

**Practical NLP (Natural Language Processing)**

**Course Number:** PYTH-236  
**Duration:** 3 days

**Overview**

This Practical NLP (Natural Language Processing) training course teaches software engineers and data scientists how to incorporate NLP into their production systems. Participants learn the foundations for each NLP concept and explore each concept’s applicability, limitations, and implementations. Throughout this NLP course, both models and theory are taught using real-world datasets and production samples.

**Prerequisites**

Students should have some experience with machine learning. Familiarity with TensorFlow, Keras, and scikit-learn is helpful but not mandatory.

**Materials**

All NLP training students receive comprehensive courseware.

**Software Needed on Each Student PC**

* Python 3.5 or later
* TensorFlow 2.0 and Keras 2.0 or later within Google Collaboratory

**Objectives**

* Demonstrate popular NLP algorithms and their applicability and limitations
* Create production systems that process text efficiently

**Outline**

* Introduction
* Representing Text for Similarity
  + Word and Character Embeddings
  + Keras Embedding Layers
  + Word2Vec, CBOW, and Skip-Gram Architecture
  + Doc2Vec and Paragraph Vectors. PV-DM and PV-DBOW Architectures
  + Ranking documents
  + Capstone: Learning to rank
* Classification in Text
  + Neural Networks
  + Backpropagation through time
  + Capstone: Sentiment Analysis
* Generating Text for Suggestions
  + Recurrent Neural Networks
  + GRU and LSTM for memory retention
  + Character and word-level text generation
  + Autocomplete and Autosuggestions
  + Capstone: Writing like Shakespeare
* Named Entity Recognition (NER) Extracting Knowledge
  + SpaCy and NER Pipelines
  + Bidirectional LSTM networks
  + Capstone: Translating
* Language Modeling with Attention
  + Self-Attention and “Hey Siri”
  + Transformers
  + Using BERT and GPT-2
  + Knowledge Distillation of language models
  + Capstone: Language models in proteins
* Hackathon
  + Distilling BERT for Detecting Toxic Comments
* Conclusion