it's the task of splitting a sequence of text into units. We have word-level tokenization, character-based tokenization, Subword level tokenization.

**Word-Level Tokenization:** It consists only of splitting a sentence by the whitespace and punctuation marks. There are plenty of libraries in Python that do this, including NLTK, SpaCy, Keras, Gensim or you can do a custom Regex.

Many logographic (character-based) languages, such as Chinese, have no space breaks between words. Tokenizing these languages requires the use of machine learning, and is beyond the scope of this article.

**Stemming and Lemmatization** both generate the root form of the inflected words. The difference is that stem might not be an actual word whereas, lemma is an actual language word.

*Stemming Example:* ponies -> poni (Poni is not a dictionary word)

*Lemmatization Example* -> pony (it is a dictionary word)

Stemming follows an algorithm with steps to perform on the words which makes it faster. Whereas, in lemmatization, you used WordNet corpus and a corpus for stop words as well to produce lemma which makes it slower than stemming. You also had to define a parts-of-speech to obtain the correct lemma.

A **contraction** is a shortened form of the word (or group of words) that omits certain letters or sounds.

Ex: We're (we are), I'd (I would)

I'd like to know who you're I would like to know who you are

**Typos:** is simply spelling mistakes, we need to remove the spelling mistakes for a given word in the document.

**Stop Words:** Stop words are commonly used words of a language that may not add much information to the task.

Ex: (to, the, a, is, are) do not relate much to document classification

## What is tokenization?

**Tokenization** is one of the first steps in NLP, and it's the task of splitting a sequence of text into units. We have word level tokenization, character-based tokenization, Subword level tokenization.

**Word-Level Tokenization:** It consists only of splitting a sentence by the whitespace and punctuation marks. There are plenty of libraries in Python that do this, including NLTK, SpaCy, Keras, Gensim or you can do a custom Regex.

Many logographic (character-based) languages, such as Chinese, have no space breaks between words. Tokenizing these languages requires the use of machine learning, and is beyond the scope of this article.

## **What are Stemming and Lemmatization?**

**Stemming** and **Lemmatization** both generate the root form of the inflected words. The difference is that stem might not be an actual word whereas, the lemma is an actual language word.

Stemming: ponies -> poni (Poni is not a dictionary word)

Lemmatization -> pony (it is a dictionary word)

Stemming follows an algorithm with steps to perform on the words which makes it faster. Whereas, in lemmatization, you used WordNet corpus and a corpus for stop words as well to produce lemma which makes it slower than stemming. You also had to define a parts-of-speech to obtain the correct lemma.

## **Suppose you are developing machine translation, which preprocessing steps you will do OR will not do?**

You don't need lemmatization/stemming, stopwords removal. I am not sure if we should not remove numbers also. Probably URL can be removed and accented words and contractions can be replaced.

## **For which task is an (MLP|RNN|CNN) useful?**

MLP for any tasks but requires a lot of computing power due to fully connected layers. So, MLP is good for small problems. RNN for the sequence to sequence or time series problems. CNN for images (object detection / classification) and sequence with fixed length (approximately 100 to 200).

## **What is the difference b/w Neural Networks and RNN**

In traditional feedforward neural networks, all the inputs, and outputs are independent of each other. RNN unit takes the current input ( $X$ ) as well as the previous input ( $A$ ) to produce output ( $H$ ) and current state ( $A$ )

The RNN can take a sequence of values of inputs and take series of values as outputs. Which opens a wide variety of applications.

When an input is singular and output is a sequence, the potential application is image captioning. A multiple-input with single output can be thought of as document classification. When input and output are sequences, it can be thinking of language translation.

## **What is the vanishing gradient problem and how it is tackled?**

Problems with RNN (Hard to train, Vanishing Gradient problem)

RNN is an extremely difficult network to train, the gradient is exponentially worst for RNN. For example, training an RNN for 100-time steps is like training a 100-layer neural network. This leads to exponentially small gradients, and decay of information through time and which does not help in updating the weights/word embeddings. One of the ways to tackle vanishing gradient is the Gated unit's LSTM, GRU.

## **How does LSTM work? Why is it useful and when?**

Use of LSTM to tackle the problem of vanishing gradient.

Long Short-Term Memory networks — usually just called LSTMs — are a special kind of RNN, capable of learning long-term dependencies. LSTM does it by keeping the useful information and forgetting which is not meaningful. In LSTMs, the information flows through a mechanism known as cell states. The addition or removal of information is controlled by gates at each cell. A value of zero means let nothing through, while a value of one means let everything through. An LSTM has three of these gates to control the cell state.

**Forget Gate:** This gate decides what information we're going to throw away from the cell state.

**Input Gate:** which outputs numbers between 0 and 1 and decides which values to update.

**Output Gate:** Finally, we need to decide what we're going to output. This output will be based on our cell state.

**Explain the following in a language, and give examples: Morphemes, lexemes, syntax, and context.**

**Phonemes:** A unit of sound in language

**Morphemes:** The smallest unit in language with meaning.

**Lexemes:** Consist of one or more morphemes, that are connected by inflections.  
(Word, Token)

**Syntax:** Set of rules that create structure

**Context:** The surroundings of a word/text, that help the meaning of that particular/word

*Sentence:* She likes Pizza

*Phonemes:* Sound notations: unit of sound

*Morphemes:* She | Likes | Pizza (There are three morphemes)

*Lexemes:* She | Likes | Pizza (There are three lexemes: Token/Words)

*Syntax:* She (Subject) likes(Verb) Pizza(Object)

*Context:* Context of likes is She and Pizza

**What is a POS tag? Make a sentence, and POS tag it.**

**PoS tagging** means assigning parts of speech to tokens. Part of speech tags like nouns, adverbs. Example: She (pronoun), likes (verb), pizza (noun)

**Describe the multi-class, multi-label, and multi-task classification problems.**

**Multi-class classification:** When a sentence/word can have multiple classes. ex: a fruit can be either an apple or a pear but not both at the same time.

**Multi-label classification:** When a sentence/word can have multiple labels. ex: a movie can have multiple genres at the same time

**Multi-task classification:** When we split the main model into sub-models that they can run simultaneously and use different loss functions. The sub-models share with each other what is learned for each task and help other tasks to perform better.

**Name two similarity measurement metrics for vectorized semantic representation, and tell which one is preferred? Why?**

Two similarity measurement metrics for vectorized semantic representation are **Euclidean** and **Cosine** distances.

Euclidean distance considers magnitude and angle, while Cosine considers the angle between two vectors. Cosine distance is more preferred than Euclidean distance since the magnitude of vectors doesn't give much information.

**What are the different types of documents, what is document classification, provide some examples?**

Types of document classification.

### **What do you understand by Gender Bias?**

Gender Bias is when we have errors in recognizing the right gender for a given instance, it mostly happens for women. For example, the AI system needs to fill in the blanks, Man is to the king, a woman is to Queen. The underlying issue arises in cases where AI fills in sentences like “Father is to doctor as a mother is to nurse.” The inherent gender bias in the remark reflects an outdated perception of women in our society that is not based on fact or equality.

AI-based Systems for gender recognition reported higher error rates for recognizing women.

#### **Best Practices for Machine-Learning Teams to Avoid Gender Bias**

- Ensure diversity in the training samples (e.g. use roughly as many female audio samples as males in your training data).
- Ensure that humans labeling the audio samples come from diverse backgrounds.

**What is semantics and how we can represent it? Specifically, discuss formal and vectorized semantic representation and give examples of each.**

Semantic approach (relating to meaning in language -> stretch the words beyond, understanding ambiguities) Word2Vec (Skip and CBOW)

Examples of formal and vectorized semantic representation.

## **What is a virtual assistant and explain briefly how does it work?**

A virtual assistant (VA) is an advanced computer program that can understand, process, learn from, and respond to voice or text inputs in natural ways. It typically combines advanced natural language processing (NLP), natural language understanding (NLU).

### **Dialog process pipeline**

1. Capture the voice | 2. Recognizing language and converting it to text (using ASR) 3. Understanding the semantics using NLU | 4. Dialog manager to search and generate results for the given query | 5. Then generating the text for the given problem using NLG | 6. Then converting the text to speech 7. Then writing the speech output to the speaker.

## **Justify why we use regularization methods, list the three major categories of regularization methods we discussed, and explain one of them in more detail.**

Regularization is a set of techniques that can prevent overfitting in neural networks. Three methods L1, L2, dropout. The main intuitive difference between the L1 and L2 regularization is that L1 regularization tries to estimate the median of the data while L2 regularization tries to estimate the mean of the

data to avoid overfitting. That's the main intuitive difference between the L1 (Lasso) and L2 (Ridge) regularization techniques. L1 regularization adds the penalty term in the cost function by adding the absolute value of weight( $W_j$ ) parameters, while L2 regularization adds the squared value of weights( $W_j$ ) in the cost function.

## **What is the difference between Syntactic and Semantic analysis, how parser can be applied as a method of syntactic analysis?**

**Syntactic analysis** (syntax) and **Semantic analysis** (semantic) are the two primary techniques that lead to the understanding of natural language. The syntax is the grammatical structure of the text, whereas semantics is the meaning being conveyed. A sentence that is syntactically correct, however, is not always semantically correct. For example, “cows flow supremely” is grammatically valid (subject-verb — adverb) but it doesn’t make any sense. parse is to “resolve a sentence into its component parts and describe their syntactic roles.”

Parsing refers to the formal analysis of a sentence by a computer into its constituents, which results in a parse tree showing their syntactic relation to one another in visual form, which can be used for further processing and understanding.

It helps in identifying parts of speech, Phrases, and relationships.

## **In human language we have polysemy and synonymy, what is the difference between them? Based on the above, what is synset in NLP?**

**Polysemy:** One word expresses multiple meanings. ( A financial bank | A river bank)

**Synonymy:** One concept is expressed by several different forms of words (author/writer, beat/hit/strike)

WordNet is the lexical database i.e. dictionary for the English language, specifically designed for natural language processing.

Synset is a special kind of simple interface that is present in NLTK to look up words in WordNet. Synset instances are the groupings of synonymous words

that express the same concept. Some of the words have only one Synset and some have several.

### **Can CNN apply for NLP or just for computers, justify your answer?**

Yes, it can be applied for NLP. CNN's are good at extracting local and position-invariant features whereas RNN's are better when classification is determined by a long-range semantic dependency rather than some local key-phrases. For tasks where feature detection in the text is more important, for example, searching for angry terms, sadness, abuses, named entities, etc. CNN's work well whereas for tasks where sequential modeling is more important, RNN's work better. Based on the above characterization, it makes sense to choose a CNN for classification tasks like sentiment classification since sentiment is usually determined by some key phrases and to choose RNNs for a sequence modeling task like language modeling or machine translation, or image captioning as it requires flexible modeling of context dependencies. RNNs usually are good at predicting what comes next in a sequence while CNNs can learn to classify a sentence or a paragraph.

A big argument for CNNs is that they are fast. Very fast. Based on computation time CNN seems to be much faster (~ 5x) than RNN. Convolutions are a central part of computer graphics and implemented on a hardware level on GPUs. Applications like text classification or sentiment analysis don't actually need to use the information stored in the sequential nature of the data.

### **Explain the main idea behind FastText word representation?**

One of the main disadvantages of Word2Vec and GloVe embedding is that they are unable to encode unknown or out-of-vocabulary words.

So, to deal with this problem Facebook proposed a model FastText. It is an extension to Word2Vec and follows the same Skip-gram and CBOW model. but unlike Word2Vec which feeds whole words into the neural network, FastText first breaks the words into several sub-words (or n-grams) and then feeds them into the neural network.

### **List some of the practical issues that NN might face and come up with some suggestions for the solution?**

## Common issues in neural network implementations:

- - - **Computationally expensive:** A neural network is also computationally expensive and time-consuming to train with traditional CPUs.
    - **Amount of Data:** Neural networks typically require much more data than traditional machine learning algorithms. Though there are some cases where neural networks perform well with a small amount of data, most of the time they don't. In this case, several simple algorithms out there like naive Bayes that deal much better with minimum data, would offer a better opportunity. Moreover, neural networks rely more on training data that leads to the problem of overfitting and generalization.
    - **Black Box:** The very most disadvantage of a neural network is its black-box nature. Because it has the ability to approximate any function, it studies its structure but doesn't give any insights on the structure of the function being approximated. So, understanding the cause of the mistake requires features that are human interpretable. This is significant because, in some domains, interpretability is critical. And that is the reason why most banks don't leverage neural networks to predict whether a person is creditworthy as they need to explain to him/her why they didn't get the loan, if not the person may feel unfairly treated.

## What is the idea behind using N-Grams in NLP?

N-gram means that how many words you are considering as a single unit when you are calculating the frequency of words. An n-gram model, instead, looks at the previous (n-1) words to estimate the next one.

## **What is the difference between tokenizing the space-delimited language and unsegmented language?**

English is a space-delimited language, it is easy to tokenize such language. Whereas in the case of unsegmented language, which is a logographic we use machine learning to separate words or tokenize.

## **What is Word Embedding (Also referred to as Feature Extraction techniques in NLP?)**

Vectorization or word embedding is nothing but the process of converting text data to numerical vectors. Later the numerical vectors are used to build various machine learning models. In a way, we say this as extracting features from text to build multiple natural languages, processing models.

## **What is popular Word Embedding Techniques**

There are numerous ways to convert the text data to numerical vectors, refer to as word embedding techniques.

Bag of words | TF-IDF | Word2vec | Glove embedding | FastText | ELMO (Embeddings for Language models).

## **Describe a language model, and explain two methods one can use to evaluate it (one extrinsic, and one intrinsic)**

A language model is a statistical model that assigns probabilities to words and sentences. Typically, we might be trying to guess the next word w in a sentence given all previous words, often referred to as the “history”.

**Extrinsic evaluation.** This involves evaluating the models by employing them in an actual task (such as machine translation) and looking at their final loss/accuracy. This is the best option as it's the only way to tangibly see how different models affect the task we're interested in. However, it can be computationally expensive and slow as it requires training a full system.

**Intrinsic evaluation.** This involves finding some metric to evaluate the language model itself, not taking into account the specific tasks it's going to be

used for. While intrinsic evaluation is not as “good” as extrinsic evaluation as a final metric, it’s a useful way of quickly comparing models. Perplexity is an intrinsic evaluation method.

## **What are context vectors, encoders, and decoders in recurrent neural networks?**

**Encoder** reads the input sequence and summarizes the information in something called the internal state vectors or context vector (in the case of LSTM these are called the hidden state and cell state vectors).

The **decoder** is an LSTM whose initial states are initialized to the final states of the Encoder LSTM. Using these initial states, the decoder starts generating the output sequence, and these outputs are also taken into consideration for future outputs.

## **Describe the model architecture of the encoder-decoder network for translation, in both phases of training and inference.**

The model takes an input sequence and outputs a sequence as output (i.e., many-to-many) is known as the sequence-to-sequence task. The sequence-to-sequence tasks are very challenging as the size of the inputs and outputs vary.

To tackle many-to-many sequence prediction problems, researchers have explored and found a way in the form of Encoder-Decoder Architecture.

**Encoder-Decoder Model:** There are three main blocks in the encoder-decoder model,  
Encoder | Hidden Vector | Decoder

The Encoder will convert the input sequence into a single-dimensional vector (hidden vector). The decoder will convert the hidden vector into the output sequence. Encoder-Decoder models are jointly trained to maximize the conditional probabilities of the target sequence given the input sequence.

Encoder | Hidden Vector | Decoder

**Encoder:** Consider the input sequence “I am a Student” to be encoded. There will be totally 4 timesteps ( 4 tokens) for the Encoder model. At each time step, the hidden state  $h$  will be updated using the previous hidden state and the current input.

**Decoder:** The input for the decoder is the final hidden vector obtained at the end of the encoder model.

What is one hot encoding? length of encoding depends upon what?

One hot encoding: It is one of the techniques to represent features. Features are represented using a binary vector. The size of the vector depends upon the vocabulary. For  $n$ -words sentences, we have  $n \times v$ , where  $v$  is the size of the vocabulary.

## What is word2vec

Word2Vec is one of the most popular techniques to learn word embeddings using a shallow neural network. Words with similar contexts are placed closed spatially. It helps achieve the semantic meaning of the word.

It can be obtained using two methods (both involving Neural Networks): Skip Gram and Common Bag of Words (CBOW)

Like we have the concept of N-gram in the bag of words, here we have a concept of Window size. The Word2vec model will capture relationships of words with the help of window size by using skip-gram and CBow methods. We use a window to keep track of the center word and the context of the center word.

## What is Word2Vec – Skip Gram?

In this method, take the center word from the window size words as an input and context words (neighbor words) as outputs. Word2vec models predict the context words of a center word using the skip-gram method. Skip-gram works well with a small dataset and identifies rare words really well.

### Skip Gram

## What is Word2Vec- CBOW (Continuous Bag of Words)

CBOW is just a reverse method of the skip-gram method. Here we are taking context words as input and predicting the center word within the window. Another difference from the skip-gram method is, It was working faster and better representations for most frequency words.

## CBOW

### **What is the difference b/w Skip-gram and CBOW**

MM

### Difference between Skip-gram and CBOW

### **What is Bag of Words: Word embedded Technique?**

This method is mostly used in language modeling and text classification tasks. In a bag of words, we perform two operations. Tokenization and Vectors Creation.

**Tokenization:** The process of dividing each sentence into words or smaller parts. Here each word or symbol is called a token. After tokenization, we will take unique words from the corpus. Here corpus means the tokens we have from all the documents we are considering for the bag of words creation.

Create vectors for each sentence: Size of the vector is equal to the number of unique words of the corpus. For each sentence we will fill each position of a vector with corresponding word frequency in a particular sentence.

Consider three below sentence:

- - - this pasta is very tasty and affordable
    - this pasta is not tasty and is affordable
    - this pasta is very very delicious

Now we will perform tokenization. Dividing sentences into words and creating a list with all unique words and also in alphabetical order. We will get the below output after the tokenization step.

[“and”, “affordable”, “delicious”, “is”, “not”, “pasta”, “tasty”, “this”, “very”]

and affordable delicious is not pasta tasty this very

this pasta is very tasty and affordable.



this pasta is not tasty and is affordable



this pasta is very very delicious.



[dataaspirant.com](http://dataaspirant.com)

| s.no | and | affordable | delicious | is | not | pasta | tasty | this | very |
|------|-----|------------|-----------|----|-----|-------|-------|------|------|
| 1    | 1   | 1          | 0         | 1  | 0   | 1     | 1     | 1    | 1    |
| 2    | 1   | 1          | 0         | 2  | 1   | 1     | 1     | 1    | 0    |
| 3    | 0   | 0          | 1         | 1  | 0   | 1     | 0     | 1    | 2    |

### Example Bag of Words: Word embedded Technique

every sentence converting into vectors. We can also find sentence similarities after converting sentences to vectors.

In the above example, we are just taking each word as a feature, another name for this is 1-gram representee, we can also take bigram words, tri-Gram words, etc..

### What is the demerit of Bag of Words Technique?

In Bag of word representation, we have more zeros in the sparse matrices. The size of the matrix will be increased based on the total number of words in the corpus. In real-world applications, the corpus will contain thousands of words. So, we need more resources to build analytics models with this type of technique for large datasets.

We need serious computing resources to build an analytical model based on the bag of words technique.

## **What is a Good NLP model?**

Best models would be able to capture 4 components:

- **Lexical approach** (relating to the words or vocabulary of a language) (covered by BOW, TF-IDF)
- **Syntactic approach** (the arrangement of words and phrases to create well-formed sentences in a language -> grammar) One hot encoding.
- **Semantic approach** (relating to meaning in language -> stretch the words beyond, understanding ambiguities) Word2Vec (Skip CBOW)
- **Contextual/Pragmatic approach** (relating proximity between words and documents)

**1. What Is Nlp?**

**2. List Some Components Of Nlp?**

**3. List Some Areas Of Nlp?**

**4. Define The Nlp Terminology?**

**5. What Is The Significance Of Tf-idf?**

**1. What Is Part Of Speech (pos) Tagging?**

**2. What Is Pragmatic Analysis In Nlp?**

**3. Explain Dependency Parsing In Nlp?**

**4. What Is Pac Learning?**

**5. What Are The Different Categories in The Sequence of Learning Process?**

**6. What Is Sequence Learning?**

**7. What are The General Principle Of An Ensemble Method?**

**8. What Is Bagging And Boosting In Ensemble Method?**

**9. What Is The Difference Between Heuristic For Rule Learning And Heuristics For Decision Trees?**

**10. What is applications of NLP ?**

**12. What is tokenization ?**

**12. What are stop words ?**

- 13. What is Noise Removal ?**
  - 14. What is Wordnet ?**
  - 15. What is NLG (Natural language Generation) ?**
  - 16. What is NLU (Natural language understanding) ?**
  - 17. What is Corpus ?**
  - 18. What is N- Gram, Unigram, Bigram and Trigram?**
  - 19. What is Language modeling ?**
  - 20. What is Latent semantic analysis ?**
  - 21. What are word embedding libraries ?**
  - 22. What is word2vec ?**
  - 23. What is Glove ?**
  - 24. What is Fasttext ?**
  - 25. What is Genism ?**
  - 26. What is text mining ?**
  - 27. What is Information Extraction ?**
  - 28. What is object standardization ? When it will be used ?**
  - 29. What is Topic Modeling ? When we will do it ?**
  - 30. What is document-term matrix ?**
- Q1. What is the difference between AI, Machine Learning and Deep Learning?**

| Artificial Intelligence                                          | Machine Learning                                                                  | Deep Learning                                                                            |
|------------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| Originated in 1950s                                              | Originated in 1960s                                                               | Originated in 1970s                                                                      |
| AI represents simulated intelligence in machine                  | ML is the practice of getting machines to make decisions without being programmed | DL is the process of using Artificial Neural Networks to solve complex problems          |
| AI is a subset of Data Science                                   | ML is a subset of AI & Data Science                                               | DL is a subset of ML, AI & Data Science                                                  |
| Aims to build machines which are capable of thinking like humans | Aims to build machines learning through data so that they can solve problems      | Aims to build neural networks that automatically discover patterns for feature detection |

*Artificial Intelligence vs Machine Learning vs Deep Learning – Artificial Intelligence Interview Questions – Edureka*

- Q2. What is Artificial Intelligence? Give an example of where AI is used on a daily basis.**

“Artificial Intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and react like humans.” “The capability of a machine to imitate the intelligent human behavior.”



*Google's Search Engine – Artificial Intelligence Interview Questions – Edureka*

**Google's Search Engine**  
One of the most popular AI Applications is the google search engine. If you open up your chrome browser and start typing something, Google immediately provides recommendations for you to choose from. The logic behind the search engine is Artificial Intelligence.

AI uses predictive analytics, NLP and Machine Learning to recommend relevant searches to you. These recommendations are based on data that Google collects about you, such as your search history, location, age, etc. Thus, Google makes use of AI, to predict what you might be looking for.

**ChatGPT Tutorial | ChatGPT Explained | What is ChatGPT ? | Edureka**



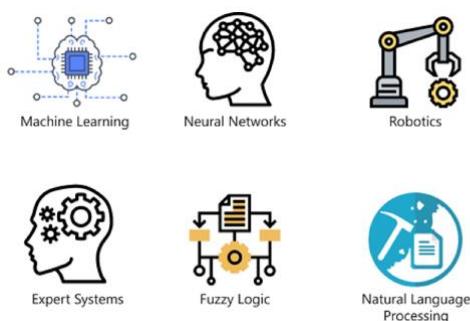
This **ChatGPT Tutorial** is intended as a Crash Course on **ChatGPT** for Beginners. **ChatGPT** has been growing in popularity exponentially. But, **ChatGPT** is still not known to many people. In this video, I aim to show you the different ways in which you can use **ChatGPT** for yourself. **ChatGPT** has been the buzzword for a while now. This yap lab was put on the throne 5 days after its release and has been changing the game ever since.

Unlock the power of language with this [ChatGPT training](#) online from Edureka.

### Q3. What are the different types of AI?

- **Reactive Machines AI:** Based on present actions, it cannot use previous experiences to form current decisions and simultaneously update their memory.  
Example: Deep Blue
- **Limited Memory AI:** Used in self-driving cars. They detect the movement of vehicles around them constantly and add it to their memory.
- **Theory of Mind AI:** Advanced AI that has the ability to understand emotions, people and other things in the real world.
- **Self Aware AI:** AIs that possess human-like consciousness and reactions. Such machines have the ability to form self-driven actions.
- **Artificial Narrow Intelligence (ANI):** General purpose AI, used in building virtual assistants like Siri.
- **Artificial General Intelligence (AGI):** Also known as strong AI. An example is the Pillo robot that answers questions related to health.
- **Artificial Superhuman Intelligence (ASI):** AI that possesses the ability to do everything that a human can do and more. An example is the Alpha 2 which is the first humanoid ASI robot.

### Q4. Explain the different domains of Artificial Intelligence.



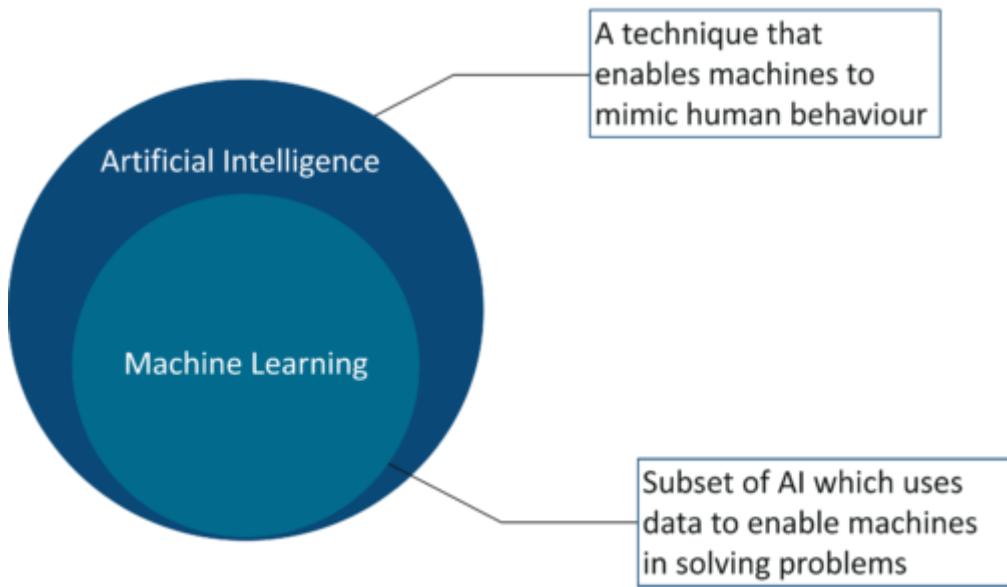
*Domains Of AI – Artificial Intelligence Interview Questions – Edureka*

- **Machine Learning:** It's the science of getting computers to act by feeding them data so that they can learn a few tricks on their own, without being explicitly programmed to do so.
- **Neural Networks:** They are a set of algorithms and techniques, modeled in accordance with the human brain. Neural Networks are designed to solve complex and advanced machine learning problems.
- **Robotics:** Robotics is a subset of AI, which includes different branches and application of robots. These Robots are artificial agents acting in a real-world environment. An AI Robot works by manipulating the objects in it's surrounding, by perceiving, moving and taking relevant actions.
- **Expert Systems:** An expert system is a computer system that mimics the decision-making ability of a human. It is a computer program that uses artificial intelligence (AI) technologies to simulate the judgment and behavior of a human or an organization that has expert knowledge and experience in a particular field.
- **Fuzzy Logic Systems:** Fuzzy logic is an approach to computing based on “degrees of truth” rather than the usual “true or false” (1 or 0) boolean logic on which the modern computer is based. Fuzzy logic Systems can take imprecise, distorted, noisy input information.
- **Natural Language Processing:** Natural Language Processing (NLP) refers to the Artificial Intelligence method that analyses natural human language to derive useful insights in order to solve problems.

