

DATA SCIENCE

INTERMEDIATE

5 DAYS

Natural Language Processing in Data Science

Build a strong foundation in Natural Language Processing by understanding core concepts, techniques, and real world applications. Learn how to design, implement, evaluate, and integrate NLP solutions using modern libraries while customizing models to meet domain specific business needs.

Why NLP?

Unlocking Language Data

By leveraging NLP techniques, you can tap into vast amounts of textual information and transform it into actionable knowledge.

Improved Customer Experience

NLP allows you to analyse customer feedback, support tickets, social media interactions, and reviews, enabling you to understand customer sentiment, preferences, and needs.

Enhanced Business Intelligence

By leveraging NLP techniques, you can gain deeper insights from text data, uncover patterns, and detect trends. This enables your business to make data-driven decisions, identify market opportunities, perform competitive analysis, and drive innovation within your business.

What you'll be able to do

- ✓ Gain a solid understanding of the fundamental concepts, techniques, and models used in Natural Language Processing.
- ✓ Learn the various applications of NLP in real-world scenarios and industry domains.
- ✓ Gain hands-on experience in applying NLP techniques such as text preprocessing, sentiment analysis, text classification etc.
- ✓ Develop practical skills in implementing NLP solutions using popular libraries and frameworks.
- ✓ Learn how to customize and fine-tune NLP models to address their organisation's specific requirements and domain-specific challenges.
- ✓ Receive training on evaluation metrics and techniques to assess the performance and quality of NLP models
- ✓ Learn how to integrate NLP capabilities into existing business workflows, systems, or applications.

Who this is for

- Data Scientists
- Software Engineers
- Data Analysts
- Business Analysts
- Data Engineers

Curriculum

01 NLP Foundations and Text Preprocessing

- Position NLP within modern data science workflows
- Clean and normalize raw text for modeling
- Handle encoding, noise, and domain-specific tokens
- Build a reusable text preprocessing pipeline

02 Tokenization, Lemmatization, and POS Tagging

- Tokenize text for analysis and modeling
- Reduce tokens with stemming/lemmatization strategies
- Apply part-of-speech tagging for syntactic signals
- Compare preprocessing choices on downstream quality

03 Named Entity Recognition and Sequence Labeling

- Identify people, organizations, and locations in text
- Frame sequence labeling tasks and evaluation
- Handle overlapping and nested entity patterns
- **Hands-on:** Extract entities from sample documents

04 Sentiment Analysis and Text Classification

- Score sentiment for reviews, tickets, or social text
- Train text classifiers for business categories
- Choose metrics for imbalanced text labels
- Ship a baseline classifier with clear error analysis

05 Topic Modeling and Text Clustering

- Extract latent topics from document collections
- Cluster similar documents for discovery and triage
- Interpret and label topic/cluster outputs
- Use results for content organization and research

06 Word Embeddings and Language Modeling

- Represent words as dense semantic vectors
- Use embeddings for similarity and feature transfer
- Introduce language modeling for context-aware text
- Select embedding approaches for domain data

07 Machine Translation and Question Answering

- Understand MT pipelines and evaluation basics
- Build QA flows over document contexts
- Measure answer relevance and faithfulness
- Identify production constraints for MT/QA systems

08 Summarization and Information Extraction

- Generate concise summaries that preserve key facts
- Extract structured fields from unstructured text
- Validate extraction precision for business systems
- Design human-in-the-loop review where needed

09 Chatbots and Language Generation

- Design conversational flows for user intents
- Generate natural language responses safely
- Ground generation in approved knowledge sources
- Test failure modes and escalation paths

10 Speech Recognition Basics

- Convert speech to text for NLP pipelines
- Assess accuracy and domain vocabulary challenges
- Connect ASR outputs to downstream NLP tasks
- Plan privacy and consent for voice data

11 Ethics in NLP

- Identify bias, privacy, and misuse risks in NLP
- Apply mitigation and review practices
- Document limitations for stakeholders
- Define responsible deployment checkpoints