### ****1. What do you understand by Natural Language Processing?****

**Natural Language Processing (NLP)** is a subfield of Artificial Intelligence (AI) that focuses on the interaction between computers and human (natural) languages. It enables machines to **read, understand, interpret, and generate human language** in a way that is both meaningful and useful.

# Common applications of NLP include:

Chatbots and virtual assistants (like Siri or ChatGPT)

Machine translation (e.g., Google Translate)

Sentiment analysis (e.g., analyzing reviews)

Speech recognition

Text summarization

### ****2. What are the steps involved in solving an NLP problem?****

The typical pipeline to solve an NLP problem includes the following steps:

**Text Collection** – Gather raw text data from documents, websites, etc.

**Text Preprocessing** – Clean and prepare the data:

Tokenization

Lowercasing

Removing punctuation, stop words

Lemmatization or stemming

**Feature Extraction** – Convert text into numerical representations:

Bag of Words (BoW)

TF-IDF (Term Frequency-Inverse Document Frequency)

Word embeddings (Word2Vec, GloVe)

**Model Building** – Use machine learning or deep learning models:

Logistic Regression, Naive Bayes, SVM

RNNs, LSTMs, Transformers

**Evaluation** – Measure model performance:

Accuracy, Precision, Recall, F1 Score

**Deployment** – Integrate the model into real-world applications.

### ****3. What is an ensemble method in NLP? With Example.****

**Ensemble methods** in NLP involve combining predictions from multiple models to improve performance and generalization. The idea is that multiple weak learners (models) when combined can produce a stronger prediction.

# **Types of ensemble methods:**

**Bagging** (e.g., Random Forest)

**Boosting** (e.g., XGBoost)

**Voting** (majority or weighted)

**Stacking** (model outputs fed into a meta-model)

**Example in NLP:**  
Suppose you're doing **sentiment analysis** on customer reviews. You can build:

Model 1: Logistic Regression

Model 2: Naive Bayes

Model 3: SVM

Now, use **majority voting** to predict sentiment:

If 2 out of 3 models say "Positive", the final ensemble result is "Positive".

✅ Ensemble methods **reduce variance, avoid overfitting**, and often yield **more accurate** results than a single model.