## **Q5. How is Machine Learning related to Artificial Intelligence?**

Artificial Intelligence is a technique that enables machines to mimic human behavior. Whereas, Machine Learning is a subset of Artificial Intelligence. It is the science of getting computers to act by feeding them data and letting them learn a few tricks on their own, without being explicitly programmed to do so.

Therefore Machine Learning is a technique used to implement Artificial Intelligence.



*Artificial Intelligence vs Machine Learning – Artificial Intelligence Interview Questions – Edureka*

Also, enroll in [Artificial Intelligence Course](#) to become proficient in this AI and ML.

## **Q6. What are the different types of Machine Learning?**

|                    | Supervised Learning                                                      | Unsupervised Learning                                          | Reinforcement Learning                                                                     |
|--------------------|--------------------------------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| Definition         | The machine learns by using labelled data                                | The machine is trained on unlabelled data without any guidance | An agent interacts with its environment by producing actions & discovers errors or rewards |
| Type of problems   | Regression & Classification                                              | Association & Clustering                                       | Reward based                                                                               |
| Type of data       | Labelled data                                                            | Unlabelled data                                                | No pre-defined data                                                                        |
| Training           | External supervision                                                     | No supervision                                                 | No supervision                                                                             |
| Approach           | Map labelled input to known output                                       | Understand patterns and discover output                        | Follow trial and error method                                                              |
| Popular algorithms | Linear regression, Logistic regression, Support Vector Machine, KNN, etc | K-means, C-means, etc                                          | Q-Learning, SARSA, etc                                                                     |

*Types Of Machine Learning – Artificial Intelligence Interview Questions – Edureka*

## **Q7. What is Q-Learning?**

*The Q-learning is a Reinforcement Learning algorithm in which an agent tries to learn the optimal policy from its past experiences with the environment. The past experiences of an agent are a sequence of state-action-rewards:*

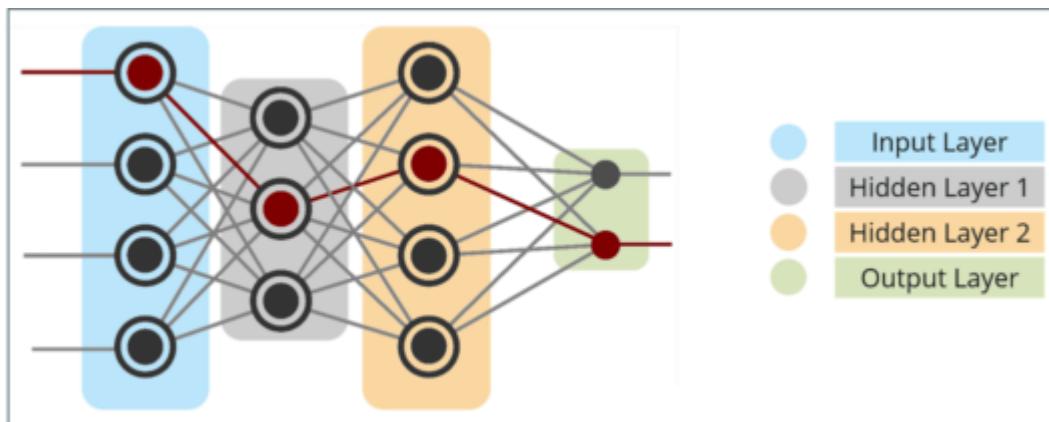
**s0** ————— **a0** ————— **r1** ————— **s1**

*What Is Q-Learning? – Artificial Intelligence Interview Questions – Edureka*

In the above state diagram, the Agent(a0) was in State (s0) and on performing an Action (a0), which resulted in receiving a Reward (r1) and thus being updated to State (s1).

### **Q8. What is Deep Learning?**

Deep learning imitates the way our brain works i.e. it learns from experiences. It uses the concepts of neural networks to solve complex problems.

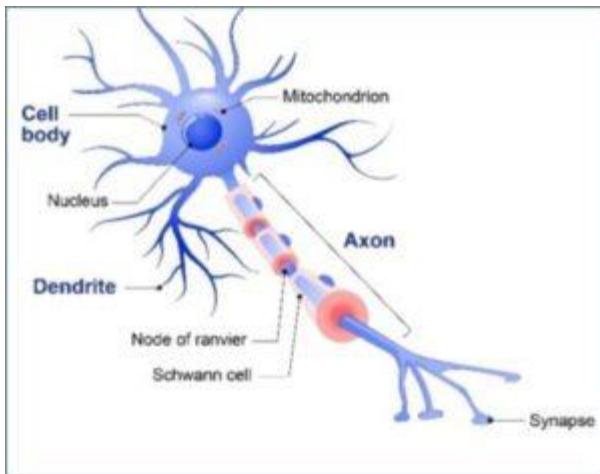


*What Is Deep Learning? – Artificial Intelligence Interview Questions – Edureka*

Any Deep neural network will consist of three types of layers:

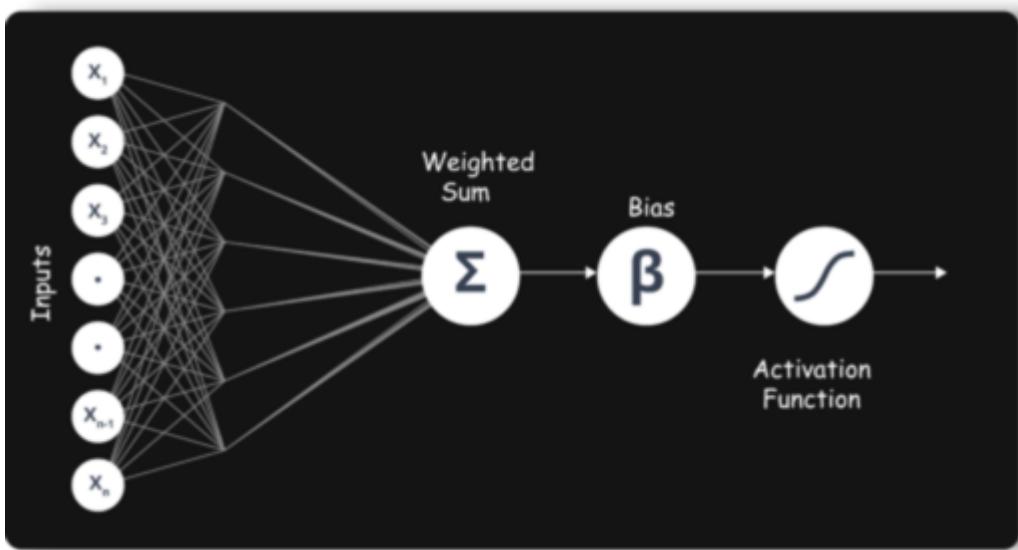
- **Input Layer:** This layer receives all the inputs and forwards them to the hidden layer for analysis
- **Hidden Layer:** In this layer, various computations are carried out and the result is transferred to the output layer. There can be n number of hidden layers, depending on the problem you're trying to solve.
- **Output Layer:** This layer is responsible for transferring information from the neural network to the outside world.

### **Q9. Explain how Deep Learning works.**



### *Biological Neurons – Artificial Intelligence Interview Questions – Edureka*

- Deep Learning is based on the basic unit of a brain called a brain cell or a neuron. Inspired from a neuron, an artificial neuron or a perceptron was developed.
- A biological neuron has dendrites which are used to receive inputs.
- Similarly, a perceptron receives multiple inputs, applies various transformations and functions and provides an output.
- Just like how our brain contains multiple connected neurons called neural network, we can also have a network of artificial neurons called perceptron's to form a Deep neural network.



### *Deep Neural Network – Artificial Intelligence Interview Questions – Edureka*

- An Artificial Neuron or a Perceptron models a neuron which has a set of inputs, each of which is assigned some specific weight. The neuron then computes some function on these weighted inputs and gives the output.

## **Q10. Explain the commonly used Artificial Neural Networks.**

### **Feedforward Neural Network**

- The simplest form of ANN, where the data or the input travels in one direction.
- The data passes through the input nodes and exit on the output nodes. This neural network may or may not have the hidden layers.

### **Convolutional Neural Network**

- Here, input features are taken in batch wise like a filter. This will help the network to remember the images in parts and can compute the operations.
- Mainly used for signal and image processing

### **Recurrent Neural Network(RNN) – Long Short Term Memory**

- Works on the principle of saving the output of a layer and feeding this back to the input to help in predicting the outcome of the layer.
- Here, you let the neural network to work on the front propagation and remember what information it needs for later use
- This way each neuron will remember some information it had in the previous time-step.

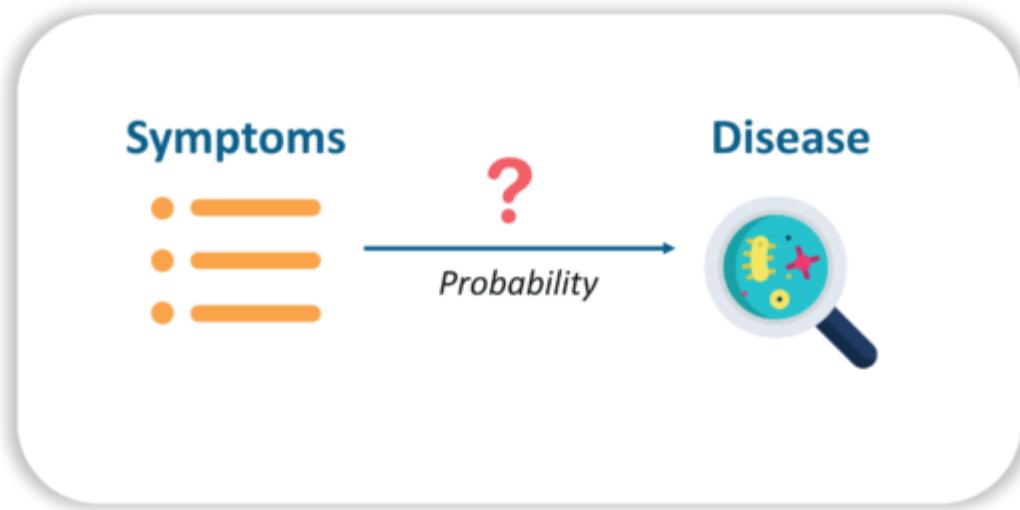
### **Autoencoders**

- These are unsupervised learning models with an input layer, an output layer and one or more hidden layers connecting them.
- The output layer has the same number of units as the input layer. Its purpose is to reconstruct its own inputs.
- Typically for the purpose of dimensionality reduction and for learning generative models of data.

## **Q11. What are Bayesian Networks?**

A Bayesian network is a statistical model that represents a set of variables and their conditional dependencies in the form of a directed acyclic graph.

On the occurrence of an event, Bayesian Networks can be used to predict the likelihood that any one of several possible known causes was the contributing factor.



### *Bayesian Network – Artificial Intelligence Interview Questions – Edureka*

For example, a Bayesian network could be used to study the relationship between diseases and symptoms. Given various symptoms, the Bayesian network is ideal for computing the probabilities of the presence of various diseases.

### **Q12. Explain the assessment that is used to test the intelligence of a machine.**

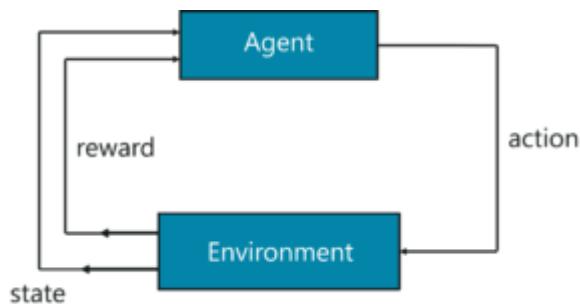
*In artificial intelligence (AI), a Turing Test is a method of inquiry for determining whether or not a computer is capable of thinking like a human being.*



### **Q1. How does Reinforcement Learning work? Explain with an example.**

Generally, a Reinforcement Learning (RL) system is comprised of two main components:

1. An agent
2. An environment



### *Reinforcement Learning – Artificial Intelligence Interview Questions – Edureka*

- The environment is the setting that the agent is acting on and the agent represents the RL algorithm.
- The RL process starts when the environment sends a state to the agent, which then based on its observations, takes an action in response to that state.
- In turn, the environment sends the next state and the respective reward back to the agent. The agent will update its knowledge with the reward returned by the environment to evaluate its last action.
- The loop continues until the environment sends a terminal state, which means the agent has accomplished all his tasks.

To understand this better, let's suppose that our agent is learning to play counterstrike. The RL process can be broken down into the below steps:



### *Counter-Strike Example – Artificial Intelligence Interview Questions – Edureka*

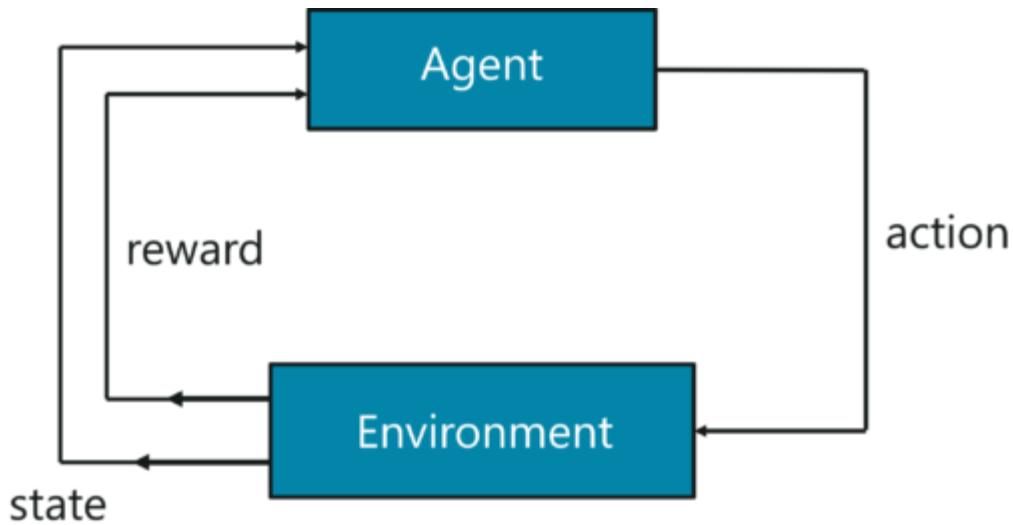
1. The RL Agent (Player1) collects state  $S^0$  from the environment (Counterstrike game)
2. Based on the state  $S^0$ , the RL agent takes an action  $A^0$ , (Action can be anything that causes a result i.e. if the agent moves left or right in the game). Initially, the action is random
3. The environment is now in a new state  $S^1$  (new stage in the game)
4. The RL agent now gets a reward  $R^1$  from the environment. This reward can be additional points or coins
5. This RL loop goes on until the RL agent is dead or reaches the destination, and it continuously outputs a sequence of state, action, and reward.

### **Q2. Explain Markov’s decision process with an example.**

The mathematical approach for mapping a solution in Reinforcement Learning is called Markov’s Decision Process (MDP).

The following parameters are used to attain a solution using MDP:

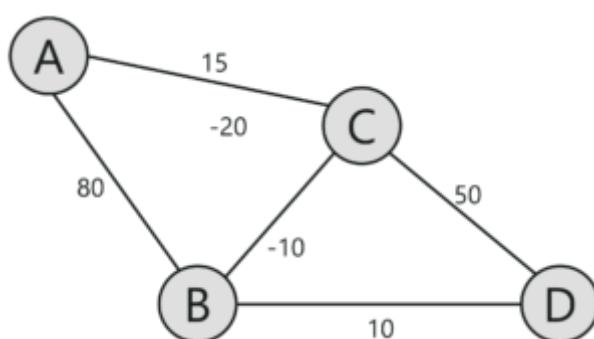
- Set of actions, A
- Set of states, S
- Reward, R
- Policy,  $\pi$
- Value, V



### *Markov's Decision Process – Artificial Intelligence Interview Questions – Edureka*

To briefly sum it up, the agent must take an action (A) to transition from the start state to the end state (S). While doing so, the agent receives rewards (R) for each action he takes. The series of actions taken by the agent, define the policy ( $\pi$ ) and the rewards collected define the value (V). The main goal here is to maximize rewards by choosing the optimum policy.

To better understand the MDP, let's solve the Shortest Path Problem using the MDP approach:



*Shortest Path Problem – Artificial Intelligence Interview Questions – Edureka*

Given the above representation, our goal here is to find the shortest path between 'A' and 'D'. Each edge has a number linked with it, this denotes the cost to traverse that edge. Now, the task at hand is to traverse from point 'A' to 'D', with minimum possible cost.

In this problem,

- The set of states are denoted by nodes i.e. {A, B, C, D}
- The action is to traverse from one node to another {A -> B, C -> D}
- The reward is the cost represented by each edge
- The policy is the path taken to reach the destination

You start off at node A and take baby steps to your destination. Initially, only the next possible node is visible to you, thus you randomly start off and then learn as you traverse through the network. The main goal is to choose the path with the lowest cost.

Since this is a very simple problem, I will leave it for you to solve. Make sure you mention the answer in the comment section.

### **Q3. Explain reward maximization in Reinforcement Learning.**

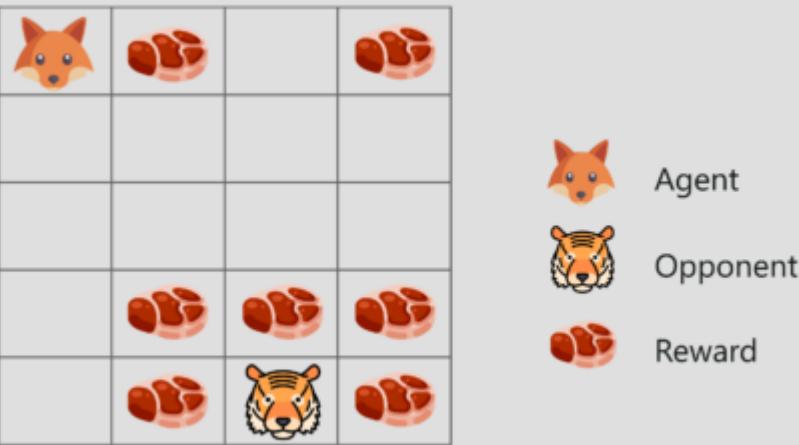
*The RL agent works based on the theory of reward maximization. This is exactly why the RL agent must be trained in such a way that, he takes the best action so that the reward is maximum.*

The collective rewards at a particular time with the respective action is written as:

$$G_t = R_{t+1} + R_{t+2} + \dots$$

*Reward Maximization Equation – Artificial Intelligence Interview Questions – Edureka*

The above equation is an ideal representation of rewards. Generally, things don't work out like this while summing up the cumulative rewards.



### *Reward Maximization – Artificial Intelligence Interview Questions – Edureka*

Let me explain this with a small game. In the figure you can see a fox, some meat and a tiger.

- Our RL agent is the fox and his end goal is to eat the maximum amount of meat before being eaten by the tiger.
- Since this fox is a clever fellow, he eats the meat that is closer to him, rather than the meat which is close to the tiger, because the closer he is to the tiger, the higher are his chances of getting killed.
- As a result, the rewards near the tiger, even if they are bigger meat chunks, will be discounted. This is done because of the uncertainty factor, that the tiger might kill the fox.

The next thing to understand is, how discounting of rewards work?

To do this, we define a discount rate called gamma. The value of gamma is between 0 and 1. The smaller the gamma, the larger the discount and vice versa.

So, our cumulative discounted rewards is:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \text{ where } \gamma \in [0, 1)$$

$$R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \dots$$

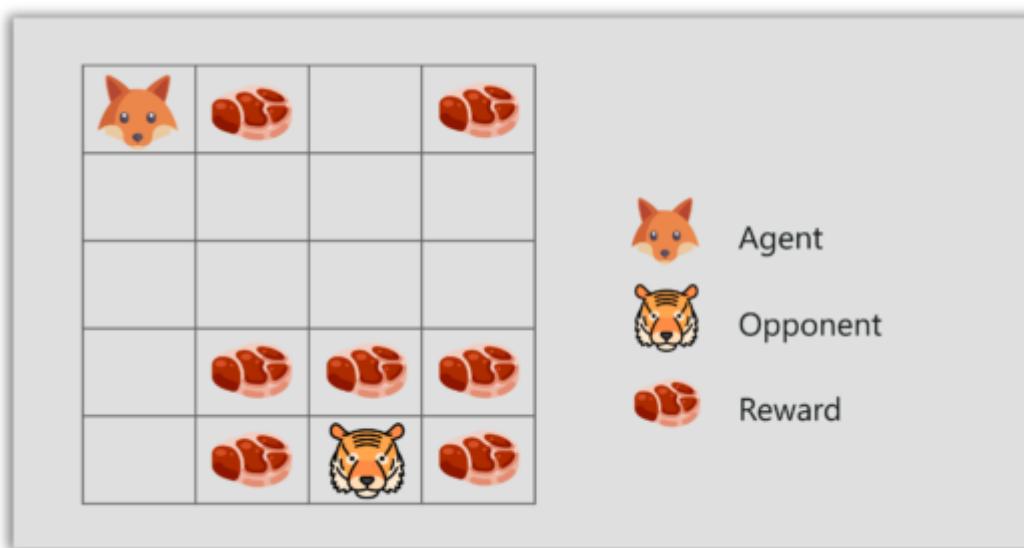
## Post Graduate Diploma in Artificial Intelligence Course

### Explore Curriculum

#### Q4. What is exploitation and exploration trade-off?

An important concept in reinforcement learning is the exploration and exploitation trade-off.

*Exploration, like the name suggests, is about exploring and capturing more information about an environment. On the other hand, exploitation is about using the already known exploited information to heighten the rewards.*



*Exploitation & Exploration – Artificial Intelligence Interview Questions – Edureka*

- Consider the fox and tiger example, where the fox eats only the meat (small) chunks close to him but he doesn't eat the bigger meat chunks at the top, even though the bigger meat chunks would get him more rewards.
- If the fox only focuses on the closest reward, he will never reach the big chunks of meat, this is called exploitation.
- But if the fox decides to explore a bit, it can find the bigger reward i.e. the big chunk of meat. This is exploration.

#### Q5. What is the difference between parametric & non-parametric models?

| Parametric model                                         | Non-parametric model                                      |
|----------------------------------------------------------|-----------------------------------------------------------|
| It uses a fixed number of parameters to build the model. | It uses flexible number of parameters to build the model. |
| Considers strong assumptions about the data.             | Considers fewer assumptions about the data.               |
| Computationally faster                                   | Computationally slower                                    |
| Require lesser data                                      | Require more data                                         |
| Example – Logistic Regression & Naïve Bayes models       | Example – KNN & Decision Tree models                      |

*Parametric vs Non Parametric model – Artificial Intelligence Interview Questions – Edureka*

## **Q6. What is the difference between Hyperparameters and model parameters?**

| Model Parameters                                                                                                         | Hyperparameters                                                                                                           |
|--------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|
| Model parameters are the features of training data that will learn on its own during training.                           | Model Hyperparameters are the parameters that determine the entire training process.                                      |
| For example, <ul style="list-style-type: none"> <li>Weights and Biases</li> <li>Split points in Decision Tree</li> </ul> | For example, <ul style="list-style-type: none"> <li>Learning Rate</li> <li>Hidden Layers</li> <li>Hidden Units</li> </ul> |
| They are internal to the model and their value can be estimated from data.                                               | They are external to the model and their value cannot be estimated from data.                                             |

*Model Parameters vs Hyperparameters – Artificial Intelligence Interview Questions – Edureka*

## **Q7. What are hyperparameters in Deep Neural Networks?**

- Hyperparameters are variables that define the structure of the network. For example, variables such as the learning rate, define how the network is trained.
- They are used to define the number of hidden layers that must be present in a network.
- More hidden units can increase the accuracy of the network, whereas a lesser number of units may cause underfitting.

## **Q8. Explain the different algorithms used for hyperparameter optimization.**

**Grid Search**  
 Grid search trains the network for every combination by using the two set of hyperparameters, learning rate and the number of layers. Then evaluates the model by using Cross Validation techniques.

**Random Search**  
 It randomly samples the search space and evaluates sets from a particular probability distribution. For example, instead of checking all 10,000 samples, randomly selected 100 parameters can be checked.

**Bayesian Optimization**  
 This includes fine-tuning the hyperparameters by enabling automated model tuning. The model used for approximating the objective function is called surrogate model (Gaussian Process). Bayesian Optimization uses Gaussian Process (GP) function to get posterior functions to make predictions based on prior functions.

## **Q9. How does data overfitting occur and how can it be fixed?**

Overfitting occurs when a statistical model or machine learning algorithm captures the noise of the data. This causes an algorithm to show low bias but high variance in the outcome.

Overfitting can be prevented by using the following methodologies:

**Cross-validation:** The idea behind cross-validation is to split the training data in order to generate multiple mini train-test splits. These splits can then be used to tune your model.

**More training data:** Feeding more data to the machine learning model can help in better analysis and classification. However, this does not always work.

**Remove features:** Many times, the data set contains irrelevant features or predictor variables that are not needed for analysis. Such features only increase the complexity of the model, thus leading to possibilities of data overfitting. Therefore, such redundant variables must be removed.

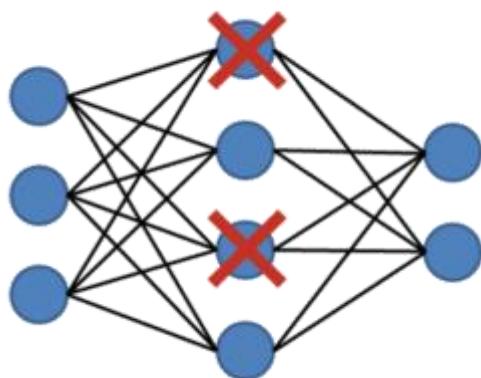
**Early stopping:** A machine learning model is trained iteratively, this allows us to check how well each iteration of the model performs. But after a certain number of iterations, the model's performance starts to saturate. Further training will result in overfitting, thus one must know where to stop the training. This can be achieved by a mechanism called early stopping.

**Regularization:** Regularization can be done in n number of ways, the method will depend on the type of learner you're implementing. For example, pruning is performed on decision trees, the dropout technique is used on neural networks and parameter tuning can also be applied to solve overfitting issues.

**Use Ensemble models:** Ensemble learning is a technique that is used to create multiple Machine Learning models, which are then combined to produce more accurate results. This is one of the best ways to prevent overfitting. An example is Random Forest, it uses an ensemble of decision trees to make more accurate predictions and to avoid overfitting.

#### **Q10. Mention a technique that helps to avoid overfitting in a neural network.**

*Dropout is a type of regularization technique used to avoid overfitting in a neural network. It is a technique where randomly selected neurons are dropped during training.*



The Dropout value of a network must be chosen wisely. A value too low will result in a minimal effect and a value too high results in under-learning by the network.

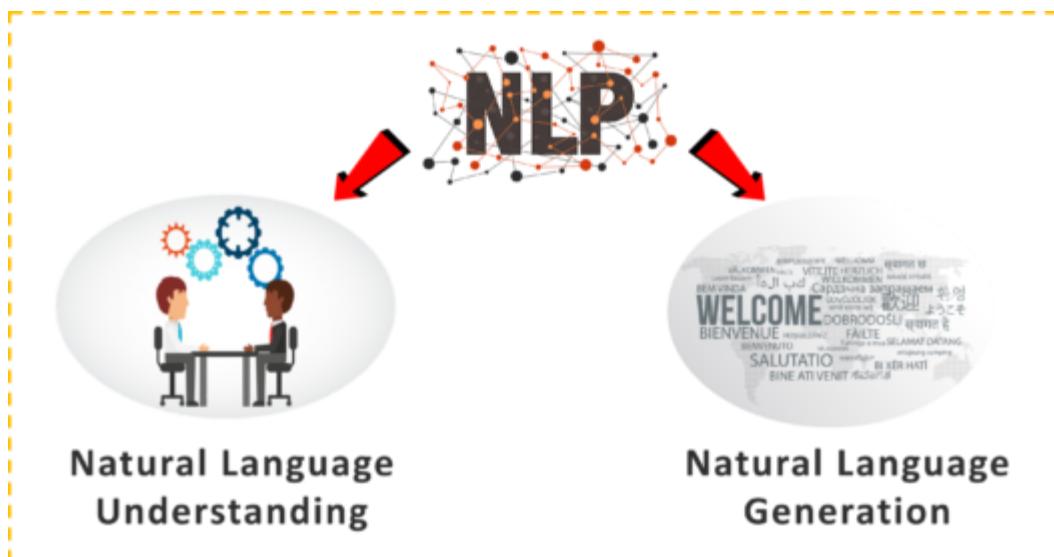
### **Q11. What is the purpose of Deep Learning frameworks such as Keras, TensorFlow, and PyTorch?**

- Keras is an open source neural network library written in Python. It is designed to enable fast experimentation with deep neural networks.
- TensorFlow is an open-source software library for dataflow programming. It is used for machine learning applications like neural networks.
- PyTorch is an open source machine learning library for Python, based on Torch. It is used for applications such as natural language processing.

