

NLP Crash Course

plugin²⁰²⁰

Organized by





Speaker

Dipanjan Sarkar

Data Science Lead | Google Dev Expert – ML | Author

Topic

NLP Crash Course - Workshop

About this Workshop

- 3 hours
- 10 mins Q & A
- 100% Hands-On Examples
- Code examples will be provided
- Try to understand key steps in the flow
- Reach out in the future for detailed queries
- Deep Dives will not be done due to time constraints

NLP Crash Course

Outline



- Introduction to NLP
 - Basics
 - POS Tagging
 - NER
 - Text Wrangling
- Text Representation
 - Traditional Statistical Models – BOW, TFIDF
 - Embedding Models
- NLP Applications
 - Text Similarity Content Recommenders
 - Topic Modeling
 - Text Summarization (Extractive & Abstractive)
 - Text Classification (ML, DNN, CNN, LSTM, BERT)
 - Multi-Task NLP (Transformers)

Introduction to NLP

1. NLP Basics

- Strings
- String Operations
- String Functions
- Regular Expressions

2. Parts of Speech (POS) Tagging

3. Named Entity Recognition (NER)

4. Text Wrangling

- Case Conversion
- Tokenization
- Removing HTML tags
- Removing Accented Characters
- Removing Special Characters
- Expanding Contractions
- Stemming
- Lemmatization
- Stopwords Removal

Text Representation

1. Traditional Statistical Models

- Bag of Words - TF
- Bag of N-Grams
- TF-IDF
- Document Similarity

2. Embedding Models

- Word2Vec
- GloVe
- FastText

NLP Applications

1. Text Similarity Content Recommenders

- TF-IDF Model based Similarity Features
- FastText Model based Similarity Features

2. Topic Modeling

- LDA Model
- LDA with MALLET
- Finding Optimal Topics
- Inference

3. Text Summarization

- Extractive Summarization with TextRank
- Extractive Summarization with Transformers
 - ✓ BERT
 - ✓ DistilBERT
- Abstractive Summarization with BART

NLP Applications

1. Text Classification

- Machine Learning Models
 - ✓ Logistic Regression
 - ✓ Random Forest
- Deep Neural Networks (Dense nets)
- Convolutional Neural Networks (CNNs)
- Long Short Term Memory Networks (LSTMs)
- Transformers
 - ✓ BERT
 - ✓ DistilBERT

2. Multi-Task NLP with Transformers

- Classification – Sentiment Analysis
- Named Entity Recognition
- Question – Answering
- Mask Filling \ Word Suggestion
- Summarization
- Translation
- Feature Extraction \ Vectorization

Workshop Resources:
https://bit.ly/plugin20_nlp



Questions

Just 10 Minutes

plugin²⁰²⁰

Thank You



LinkedIn

<https://www.linkedin.com/in/dipanzan>



GitHub

<https://github.com/dipanjanS>



Medium

<https://medium.com/@dipanzan.sarkar>