# Natural language process

**Module 1: Introduction to NLP**

- What is NLP?
- Applications and real-world use cases
- Challenges in NLP (ambiguity, context, sarcasm, etc.)
- Overview of NLP pipelines (from text to insights)
- Brief history and evolution (rule-based → ML → deep learning)

**Module 2: Text Preprocessing**

- Tokenization (word, sentence, subword)
- Text normalization (lowercasing, stemming, lemmatization)
- Stop word removal
- Regular expressions in text processing
- N-grams and sliding window techniques
- Handling special characters, emojis, spelling errors

**Module 3: Text Representation**

- Bag-of-Words (BoW)
- Term Frequency–Inverse Document Frequency (TF-IDF)
- One-hot encoding
- Word embeddings:
    - Word2Vec (CBOW, Skip-gram)
    - GloVe
    - FastText
- Contextual embeddings:
    - ELMo
    - BERT and Transformer-based embeddings

**Module 4: Syntax and Structure**

- Parts of Speech (POS) tagging
- Chunking and parsing (shallow parsing, constituency & dependency parsing)
- Named Entity Recognition (NER)
- Syntax trees and grammar parsing
- Syntax-based feature extraction

**Module 5: Semantic Analysis**

- Word sense disambiguation
- Coreference resolution
- Semantic role labeling
- Topic modeling (LDA, NMF)
- Semantic similarity (cosine, Euclidean, etc.)
- Knowledge graphs and ontologies (brief intro)

**Module 6: Text Classification and Sentiment Analysis**

- Sentiment analysis techniques
- Rule-based vs ML-based classifiers
- Naive Bayes, Logistic Regression, SVM for NLP
- Deep learning for classification (CNNs, RNNs)
- Evaluation metrics: Precision, Recall, F1, Confusion matrix

**Module 7: Sequence Modeling**

- Introduction to sequential data
- Recurrent Neural Networks (RNNs)
- Long Short-Term Memory (LSTM) and GRU
- Attention mechanisms
- Bidirectional RNNs
- Applications: POS tagging, NER, translation, etc.

**Module 8: Transformer Models**

- Introduction to Transformers
- Self-attention and multi-head attention
- Encoder-decoder architecture
- Pretrained models: BERT, GPT, RoBERTa, T5, XLNet
- Fine-tuning transformer models for NLP tasks

**Module 9: NLP Applications**

- Question answering systems
- Machine translation (Seq2Seq, Transformer-based)
- Summarization (extractive and abstractive)
- Text generation (GPT-based)
- Chatbots and conversational agents
- Speech-to-text and text-to-speech (intro)

**Module 10: Advanced NLP Topics**

- Transfer learning in NLP
- Zero-shot and few-shot learning
- Multilingual NLP
- Prompt engineering for LLMs
- Retrieval-Augmented Generation (RAG)
- Ethics in NLP: bias, fairness, toxicity detection

**Module 11: Practical Tools & Libraries**

- **NLTK**, **spaCy**, **TextBlob** – for traditional NLP
- **scikit-learn** – for ML-based pipelines
- **Gensim** – for topic modeling & word embeddings
- **Transformers (Hugging Face)** – for modern deep learning models

- **Flair**, **Stanza**, **OpenNLP**
- **Streamlit/Gradio** – for building NLP app frontends

**Module 12: Evaluation, Deployment & Real-world Systems**

- Evaluating NLP systems (BLEU, ROUGE, METEOR, perplexity)
- Human vs automatic evaluation
- Deploying NLP models (Flask, FastAPI, Docker)
- Building APIs for NLP tasks
- Monitoring, updating, and scaling NLP systems

==========

**END**

==========