### **Q12. Differentiate between NLP and Text mining.**

| Text Mining                                                                                                             | Natural Language Processing                                                                                                                        |
|-------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| Aim of text mining is to extract useful insights from structured & un-structured text.                                  | Aim of NLP is to understand what is conveyed in speech.                                                                                            |
| Text Mining can be done using text processing languages like Perl, statistical models, etc.                             | NLP can be achieved using advanced machine learning models, deep neural networks, etc.                                                             |
| Outcome: <ul style="list-style-type: none"><li>• Frequency of words</li><li>• Patterns</li><li>• Correlations</li></ul> | Outcome: <ul style="list-style-type: none"><li>• Semantic meaning of text</li><li>• Sentimental analysis</li><li>• Grammatical structure</li></ul> |

### **Q13. What are the different components of NLP?**



### *Components Of NLP – Artificial Intelligence Interview Questions – Edureka*

Natural Language Understanding includes:

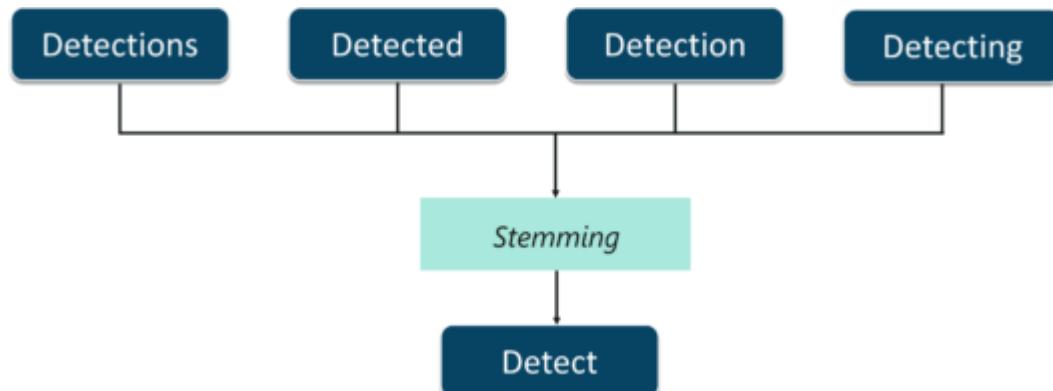
- Mapping input to useful representations
- Analyzing different aspects of the language

Natural Language Generation includes:

- Text Planning
- Sentence Planning
- Text Realization

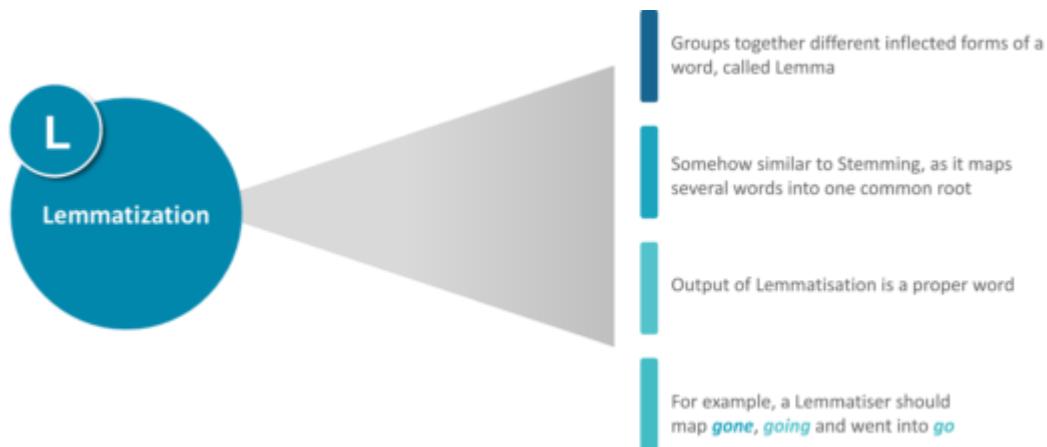
### **Q14. What is Stemming & Lemmatization in NLP?**

Stemming algorithms work by cutting off the end or the beginning of the word, taking into account a list of common prefixes and suffixes that can be found in an inflected word. This indiscriminate cutting can be successful on some occasions, but not always.

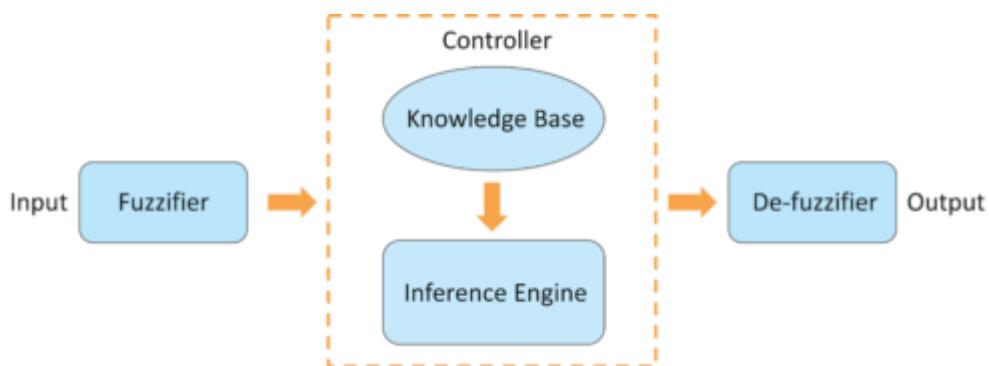


## *Stemming – Artificial Intelligence Interview Questions – Edureka*

Lemmatization, on the other hand, takes into consideration the morphological analysis of the words. To do so, it is necessary to have detailed dictionaries which the algorithm can look through to link the form back to its lemma.



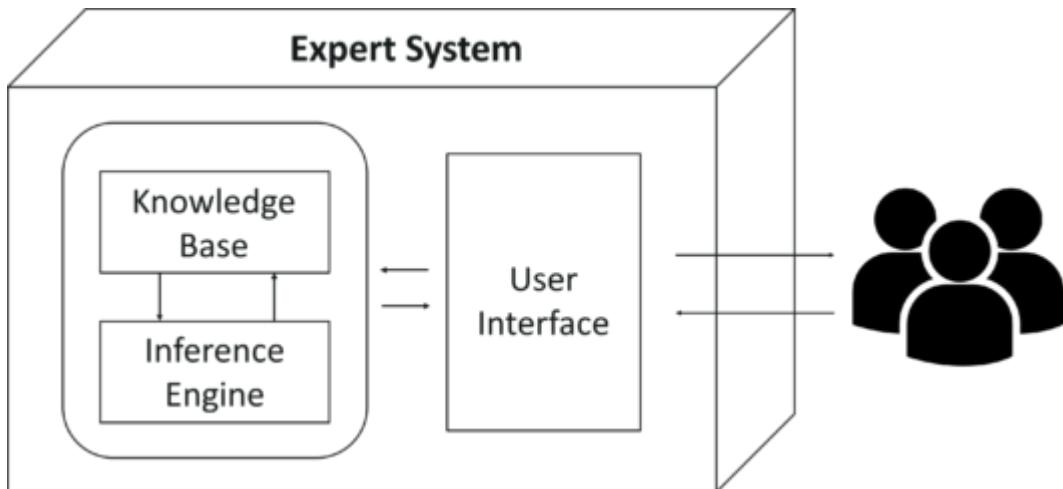
## **Q15. Explain Fuzzy Logic architecture.**



## *Fuzzy Logic Architecture – Artificial Intelligence Interview Questions – Edureka*

- Fuzzification Module** – The system inputs are fed into the Fuzzifier, which transforms the inputs into fuzzy sets.
- Knowledge Base** – It stores analytic measures such as IF-THEN rules provided by experts.
- Inference Engine** – It simulates the human reasoning process by making fuzzy inference on the inputs and IF-THEN rules.
- Defuzzification Module** – It transforms the fuzzy set obtained by the inference engine into a crisp value.

## **Q16. Explain the components of Expert Systems.**

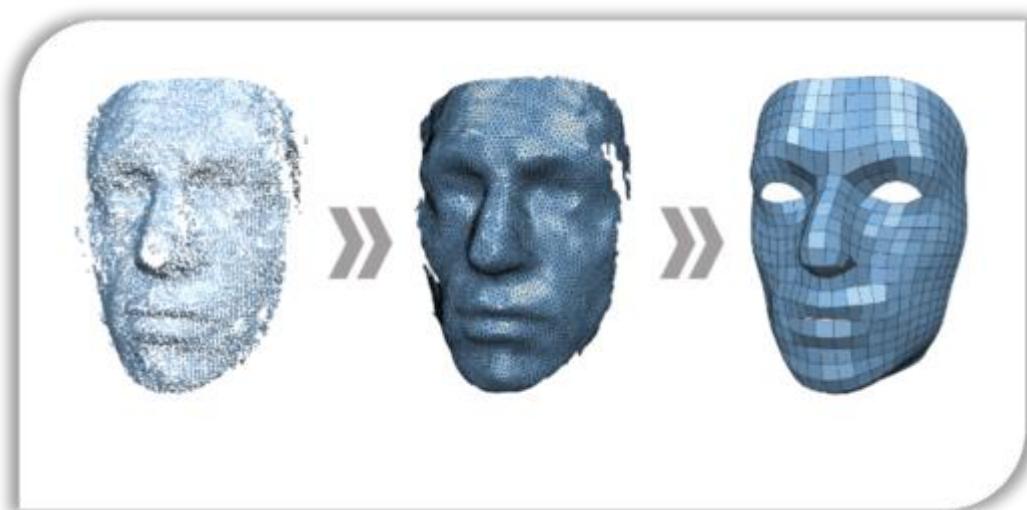


*Expert Systems – Artificial Intelligence Interview Questions – Edureka*

- **Knowledge Base**  
It contains domain-specific and high-quality knowledge.
- **Inference Engine**  
It acquires and manipulates the knowledge from the knowledge base to arrive at a particular solution.
- **User Interface**  
The user interface provides interaction between the user and the Expert System itself.

### **Q17. How is Computer Vision and AI related?**

Computer Vision is a field of Artificial Intelligence that is used to obtain information from images or multi-dimensional data. Machine Learning algorithms such as K-means is used for Image Segmentation, Support Vector Machine is used for Image Classification and so on.



## *Computer Vision And AI – Artificial Intelligence Interview Questions – Edureka*

Therefore Computer Vision makes use of AI technologies to solve complex problems such as Object Detection, Image Processing, etc.

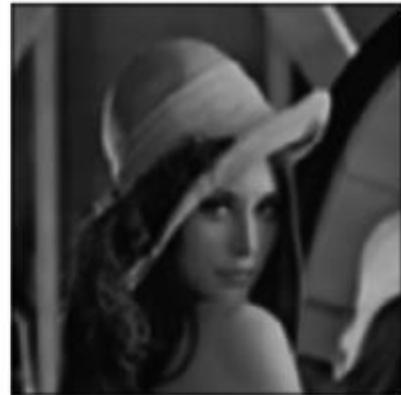
### **Q18. Which is better for image classification? Supervised or unsupervised classification? Justify.**

- In supervised classification, the images are manually fed and interpreted by the Machine Learning expert to create feature classes.
- In unsupervised classification, the Machine Learning software creates feature classes based on image pixel values.

Therefore, it is better to choose supervised classification for image classification in terms of accuracy.

### **Q19. Finite difference filters in image processing are very susceptible to noise. To cope up with this, which method can you use so that there would be minimal distortions by noise?**

Image Smoothing is one of the best methods used for reducing noise by forcing pixels to be more like their neighbors, this reduces any distortions caused by contrasts.



## *Image Smoothing – Artificial Intelligence Interview Questions – Edureka*

### **Q20. How is Game theory and AI related?**

*“In the context of artificial intelligence(AI) and deep learning systems, game theory is essential to enable some of the key capabilities required in multi-agent*

*environments in which different AI programs need to interact or compete in order to accomplish a goal.”*

*Game Theory And AI – Artificial Intelligence Interview Questions – Edureka*

## **Q21. What is the Minimax Algorithm? Explain the terminologies involved in a Minimax problem.**

*Minimax is a recursive algorithm used to select an optimal move for a player assuming that the other player is also playing optimally.*

A game can be defined as a search problem with the following components:

- **Game Tree:** A tree structure containing all the possible moves.
- **Initial state:** The initial position of the board and showing whose move it is.
- **Successor function:** It defines the possible legal moves a player can make.
- **Terminal state:** It is the position of the board when the game ends.
- **Utility function:** It is a function which assigns a numeric value for the outcome of a game.

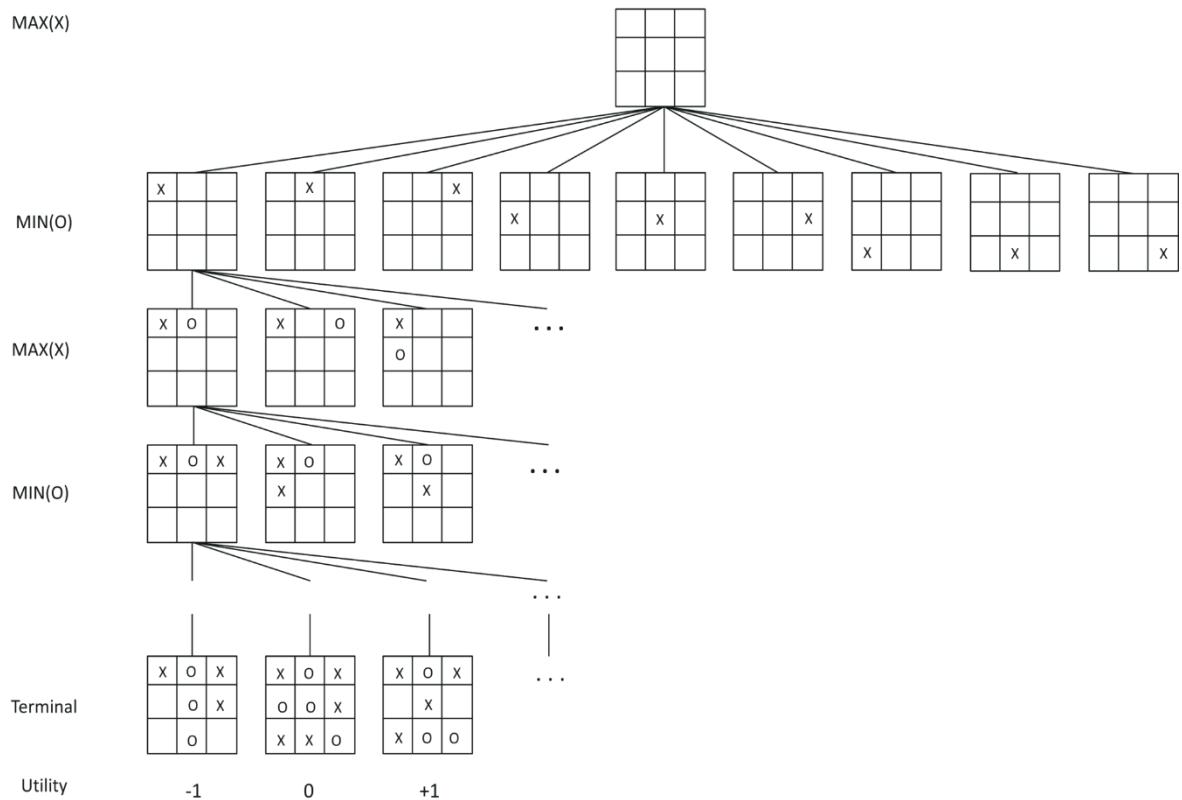
## **Q1. Show the working of the Minimax algorithm using Tic-Tac-Toe Game.**

There are two players involved in a game:

- **MAX:** This player tries to get the highest possible score
- **MIN:** MIN tries to get the lowest possible score

The following approach is taken for a Tic-Tac-Toe game using the Minimax algorithm:

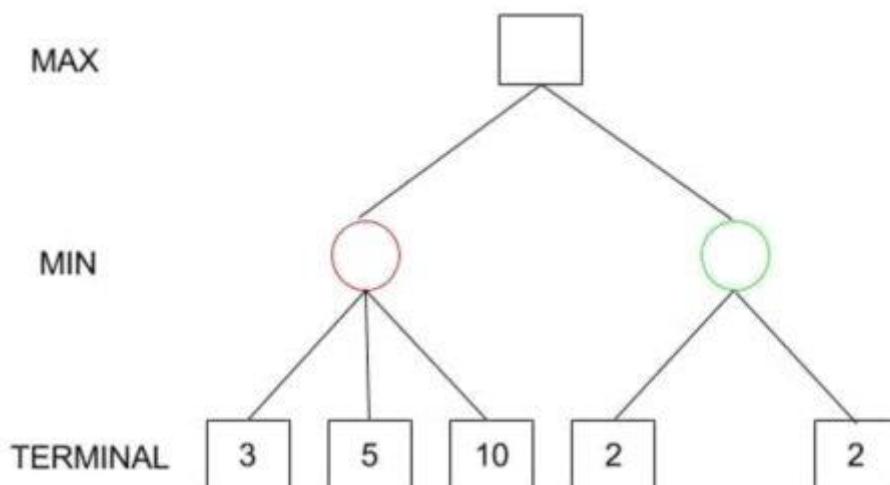
**Step 1:** First, generate the entire game tree starting with the current position of the game all the way up to the terminal states.



*Tic-Tac-Toe – Artificial Intelligence Interview Questions – Edureka*

**Step 2:** Apply the utility function to get the utility values for all the terminal states.

**Step 3:** Determine the utilities of the higher nodes with the help of the utilities of the terminal nodes. For instance, in the diagram below, we have the utilities for the terminal states written in the squares.



## Tic-Tac-Toe – Artificial Intelligence Interview Questions – Edureka

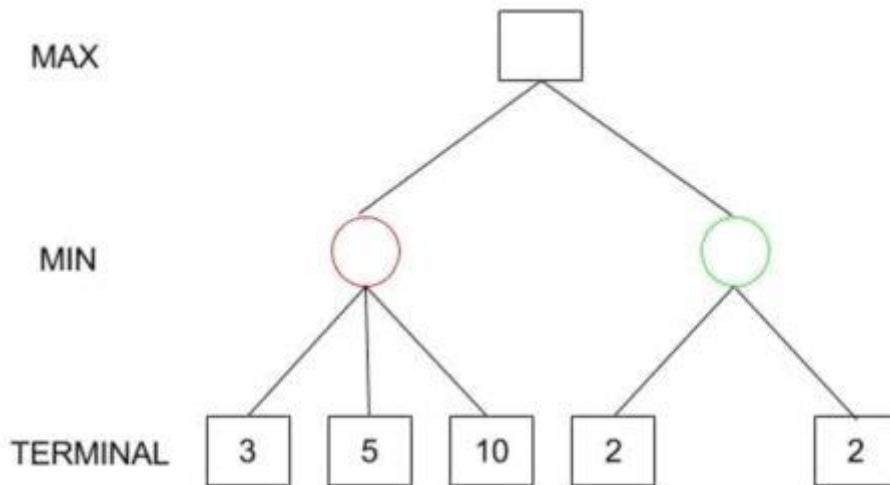
Let us calculate the utility for the left node(red) of the layer above the terminal:

$\text{MIN}\{3, 5, 10\}$ , i.e. 3.

Therefore, the utility for the red node is 3.

Similarly, for the green node in the same layer:

$\text{MIN}\{2,2\}$ , i.e. 2.



## Tic-Tac-Toe – Artificial Intelligence Interview Questions – Edureka

**Step 4:** Calculate the utility values.

**Step 5:** Eventually, all the backed-up values reach to the root of the tree. At that point, MAX has to choose the highest value: i.e.  $\text{MAX}\{3,2\}$  which is 3.

Therefore, the best opening move for MAX is the left node(or the red one). To summarize,

$$\begin{aligned} \text{Minimax Decision} &= \text{MAX}\{\text{MIN}\{3,5,10\}, \text{MIN}\{2,2\}\} \\ &= \text{MAX}\{3,2\} \\ &= 3 \end{aligned}$$

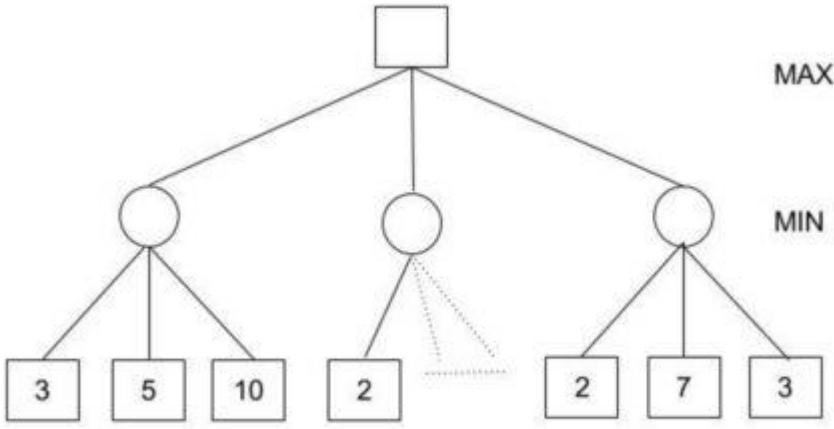
**Q2. Which method is used for optimizing a Minimax based game?**

**Alpha-beta**

If we apply alpha-beta pruning to a standard minimax algorithm, it returns the

**Pruning**

same move as the standard one, but it removes all the nodes that are possibly not affecting the final decision.



### *Alpha-beta Pruning – Artificial Intelligence Interview Questions – Edureka*

1. What do you mean by natural language and computer language?

Natural language is a human language such as Hindi, English, and German. Designing computers that can understand natural languages is a challenging problem.

Computer language is a set of instructions that used to produce the desired output such as C, C++, Python, Julia, and Scala.

2. What is the difference between NLP and NLU?

Natural language processing is one of the components of text mining. NLP helps identified sentiments, finding entities in the sentence, and category of blog/article. Text Mining is about exploring large textual data and find patterns. It includes finding frequent words, the length of the sentence, and the presence/absence of specific words.

NLP is a combination of NLU(Natural Language Understanding) and NLG (Natural Langauge Generation). NLU is used to understand the meaning of a given input text. It understands based on the grammar and context of the text. NLU focuses on sentiment, semantics, context, and intent. For example, the questions “what’s the weather like outside?” and “how’s the weather?” are both asking the same thing. NLG generates text based on structured data. It takes data from a search result and returns it into understandable language.

3. What is sentiment analysis?

The business organization wants to understand the opinion of customers and the public. For example, what went wrong with their latest products? what users

and the general public think about the latest feature? Quantifying the user's content, idea, belief, and opinion are known as sentiment analysis. It is not only limited to marketing, but it can also be utilized in politics, research, and security. The sentiment is more than words, it is a combination of words, tone, and writing style.

#### 4. What is tokenization?

Tokenization is the process of splitting text into small pieces, called tokens such as words, or sentences. It ignores characters like punctuation marks (., “ ‘) and spaces. Word tokenization is breaking up the text into individual words and Sentence tokenization is breaking up the text into individual sentences.

#### 5. What is n-gram model?

The Bag-of-words model(BoW ) is the simplest way of extracting features from the text. BoW converts text into the matrix of the occurrence/frequency of words within a document. This model concerns whether given words occurred or not in the document. It can be a combination of two or more words, which is called the bigram or trigram model and the general approach is called the n-gram model. n-gram creates a matrix is known as the Document-Term Matrix(DTM) or Term-Document Matrix(TDM).

|        | Tweet 1 | Tweet 2 | Tweet 3 | ... | Tweet N |
|--------|---------|---------|---------|-----|---------|
| Term 1 | 0       | 0       | 0       | 0   | 0       |
| Term 2 | 1       | 1       | 0       | 0   | 0       |
| Term 3 | 1       | 0       | 0       | 0   | 0       |
| ...    | 0       | 0       | 3       | 1   | 1       |
| Term M | 0       | 0       | 0       | 1   | 0       |

Term Document Matrix (TDM)

|         | Term 1 | Term 2 | Term 3 | ... | Term M |
|---------|--------|--------|--------|-----|--------|
| Tweet 1 | 0      | 1      | 1      | 0   | 0      |
| Tweet 2 | 0      | 1      | 0      | 0   | 0      |
| Tweet 3 | 0      | 0      | 0      | 3   | 0      |
| ...     | 0      | 0      | 0      | 1   | 1      |
| Tweet N | 0      | 0      | 0      | 1   | 0      |

Document Term Matrix (DTM)

#### 6. What is TF-IDF?

TF-IDF(Term Frequency-Inverse Document Frequency) normalizes the document term matrix. In Term Frequency(TF), you just count the number of words that occurred in each document. IDF(Inverse Document Frequency) measures the amount of information a given word provides across the document. IDF is the logarithmically scaled inverse ratio of the number of documents that contain the word and the total number of documents. TF-IDF is the multiplication of TF and IDF.

$$\text{idf}(W) = \log \frac{\#\text{(documents)}}{\#\text{(documents containing word } W)}$$

#### 7. What is stemming and Lemmatization?

Stemming involves simply lopping off easily-identified prefixes and suffixes to produce what's often the simplest version of a word. *Connection*, for example, would have the *-ion* suffix removed and be correctly reduced to *connect*.

Lemmatization is a way of dealing with the fact that while words like *connect*, *connection*, *connecting*, *connected*, etc. aren't exactly the same, they all have the same essential meaning: *connect*.

Lemmatization looks at words and their roots (called lemma) as described in the dictionary. It is more precise than stemming. Stemming reduces word-forms to (pseudo)stems, whereas lemmatization reduces the word-forms to linguistically valid lemmas. The word "better" has "good" as its lemma. This link is missed by stemming, as it requires a dictionary look-up.

## 8. What is PoS Tagging?

A word's part of speech defines its function within a sentence. A noun, for example, identifies an object. An adjective describes an object. A verb describes the action. Identifying and tagging each word's part of speech in the context of a sentence is called Part-of-Speech Tagging, or POS Tagging.

## 9. What is named entity recognition?

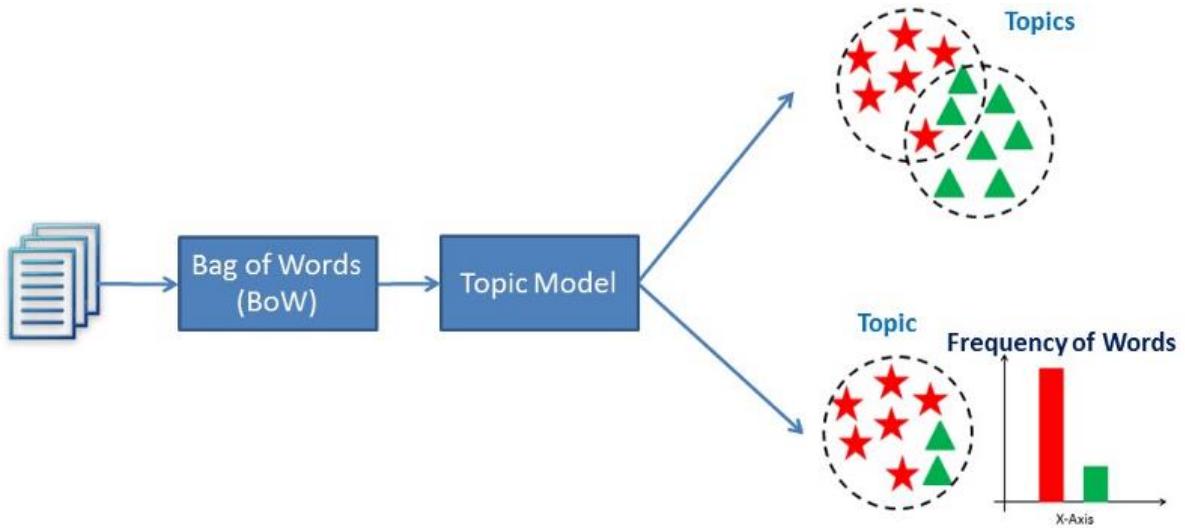
Named Entity recognition, also called entity detection, is a more advanced form of language processing that identifies important elements like places, people, organizations, and languages within an input string of text. This is really helpful for quickly extracting information from text since you can quickly pick out important topics or identify key sections of text.

