



# **HANDS ON - NATURAL LANGUAGE PROCESSING(DOCUMENT CLASSIFICATION)**

# SUMMARY

## 1. What is NLP?

## 2. Stages in NLP

Paragraph

Sentence

Words

## 3. Word Embedding Techniques

- ▶ Bag of Words (BOW)
- ▶ Term Frequency-Inverse Document Frequency (TF-IDF)
- ▶ Word2Vec

# SUMMARY

## 1. What is NLP?

## 2. Stages in NLP

Paragraph

Sentence

Words

- ▶ Tokenization: Break sentences into words
- ▶ Vectorization - Convert words into vectors to find relation between them

Lexical  
Analysis

Syntactic  
Analysis

Semantic  
Analysis

- ▶ Bag of Words (BOW)
- ▶ Term Frequency-Inverse Document Frequency (TF-IDF)
  - ▶ Filtration of spam mails (Document Classification)