New York City GPE on Tuesday DATE declared a public health emergency and ordered mandatory measles vaccinations amid an outbreak, becoming the latest national flash point over refusals to inoculate against dangerous diseases. At least 285 CARDINAL people have contracted measles in the city since September DATE, mostly in Brooklyn GPE 's Williamsburg GPE neighborhood. The order covers four CARDINAL Zip codes there, Mayor Bill de Blasio PERSON (D) said Tuesday DATE . The mandate orders all unvaccinated people in the area, including a concentration of Orthodox Jews NORP , to receive inoculations, including for children as young as 6 months old DATE . Anyone who resists could be fined up to \$1,000 MONEY .

## 10. What is topic modeling?

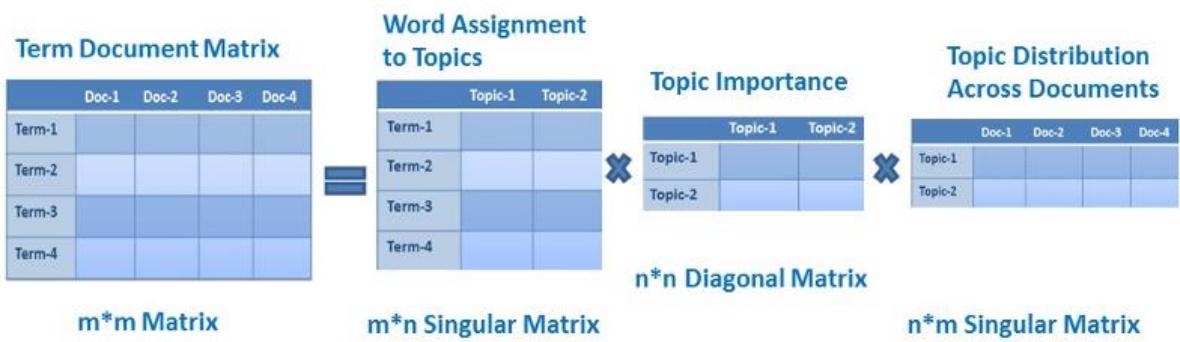
Topic modeling is a text mining technique that provides methods to identify co-occurring keywords to summarize large collections of textual information. It helps in discovering hidden topics in the document, annotate the documents with these topics, and organize a large amount of unstructured data.

There is a possibility that a single document can associate with multiple themes. for example, group words such as ‘patient’, ‘doctor’, ‘disease’, ‘cancer’, ad ‘health’ will represent the topic ‘healthcare’. Topic Modelling is a different game compared to rule-based text searching that uses regular expressions.



## 11. How Latent Semantic Indexing works?

LSA (Latent Semantic Analysis) also known as LSI (Latent Semantic Index) LSA uses a bag of word(BoW) model, which results in the term-document matrix (occurrence of terms in a document). rows represent terms and columns represent documents.LSA learns latent topics by performing a matrix decomposition on the document-term matrix using Singular value decomposition. LSA is typically used as a dimension reduction or noise-reducing technique. LSA uses SVD(Singular Value Decomposition) for factorizing the Term document matrix(TDM) into 3 matrices singular, diagonal, and singular matrix.



## 12. How LDA works?

Latent Dirichlet Allocation(LDA) is an unsupervised algorithm for topic modeling. It helps users to detect the relationship between words and discover groups in those words. The core idea behind the topic modeling is that each document is represented by the distribution of topics and the topic is represented

by the distribution of words. In other words, we can say first we connect words to the topic, and then topics will be connected by each document.

It is a kind of “generative probabilistic model”. LDA consists of two tables or matrices. The first table describes the probability of selecting a particular part when sampling a particular topic. The second table describes the chance of selecting a particular topic when sampling a particular document or composite.

### 13. Explain the masked language model?

Masked language modeling is an example of autoencoding language modeling. It is a kind of fill-in-the-blank task. It uses context words surrounded by mask tokens and predicts what word should be in the place of the mask. Masked language modeling is useful when trying to learn deep representations. BERT (Bidirectional Encoder Representations from Transformers) pre-trained model used the Masked Language model for model training.

### 14. what is perplexity?

In **natural language processing**, **perplexity** is a way of evaluating language models. A language model is a probability distribution over entire sentences or texts [[Wikipedia](#)]. Perplexity is a way to express a degree of confusion a model has in predicting. Perplexity is the exponentiation of the entropy. Low perplexity is good and high perplexity is bad since the perplexity is the exponentiation of the entropy.

### 15. How do you preprocess text in NLP?

Text Preprocessing involves the cleaning, normalizing, and noise removal from the given input text. text preprocessing involves the following steps:

- Lowercase all texts.
- Word and Sentence Tokenization
- Removing stopwords
- Handling contractions
- Removing punctuations, numbers, special characters, and extra whitespaces.
- Stemming and Lemmatization

### 16. What is Word Embeddings? How it is generated?

Word embedding converts text data (e.g words, sentences, and paragraphs) into some kind of vector representation. A word having the same meaning has a similar representation. Word embedding uses an embedding layer to learn vector representation of the given text using Backpropagation. Some examples of word embedding are Word2Vec, Sent2Vec, and Doc2Vec.

There are two models for learning word2vec models:

- The continuous Bag-of-Words (CBOW) model learns the embedding by predicting the current word based on its context.

- The continuous skip-gram model learns by predicting the surrounding words given a current word.

Word embeddings use very shallow Language Models and do not consider the context of the word into account.

## 17. How LSTM works?

LSTM stands for Long Short term memory. It is an extension in RNN(Recurrent Neural Networks). It mitigates the problem of vanishing gradient descent problem of RNN. In RNN, the neural network weights become very smaller or close to zero as we move backward in the network. Due to this problem neurons in the earlier layers learn very slow. This is also the reason that we don't use the sigmoid and Tanh activation function. Nowadays, mostly we are using ReLu(Rectified Linear Unit) in hidden layers and sigmoid at the output layer.

LSTM uses gates to overcome this problem. It uses three gates: Input, Output, and Forget gate. Here, the gates control whether to store or delete (remember or forget) the information.

- The **input gate** is responsible for adding useful information to the cell state.
- The **Forget gate** is responsible for removing information from the cell state.
- The **output gate** is responsible for selecting useful information from the current cell state.

## 18. How do you extract features in NLP?

In-Text Classification Problem, we have a set of texts and their respective labels. but we directly can't use text for our model. you need to convert this text into some numbers or vectors of numbers. There are various ways to convert your text into some vector:

- BoW(Bag of words) or N-gram model
- TF-IDF
- Word Embeddings: Word2Vec, Sent2Vec, Doc2Vecs
- Advanced word embeddings such as Elmo, ULMFit, USE, GPT, BERT
- Transformers

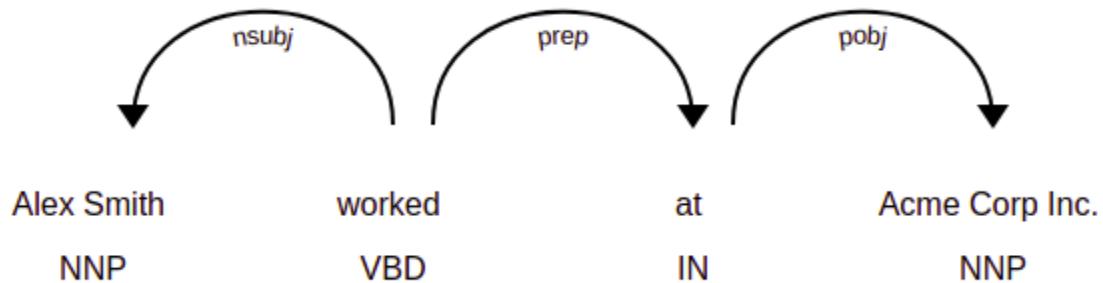
## 19. What are some real-life NLP applications?

NLP offers the following real-life applications:

- Google Translation
- Spell and grammar correction
- Chatbots
- Summarizing news article

## 20. What is parsing?

Parsing helps us to understand the structure of text data. It has capability to analyze any sentence using the parse tree and verify the grammar of a sentence. SpaCy library implemented the dependency parsing.



## 21. How do you find similarities between the two sentences?

We can find a similarity between two texts using Cosine and Jaccard similarity.  
Table 1. The three similarity coefficients

| Similarity Coefficient<br>(X,Y) | Actual Formula                                 |
|---------------------------------|------------------------------------------------|
| Dice Coefficient                | $2 \frac{ X \cap Y }{ X  +  Y }$               |
| Cosine Coefficient              | $\frac{ X \cap Y }{ X ^{1/2} \cdot  Y ^{1/2}}$ |
| Jaccard Coefficient             | $\frac{ X \cap Y }{ X  +  Y  -  X \cap Y }$    |

## 22. What are regular expressions?

Regular Expressions or Regex are searching patterns that used to search in textual data. It is helpful in extracting email addresses, hashtags, and phone numbers.

## 23. What is the major difference between CRF (Conditional Random Field) and HMM (Hidden Markov Model)?

- CRF is Discriminative whereas HMM is a Generative model.
- Conditional Random Fields is a sequential extension of the Maximum Entropy Model.
- CRF models are more powerful than HMMs due to their application of feature functions.
- Both HMMs and linear CRFs are typically trained with Maximum Likelihood techniques with gradient descent, Quasi-Newton methods, or for HMMs with Expectation-Maximization techniques. If the optimization problems are convex, these methods all yield the optimal parameter set.

#### 24. What is BiLSTM?

BiLSTM is a Bidirectional LSTM that processes signals in both forward and backward directions. It is a sequential processing model that consists of two LSTM. This bidirectional nature helps us to increase the more contextual information for the algorithm. It shows quite good results when it understands the context better.

#### 25. What is Syntactic and Semantic Analysis?

Syntactic analysis is analyzing the sentence and understand the grammar rules and order of sentence. It includes the following tasks parsing, word segmentation, morphological segmentation, stemming, and lemmatization.

Semantic analysis analyses the meaning and interpretation of a text. It includes the following tasks such as Named entity recognition, word sense ambiguation, and Natural language generation.

*Q1. Explain in details about Natural Processing language (NLP), which is currently one of the key artificial language learning processes that have been started in the industry?*

#### **Answer:**

Natural Language Processing (NLP) is designed for understanding and analyzing the natural languages automatic way, and export data or possibly require information from those available data. NLP has defined an algorithm that helps mainly with machine learning. This kind of **machine learning algorithm** actually helps for understanding analyzing some of the natural languages.

*Q2. There have some various common elements of natural language processing. Those elements are essential for understanding NLP properly. Can you please explain the same in detail with an example?*

**Answer:**

There have a lot of components normally using natural language processing (NLP). Some of the major components are explained below:

- **Extraction of Entity:** It identifies and extracts some critical data from the available information, which helps to segment of provided sentence on identifying each entity. It can help identify one human that it's fictional or real, the same kind of reality identification for any organization, events, or any geographic location, etc.
- **The analysis in a syntactic way:** it mainly helps for maintaining ordering properly of the available words.
- **An analysis in a programmatic way:** It is one of the key processes of NLP. It helps for extracting data from the specifically available text in natural languages.

*Q3. Explain details about varieties areas available in processing natural languages smartly, whether we know impacted areas are very small as this processing started very recently?*

**Answer:**

Natural language processing (NLP) can have an implementation in various areas of the current industry environment. Some of the key areas are explaining below:

- An analysis was done semantically.
- Summarize natural language information automatically.
- Classification of varieties text is written in natural language.
- The ready answer to some common questions

We can give some key examples of real-life where natural language processing (NLP) used broadly. Examples are Google Assistance, IOS Siri, or Amazon echo.

*Q4. In the case of processing natural language, we normally mentioned one common terminology, NLP and binding every language with the same terminology properly. Please explain in details about this NLP terminology with an example?*

**Answer:**

These are the basic NLP Interview Questions asked in an interview. There have several factors available in case of explaining natural language processing.

Some of the key factors are given below:

- **Vectors and Weights:** Google Word vectors, length of TF-IDF, varieties documents, word vectors, TF-IDF.
- **Structure of Text:** Named Entities, tagging of part of speech, identifying the sentence's head.
- **Analysis of sentiment:** Know about the sentiment features, entities available for the sentiment, sentiment common dictionary.

- **Classification of Text:** Learning supervising, set off a train, a set of validation in Dev, Set of define test, a feature of the individual text, LDA.
- **Reading of Machine Language:** Extraction of the possible entity, linking with an individual entity, DBpedia, some libraries like Pikes or FRED.

*Q5. One another very common terminology used in the case of natural learning processing is called TF-IDF. Please explain in details on the understanding of TFIDF properly and come with some example?*

**Answer:**

TF-IDF or tf-IDF is basically stood for some critical frequency of term or some inverse frequency of a specific document. TF-IDF is basically using for identifying some of the keywords from an entire document written in natural language. It mainly involves retrieving information from the critical document by using some statistical numeric data for identifying some of the keywords and mentioning how much important that word is specifically in the collection of multiple documents or in the set of collections.

*Q6. There are several tagging using for processing natural languages. In all those tagging parts of speech (POS), tagging is one of our industry's popular ones. Please explain in detail about the part of speech (POS) tagging and how it can be used properly?*

**Answer:**

Part of speech tagger is a very interesting and most important tool for properly processing natural language. This part of speech (POS) tagger is a normal tool or software which helps for reading some critical text independent of any

languages, then assign entire sentence in part of speech for each word or some other tokenization logic define in the software, such as adjective, verb or noun etc.

It normally holds some specific algorithm that helps to label some of the terms in the entire text body. It has some varieties of categories which are more complex than define above utility. The above define functionality is one of the very basic features of the POS tag.

*Q7. As the analysis is one of the critical requirements of natural language processing (NLP), we can follow several analysis approaches for understanding NLP properly. In between all those, one of the key analysis called Pragmatic Analysis. Please explain about Pragmatic analysis in details?*

**Answer:**

Pragmatic analysis is one of the critical analysis defines in NLP. It is mainly handling some knowledge which is belonging to the outside world. That means some of the knowledge which always be external for some define documents or already queries. This kind of analysis mainly concentrates on the critical interpretation of some specific word and tries to understand its actual meaning. For doing this kind of analysis, real-world knowledge is very much required.

*Q8. Again, as NLP is used for multiple language processing smartly and interacting with a computer system based on proper language understanding, one of the key parsings normally used by NLP is dependency parsing. Please explain dependency parsing in details with proper explanation?*

**Answer:**

Dependency parsing is actually known in the industry as syntactic parsing. It is

doing one of NLP processing's critical tasks, identifying or recognizing some of the sentences and then assigning those in some define a syntactic structure for understanding properly. One of the popular syntactic structures is parsed tree define with some parsing algorithm.

*Q9. One of the very basic requirement of NLP is keyword normalization. There have normally two process or techniques followed by NLP for handling proper keyword normalization. Please explain in details about keyword normalization and which techniques can be followed for the same.*

**Answer:**

This is the most asked NLP Interview Question in an interview. There has two key normalization processes in NLP, which help for keyword normalization. These two processes are Stemming and Lemmatization.

*Q10. There have some classification model define in NLP. What kind of features can be followed by NLP for improving accuracy in the classification model?*

**Answer:**

There have several classifications followed by NLP, explaining the same below:

- Counting frequency of defined terms.
- Notation of vector for every sentence.
- Part of Speech (POS) tagging.
- Grammatical dependency or some defined dictionary or library.

*1. What are some most common areas of usage of Natural Language Processing?*

- Semantic Analysis
- Text classification

- Automatic summarization
- Questioning Answering

*2. What are some of the major components of Natural Language Processing?*

- Entity extraction: It is used for segmenting a sentence to identify and extract entities, such as a person (real or fictional), organization, geographies, events, etc. 85
- Pragmatic analysis: Pragmatic analysis extracts information from the input text. It is part of the process of data extraction.
- Syntactic analysis: Syntactic analysis is used for the proper ordering of words.

*3. What do you understand by NLTK in Natural Language Processing?*

- In Natural Language Processing, NLTK stands for Natural Language Toolkit. It is a Python library used to process data in human spoken languages.
- NLTK facilitates developers to apply parsing, tokenization, lemmatization, stemming techniques, and more to understand natural languages. It is also used for categorizing text, parsing linguistic structure, analyzing documents, etc.

*4. What are the most used Natural Language Processing Terminologies?*

- Preprocessing: This is a method used to remove unwanted text or noise from the given text and make it “clean.” It is the first step of any NLP task.
- Documents: Documents are the body of text and are collectively used to form a corpus.
- Corpus, or Corpora (Plural): It is a collection of text of similar type, for example, movie reviews, social media posts, etc.
- Vocabulary: It is a group of terms used in a text or speech.
- Out of Vocabulary: It specifies the terms not included in the vocabulary. We put the terms created during the model’s training in this category.
- Tokenization: It is used in NLP to break down large sets of text into small parts for easy readability and understanding. Here, the small parts are referred to as ‘text’ and provide a piece of meaningful information.
- N-grams: It specifies the continuous sequence (similar to the power set in number theory) of n-tokens of a given text.
- Parts of Speech (POS): It specifies the word’s functions, such as a noun, verb, etc.
- Parts of Speech Tagging: It is the process of tagging words in the sentences into different parts of speech.

*5. What is the difference between formal and natural languages?*

- The main difference between a formal language and a natural language is that a formal language is a collection of strings.

- Each string contains symbols from a finite set called alphabets. On the other hand, a natural language is a language that humans use to speak. This is completely different from a formal language as it contains fragments of words and pause words like uh, um, et

6. *What is the use of TF-IDF?*

- In Natural language Processing, tf-idf, TF-IDF, or TFIDF stands for Term Frequency-Inverse Document Frequency.
- It is a numerical statistic used to specify how important a word is to a document in a collection or the collection of a set.

7. *What is the full form of NLP?*

- NLP stands for “Natural Language Processing”.
- NLP is a field of computer science that deals with communication between computer systems and humans.
- This technique uses Artificial Intelligence and Machine Learning to create automated software that helps understand the human spoken languages and extract useful information from the data gathered from the audio.

8. *What are the tools used for training NLP models?*

- The most common tools used for training NLP models are NLTK, spaCY, PyTorch-NLP, openNLP, etc.

9. *What is Bag of Words in Natural Language Processing?*

- Bag of Words is a commonly used model in Natural Language Processing that depends on word frequencies or occurrences to train a classifier. This model creates an occurrence matrix for documents or sentences without depending on their grammatical structure or word order.

10. *What do you understand by Dependency Parsing in Natural Language Processing?*

- In Natural Language Processing, Dependency Parsing is a process of assigning syntactic structure to a sentence and identifying its dependency parses.
- This is an important process to understand the correlations between the “head” words in the syntactic structure. That’s why it is also known as syntactic parsing.
- The process of dependency parsing becomes a little complex if there are more sentences that have more than one dependency parses. Multiple parse trees are known as ambiguities.

11. *What do you understand by semantic analysis?*

- Semantic analysis is a process that makes a machine understand the meaning of a text. It uses several algorithms to interpret the words in sentences. It is also used to understand the structure of a sentence.

*12. What are the stop words in Natural Language Processing?*

- In Natural Language Processing, stop words are regarded as useless data for a search engine. It includes the words like articles, prepositions, was, were, is, am, the, a, an, how, why, and many more.
- The algorithm used in Natural Language Processing eliminates the stop words to understand and analyze the meaning of the sentences. Eliminating the stop words is one of the most important tasks for search engines to process data.

*13. What do you understand by information extraction?*

In Natural Language Processing, information extraction is a technique of automatically extracting structured information from unstructured sources to get useful information. It extracts information such as attributes of entities, the relationship between different entities, and more.

*14. What is NES in Natural Language Processing?*

- NES is an acronym that stands for Name Entity Recognition.
- It is used in Natural Language Processing and is most commonly known as NER.
- It is the process of identifying specific entities in a text document that is more informative and have a unique context.

*15. What is pragmatic ambiguity in NLP?*

- Pragmatic ambiguity is used to specify words with more than one meaning, and they can be used in any sentence depending on the context.
- In pragmatic ambiguity, words have multiple interpretations.
- Pragmatic ambiguity occurs when the meaning of the words is not specific.

*16. What are the techniques used for semantic analysis?*

- **Named entity recognition:** This technique is used to specify the process of information retrieval that helps identify the entities like the name of a person, organization, place, time, emotion, etc.
- **Natural language generation:** This technique specifies a process used by the software to convert the structured data into human spoken languages. By using natural language generation, organizations can automate content for custom reports.
- **Word sense disambiguation:** It technique is used to identify the sense of a word used in different sentences.

*17. What are the various models of information extraction?*

- Sentiment Analysis Module
- Tagger Module
- Relation Extraction Module

- Fact Extraction Module
- Entity Extraction Module
- Network Graph Module
- Document Classification and Language Modeling Module

*18. What are the most commonly used models to reduce data dimensionality in NLP?*

- The most commonly used models to reduce the dimensionality of data in NLP are TF-IDF, Word2vec/Glove, LSI, Topic Modelling, Elmo Embeddings, etc.

*19. What is language modeling in NLP?*

- In Natural Language Processing, language modeling creates a probability distribution of a sequence of words.
- It provides probability to all the words present in that sequence.

*20. What is Lemmatization in Natural Language Processing?*

- Lemmatization is a process of doing things properly using a vocabulary and morphological analysis of words.
- It is mainly used to remove the inflectional endings only and return the base or dictionary form of a word, known as the lemma. It is just like cutting down your beard or shaving to get the original shape of your face.

*21. What do you understand by MLM in Natural Language Processing?*

- In Natural Language Processing, MLM is a term that stands for Masked Language Model.
- It helps learners understand deep representations in downstream tasks by taking the output from the corrupt input.

*22. Which NLP techniques use a lexical knowledge base to obtain the correct base form of the words?*

- The NLP techniques that use a lexical knowledge base to obtain the correct base form of the words are Lemmatization and stemming.

*23. What is the difference between Stemming and Lemmatization in NLP?*

| <b>Stemming</b>                                                                                                                                                                    | <b>Lemmatization</b>                                                                          |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| Stemming is the process of extracting the base form of a word by removing the affixes from them.                                                                                   | Lemmatization is a more advanced process and looks beyond word reduction, just like stemming. |
| Stemming is not as efficient as Lemmatization. This method works fairly well in most cases, but unfortunately, English has many exceptions requiring a more sophisticated process. | Lemmatization is more efficient than Stemming as it works well in exceptional words.          |

*24. What is Stemming in Natural Language Processing?*

- Stemming is a process of extracting the base form of a word by removing the affixes from them.
- It is just like cutting down the branches of a tree to its stems.

*25. What is Latent Semantic Indexing?*

- LSI or Latent Semantic Indexing is a mathematical technique used in Natural Language Processing. This technique is used to improve the accuracy of the information retrieval process. The LSI algorithm is designed to allow machines to detect the latent correlation between semantics.

*26. What is tokenization in Natural Language Processing?*

- In Natural Language Processing, tokenization is a method of dividing the text into various tokens.
- These tokens are the form of the words, just like a word forms into a sentence.
- In NLP, the program computers process large amounts of natural language data.

*27. What is the key difference between dependency parsing and shallow parsing?*

- The key difference between dependency parsing and shallow parsing is that dependency parsing is the process of finding a relation between all the different words.

*28. What are the key differences between NLP and NLU?*

| <b>NLP</b>                                                                                | <b>NLU</b>                                                                                                         |
|-------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| NLP is a short form of Natural Language Processing                                        | NLU is a short form of Natural Language Understanding                                                              |
| It includes all the techniques required for the interaction between computers and humans. | It converts the uncategorized input data into a structured format and allows the computers to understand the data. |

*29. What are the best open sources of NLP Tools available in the market?*

- SpaCy
- TextBlob
- Textacy
- Natural language Toolkit (NLTK)
- Retext
- NLP.js
- Stanford NLP
- CogcompNLP

*30. What are some open-source libraries used in NLP?*

- Some popular open-source libraries used in NLP are NLTK (Natural Language ToolKit), SciKit Learn, Textblob, CoreNLP, spaCY, Gensim, etc.

**1) What do you know about NLP?**

**Ans:**

NLP stands for Natural Language Processing. It deals with making a machine understand the way human beings read and write in a language. This task is achieved by designing algorithms that can extract meaning from large datasets in audio or text format by applying machine learning algorithms.

**2) Give examples of any two real-world applications of NLP?**

**Ans:**

**Spelling/Grammar Checking Apps:**

The mobile applications and websites that offer users correct grammar mistakes in the entered text rely on NLP algorithms. These days, they can also recommend the following few words that the user might type, which is also because of specific NLP models being used in the backend.

**ChatBots:**

Many websites now offer customer support through these virtual bots that chat with the user and resolve their problems. It acts as a filter to the issues that do not require an interaction with the companies' customer executives.

**3) What is tokenization in NLP?**

**Ans:**

Tokenization is the process of splitting running text into words and sentences.



#### 4) What is the difference between a formal language and a natural language?

Ans:

| Formal Language                                                                                                      | Natural Language                                                                                                                                                           |
|----------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A formal language is a collection of strings, where each string contains symbols from a finite set called alphabets. | A natural language is a language that humans speak. It is usually a lot different from a formal language. These typically contain fragments of words and like uh, um, etc. |

#### 5) What is the difference between stemming and lemmatization?

Ans:

Both stemming and lemmatization are keyword normalization techniques aiming to minimize the morphological variation in the words they encounter in a sentence. But, they are different from each other in the following way.

| Stemming                                                                                               | Lemmatization                                                                             |
|--------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| This technique involves removing the affixes added to a word and leaving us with the rest of the word. | Lemmatization is the process of converting a word into its lemma from its inflected form. |
| Example: 'Caring' → 'Car'                                                                              | Example: 'Caring' → 'Care'                                                                |

#### 6) What is NLU?

**Ans:**

NLU stands for Natural Language Understanding. It is a subdomain of NLP that concerns making a machine learn the skills of reading comprehension. A few applications of NLU include Machine translation (MT), Newsgathering, and Text categorization. It often goes by the name Natural Language Interpretation (NLI) as well.

**7) List the differences between NLP and NLU?**

**Ans:**

| <b>Natural Language Processing</b>                                                                                                                                                                                          | <b>Natural Language Understanding</b>                                                                                                                                               |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| NLP is a branch of AI that deals with designing programs for machines that will allow them to process the language that humans use. The idea is to make machines imitate the way humans utilize language for communication. | In NLU, the aim is to improve a computer's ability to understand and analyze human language. This aim is achieved by transforming unstructured data into a machine-readable format. |

**8) What do you know about Latent Semantic Indexing (LSI)?**

**Ans:**

LSI is a technique that analyzes a set of documents to find the statistical coexistence of words that appear together. It gives an insight into the topics of those documents.

LSI is also known as Latent Semantic Analysis.

**9) List a few methods for extracting features from a corpus for NLP.**

**Ans:**

- 1. Bag-of-Words
- 2. Word Embedding

**10) What are stop words?**

**Ans:**

Stop words are the words in a document that are considered redundant by NLP engineers and are thus removed from the document before processing it. Few examples are 'is', 'the', 'are', 'am'.

**11) What do you know about Dependency Parsing?**

**Ans:**

Dependency parsing is a technique that highlights the dependencies among the words of a sentence to understand its grammatical structure. It examines how the words of a sentence are linguistically linked to each other. These links are called dependencies.

**12) What is Text Summarization? Name its two types?**

**Ans:**

- Extraction-based Summarization
- Abstraction-based Summarization

**13) What are false positives and false negatives?**

**Ans:**

If a machine learning algorithm falsely predicts a negative outcome as positive, then the result is labeled as a false negative.

And, if a machine learning algorithm falsely predicts a positive outcome as negative, then the result is labeled as a false positive.

**14) List a few methods for part-of-speech tagging?**

**Ans:**

Rule-based tagging, HMM-tagging, transformation-based tagging, and memory-based tagging.

**15) What is a corpus?**

**Ans:**

'Corpus' is a Latin word that means 'body.' Thus, a body of the written or spoken text is called a corpus.

**16) List a few real-world applications of the n-gram model?**

**Ans:**

- Augmentive Communication
- Part-of-speech Tagging

- Natural language generation
- Word Similarity
- Authorship Identification
- Sentiment Extraction
- Predictive Text Input

**17) What does TF\*IDF stand for? Explain its significance?**

**Ans:**

TF\*IDF stands for Term-Frequency/Inverse-Document Frequency. It is an information-retrieval measure that encapsulates the semantic significance of a word in a particular document N, by degrading words that tend to appear in a variety of different documents in some huge background corpus with D documents.

Let  $n_w$  denote the frequency of a word w in the document N, m represents the total number of documents in the corpus that contain w. Then, TF\*IDF is defined as

$$\text{TF*IDF}(w) = n_w \times \log n_m$$

**18) What is perplexity in NLP?**

**Ans:**

It is a metric that is used to test the performance of language models. Mathematically, it is defined as a function of the probability that the language model represents a test sample. For a test sample  $X = x_1, x_2, x_3, \dots, x_n$ , the perplexity is given by,

$$PP(X) = P(x_1, x_2, \dots, x_N)^{-1/N}$$

where N is the total number of word tokens. Higher the perplexity, lesser is the information conveyed by the language model.

**19) Which algorithm in NLP supports bidirectional context?**

**Ans:**

Naive Bayes is a classification machine learning algorithm that utilizes Baye's Theorem for labeling a class to the input set of features. A vital

element of this algorithm is that it assumes that all the feature values are independent.

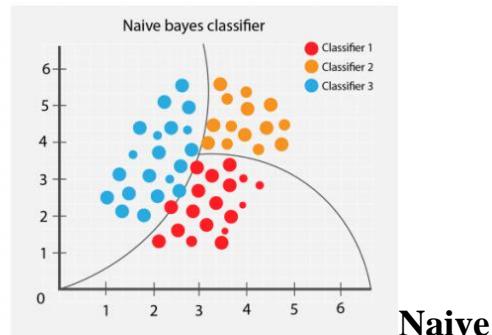
# Naive Bayes

In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$



## Bayes Algorithm

### 20) What is Part-of-Speech tagging?

Ans:

Part-of-speech tagging is the task of assigning a part-of-speech label to each word in a sentence. A variety of part-of-speech algorithms are available that contain tagsets having several tags between 40 and 200.

### 21) What is the bigram model in NLP?

Ans:

A bigram model is a model used in NLP for predicting the probability of a word in a sentence using the conditional probability of the previous word. For calculating the conditional probability of the previous word, it is crucial that all the previous words are known.

### 22) What is the significance of the Naive Bayes algorithm in NLP?

Ans:

The Naive Bayes algorithm is widely used in NLP for various applications. For example: to determine the sense of a word, to predict the tag of a given text, etc.

### 23) What do you know about the Masked Language Model?

Ans:

The Masked Language Model is a model that takes a sentence with a few hidden (masked) words as input and tries to complete the sentence by correctly guessing those hidden words.

**24) What is the Bag-of-words model in NLP?**

**Ans:**

Bag-of-words refers to an unorganized set of words. The Bag-of-words model in NLP is a model that assigns a vector to a sentence in a corpus. It first creates a dictionary of words and then produces a vector by assigning a binary variable to each word of the sentence depending on whether it exists in the bag of words or not.

**25) Briefly describe the N-gram model in NLP?**

**Ans:**

N-gram model is a model in NLP that predicts the probability of a word in a given sentence using the conditional probability of n-1 previous words in the sentence. The basic intuition behind this algorithm is that instead of using all the previous words to predict the next word, we use only a few previous words.

**26) What do you understand by word embedding?**

**Ans:**

In NLP, word embedding is the process of representing textual data through a real-numbered vector. This method allows words having similar meanings to have a similar representation.

**27) What is an embedding matrix?**

**Ans:**

A word embedding matrix is a matrix that contains embedding vectors of all the words in a given text.

**28) List a few popular methods used for word embedding.**

**Ans:**

Following are a few methods of word embedding.

- Embedding Layer
- Word2Vec
- Glove

**29) How will you use Python's concordance command in NLTK for a text that does not belong to the package?**

**Ans:**

The concordance() function can easily be accessed for a text that belongs to the NLTK package using the following code:

1. `>>>from nltk.book import *`
2. `>>>text1.concordance("monstrous")`

However, for a text that does not belong to the NLTK package, one has to use the following code to access that function.

1. `>>>import nltk.corpus`
2. `>>>from nltk.text import Text`
3. `>>>NLTKtext` =  
`Text(nltk.corpus.gutenberg.words('Your_file_name_here.txt'))`
4. `>>>NLTKtext.concordance('word')`

Here, we have created a Text object to access the concordance() function. The function displays the occurrence of the chosen word and the context around it.

**30) Write the code to count the number of distinct tokens in a text?**

**Ans:**

`len(set(text))`

**31) What are the first few steps that you will take before applying an NLP machine-learning algorithm to a given corpus?**

**Ans:**

- Removing white spaces
- 
- Removing Punctuations
- Converting Uppercase to Lowercase
- Tokenization

- Removing Stopwords
- Lemmatization

**32) For correcting spelling errors in a corpus, which one is a better choice: a giant dictionary or a smaller dictionary, and why?**

**Ans:**

Initially, a smaller dictionary is a better choice because most NLP researchers feared that a giant dictionary would contain rare words that may be similar to misspelled words. However, later it was found (Damerau and Mays (1989)) that in practice, a more extensive dictionary is better at marking rare words as errors.

**33) Do you always recommend removing punctuation marks from the corpus you're dealing with? Why/Why not?**

**Ans:**

No, it is not always a good idea to remove punctuation marks from the corpus as they are necessary for certain NLP applications that require the marks to be counted along with words.

For example: Part-of-speech tagging, parsing, speech synthesis.

**34) List a few libraries that you use for NLP in Python?**

**Ans:**

NLTK, Scikit-learn, GenSim, SpaCy, CoreNLP, TextBlob.

**35) Suggest a few machine learning/deep learning models that are used in NLP?**

**Ans:**

Support Vector Machines, Neural Networks, Decision Tree, Bayesian Networks.

**36) Which library contains the Word2Vec model in Python?**

**Ans:**

GenSim

**37) What are homographs, homophones, and homonyms?**

**Ans:**

| <b>Homographs</b>                                                                                | <b>Homophones</b>                                                    | <b>Homonyms</b>                                                                           |
|--------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| “Home”=same<br>“graph”=write                                                                     | “Home”=same<br>“phone”=sound                                         | “Homo”=same,<br>“onym” = name                                                             |
| These are the words that have the same spelling but may or may not have the same pronunciations. | These are the words that sound similar but have different spellings. | These are the words that have the same spelling and pronunciation but different meanings. |
| To live a life, airing a show live                                                               | Eye, I                                                               | River Bank, Bank Account                                                                  |

**38) Is converting all text in uppercase to lowercase always a good idea?  
Explain with the help of an example?**

**Ans:**

No, for words like The, the, THE, it is a good idea as they all will have the same meaning. However, for a word like brown which can be used as a surname for someone by the name Robert Brown, it won't be a good idea as the word ‘brown’ has different meanings for both the cases. We, therefore, would want to treat them differently. Hence, it is better to change uppercase letters at the beginning of a sentence to lowercase, convert headings and titles to which are all in capitals to lowercase, and leave the remaining text unchanged.

**39) What is a hapax/hapax legomenon?**

**Ans:**

The rare words that only occur once in a sample text or corpus are called hapaxes. Each one of them is called an hapax or hapax legomenon (greek for ‘read-only once’). It is also called a singleton.

**40) Is tokenizing a sentence based on white-space ‘ ‘ character sufficient? If not, give an example where it may not work.**

**Ans:**

Tokenizing a sentence using the white space character is not always sufficient.

Consider the example,

“ One of our users said, ‘I love Dezyre’s content’.”

Tokenizing purely based on white space would result in the following words:

‘I said, content’.

#### **41) What is a collocation?**

Ans:

A collocation is a group of two or more words that possess a relationship and provide a classic alternative of saying something. For example, ‘strong breeze’, ‘the rich and powerful’, ‘weapons of mass destruction’.

#### **42) List a few types of linguistic ambiguities.**

Ans:

**Lexical Ambiguity:** This type of ambiguity is observed because of homonyms and polysemy in a sentence.

**Syntactic Ambiguity:** A syntactic ambiguity is observed when based on the sentence’s syntax, more than one meaning is possible.

**Semantic Ambiguity:** This ambiguity occurs when a sentence contains ambiguous words or phrases that have ambiguous

#### **43) Differentiate between orthographic rules and morphological rules with respect to singular and plural forms of English words.**

Ans:

##### **Orthographical Rules**

##### **Morphological Rules**

|                                                                                                                                                                                                                                       |                                                                                                                                                              |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| These are the rules that contain information for extracting the plural form of English words that end in ‘y’. Such words are transformed into their plural form by converting ‘y’ into ‘ies’ and adding the letters ‘es’ as suffixes. | These rules contain information for words like fish; there are null plural forms. And words like goose have their plural generated by a change of the vowel. |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|

#### **44) Calculate the Levenshtein distance between two sequences ‘intention’ and ‘execution’.**

Ans:



### Intention to Execution

The image above can be used to understand the number of editing steps it will take for the word intention to transform into execution.

- The first step is deletion (d) of 'I.'
- The next step is to substitute (s) the letter 'N' with 'E.'
- Replace the letter 'T' with 'X.'
- The letter E remains unchanged, and the letter 'C' is inserted (i).
- Substitute 'U' for the letter 'N.'

Thus, it will take five editing steps for transformations, and the Levenshtein distance is five.

### 45) What are the full listing hypothesis and minimum redundancy hypothesis?

Ans:

**Full Listing Hypothesis:** This hypothesis suggests that all humans perceive all the words in their memory without any internal morphological structure. So, words like tire, tiring, tired are all stored separately in the mental lexicon.

**Minimum Redundancy Hypothesis:** This hypothesis proposes that only the raw form of the words (morphemes) form the part of the mental lexicon. When humans process a word like tired, they recall both the morphemes (tire-d).

### 46) What is sequence learning?

**Ans:**

Sequence learning is a method of learning where both input and output are sequences.

**47) What is NLTK?**

**Ans:**

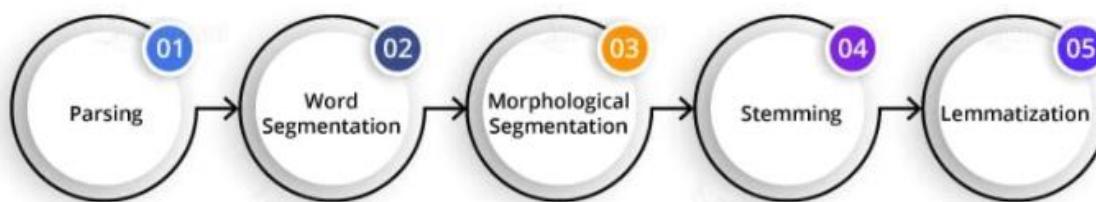
NLTK is a Python library, which stands for Natural Language Toolkit. We use NLTK to process data in human spoken languages. NLTK allows us to apply techniques such as parsing, tokenization, lemmatization, stemming, and more to understand natural languages. It helps in categorizing text, parsing linguistic structure, analyzing documents, etc.

**48) What is Syntactic Analysis?**

**Ans:**

**Syntactic analysis** is a technique of analyzing sentences to extract meaning from it. Using syntactic analysis, a machine can analyze and understand the order of words arranged in a sentence. NLP employs grammar rules of a language that helps in the syntactic analysis of the combination and order of words in documents.

### Syntactic Analysis

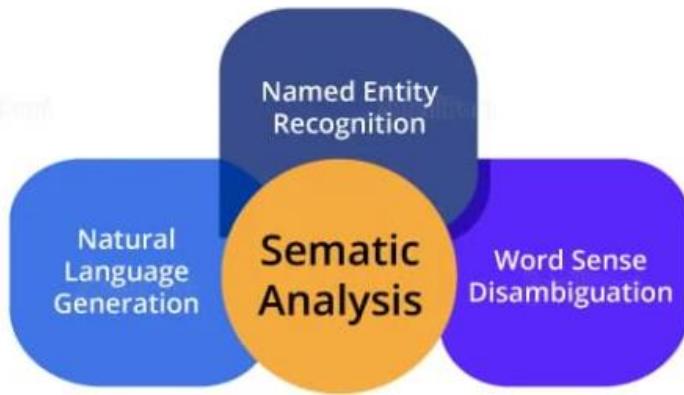


**Syntactic analysis**

**49) What is Semantic Analysis?**

**Ans:**

Semantic analysis helps make a machine understand the meaning of a text. It uses various algorithms for the interpretation of words in sentences. It also helps understand the structure of a sentence.



## 50) List the components of Natural Language Processing?

**Ans:**

The major components of NLP are as follows:



## Glossary in NLP

## 51) What is Latent Semantic Indexing (LSI)?

**Ans:**

Latent semantic indexing is a mathematical technique used to improve the accuracy of the information retrieval process. The design of LSI algorithms allows machines to detect the hidden (latent) correlation between semantics (words). To enhance information understanding, machines generate various concepts that associate with the words of a sentence.

The technique used for information understanding is called singular value decomposition. It is generally used to handle static and unstructured data. The matrix obtained for singular value decomposition contains rows for words and columns for documents. This method best suits to identify components and group them according to their types.

The main principle behind LSI is that words carry a similar meaning when used in a similar context. Computational LSI models are slow in

comparison to other models. However, they are good at contextual awareness that helps improve the analysis and understanding of a text or a document.

## 52) What is Regular Grammar?

**Ans:**

Regular grammar is used to represent a regular language:

A regular grammar comprises rules in the form of  $A \rightarrow a$ ,  $A \rightarrow aB$ , and many more. The rules help detect and analyze strings by automated computation.

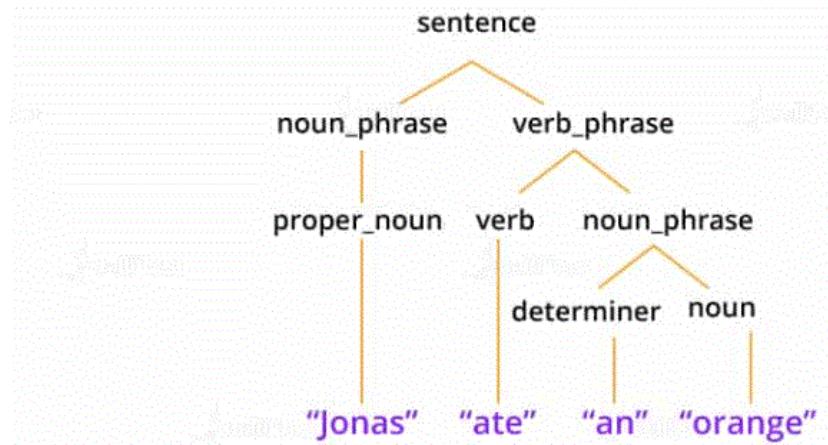
Regular grammar consists of four tuples:

- ‘N’ is used to represent the non-terminal set.
- ‘ $\Sigma$ ’ represents the set of terminals.
- ‘P’ stands for the set of productions.
- ‘ $S \in N$ ’ denotes the start of non-terminal.

## 53) What is Parsing in the context of NLP?

**Ans:**

Parsing in NLP refers to the understanding of a sentence and its grammatical structure by a machine. Parsing allows the machine to understand the meaning of a word in a sentence and the grouping of words, phrases, nouns, subjects, and objects in a sentence. Parsing helps analyze the text or the document to extract useful insights from it. To understand parsing, refer to the below diagram:



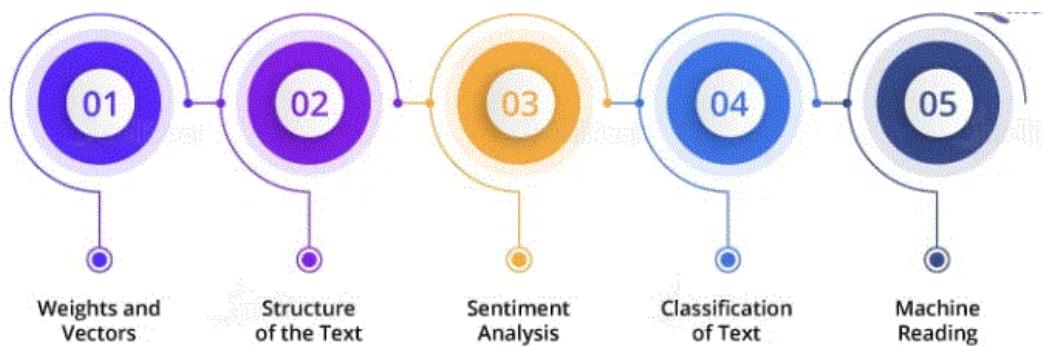
**Parsing context in the NLP**

#### 54) Define the terminology in NLP?

**Ans:**

This is one of the most often asked NLP interview questions.

The interpretation of Natural Language Processing depends on various factors, and they are:

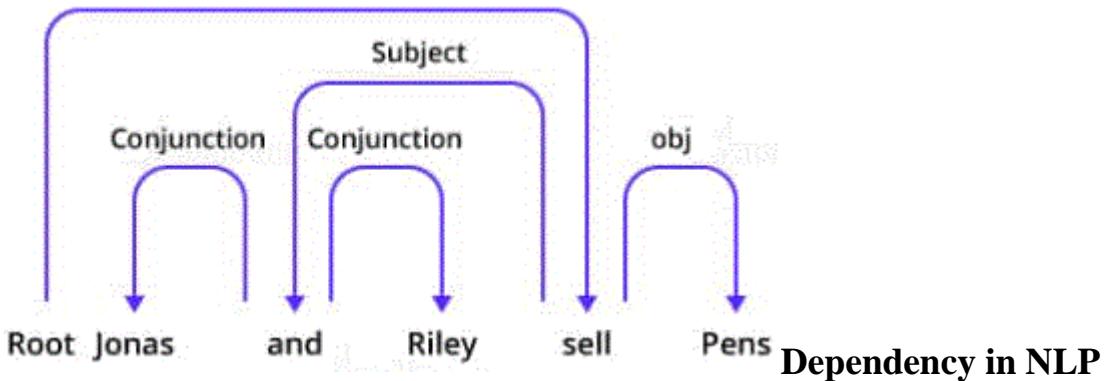


**Terminology in NLP**

#### 55) Explain Dependency Parsing in NLP?

**Ans:**

Dependency parsing helps assign a syntactic structure to a sentence. Therefore, it is also called syntactic parsing. Dependency parsing is one of the critical tasks in NLP. It allows the analysis of a sentence using parsing algorithms. Also, by using the parse tree in dependency parsing, we can check the grammar and analyze the semantic structure of a sentence.



## 56) What is the difference between NLP and NLU?

**Ans:**

The below table shows the difference between NLP and NLU:

| Natural Language Processing (NLP)                                                                                                                                                                                                                       | Natural Language Understanding (NLU)                                                                                                                                                                                                                    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Used to create systems capable of establishing communication between humans and computers</p> <p>Takes care of all the techniques required for the interaction between machines and humans</p> <p>Provides techniques to analyze 'What is said?'</p> | <p>Provides techniques to solve complicated problems related to machine understanding</p> <p>Helps convert the unorganized input data into a structured format to allow the machine to understand the data</p> <p>Helps understand 'What is meant?'</p> |

## NLP Vs NLU

## 57) What is Pragmatic Analysis?

**Ans:**

Pragmatic analysis is an important task in NLP for interpreting knowledge that is lying outside a given document. The aim of implementing pragmatic analysis is to focus on exploring a different aspect of the document or text in a language. This requires a comprehensive knowledge of the real world. The pragmatic analysis allows software applications for the critical interpretation of the real-world data to know the actual meaning of sentences and words.

### Example:

Consider this sentence: 'Do you know what time it is?'

This sentence can either be asked for knowing the time or for yelling at someone to make them note the time. This depends on the context in which we use the sentence.

## 58) What are unigrams, bigrams, trigrams, and n-grams in NLP?

**Ans:**

When we parse a sentence one word at a time, then it is called a unigram. The sentence parsed two words at a time is a bigram.

When the sentence is parsed three words at a time, then it is a trigram. Similarly, n-gram refers to the parsing of n words at a time.

Example: To understand unigrams, bigrams, and trigrams, you can refer to the below diagram:



## Unigrams, Bigrams, Trigrams, and n-grams in NLP

Therefore, parsing allows machines to understand the individual meaning of a word in a sentence. Also, this type of parsing helps predict the next word and correct spelling errors.

## 59) What are the steps involved in solving an NLP problem?

**Ans:**

Below are the steps involved in solving an NLP problem:

- Gather the text from the available dataset or by web scraping
- Apply stemming and lemmatization for text cleaning
- Apply feature engineering techniques
- Embed using word2vec
- Train the built model using neural networks or other Machine Learning techniques

- Evaluate the model's performance
- Make appropriate changes in the model
- Deploy the model

## 60) What is Feature Extraction in NLP?

### Ans:

Features or characteristics of a word help in text or document analysis. They also help in sentiment analysis of a text. Feature extraction is one of the techniques that are used by recommendation systems. Reviews such as ‘excellent,’ ‘good,’ or ‘great’ for a movie are positive reviews, recognized by a recommender system. The recommender system also tries to identify the features of the text that help in describing the context of a word or a sentence. Then, it makes a group or category of the words that have some common characteristics. Now, whenever a new word arrives, the system categorizes it as per the labels of such groups.

## 61) What is precision and recall?

### Ans:

The metrics used to test an NLP model are precision, recall, and F1. Also, we use accuracy for evaluating the model's performance. The ratio of prediction and the desired output yields the accuracy of the model.

Precision is the ratio of true positive instances and the total number of positively predicted instances.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$= \frac{\text{True Positive}}{\text{Total Predicted Positive}}$$

**Recall** is the ratio of true positive instances and the total actual positive instances.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$= \frac{\text{True Positive}}{\text{Total Actual Positive}}$$

## Precision in NLP

### 62) What is F1 score in NLP?

**Ans:**

F1 score evaluates the weighted average of recall and precision. It considers both false negative and false positive instances while evaluating the model. F1 score is more accountable than accuracy for an NLP model when there is an uneven distribution of class. Let us look at the formula for calculating F1 score:

## Precision in NLP

### 63) How to tokenize a sentence using the nltk package?

**Ans:**

Tokenization is a process used in NLP to split a sentence into tokens. Sentence tokenization refers to splitting a text or paragraph into sentences.

For tokenizing, we will import sent\_tokenize from the nltk package:

```
from nltk.tokenize import sent_tokenize<>
```

We will use the below paragraph for sentence tokenization:

Para = “Hi Guys. Welcome to ACTE. This is a blog on the NLP interview questions and answers.”

```
sent_tokenize(Para)
```

**Output:**

1. [ ‘Hi Guys.’ ,
2. ‘Welcome to ACTE. ‘ ,
3. ‘This is a blog on the NLP interview questions and answers. ‘ ]

Tokenizing a word refers to splitting a sentence into words.

Now, to tokenize a word, we will import word\_tokenize from the nltk package.

```
from nltk.tokenize import word_tokenize
```

Para = “Hi Guys. Welcome to ACTE. This is a blog on the NLP interview questions and answers.”

```
word_tokenize(Para)
```

**Output:**

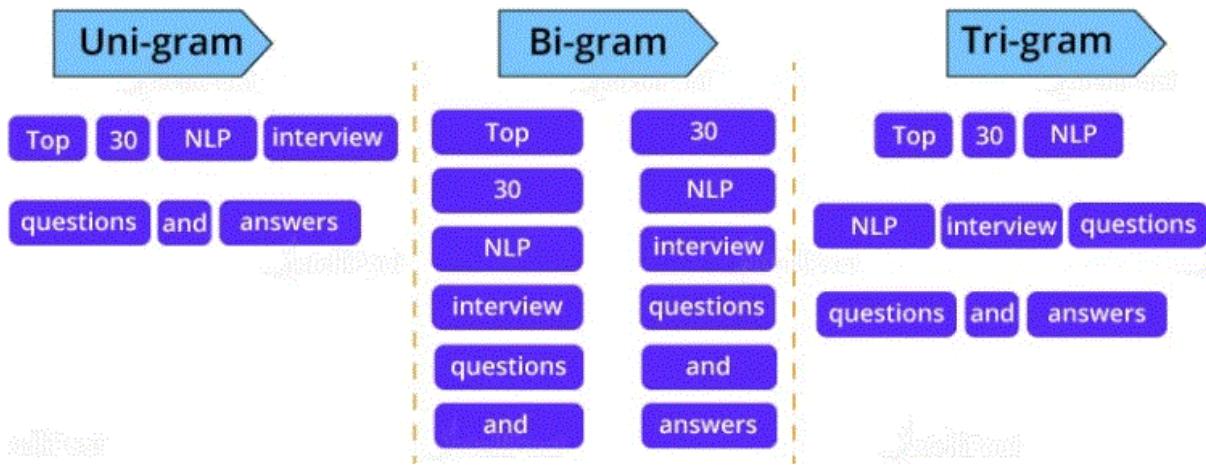
```
[ ‘Hi’ , ‘Guys’ , ‘. ‘ , ‘Welcome’ , ‘to’ , ‘ACTE’ , ‘. ‘ , ‘This’ , ‘is’ ,
```

**64) Explain how we can do parsing?**

**Ans:**

Parsing is the method to identify and understand the syntactic structure of a text. It is done by analyzing the individual elements of the text. The machine parses the text one word at a time, then two at a time, further three, and so on.

- When the machine parses the text one word at a time, then it is a unigram.
- When the text is parsed two words at a time, it is a bigram.
- The set of words is a trigram when the machine parses three words at a time.



## Parsing in NLP

### 65) Explain Stemming with the help of an example?

**Ans:**

In Natural Language Processing, stemming is the method to extract the root word by removing suffixes and prefixes from a word.

For example, we can reduce ‘stemming’ to ‘stem’ by removing ‘m’ and ‘ing.’

We use various algorithms for implementing stemming, and one of them is PorterStemmer.

First, we will import PorterStemmer from the nltk package.

```
from nltk.stem import PorterStemmer
```

Creating an object for PorterStemmer

1. *pst=PorterStemmer()*
2. *pst.stem("running"), pst.stem("cookies"), pst.stem("flying")*

**Output:**

```
(‘run’, ‘cooki’, ‘fly’ )
```

### 66) Explain Lemmatization with the help of an example?

**Ans:**

We use stemming and lemmatization to extract root words. However, stemming may not give the actual word, whereas lemmatization generates a meaningful word.

In lemmatization, rather than just removing the suffix and the prefix, the process

tries to find out the root word with its proper meaning.

**Example:** ‘Bricks’ becomes ‘brick,’ ‘corpora’ becomes ‘corpus,’ etc.

Let’s implement lemmatization with the help of some nltk packages.

First, we will import the required packages.

- *from nltk.stem import wordnet*
- *from nltk.stem import WordnetLemmatizer*

Creating an object for WordnetLemmatizer()

- *lemma=WordnetLemmatizer()*
- *list = [“Dogs”, “Corpora”, “Studies”]*
- *for n in list:*
- *print(n + “:” + lemma.lemmatize(n))*

**Output:**

- *Dogs: Dog*
- *Corpora: Corpus*
- *Studies: Study*

## 67) Explain Named Entity Recognition by implementing it?

**Ans:**

Named Entity Recognition (NER) is an information retrieval process. NER helps classify named entities such as monetary figures, location, things, people, time, and more. It allows the software to analyze and understand the meaning of the text. NER is mostly used in NLP, Artificial Intelligence, and Machine Learning. One of the real-life applications of NER is chatbots used for customer support. Let’s implement NER using the spacy package.

Importing the spacy package:

1. *import spacy*
2. *nlp = spacy.load(‘en\_core\_web\_sm’)*
3. *Text = “The head office of Google is in California”*
4. *document = nlp(text)for ent in document.ents:*
5. *print(ent.text, ent.start\_char, ent.end\_char, ent.label\_)*

**Output:**

1. *Office 9 15 Place*
2. *Google 19 25 ORG*
3. *California 32 41 GPE*

**68) How to check word similarity using the spacy package?****Ans:**

To find out the similarity among words, we use word similarity. We evaluate the similarity with the help of a number that lies between 0 and 1. We use the spacy library to implement the technique of word similarity.

1. *import spacy*
2. *nlp = spacy.load('en\_core\_web\_md')*
3. *print("Enter the words")*
4. *input\_words = input()*
5. *tokens = nlp(input\_words)*
6. *for i in tokens:*
7. *print(i.text, i.has\_vector, i.vector\_norm, i.is\_oov)*
8. *token\_1, token\_2 = tokens[0], tokens[1]*
9. *print("Similarity between words:", token\_1.similarity(token\_2))*

**Output:**

1. *hot True 5.6898586 False*
2. *cold True 6.5396233 False*
3. *Similarity: 0.597265*



**Get JOB Oriented Natural Language Processing (NLP) Training for Beginners  
By MNC Experts**

- *Instructor-led Sessions*

- *Real-life Case Studies*
- *Assignments*

### **Explore Curriculum**

**69) List some open-source libraries for NLP?**

**Ans:**

The popular libraries are NLTK (Natural Language ToolKit), SciKit Learn, Textblob, CoreNLP, spaCY, Gensim.



**70) Explain the masked language model?**

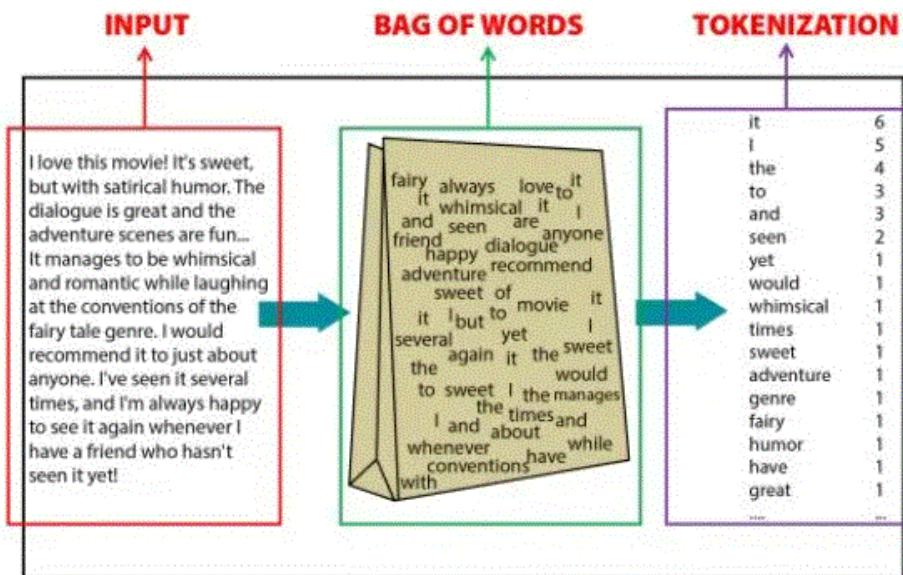
**Ans:**

Masked modeling is an example of autoencoding language modeling. Here the output is predicted from corrupted input. By this model, we can predict the word from other words present in the sentences.

**71) What is the bag of words model?**

**Ans:**

The Bagofwords model is used for information retrieval. Here the text is represented as a multiset, i.e., a bag of words. We don't consider grammar and word order, but we surely maintain the multiplicity:



**Bag of words model**

**72) What are the features of the text corpus in NLP?**

**Ans:**

The features of text corpus are:

- Word count
- Vector notation
- Part of speech tag
- Boolean feature
- Dependency grammar

**73) What is named entity recognition (NER)?**

**Ans:**

Named Entity Recognition is a part of information retrieval, a method to locate and classify the entities present in the unstructured data provided and convert them into predefined categories.

**74) What is an NLP pipeline, and what does it consist of?**

**Ans:**

Generally, NLP problems can be solved by navigating the following steps (referred as a pipeline):

- Gathering text, whether it's from web scraping or the use of available datasets

- Cleaning text (through the processes of stemming and lemmatization)
- Representation of the text (bag-of-words method)
- Word embedding and sentence representation (Word2Vec, SkipGram model)
- Training the model (via neural nets or regression techniques)
- Evaluating the model
- Adjusting the model, as needed
- Deploying the model

### **75) What is a “stop” word?**

**Ans:**

Articles such as “the” or “an,” and other filler words that bind sentences together (e.g., “how,” “why,” and “is”) but don’t offer much additional meaning are often referred to as “stop” words. In order to get to the root of a search and deliver the most relevant results, search engines routinely filter out stop words.

### **76) What is “term frequency-inverse document frequency”?**

**Ans:**

Term frequency-inverse document frequency (TF-IDF) is an indicator of how important a given word is in a document, which helps identify key words and assist with the process of feature extraction for categorization purposes. While “TF” identifies how frequently a given word or phrase (“W”) is used, “IDF” measures its importance within the document. The formulas to answer this NLP interview question are as follows:

$$TF(W) = \text{Frequency of } W \text{ in a document} / \text{Total number of terms in the document}$$

$$IDF(W) = \log_e (\text{Total number of documents} / \text{Number of documents having the term } W)$$

Using these formulas, you can determine just how important a given word or phrase is within a document. If the TF-IDF is high, then the frequency of that term is lower; if the TF-IDF is low, then its frequency is higher. Search engines use this to help them rank sites.

**77) What is perplexity? What is its place in NLP?**

**Ans:**

Perplexity is a way to express a degree of confusion a model has in predicting. More entropy = more confusion. Perplexity is used to evaluate language models in NLP. A good language model assigns a higher probability to the right prediction.

**78) What is the problem with ReLu?**

**Ans:**

- Exploding gradient(Solved by gradient clipping)
- Dying ReLu — No learning if the activation is 0 (Solved by parametric relu)
- Mean and variance of activations is not 0 and 1.(Partially solved by subtracting around 0.5 from activation. Better explained in fastai videos)

**79) What is the difference between learning latent features using SVD and getting embedding vectors using deep network?**

**Ans:**

SVD uses linear combination of inputs while a neural network uses non-linear combination.

**80) What are the different types of attention mechanism?**

**Ans:**

| Name                   | Alignment score function                                                                                                                                                                                        | Citation     |
|------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| Content-base attention | $\text{score}(s_t, \mathbf{h}_i) = \text{cosine}[s_t, \mathbf{h}_i]$                                                                                                                                            | Graves2014   |
| Additive(*)            | $\text{score}(s_t, \mathbf{h}_i) = \mathbf{v}_a^\top \tanh(\mathbf{W}_a[s_t; \mathbf{h}_i])$                                                                                                                    | Bahdanau2015 |
| Location-Base          | $\alpha_{t,i} = \text{softmax}(\mathbf{W}_a s_t)$<br>Note: This simplifies the softmax alignment to only depend on the target position.                                                                         | Luong2015    |
| General                | $\text{score}(s_t, \mathbf{h}_i) = s_t^\top \mathbf{W}_a \mathbf{h}_i$<br>where $\mathbf{W}_a$ is a trainable weight matrix in the attention layer.                                                             | Luong2015    |
| Dot-Product            | $\text{score}(s_t, \mathbf{h}_i) = s_t^\top \mathbf{h}_i$                                                                                                                                                       | Luong2015    |
| Scaled Dot-Product(^)  | $\text{score}(s_t, \mathbf{h}_i) = \frac{s_t^\top \mathbf{h}_i}{\sqrt{n}}$<br>Note: very similar to the dot-product attention except for a scaling factor; where n is the dimension of the source hidden state. | Vaswani2017  |

(\*) Referred to as “concat” in Luong, et al., 2015 and as “additive attention” in Vaswani, et al., 2017.

(^) It adds a scaling factor  $1/\sqrt{n}$ , motivated by the concern when the input is large, the softmax function may have an extremely small gradient, hard for efficient learning.

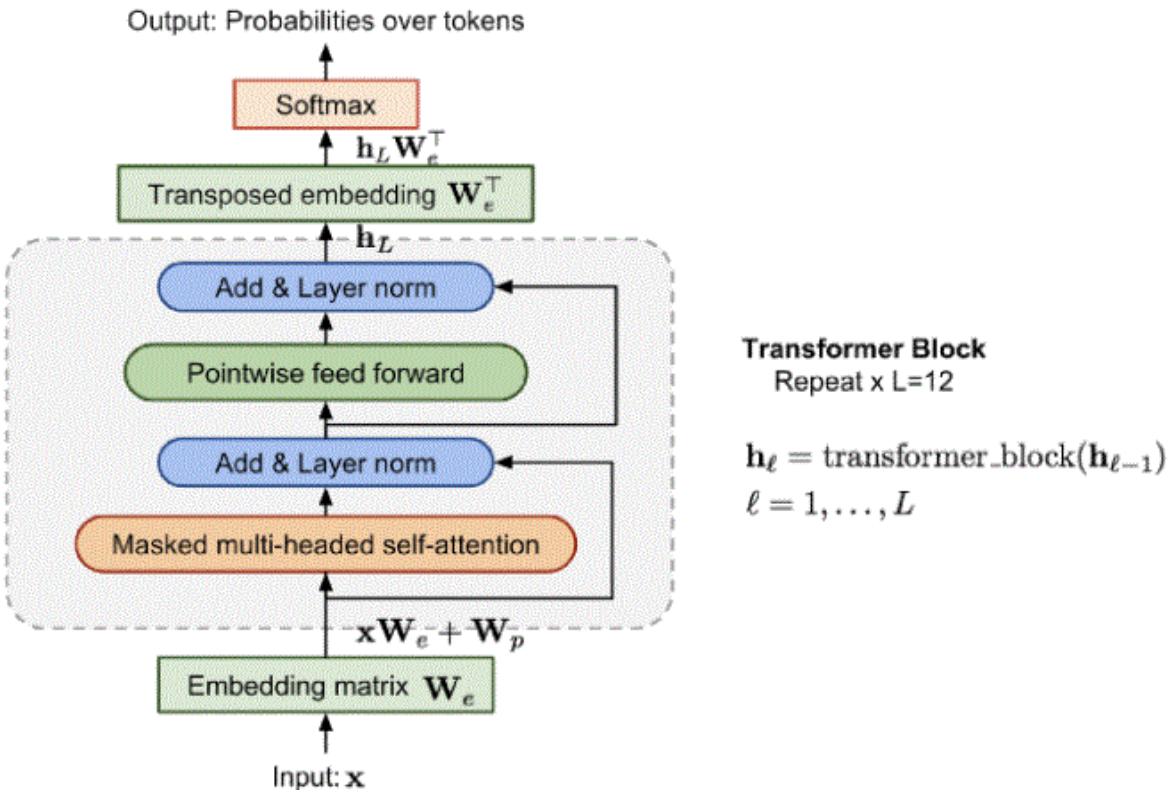
Here are a summary of broader categories of attention mechanisms:

| Name              | Definition                                                                                                                                                                                        | Citation             |
|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| Self-Attention(&) | Relating different positions of the same input sequence. Theoretically the self-attention can adopt any score functions above, but just replace the target sequence with the same input sequence. | Cheng2016            |
| Global/Soft       | Attending to the entire input state space.                                                                                                                                                        | Xu2015               |
| Local/Hard        | Attending to the part of input state space; i.e. a patch of the input image.                                                                                                                      | Xu2015;<br>Luong2015 |

## Attention Mechanism in NLP

81) Why does the transformer block have LayerNorm instead of BatchNorm?

Ans:



*LayerNorm instead of BatchNorm*

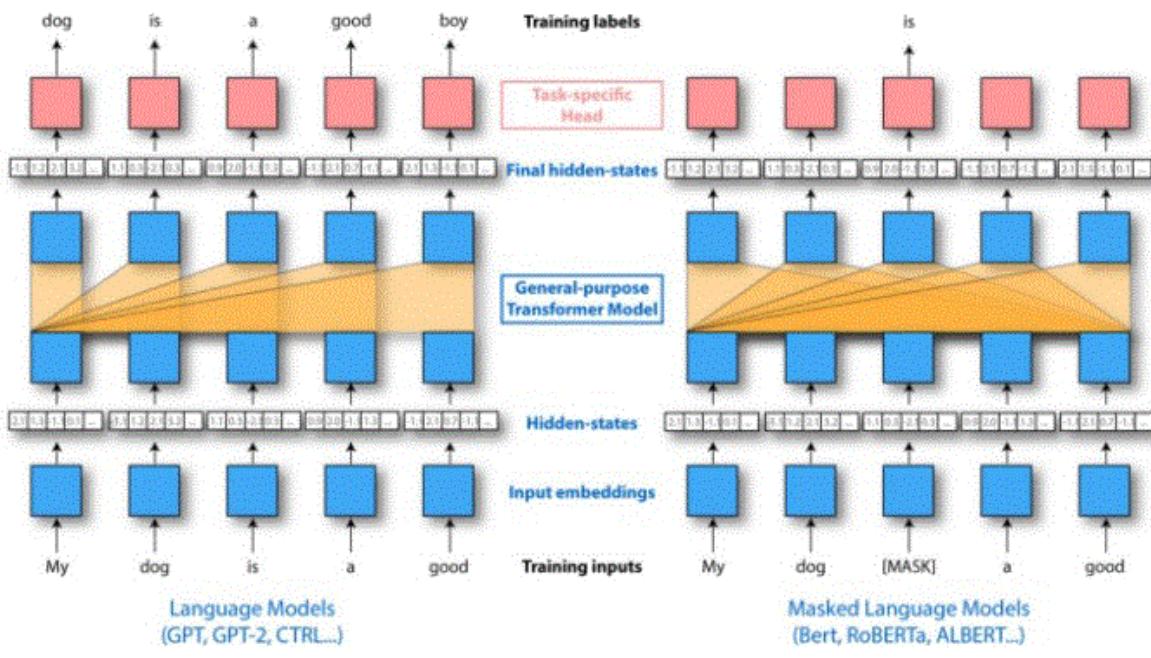
**82) What are the tricks used in ULMFiT? (Not a great questions but checks the awareness)**

**Ans:**

- LM tuning with task text
- Weight dropout
- Discriminative learning rates for layers
- Gradual unfreezing of layers
- Slanted triangular learning rate schedule

**83) What are the differences between GPT and BERT?**

**Ans:**



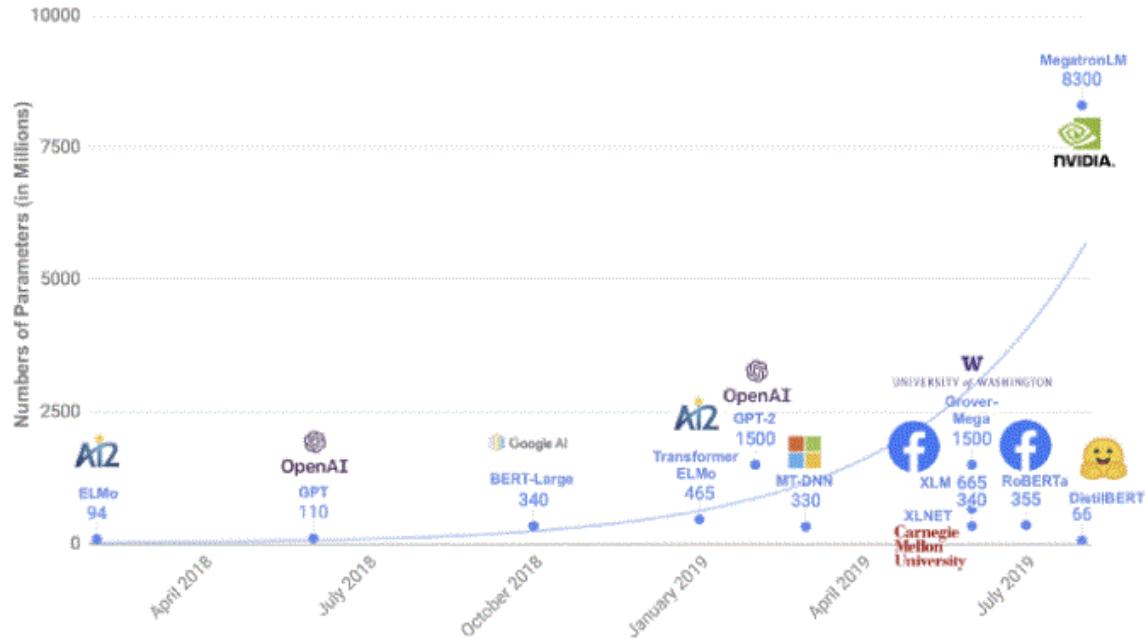
Thomas Wolf

## GPT and BERT

- GPT is not bidirectional and has no concept of masking
- BERT adds next sentence prediction task in training and so it also has a segment embedding

84) Given this chart, will you go with a transformer model or LSTM language model?

Ans:



### *LSTM language model*

**85) What is Language modeling ?**

**Ans:**

A statistical language model is a probability distribution over sequences of words. Given such a sequence, say of length m, it assigns a probability to the whole sequence. The language model provides context to distinguish between words and phrases that sound similar.

**86) What is Latent semantic analysis ?**

**Ans:**

Latent semantic analysis is a technique in natural language processing, in particular distributional semantics, of analyzing relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms.

**87) What is Fasttext?**

**Ans:**

FastText is a library for learning of word embeddings and text classification created by Facebook's AI Research lab. The model allows to create an unsupervised learning or supervised learning algorithm for obtaining vector representations for words.

## **88) Explain parts of speech tagging (POS tagging)?**

**Ans:**

**Example:-**

1. *import nltk*
2. *from nltk.corpus import stopwords*
3. *from nltk.tokenize import word\_tokenize, sent\_tokenize*
4. *stop\_words = set(stopwords.Words('english'))*
5. *txt = "A, B, C are longtime classmates."*
1. *## Tokenized via sent\_tokenize*
2. *tokenized\_text = sent\_tokenize(txt)/li>*
1. *## Using word\_tokenizer to identify a string's words and punctuation then removing stop words*
2. *for n in tokenized\_text:*
3. *wordsList = nltk.word\_tokenize(i)*
4. *wordsList = [w for w in wordsList if not w in stop\_words]*
1. *## Using POS tagger*
2. *tagged\_words = nltk.pos\_tag(wordsList)*
3. *print(tagged\_words)*

**Output:-**

1. *[('A', 'NNP'), ('B', 'NNP'), ('C', 'NNP'), ('longtime', 'JJ'), ('classmates', 'NNS')]*

## **89) Define and implement named entity recognition?**

**Ans:**

For retrieving information and identifying entities present in the data for instance location, time, figures, things, objects, individuals, etc. NER (named entity recognition) is used in AI, NLP, machine learning, implemented for making the software understand what the text means. Chatbots are a real-life example that makes use of NER.

1. *import spacy*
2. *nlp = spacy.load('en\_core\_web\_sm')*
3. *Text = "The head office of Tesla is in California"*
4. *document = nlp(text)*

5. *for ent in document.ents:*
6. *print(ent.text, ent.start\_char, ent.end\_char, ent.label\_)*

**output**

1. *Office 9 15 Place*
2. *Tesla 19 25 ORG*
3. *California 32 41 GPE*

**90) What is the significance of TF-IDF?**

**Ans:**

tf-idf or TFIDF stands for term frequency-inverse document frequency. In information retrieval TFIDF is a numerical statistic that is intended to reflect how important a word is to a document in a collection or in the collection of a set.

**91) What is Named Entity Recognition(NER)?**

**Ans:**

Named entity recognition is a method to divide a sentence into categories. Neil Armstrong of the US had landed on the moon in 1969 will be categorized as Neil Armstrong- name; The US – country; 1969 – time(temporal token).

The idea behind NER is to enable the machine to pull out entities like people, places, things, locations, monetary figures, and more.

**92) Explain briefly about word2vec?**

**Ans:**

Word2Vec embeds words in a lower-dimensional vector space using a shallow neural network. The result is a set of word-vectors where vectors close together in vector space have similar meanings based on context, and word-vectors distant to each other have differing meanings.

**For example,** apple and orange would be close together and apple and gravity would be relatively far. There are two versions of this model based on skip-grams (SG) and continuous-bag-of-words (CBOW).

**93) What are the 5 steps in NLP?**

**Ans:**

The five phases of NLP involve lexical (structure) analysis, parsing, semantic analysis, discourse integration, and pragmatic analysis. Some well-known application areas of NLP are Optical Character Recognition (OCR), Speech Recognition, Machine Translation, and Chatbots.

**94) What is the main challenges of NLP?**

**Ans:**

There are enormous ambiguity exists when processing natural language. 4. Modern NLP algorithms are based on machine learning, especially statistical machine learning.

**95) Which NLP model gives the best accuracy amongst the following?**

**Ans:**

Naive Bayes is the most precise model, with a precision of 88.35%, whereas Decision Trees have a precision of 66%.

**96) How many components of NLP are there?**

**Ans:**

Five main Component of Natural Language processing in AI are: Morphological and Lexical Analysis. Syntactic Analysis. Semantic Analysis.

**97) What is NLP example?**

**Ans:**

Email filters are one of the most basic and initial applications of NLP online. It started out with spam filters, uncovering certain words or phrases that signal a spam message. The system recognizes if emails belong in one of three categories (primary, social, or promotions) based on their contents.

**98) Who uses NLP?**

**Ans:**

Interest in NLP grew in the late 1970s, after Bandler and Grinder began marketing the approach as a tool for people to learn how others achieve success. Today, NLP is used in a wide variety of fields, including

counseling, medicine, law, business, the performing arts, sports, the military, and education.

## **99) Why is neurolinguistics important?**

### **Ans:**

Neurolinguistics is important because it studies the mechanisms in the brain that control acquisition, comprehension, and production of language.

#### **1. What do you mean by NLP?**

Natural Language Processing is a programmed way to understand or consider the natural languages and remove necessary information from such data by applying machine learning Algorithms.

#### **2. Tell me the significance of TF-IDF?**

TFIDF refers to term frequency opposite document occurrence. In information retrieval, TFIDF is an arithmetical statistic that is planned to reproduce how significant a word is to a text in a compilation.

#### **3. What do you mean by tokenization in NLP?**

Natural Language Processing aims to plan computers to route large amounts of natural language data. Tokenization in NLP means the technique of dividing the text into a variety of tokens. You can think of a coupon in the shape of the word. Just like a word forms into a sentence.

#### **4. Define Pragmatic Analysis?**

The pragmatic analysis is a significant task in NLP for interpreting knowledge that is laying exterior a given document. The plan of implementing pragmatic analysis is to spotlight on exploring a diverse aspect of the document or text in a language. The pragmatic analysis permits software applications for the serious interpretation of the real-world data to know the definite meaning of sentences and words.

#### **5. Tell me the steps involved in solving an NLP Problem?**

The following steps involved are:

- Gather the text from the obtainable dataset or by web scraping
- Apply stemming and lemmatization for text crackdown

- Apply characteristic engineering techniques
- Embed using word2vec
- Train the built model using neural networks
- Assess the model's performance
- Make suitable changes in the model.
- Deploy the replica

## **6. What is the F1 score in NLP?**

F1 score evaluates the subjective standard of recall and precision. It considers both false unconstructive and false constructive instances while evaluating the model. F1 score is more answerable than accurateness for an NLP model when there is a rough allocation of class.

## **7. What do you mean by Regular Grammar?**

A regular grammar comprises a system in the shape of  $A \rightarrow a$ ,  $A \rightarrow aB$  and many more. The rules help sense and examine strings by automatic computation.

## **8. Define dependency parsing in NLP?**

Dependency parsing is single of the serious tasks in NLP. It allows the examination of a ruling using parsing algorithms. Also, by using the parse tree independence parsing, we can ensure the grammar and examine the semantic arrangement of a sentence.

## **9. Tell me some areas of NLP?**

Some areas of NLP are:

- Question Answering
- Semantic Analysis
- Text classifications
- Automation Summarization

## **10. Define Lemmatization in NLP?**

Lemmatization usually means to do the things suitably with the use of vocabulary and morphological investigation of words. In this method, the

endings of the words are detached to revisit the base word, which is also known as Lemma.

### **11. Name some several classifications model to define in NLP?**

Some several classification models:

- Part of speech tagging
- Counting frequency of define terms
- Notation of vector for every sentence
- Grammatical dependency

### **12. What are bigrams, unigrams and n-grams in NLP?**

When we parse a ruling one word at a time, then it is called a unigram. The ruling parsed two words at a time are a bigram. When the ruling is parsed three words at a time, then it is a trigram. Likewise, n-gram refers to the parsing of n languages at a time.

### **13. Define Regular Expressions?**

A regular expression is used to contest and tag words. It consists of a sequence of characters for identical strings. Imagine, if X and Y are regular expressions, then the next is true for them:

- If {?} is a habitual language, then ? is a regular expression for it.
- If X and Y are habitual expressions, then X + Y is also a regular expression within the language {X, Y}.
- If X and Y are regular expressions, then the concatenation of X and Y (X.Y) is a usual expression.
- If X is a regular expression, then X\* (X occurring numerous times) is also a regular expression.

### **14. Which one of the following is not a pre-processing technique in NLP?**

- Stemming and Lemmatization
- converting to lowercase
- removing punctuations
- removal of stop words

- Sentiment analysis

Sentiment Analysis is not a pre-processing procedure. It is done after pre-dispensation and is an NLP use case. All other scheduled ones are used as part of declaration pre-processing.

### **15. Name some popular models other than a bag of words?**

Indexing, Latent semantic word2vec.

### **16. Where NER can be used?**

Scanning documents for categorization, customer support (chatbots, understanding feedback) and unit recognition in molecular biology (names of genes etc.)

### **17. Define parsing in the context of NLP?**

Parsing a document means to function out the grammatical constitution of sentences, for example, which groups of words go mutually (as “phrases”) and which words are the topic or object of a verb. Probabilistic parsers use the information of language gained from hand-parsed sentences to try to create the most likely examination of new sentences.

### **18. What does NLP pipeline consist of?**

It consists of the followings ways:

- Feature generation (Bag of words)
- Text cleaning (stemming, lemmatization)
- Embedding and ruling representation (word2vec)
- Text gathering (web scraping or available datasets)
- Model assessment
- Making adjustments to the model
- Training the model by leveraging neural nets or weakening techniques

### **19. How to check word similarity using the spacy package?**

To find out the comparison among words, we use word likeness. We appraise the similarity with the assist of a number that lies between 0 and 1.

```

import spacy

nlp = spacy.load('en_core_web_md'

print("Enter the words")

input_words = input()

tokens = nlp(input_words)

for i in tokens:

    print(i.text, i.has_vector, i.vector_norm, i.is_oov)

token_1, token_2 = tokens[0], tokens[1]

print("Similarity between words:", token_1.similarity(token_2))

```

Output:

*hot True 5.6898586 False*

*cold True 6.5396233 False*

*Similarity: 0.597265*

## 20. Define Named entity recognition?

Named-entity recognition is the process of extracting information. It arranges and classifies named unit in the shapeless text in diverse categories like locations, time expressions, percentages, and monetary values. It allows the users to appreciate the subject of the text properly.

## Q1. What are the different types of activation functions?

This is an important Deep Learning coding interview question. You must know the following types of activation functions:

- Sigmoid function:** It is a nonlinear function in an ANN that is mostly used in feedforward neural networks. It's a differentiable real function with positive derivatives and a certain degree of smoothness. It is written as: {"detectHand":false}.
- Hyperbolic tangent function (Tanh):** It is a smoother, zero-centered function (range of -1 to +1). The output is represented by: {"detectHand":false}. The primary advantage of this function is that it gives a zero-centered output that helps in backpropagation.
- Softmax function:** It is used to generate probability distribution from a vector of real numbers. This function returns the output between 0 and 1, with the sum of probabilities equals to 1. This is written as: {"detectHand":false}. It is used in multi-class models, returning probabilities of each class, with the target having the highest probability.
- Softsign function:** It is commonly used in regression computation issues and text-to-speech applications. It's a quadratic polynomial, written as: {"detectHand":false}.
- Rectified linear unit of function:** It outperforms other AFs in generalization and performance. The function is roughly linear and preserves the features of linear models, making gradient-descent approaches easier to optimize. It is written as: {"detectHand":false}.
- Exponential linear unit of function:** The major advantage of this function is that it can solve the vanishing gradient problem by employing identity for positive values and boosting the model's learning properties. It is represented by: {"detectHand":false}.

## Q2. How does recurrent neural network backpropagation vary from artificial neural network backpropagation?

Each node in a recurrent neural network has an additional loop. This makes it different from artificial neural network propagation. This loop incorporates a temporal component into the network. The main advantage of recurrent neural networks is that they allow for sequential data information. This is usually impossible with a generic artificial neural network.

## Q3. Can a deep learning model be solely built on linear regression?

If you are well-versed in Deep Learning, you can answer these types of Deep Learning interview questions with ease.

A deep learning model may be solely built on linear regression. However, the problem should be represented by a linear equation, which does not boost the

machine learning model's predictive capacity due to the addition of nodes. Hence, building a deep learning model solely on linear regression creates no spectacular results.

#### **Q4. What is a computational graph in Deep Learning?**

This is one of the important topics asked in Deep Learning interview questions.

A computational graph is a series of operations performed to take inputs and arrange them as nodes in a graph. It is a way of implementing mathematical calculations into a graph. This way, it will help in parallel processing and provide high performance in terms of computational capability.

#### **Q5. What are the types of autoencoders, and where are they used?**

This is a commonly asked in Deep Learning interview question. You must have a sound understanding of what autoencoders are to answer this.

Autoencoders are used worldwide. Some of the popular usages of autoencoders are:

1. Adding color to black-white images
2. Removing noise from images
3. Dimensionality reduction
4. Feature removal and variation

You must know there are four types of autoencoders. They are:

1. Deep autoencoders
2. Convolutional autoencoders
3. Sparse autoencoders
4. Contractive autoencoders

Recommended Reading: [Google Machine Learning Engineer Interview Prep](#)

### **Deep Learning NLP Interview Questions**

To excel at ML or data science interviews, you must have profound knowledge of natural language processing (NLP) in Deep Learning. Following are a few questions that you must practice to nail your Deep Learning interview at FAANG and tier-1 tech companies.

## **Q1. What do you understand about text normalization in NLP?**

These types of Deep Learning interview questions test your fundamental knowledge of the subject.

When developing NLP tools to work with exceptional data, it's beneficial to attain a canonical representation of textual content. This is known as textual normalization. Textual normalization captures different kinds of variations into one representation.

## **Q2. Do you know what feature engineering is?**

When you employ machine learning methods to complete your modeling, you need to input pre-processed text into an NLP algorithm. This set of strategies used for this process is known as feature engineering or feature extraction. The main purpose of feature extraction is to convert the text's qualities into a numeric vector that NLP algorithms can understand. This stage is known as text representation.

## **Q3. Explain TF-IDF in NLP.**

TF-IDF is known as Term-Frequency-Inverse Document Frequency. It helps you get the importance of a particular word relative to other words in the corpus. It converts words into vectors and adds semantic information, resulting in weighted unusual words. These words can be utilized in various NLP applications. Moreover, it's a common scoring metric in information retrieval and summarization.

## **Q4. What do you understand about POS tagging?**

A part-of-speech (POS) tagger reads the text in a language and assigns speed parts to each word, such as noun, verb, adverb, and others. POS taggers employ an algorithm to label terms in text bodies. These labels create various complex categories with tags like "noun plural" or other complicated labels.

## **Q5. What is the difference between NLP and NLU?**

This is one of the most asked Deep Learning interview questions. The differences between NLP and NLU are:

| Natural Language Understanding (NLU)                                                       | Natural Language Programming (NLP)                                                          |
|--------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| Aids in solving AI's complex problems.                                                     | A system that manages end-to-end conversations between computers and people simultaneously. |
| Allows machines to interpret unstructured input by transforming them into structured text. | Humans and machines are involved in NLP.                                                    |
| Concentrates on extracting meaning and context.                                            | Focuses on interpreting language in its most literal sense.                                 |
| Helps machines deduce the meaning behind the language content.                             | Can parse text-based on grammar, typography, structure, and point of view.                  |

Recommended Reading: **Amazon Machine Learning Engineer Interview Prep**

### Deep Learning Computer Vision Interview Questions

If you are applying for a role of a Computer Vision Engineer in any top company, you must practice the following Deep Learning computer vision interview questions to uplevel your preparation:

#### **Q1. What are the features detected by the initial layers of a neural network used for computer vision? How is it different from what is detected by the later neural network layers?**

Neural network's earlier layers detect simple features of an image (for example, edges or corners). As you go deeper, the features become increasingly complex, detecting patterns and shapes in the neural network. The later layers can detect intricate patterns, such as complete objects.

#### **Q2. How will you address the edge pixels issue during convolutional operation?**

You can use padding to address the issue of filter or kernel extracting information from the edge pixels less compared to the central pixel. Padding is the addition of one or more rows or columns of pixels along the boundary of the image.

It forms the new pixels of the picture. Therefore, it results in insufficient extraction of information from the original edge pixels. It also prevents the shrinking of an image due to the convolution operations.

**Q3. You are given a 5x5 image with a 3x3 filter and a padding p = 1. What will be the resultant image's size if a convolutional stride of s = 2 is used?**

You should know that for an  $n \times n$  image with an  $f \times f$  filter, padding  $p$ , and stride length  $s$ , resultant image's size after convolution has the shape  $n + 2p - fs + 1 \times n + 2p - fs + 1$ . Therefore, per the data provided, the resulting size of the image will be  $((5 + 2 * 1 - 3) / 2) + 1 \times ((5 + 2 * 1 - 3) / 2) + 1 = 3 \times 3$ .

**Q4. What will be the resultant image size for an RGB image of 10x10x3 convolved with a 3x3 filter?**

The convolution operation is not possible for such dimensions of an RGB image. The third dimension (number of channels) should be the same to achieve convolution. However, if the 10x10x3 image is convolved in a 3x3x3 filter, the dimensions of the resultant image will be 4x4.

**Q5. How many parameters need to be learned in pooling layers?**

The pooling layer contains hyperparameters describing the filter size and the stride length. These parameters are set and work as a fixed computation. Hence, no parameters are to be learned in the pooling layers.

1. What is an ensemble method in NLP?
2. State the steps to build a text classification system.
3. How is parsing done in NLP?
4. Differentiate between deep learning and machine learning.
5. What is a bag of words (BOW)?
6. What is Latent Semantic Indexing (LSI) in NLP?
7. What are some metrics on which NLP models are evaluated?
8. Explain the pipeline for information extraction.
9. What do you understand about autoencoders?
10. Explain the meaning of masked language modeling.
11. Explain pragmatic analysis in NLP.
12. What is the meaning of N-gram in NLP?

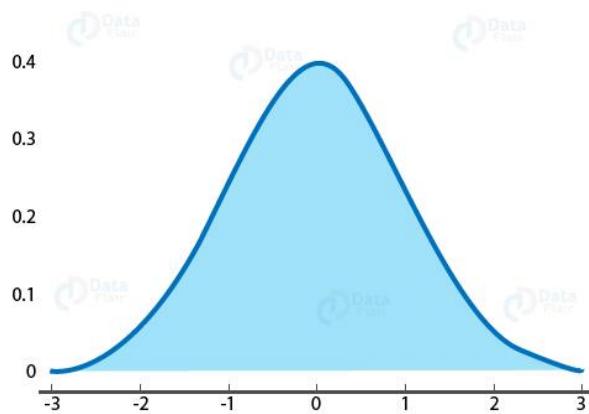
- 13.What do you mean by perplexity in NLP?
- 14.Explain why the inputs in computer vision problems can get huge.  
Provide a solution to overcome this challenge.
- 15.What should the padding be for a 10x10 image used with a 5x5 filter to get an image of the same size as the original image?
- 16.What method can be used to evaluate an object localization model? How does it work?
- 17.How will you use IoU for resolving the issue of multiple detections of the same object?
- 18.Give us an example of a scenario that would require the use of anchor boxes.
- 19.How is the Siamese Network beneficial in addressing the one-shot learning problem?
- 20.What purpose does grayscaling serve?
- 21.Explain translational equivariance.
- 22.Explain the object detection algorithm YOLO.
- 23.What do you know about dropouts?
- 24.Explain exploding and vanishing gradients.
- 25.Differentiate between bias and variance in the context of deep learning models. How can you achieve a balance between the two?
- 26.According to you, which one is more powerful — a two-layer neural network without any activation function or a two-layer decision tree?
- 27.While building a neural network architecture, how will you decide how many neurons and hidden layers should the neural network have?
- 28.What is an activation function? What is the use of an activation function?
- 29.What deep learning algorithm works best for face detection?
- 30.What is Stochastic Gradient Descent and how is it different from Batch Gradient Descent?
- 31.Explain how you would fix the constant validation accuracy in a Convolutional Neural Network (CNN)?
- 32.What are the differences between a shallow network and a deep network.
- 33.What is a tensor in deep learning?
- 34.What are the advantages of transfer learning?

35. Difference between multi-class and multi-label classification problems.
36. What are the different techniques to achieve data normalization?
37. What are Forward and Back Propagation in the context of deep learning?
38. List the different types of deep neural networks.

### **Q.1 What do you understand by the term Normal Distribution?**



#### Normal Distribution



**Normal Distribution** is also known as Gaussian Distribution. It is a type of probability distribution that is symmetric about the mean. it shows that the data is closer to the mean and the frequency of occurrences in data are far from the mean.

### **Q.2 How will you explain linear regression to a non-tech person?**

Linear Regression is a statistical technique of measuring the linear relationship between the two variables. By linear relationship, we mean that an increase in a variable would lead to increase in the other variable and a decrease in one variable would lead to attenuation in the second variable as well. Based on this linear relationship, we establish a model that predicts the future outcomes based on an increase in one variable.

### **Q.3 How will you handle missing values in data?**

There are several ways to handle missing values in the given data-

- Dropping the values
- Deleting the observation (not always recommended).
- Replacing value with the mean, median and mode of the observation.
- Predicting value with regression
- Finding appropriate value with clustering

### **Q.4 How will you verify if the items present in list A are present in series B?**

We will use the `isin()` function. For this, we create two series `s1` and `s2` –

```
s1 = pd.Series([1, 2, 3, 4, 5])
```

```
s2 = pd.Series([4, 5, 6, 7, 8])
s1[s1.isin(s2)]
```

### **Q.5 How to find the positions of numbers that are multiples of 4 from a series?**

For finding the multiples of 4, we will use the `argwhere()` function. First, we will create a list of 10 numbers –

```
s1 = pd.Series([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])
np.argwhere(ser % 4==0)
```

**Output > [3], [7]**

### **Q.6 How are KNN and K-means clustering different?**

Firstly, KNN is a supervised learning algorithm. In order to train this algorithm, we require labeled data. K-means is an unsupervised learning algorithm that looks for patterns that are intrinsic to the data. The K in KNN is the number of nearest data points. On the contrary, the K in K-means specify the number of centroids.

### **Q.7 Can you stack two series horizontally? If so, how?**

Yes, we can stack the two series horizontally using `concat()` function and setting `axis = 1`.

```
df = pd.concat([s1, s2], axis=1)
```

### **Q.8 How can you convert date-strings to timeseries in a series?**

**Input:**

```
s = pd.Series(['02 Feb 2011', '02-02-2013', '20160104', '2011/01/04', '2014-12-05', '2010-06-06T12:05'])
```

To solve this, we will use the `to_datetime()` function.

```
pd.to_datetime(s)
```

### **Q.9 Python or R – Which one would you prefer for text analytics?**



## Difference Between R and Python



| Features             | R                                              | Python                                                                            |
|----------------------|------------------------------------------------|-----------------------------------------------------------------------------------|
| Scope                | Used mainly for statistical modeling           | Used for a variety of purposes like web-application development and data analysis |
| Used By              | Statisticians, Analyst & Data Scientist        | Developer, Data Engineers & Data Scientist                                        |
| Suitable For         | People with no prior experience in programming | Newbies to experienced IT professionals                                           |
| Package Distribution | CRAN                                           | PyPi                                                                              |
| Visualization Tools  | ggplot2, plotly, ggiraph                       | Matplotlib, bokkeh, seaborn                                                       |

Both Python and R provide robust functionalities for working with text data. R provides extensive text analytics libraries but its data mining libraries are still in a nascent stage. Python is best suited for enterprise level and for increasing software productivity. For handling unstructured data, R provides a vast variety of support packages. Python is best apt at handling colossal data while R has memory constraints and is slower in response to large data. Therefore, the preference for using Python or R depends on the area of functionality and usage.

***Revise Python vs R to frame the answer of this data science interview question***  
**Q.10 Explain ROC curve.**

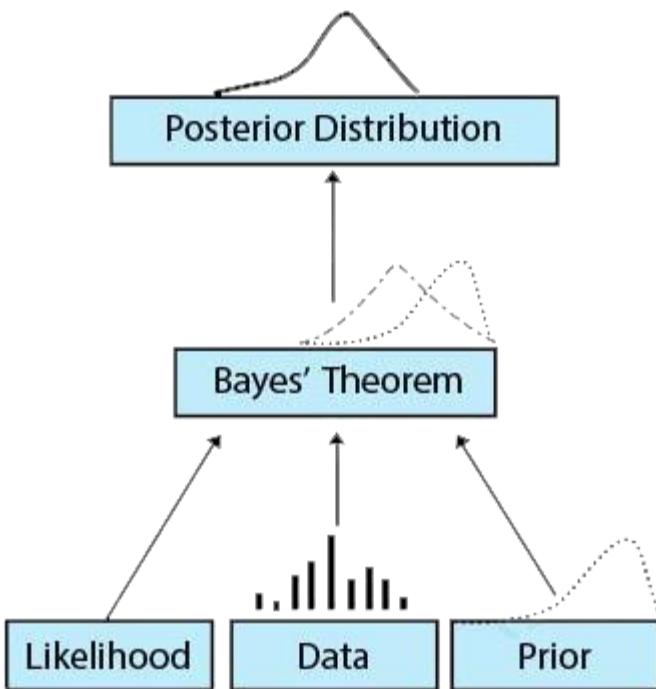
Receiver Operating Characteristic is a measurement of the True Positive Rate (TPR) against False Positive Rate (FPR). We calculate True Positive (TP) as  $TPR = TP / (TP + FN)$ . On the contrary, false positive rate is determined as  $FPR = FP / (FP + TN)$  where where TP = true positive, TN = true negative, FP = false positive, FN = false negative.

**Q.11 How is AUC different from ROC?**

AUC curve is a measurement of precision against the recall. Precision =  $TP / (TP + FP)$  and  $TP / (TP + FN)$ . This is in contrast with ROC that measures and plots True Positive against False positive rate.

**Q.12 Why is Naive Bayes referred to as Naive?**

Ans. In *Naive Bayes*, the assumptions and probabilities that are computed of the features are independent of each other. It is the assumption of feature independence that makes Naive Bayes, “Naive”.



### **Q.13 How will you create a series from a given list in Pandas?**

We will use the list to the Series() function.

```
ser1 = pd.Series(mylist)
```

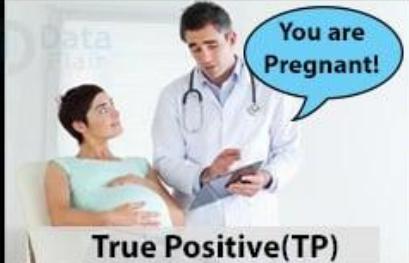
### **Q.14 Explain bias, variance tradeoff.**

Bias leads to a phenomenon called underfitting. This is caused by the introduction of error due to the oversimplification of the model. On the contrary, variance occurs due to complexity in the machine learning algorithm. In variance, the model also learns noise and other distortions that affect the overall performance of it. If you increase the complexity of your model, then the error will go down due to reduction in bias. However, after a certain point, the error will increase due to increasing complexity and addition of noise. This is known as bias-variance tradeoff. A good machine learning algorithm should possess low bias and low variance.

### **Q.15 What is a confusion matrix?**

A confusion matrix is a table that delineates the performance of a supervised learning algorithm. It provides a summary of prediction results on a classification problem. With the help of confusion matrix, you can not only find the errors made by the predictor but also the type of errors.

## Type I and Type II Errors

|                        | Actually Pregnant                                                                                                  | Actually Not Pregnant                                                                                               |
|------------------------|--------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| Predicted Pregnant     |  <p><b>True Positive(TP)</b></p>  |  <p><b>False Positive(FP)</b></p> |
| Predicted Not Pregnant |  <p><b>False Negative(FN)</b></p> |  <p><b>True Negative(TN)</b></p>  |

## Confusion Matrix

**Q.16 What is SVM? Can you name some kernels used in SVM?**

**SVM stands for support vector machine.** They are used for classification and prediction tasks. SVM consists of a separating plane that discriminates between the two classes of variables. This separating plane is known as hyperplane.

Some of the kernels used in SVM are –

- Polynomial Kernel
- Gaussian Kernel
- Laplace RBF Kernel
- Sigmoid Kernel
- Hyperbolic Kernel

**Support Vector Machine – Important topic for data science interview**

**Q.17 How is Deep Learning different from Machine Learning?**

Deep Learning is an extension of Machine Learning. It is a special area within ML that is about developing algorithms that simulate human nervous system.

Deep Learning involves neural networks which are trained over large datasets to understand the patterns and then perform classification and prediction. ***Check out the detailed comparison of Deep Learning vs Machine Learning in easy steps***

## Deep Learning Vs Machine Learning

| Factors               | Deep Learning                           | Machine Learning            |
|-----------------------|-----------------------------------------|-----------------------------|
| Data Requirement      | Requires large data                     | Can train on lesser data    |
| Accuracy              | Provides high accuracy                  | Gives lesser accuracy       |
| Training Time         | Takes longer to train                   | Takes less time to train    |
| Hardware Dependency   | Requires GPU to train properly          | Trains on CPU               |
| Hyperparameter Tuning | Can be tuned in various different ways. | Limited tuning capabilities |

### Q.18 How can you compute significance using p-value?

After a hypothesis test is conducted, we compute the significance of the results. The p-value is present between 0 and 1. If the p-value is less than 0.05, then it means that we cannot reject the null hypothesis. However, if it is greater than 0.05, then we reject the null hypothesis.

### Q.19 Why don't gradient descent methods always converge to the same point?

This is because, in some cases, they reach to local or local optima point. The methods don't always achieve global minima. This is also dependent on the data, the descent rate and origin point of descent.

### Q.20 Explain A/B testing.

To perform a hypothesis testing of a randomized experiment with two variables A and B, we make use of A/B testing. A/B testing is used to optimize web-pages based on user preferences where small changes are added to web-pages that are delivered to a sample of users. Based on their reaction to the web-page and reaction of the rest of the audience to the original page, we can carry out this statistical experiment.

### Q.21 What is box cox transformation?

In order to transform the response variable so that the data meets its required assumptions, we make use of Box Cox Transformation. With the help of this technique, we can transform non-normal dependent variables into normal shapes. We can apply a broader number of tests with the help of this transformation.

### Q.22 What is meant by ‘curse of dimensionality’? How can we solve it?

While analyzing the dataset, there are instances where the number of variables or columns are in excess. However, we are required to only extract significant

variables from the group. For example, consider that there are a thousand features. However, we only need to extract handful of significant features. This problem of having numerous features where we only need a few is called ‘curse of dimensionality’.

There are various algorithms for dimensionality reduction like PCA (Principal Component Analysis).

### **Q.23 What is the difference between recall and precision?**

Recall is the fraction of instances that have been classified as true. On the contrary, precision is a measure of weighing instances that are actually true. While recall is an approximation, precision is a true value that represents factual knowledge.

### **Q.24 What is pickle module in Python?**

For serializing and de-serializing an object in Python, we make use of pickle module. In order to save this object on drive, we make use of pickle. It converts an object structure into character stream.

*Learn everything about Pickle module in Python*

### **Q.25 What are the different forms of joins in a table?**

Some of the different joins in a table are –

- Inner Join
- Left Join
- Outer Join
- Full Join
- Self Join
- Cartesian Join

### **Q.26 List differences between DELETE and TRUNCATE commands.**

DELETE command is used in conjunction with WHERE clause to delete some rows from the table. This action can be rolled back.

However, TRUNCATE is used to delete all the rows of a table and this action cannot be rolled back.

### **Q.27 Can you tell some clauses used in SQL?**

Some of the commonly used *clauses in SQL* are –

- WHERE
- GROUP BY
- ORDER BY
- USING

### **Q.28 How will you get second highest salary of an employee emp from employee\_table?**

In order to get the second highest salary of an employee, we will use the following query –

```
SELECT TOP 1 salary  
FROM(  
SELECT TOP 2 salary  
FROM employee_table  
ORDER BY salary DESC) AS emp  
ORDER BY salary ASC;
```

*According to many data scientist, this question is considered as the most asked data science interview question.*

### **Q.29 What is a foreign key?**

A foreign key is a special key that belongs to one table and can be used as a primary key of another table. In order to create a relationship between the two tables, we reference the foreign key with the primary key of the other table.

### **Q.30 What do you mean by Data Integrity?**

With data integrity, we can define the accuracy as well as the consistency of the data. This integrity is to be ensured over the entire life-cycle.

### **Q.31 How is SQL different from NoSQL?**

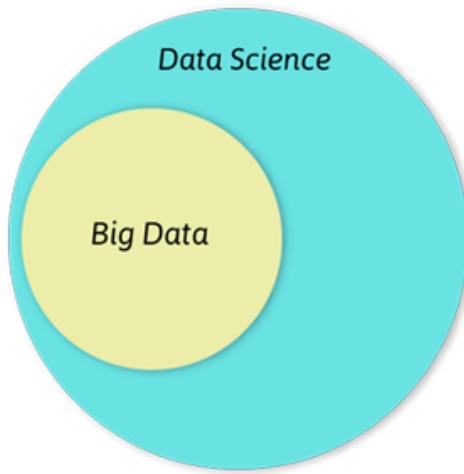
SQL deals with *Relational Database Management Systems* or RDBMS. This type of database stores structured data that is organized in rows and columns, that is, in a table. However, NoSQL is a query language that deals with Non-Relational Database Management Systems. The data present here is unstructured. Structured data is mostly generated from services, gadgets and software systems. However, unstructured data, which is increasing day by day, is generated from users directly.

### **Q.32 Can you tell me about some NoSQL databases?**

Some of the popular NoSQL databases are Redis, MongoDB, Cassandra, HBase, Neo4j etc.

### **Q.33 How is Hadoop used in Data Science?**

Hadoop provides the data scientists the ability to deal with large scale unstructured data. Furthermore, various new extensions of Hadoop like Mahout and PIG provide various features to analyze and implement machine learning algorithms on large scale data. This makes Hadoop a comprehensive system that is capable of handling all forms of data, making it an ideal suite for data scientists.



### **Q.34 How can you select an ideal value of K for K-means clustering?**

There are several methods like the elbow method and kernel method to find the number of centroids in the given cluster. However, to ascertain an approximate number of centroids quickly, we can also take the square root of the number of data points divided by two. While this technique is not entirely accurate but is fast as compared to the previously mentioned techniques.

### **Q.35 Define underfitting and overfitting.**

Most statistics and ML projects need to fit a model on training data to be able to create predictions. There can be two problems while fitting a model- overfitting and underfitting.

- Overfitting is when a model has random error/noise and not the expected relationship. If a model has a large number of parameters or is too complex, there can be overfitting. This leads to bad performance because minor changes to training data highly changes the model's result.
- Underfitting is when a model is not able to understand the trends in the data. This can happen if you try to fit a linear model to non-linear data. This also results in bad performance.

### **Q.36 What are univariate, bivariate and multivariate analysis?**

Three types of analysis are univariate, bivariate and multivariate.

- Univariate analysis includes descriptive statistical analysis techniques which you can differentiate on the basis of how many variables are involved. Some pie charts can have a single variable.
- Bivariate analysis explains the difference between two variables at one time. This can be analyzing sale volume and spending volume using a scatterplot.
- Multivariate analysis has more than two variables and explains effects of variables on responses.

**Q1:** What are the steps involved in pre-processing text data?

Let's say we have a collection of English text documents (often referred to as a **corpus**). Here are some of the common methods to pre-process the text data before modeling:

1. Change everything into lower case, e.g., **Everyday** => everyday;
2. Tokenization, e.g., we can split words using space in English to extract every single word; we can also extract every two continuous words (bigram) or even more (less often).
3. Remove punctuations and stopwords, e.g., 'a, an, the, @, !';
4. Remove rare tokens, tokens that appear in less than **1%** of the corpus, for example;
5. Remove tokens that appear too often, e.g., more than 90% of the documents;
6. Apply **stemming**: e.g., playing, played, plays => play;
7. Apply **lemmatization**: e.g., win, won => win;
8. For all unique tokens, we create a word dictionary;
9. For a traditional **bag of words** model, we can then count the frequency of each token's appearance in the document and represent it with a word-frequency count vector, where each dimension is a unique word from the dictionary, or we can use an algorithm such as **TF-IDF** instead of simply counting a word.
10. Nowadays, with the popularity of the word2vec method, we can also use a pre-trained word2vec model and look for the word embeddings to represent each unique token and average them out to represent the document. Of course, you can pre-train your own word2vec model instead of using an existing one.

Q2: What does word embedding mean?

Word embedding is just a fancy way to say transforming a word into numerical representations.

Q3: How to get the word embedding from a Word2Vec model?

The word embedding for each word is its corresponding embedding layer.

A sentence embedding can be represented by taking the average of its word embeddings.

Q4: Explain the Word2Vec model and how to train a Word2Vec model?

The word2vec algorithm uses a shallow neural network model to find semantic representations of words based on their co-occurrences in a sentence.

After the neural network is trained, the corresponding embedding layers are extracted for each word in a low dimensional space, where related words are relatively closer than unrelated words.

For example: the ***king*** will be closer to the ***queen*** compared to the ***king*** with the ***apple***.

There are two ways to train a Word2Vec model:

1. **CBOW** (continuous bag of words). The goal is to predict a word based on its surrounding words.
2. **Skip-gram**, where a word is used as an input to predict its neighboring words.

The Word2Vec model is a very shallow neural network:

- The words are first represented as a one-hot encoding using a dictionary, they are then passed as an input to an embedding layer (initialized with some random weights).
- The word embeddings are then passed to another layer where we average out the word embeddings.
- We then pass those outputs to a dense SoftMax layer for outputs.

The corresponding embedding layer will be the word embeddings we can then use for further tasks, such as computing the similarity of two words/texts.

Q5: What are some of the common methods to compute the similarities between two texts?

Computing similarity between two texts is one of the most common use cases in NLP and information retrieval.

There are two groups of similar metrics that are being used:

1. Traditional lexical similarity that is based on some form of overlappings between two texts.

In this case, we can use cosine similarity (after converting them into a word frequency vector) or Jaccard similarity.

2. Most recently semantically based, we can compare the meanings of two texts.

For embedding-based methods, the most common method is first to convert two texts into a vector, then use cosine to determine their similarity

*1. Briefly explain what the BERT model is.*

Answer: BERT, which stands for Bidirectional Encoder Representations from Transformers, is a language representation model that aims at tackling various NLP tasks, such as question answering, language inference, and text summarization.

*2. How is the BERT model different from other language representation models?*

Answer: The BERT model pretrains deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. The BERT model is pretrained on two relatively generic tasks, Masked Language Modelling (MLM), and Next Sentence Prediction. After the pretraining process, which happens to be very computationally expensive, the BERT model can be fine tuned with much fewer resources (just one extra output layer) in order to create state-of-the-art models that are fit to tackle a variety of NLP tasks, which is not often seen in other language representation models.

*3. How do BERT's results on different NLP tasks compare to those of the best models that came before it?*

Answer: BERT obtains state-of-the-art results on *eleven* different NLP tasks, and in some areas, produces not just incrementally, but significantly better results. For example, it pushes the GLUE score to 80.5%, which is a massive 7.7% absolute improvement when compared to previous state-of-the-art models.

*4. What are downstream tasks in the context of NLP?*

Answer: In NLP, downstream tasks represent the problem that we actually wish to solve. This is noteworthy since a lot of steps in NLP tasks, such as the pretraining of models using general data, do not necessarily represent the problem that we wish to solve. The pretraining phase simply allows our model to learn. The fine-tuning phase, where we use relevant data on our pre-trained model in order to obtain meaningful results, is an example of a downstream task.

*5. What are the two existing strategies for applying pre-trained language representations to downstream tasks?*

Answer: The two strategies for applying pre-trained language representations to downstream tasks are: fine-tuning, and feature extraction. In fine-tuning, we take a pretrained model and update all of the model's parameters in order for it to be able to perform the task that we want it to. The feature extraction approach uses task-specific architectures that include the pre-trained representations as additional features.

*6. How is BERT pretrained?*

Answer: BERT is pretrained on Masked Language Modelling (MLM) and Next Sentence Prediction. The masked language model randomly masks some of the tokens from the input, and our objective is to accurately predict what these masked tokens are, based only on their context. This rids us of the unidirectionality constraint, as the MLM objective enables the representation to fuse the left and the right context, which allows us to pretrain a deep bidirectional transformer. The Next Sentence Prediction task trains our model to predict whether or not a sentence comes after another sentence by studying the longer-term dependencies between sentences.

*7. What are context-free models?*

Answer: Context-free models, such as GloVe, generate a single word embedding representation for each word in the vocabulary. This means that certain words that have multiple meanings, will have the same inherent representation in very different sentences. For example, 'parks' will have the same context-free representation in 'he parks his car' and 'she likes going to water parks'.

*8. What are contextual models?*

Answer: Contextual models generate a representation of each word based on the all of the surrounding words in a sentence. This has many implications, one of which makes it possible for words that have multiple meanings, or homonyms, to be consistently interpreted correctly in contextual models, since the model interprets each word based on the context of its use.

*9. What are the different types of contextual models? Is BERT a contextual model or a context-free model?*

Answer: The different types of contextual models are:

1. Unidirectional Contextual Models

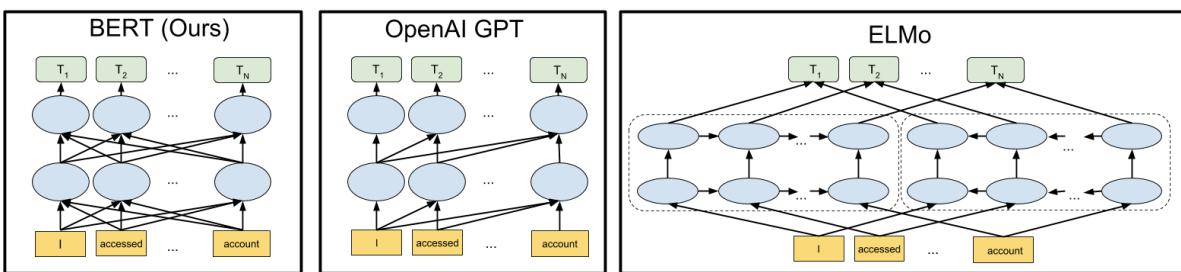
## 2. Bidirectional Contextual Models.

Unidirectional Contextual Models represent words based on all of their *previous context*. This means that each word is contextualized with respect to only the words before it, and not after it.

For example:

If we consider the sentence- 'he went to the bank to withdraw money', the word 'bank' is contextualized based on the sequence of words- 'he went to the', and *not* 'to withdraw money'. As we can see, this can lead to important information being omitted. Bidirectional Contextual Models, however, tackle this issue by contextualizing words with respect to their previous context as well as next context. Thus, if we consider the above sentence once again, a Bidirectional Contextual Model would represent 'bank' based on 'he went to the....to withdraw money'.

BERT is a Bidirectional Contextual Model, and this is why it is considered a deeply bidirectional model.



## 10. What are Transformers?

Answer: Transformers are a type of neural network architecture. They avoid using the principle of recurrence, and work entirely on an attention mechanism to draw global dependencies between the input and the output. Transformers allow for much more parallelization than sequential models, and can achieve very high translation quality even after being trained only for short periods of time. They can also be trained on very large amounts of data without as much difficulty. Transformer models, such as BERT, are used for various Natural Language Processing (NLP) tasks, such as paraphrasing, summarization, and abstractive question answering.

## 11. What is Self-Attention in the context of Transformers?

Answer: Traditional Attention is something that was used in combination with Recurrent Neural Networks (RNNs) in order to improve their performance while carrying out Natural Language Processing (NLP) tasks. Self-Attention, on the other hand, was introduced as a replacement to RNNs entirely. Instead of RNNs being used in the encoder and decoder networks, Attention methods were adopted instead, and these methods happen to be much faster overall.

*12. Highlight the properties of a Transformer's encoder and decoder.*

Answer: Encoders are Bidirectional Self-Attentive models. This means that each token in a sentence is attended to after taking into consideration all of its surrounding tokens.

$$\max_{\theta} \log p_{\theta}(\bar{\mathbf{x}} \mid \hat{\mathbf{x}}) \approx \sum_{t=1}^T m_t \log p_{\theta}(x_t \mid \hat{\mathbf{x}})$$

Decoders, on the other hand, are Unidirectional Self-Attentive models. Here, only the tokens that appear before a particular token are considered while attending to that token.

$$\max_{\theta} \log p_{\theta}(\mathbf{x}) = \sum_{t=1}^T \log p_{\theta}(x_t \mid \mathbf{x}_{<t})$$

*13. What part of BERT's architecture gives it Bidirectionality?*

Answer: The Transformer Encoder, which is a Bidirectional Self-Attentive model, gives BERT its Bidirectionality. This is because every token in a sentence is attended to after considering its context with respect to every other token in that sentence.

*14. What is BART? How is it different from BERT?*

Answer: Bidirectional Auto Regressive Transformer, or BART, is a sequence-to-sequence de-noising auto encoder. Its architecture comprises of a Bidirectional Encoder (such as in BERT), which uses more complicated masking mechanisms during the pretraining phase, as opposed to BERT's simpler token masking approach. It also uses an Auto Regressive decoder (such as in the GPT-2 model).

*15. Explain the Masked Language Model (MLM) task.*

Answer: Masked Language Model (MLM) represents the task of inputting a sentence to our model, such that it can optimize its weights and output the same sentence. The only catch is that we mask some of the tokens in the sentence. This process allows BERT to contextualize the remaining words in the sentence to try and uncover the mask, essentially filling the blank. This task improves the model's ability to comprehend the style and stylistic patterns of the language being used.

*16. Explain the Next Sentence Prediction (NSP) task.*

Answer: Next Sentence Prediction (NSP) involves giving BERT two sentences, say sentence 1 and sentence 2, and asking it whether sentence 2 comes after sentence 1. If the two sentences have very few logical similarities, then sentence 2 will most likely not follow sentence 1. MLM and NSL are the two tasks that BERT is trained on.

*17. Which BERT process is more computationally expensive: pretraining or fine tuning?*

Answer: Compared to pretraining, fine tuning is relatively inexpensive, as the transformer's self attention mechanism allows for BERT to model many downstream tasks. The entire process can take place in as little as a few hours on a standard GPU.

*18. Briefly discuss the text corpora that were used to pretrain BERT.*

Answer: BooksCorpus, which contains close to 800 million words, and the English Wikipedia, which contains over 3 billion words, were used to pretrain BERT. These corpora were chosen because they contain long and contiguous sentences, as opposed to shuffled sentences.

*19. What is Natural Language Understanding (NLU)?*

Answer: Natural Language Understanding (NLU) is a domain of Artificial Intelligence and Machine Learning that makes it possible for software systems to comprehend input (text or speech) in the form of sentences. This allows for humans to interact with computers without needing to use formalized syntaxes.

*20. What is GLUE in the context of BERT?*

Answer: General Language Understanding Evaluation (GLUE), is a collection of a broad range of Natural Language Understanding (NLU) tasks. Models such as BERT are evaluated on these tasks in order to determine how well they perform in different scenarios. BERT performs remarkably well in these tasks.

*21. What are some GLUE tasks?*

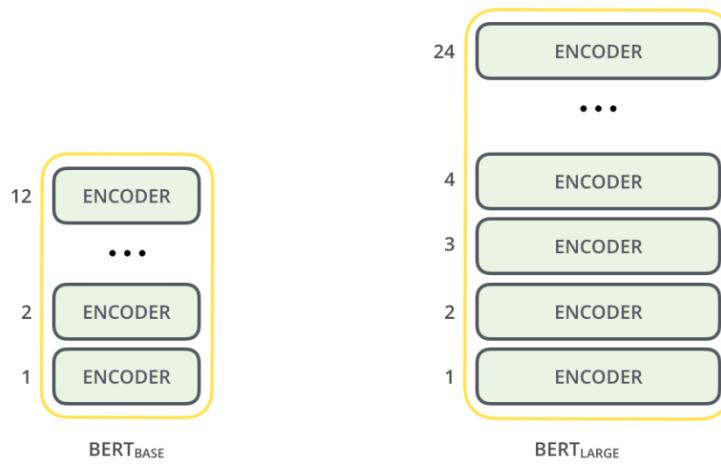
Answer: Some GLUE tasks are-

- CoLA (Corpus of Linguistic Acceptability): Determining whether or not a sentence is grammatically correct.
- SST-2 (Stanford Sentiment Treebank): Predicting the sentiment of a sentence.

- MRPC (Microsoft Research Paraphrase Corpus): Determining whether or not two sentences are semantically equivalent.
- QQP (Quora Question Pairs2): Determining whether or not two questions are semantically equivalent.

## 22. What is the difference between BERT Base and BERT Large?

Answer: The difference between BERT Base and BERT Large lies in the number of encoder layers that they possess. BERT Base contains 12 encoder layers stacked one on top of the other, whereas BERT Large contains 24 encoder layers stacked one on top of the other. This difference makes BERT Large the better performer, but it also the need for more computational power as well as memory.



## 23. What is RoBERTa?

Answer: A Robustly Optimized BERT Pretraining Approach, or RoBERTa, was worked on by Facebook AI and is a slightly modified version of BERT with better optimization for pretraining. While RoBERTa's architecture is almost the same as BERT's architecture, a few notable changes are as follows:

1. The Next Sentence Prediction (NSP) objective was removed from consideration, since the original paper's authors observed equal or slightly better performance on downstream tasks with the NSP loss removed.
2. The authors trained the model with bigger batch sizes and longer sequences, which aids the Masked Language Modelling objective, leading to better end-task accuracy.

3. Instead of obtaining a single static mask during the preprocessing of data (BERT), the data is duplicated and masked on ten different occasions, with different masking patterns on each occasion over multiple epochs.
4. Aside from the English Wikipedia and BooksCorpus, additional datasets were used to pretrain RoBERTa. In total, RoBERTa was pretrained on 160 gigabytes of uncompressed text!

All of these changes brought about an increase in overall performance and efficiency.

#### *24. What is Knowledge Distillation?*

Answer: Knowledge Distillation is the process of training a smaller model to replicate the results produced by a much larger model. This is done to reduce pretraining time.

#### *25. What is DistilBERT?*

Answer: DistilBERT is a compressed (or distilled) version of BERT, that performs very similarly to BERT Base (95% of the performance is retained when evaluated using GLUE), but has HALF as many parameters. A few techniques that were used to pretrain RoBERTa (such as removing the NSP objective and using much larger batch sizes/sequences) are also used to pretrain DistilBERT, thus making it all the more efficient. This method of compression can be applied to other models, such as GPT2, as well!

#### *26. What is ALBERT?*

Answer: A Lite BERT, or ALBERT, is an upgrade to BERT, which boasts optimized performance with significantly fewer parameters. A few notable changes that were made while designing ALBERT are as follows:

1. Most language models, including BERT, have different parameters for different layers. This is not ideal, since it causes the model to learn similar operations at different layers, using different parameters. ALBERT, however, uses parameter sharing. Thus, the same layer is applied one on top of the other. This leads to a slight reduction in accuracy, but also results in a significant decrease in the number of parameters required for ideal functionality.
2. While Next Sentence Prediction (NSP) loss was used to pretrain BERT, a more complicated loss measure known as Sentence Order Prediction (SOP), which is based on the coherence of sentences, is used in ALBERT.

3. In BERT, the input layer embeddings and the hidden layer embeddings have the same size. In ALBERT, however, the embedding matrix is split in a way such that the input layer has relatively lower dimensionality, and the hidden layer has relatively higher dimensionality. This factorization of the embedding matrix brings about an 80% drop in the number of parameters, with only a small drop in performance.

27. *What are some alternatives to BERT that can be used for Natural Language Understanding tasks?*

Answer: Some alternatives to BERT are:

1. GPT-3 by OpenAI: GPT-3 is a large scale transformer based model, and performs incredibly well at tasks such as translation and question answering. Having been trained on 175 billion parameters, it is over four hundred times bigger than in size than even BERT Large. One major drawback, however, is that while BERT is open sourced, GPT-3 is not.
2. Text-To-Text Transfer Transformer (T5): T5 reframes all NLP tasks into a unified format wherein the input and output are always text strings, as opposed to BERT. It excels at various Natural Language Processing (NLP) tasks, such as translation, question answering and classification.
3. XLNet: XLNet is an extension of TransformerXL, pretrained using an auto regressive method. This model is particularly good at tasks involving long contexts. Under suitable circumstances, it outperforms BERT in twenty tasks, often by a notable margin.

1. **Name two popular applications of Natural Language Processing.**

NLP has many real-life applications, two of the most popular ones are:

- **Chatbots:** Companies have begun to use chatbots for 24/7 service to give better customer assistance. Customers' fundamental questions are answered via chatbots. If a chatbot is unable to handle a client's query, it sends it to the support staff while continuing to engage the consumer. It gives clients the impression that the customer service team is responding fast.

Companies have been able to establish pleasant relationships with customers thanks to chatbots. Natural Language Processing is the only way to make it happen.

- **Google Translate:** One of the most well-known uses of Natural Language Processing is Google Translate. It assists in the translation of written or spoken phrases into any language. We may also use Google Translate to determine the right pronunciation and meaning of a word. It achieves success in translating sentences into multiple languages by employing sophisticated Natural Language Processing methods.

## 2. What is NLTK?

Natural Language Toolkit (NLTK) is a python library that processes natural language and extracts data from it for computers. To comprehend natural languages, we may use NLTK to perform techniques like parsing, tokenization, lemmatization, stemming, and more. It aids in text categorization, linguistic structure parsing, document analysis, and other tasks.

Some of the most common NLTK packages are DefaultTagger, UnigramTagger, treebank, wordnet, patterns, SequentialBackoffTagger, and so much more.

## 3. What does an NLP pipeline consist of?

NLP uses pipelines to understand the natural language of humans and the following are the processes of an NLP pipeline:

- Text gathering(web scraping or available datasets)
- Text cleaning(stemming, lemmatization)
- Feature generation (Bag of words)
- Embedding and sentence representation(word2vec)
- Training the model by leveraging neural nets or regression techniques
- Model evaluation
- Making adjustments to the model
- Deployment of the model.

#### **4. What is the process of feature extraction in NLP?**

In any sentence, the features are used to conduct semantic analysis or document classification. A typical paradigm for feature creation is the bag of words. A phrase can be tokenized, and then a group or category can be created from these individual words, which can then be further examined or exploited for specific features (number of times a certain word appears, etc).

Other than the bag of words, latent semantic indexing and word2vec are also popular models for feature extraction in NLP.

#### **5. What is Syntactic Analysis?**

Syntactic analysis is a method of examining sentences in order to determine their meaning. A machine can examine and comprehend the order of words in a phrase using syntactic analysis. NLP uses a language's grammar rules to aid in the syntactic analysis of word combinations and order in documents.

## 6. **What are the techniques used for syntactic analysis?**

The following diagram shows the techniques of syntactic analysis:

- **Parsing:** Parsing is used to decide the structure of the text in any document and analyze it on the basis of the grammar used.
- **Word Segmentation:** In the second step, the text is segregated into smaller units.
- **Morphological segmentation:** The goal of morphological segmentation is to deconstruct words into their simplest form.
- **Stemming:** It removes the suffix from any word to obtain the root word.
- **Lemmatization:** It allows you to mix words with suffixes without changing their meaning.

## 7. **What do you mean by LSI?**

The mathematical method of latent semantic indexing is used to increase the accuracy of the information retrieval process. Machines can identify the hidden (latent) link between meanings thanks to the design of LSI algorithms (words). Machines produce numerous ideas that are associated with the words in a phrase to improve information comprehension.

Singular value decomposition is a technique used to interpret information.

It is commonly used to manage both structured and unstructured data.

Singular value decomposition is a technique used to interpret information.

It is commonly used to manage both structured and unstructured data.

## 8. What is the significance of TF-IDF?

TFIDF (term frequency-inverse document frequency) stands for term frequency-inverse document frequency. TFIDF is a numerical statistic used in information retrieval to indicate how significant a word is to a document in a collection or group of collections.

## 9. What is Lemmatization?

Lemmatization is a term that refers to doing things correctly using vocabulary and morphological examination of words. The ends of the words are eliminated in this procedure to restore the base word, which is also known as Lemma.

As a result, the major goal of Lemmatization and stemming is to discover and return the sentence's root words in order to investigate different extra information.

## 10. What is Regular Grammar?

A regular language is represented by regular grammar. Regular grammar has rules like  $A \rightarrow a$ ,  $A \rightarrow aB$ , and many others. The rules automate the detection and analysis of strings. There are four tuples in regular grammar:

- The non-terminal set is denoted by the letter 'N.'
- The phrase " refers to a group of terminals.
- The letter 'P' stands for the production set.
- The commencement of non-terminal is indicated by the letters'  $\in N'$ .

## 11. What is the difference between regular grammar and regular expression?

Regular Grammars have four tuples  $(N, P, S \in N)$ . N indicates for the non-terminals' sets, T stands for the terminals' set, P stands for the set of productions to alter the start symbol, P has its productions from one of the kinds, and finally, S stands for the start non-terminal.

Regular expressions, on the other hand, are a set of characters that define a search pattern and are commonly used in pattern matching or string matching.

## 12. What are the terminologies in NLP?

The following are the terminologies in NLP:

### 1. Weights and Vectors

- Use of TF-IDF for information retrieval

- Length (TF-IDF and doc)
- Google Word Vectors
- Word Vectors

## *2. Structure of the Text*

- POS tagging
- Head of the sentence
- Named Entity Recognition (NER)

## *3. Sentiment Analysis*

- Knowledge of the characteristics of sentiment
- Knowledge about entities and the common dictionary available for sentiment analysis

## *4. Classification of Text*

- Supervised learning algorithm
- Training set
- Validation set
- Test set
- Features of the text
- LDA

## ***5. Machine Reading***

- Removal of possible entities
- Joining with other entities
- DBpedia

## **13. What is the main difference between NLP and NLU?**

The difference between Natural Language Processing and Natural Language Understanding is as follows:

| <b>Natural Language Processing</b>                                                                     | <b>Natural Language Understanding</b>                                                              |
|--------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
| NLP is used to produce technologies that help in better communication between humans and computers.    | NLU techniques are used to solve complex programs that are related to machine understanding.       |
| NLP takes care of all the processes that are required for the interaction between computers and humans | NLU helps in converting the unorganized data into structured data, for the machines to understand. |

## **14. What is tokenization in NLP?**

The goal of natural language processing is to teach computers how to analyze huge quantities of data in natural language. In NLP, tokenization refers to the process of breaking down a text into individual tokens.

A token in the shape of the word can be imagined. A sentence is formed in the same way that a word is formed. Splitting the text into minimum units is a key step in NLP.

## **15. What is Pragmatic Analysis and Pragmatic Ambiguity?**

In NLP, pragmatic analysis is a crucial job for understanding knowledge that exists outside of a given document. The goal of using pragmatic analysis is to concentrate on a specific feature of a document or text in a language. This necessitates a thorough understanding of the real world. The pragmatic analysis helps software programs to know the true meaning of phrases and words through critical interpretation of real-world data.

Multiple descriptions of a word or a sentence are referred to as pragmatic ambiguity. When the meaning of a statement is unclear, it is called ambiguity. The meanings of the sentence's words may vary.

As a result, understanding the meaning of a sentence becomes a difficult challenge for a computer in practice. As a result, pragmatic uncertainty emerges.

### **1. How can you assess a good logistic model?**

- Likelihood Ratio Test and Pseudo R<sup>2</sup>.
- Hosmer-Lemeshow, Wald Test.
- Variable Importance, Classification Rate.
- ROC Curve, K-Fold Cross-Validation.

### **2. What are various steps involved in an analytics project?**

- Find an Interesting Topic followed by obtain and understand Data.
- Data Preparation and data modelling.
- Model Evaluation.

- Deployment and Visualization.

### **3. During analysis, how do you treat missing values?**

- Deleting Rows with missing values.
- Impute missing values for continuous variable and categorical variable.
- Other Imputation Methods.
- Using Algorithms that support missing values with prediction of missing values.

### **4. Explain about the box cox transformation in regression models**

A Box Cox transformation is a transformation of non-normal dependent variables into a normal shape. Normality is an important assumption for many statistical techniques; if your data isn't normal, applying a Box-Cox means that you are able to run a broader number of tests.

### **5. Can you use machine learning for time series analysis?**

Time series forecasting is an important area of machine learning that is often neglected. It is important because there are so many prediction problems that involve a time component. Standard definitions of time series, time series analysis, and time series forecasting.

### **6. Write a function that takes in two sorted lists and outputs a sorted list that is their union.**

- Take in the number of elements for the first list and store it in a variable.
- Take in the elements of the list one by one.
- Similarly, take in the elements for the second list also.
- Merge both the lists using the ‘+’ operator and then sort the list.
- Display the elements in the sorted list.
- Exit.

### **7. What is Regularization and what kind of problems does regularization solve?**

Overfitting is a phenomenon that occurs when a Machine Learning model is constraint to training set and not able to perform well on unseen data. Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

### **8. What is multicollinearity and how you can overcome it?**

Multicollinearity occurs when independent variables in a regression model are correlated. This correlation is a problem because independent variables should be independent. If the degree of correlation between variables is high enough, it can cause problems when you fit the model and interpret the results.

## **9. What is the curse of dimensionality?**

The curse of dimensionality refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces that do not occur in low-dimensional settings such as the three-dimensional physical space of everyday experience.

## **10. How do you decide whether your linear regression model fits the data?**

- Make sure the assumptions are satisfactorily met.
- Examine potential influential point, the change in R<sup>2</sup> and Adjusted R<sup>2</sup> statistics.
- Check necessary interaction and apply the model to another data set and check its performance.

## **11. What is Data Science?**

Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data.

## **12. What is the Law of Large Numbers?**

The law of large numbers is a theorem from probability and statistics that suggests that the average result from repeating an experiment multiple times will better approximate the true or expected underlying result. All sample observations for an experiment are drawn from an idealized population of observations.

## **13. How Machine Learning Is Deployed In Real World Scenarios?**

Deployment is the method by which you integrate a machine learning model into an existing production environment to make practical business decisions based on data. It is one of the last stages in the machine learning life cycle and can be one of the most cumbersome.

## **14. What is collaborative filtering?**

Collaborative filtering uses a large set of data about user interactions to generate a set of recommendations. The idea behind collaborative filtering is that users

with similar evaluations of certain items will enjoy the same things both now and in the future. User preference data can also be gathered implicitly.

## **15. What are the important libraries of Python that are used in Data Science?**

- TensorFlow, NumPy.
- SciPy, Matplotlib.
- Pandas, Keras.
- SciKit-Learn, Statsmodels.

**Free PDF : Get our updated Data Science Course Content pdf**[DOWNLOAD NOW](#)

## **16. What is the difference between squared error and absolute error?**

The magnitude of the difference between the individual measurement and the true value of the quantity is called the absolute error of the measurement. The arithmetic mean of all the absolute error is taken as the mean absolute error of the value of the physical quantity.

## **17. What is Machine Learning?**

Machine-learning algorithms use statistics to find patterns in massive amounts of data. And data, here, encompasses a lot of things—numbers, words, images, clicks, what have you. If it can be digitally stored, it can be fed into a machine-learning algorithm.

## **18. How are confidence intervals constructed and how will you interpret them?**

A confidence interval displays the probability that a parameter will fall between a pair of values around the mean. Confidence intervals measure the degree of uncertainty or certainty in a sampling method. They are most often constructed using confidence levels of 95% or 99%.

## **19. How will you explain logistic regression to an economist, physician scientist and biologist?**

Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables

## **20. How can you overcome Overfitting?**

- Cross-validation and train with more data.
- Remove features and early stopping.
- Regularization, ensembling.

## **21. Differentiate between wide and tall data formats?**

Wide data has a column for each variable whereas long format data has a column for possible variable types & a column for the values of those variables.

## **22. Is Naïve Bayes bad? If yes, under what aspects.**

One of the disadvantages of Naïve-Bayes is that if you have no occurrences of a class label and a certain attribute value together then the frequency-based probability estimate will be zero. And this will get a zero when all the probabilities are multiplied.

## **23. How would you develop a model to identify plagiarism?**

- Tokenize the document.
- Remove all the stop words using NLTK library.
- Use GenSim library and find the most relevant words, line by line. This can be done by creating the LDA or LSA of the document.
- Use Google Search API to search for those words.

## **24. How will you define the number of clusters in a clustering algorithm?**

The optimal number of clusters can be defined as follow: Compute clustering algorithm (e.g., k-means clustering) for different values of k. For each k, calculate the total within-cluster sum of square (wss). Plot the curve of wss according to the number of clusters k.

## **25. Is it possible to perform logistic regression with Microsoft Excel?**

To activate the Logistic regression dialog box, start XLSTAT then select the XLSTAT / Modeling data / Logistic regression function. When you click on the button, the Logistic regression dialog box appears. Select the data on the Excel sheet.

## **26. What are Eigenvalue and Eigenvector?**

Eigenvectors are unit vectors which mean that their length or magnitude is equal to 1. They are often referred to as right vectors which simply mean a column vector whereas eigenvalues are coefficients applied to eigenvector give the vectors their length or magnitude.

## **27. Compare Sas, R, And Python Programming?**

All big IT organizations choose SAS as their data analytics tools. As R is very good with heavy calculations, it is largely used by statisticians and researchers. Startups prefer Python over the other two due to its lightweight nature, large community, and deep learning capabilities.

## **28. How regularly must an algorithm be updated?**

Algorithm can be updated regularly based on its need, usage, and market growth. For example, Google is reported to change its search algorithm around 500 to 600 times each year.

## **29. What is the goal of A/B Testing?**

The goal of A/B testing is to find the best performing content for a specific goal (or goals). Choosing the goal of your test should be part of your test development process. Most marketers focus on improving one of a few different key performance indicators (KPIs)

## **30. What are the feature vectors?**

The feature vector is an n-dimensional vector of numerical features that represent some object. Many algorithms in machine learning require a numerical representation of objects, since such representations facilitate processing and statistical analysis.

