



What is the Natural Language Processing (NLP) starter kit?

This starter kit is a concoction of all the best resources that we have developed at Analytics Vidhya on NLP through our combined years of experience in the field. In this, you will find resources including articles and free courses that we have created for you to have the best learning experience.

Why the starter kit?

Anyone with a little experience in machine learning knows that NLP is the hottest field to get into at the moment. The amount of text data being generated around the world is staggering. We cannot even comprehend the volume and velocity of this data! From your Facebook status and LinkedIn posts to online articles and product reviews, there is no shortage of text data around us.

But getting started in NLP is not an easy task. There is so much to learn in NLP from text classification, text summarization, information extraction, information retrieval, and so much more. Given such a steep learning curve for NLP, many learners have no clue where to start! That is why we have created this starter kit with one objective – To get your NLP journey started on the right note.

This kit should act as your guide to learn the fundamentals of NLP the right way and by the end of it you will get a thorough understanding of the NLP field so that you can venture out, and explore and experiment the wonderful field of NLP by yourself!

Prerequisites?

You need to be aware of the basic machine learning algorithms. If you don't feel confident in that, don't worry. We have a separate dedicated starter kit on Machine learning which you can explore before jumping on to NLP.

Who is this for?

Anyone with an experience in Machine learning.

This guide is created by actively keeping in mind its target audience. Therefore, it should be a smooth learning experience for you if you are an absolute beginner in the field of NLP. But rest assured as, although you start as a beginner, but you don't leave as one!

How to go about using this starter kit?

We have broken down this kit into several chronological steps. Each step has its own focus area and builds on the fundamental concepts taught in the previous steps.

Step 0 – Python for NLP

This will act as a masterclass on Python for NLP. The course is designed in such a way that anyone can quickly grasp the essential Python skills needed in any NLP/ML project, even if you are complete beginner in programming. Using just these skills, you will be able to implement your own ML models.

Learning objective:

- Introduction to Python
- Understanding the fundamentals like Python operators, variables and data types
- Learning conditional statements and Python data structures
- Introduction to standard libraries
- Learning to read CSV files and creating dataframes
- Manipulating and Visualizing data
- Understanding regular expressions

Module folder: Step 0 - Python for NLP

Step 1 - What is Natural Language Processing (NLP)?

We will start on the right foot by introducing the fundamentals of NLP. You will get to know what NLP really is, what it entails, the different kinds of NLP techniques that are present, and some of the applications of NLP. This will give you a broad idea about the field and will get you excited about what lies ahead.

Learning objective

- Understanding the basic definition of NLP
- Applications of NLP
- Learning basic preprocessing in NLP

Module folder: Step 1 - What is NLP

Step 2 – Learning to handle text data using Python

NLP is all about handling your text data well. Therefore, when learning NLP, your initial focus should be on learning different ways to leverage the text data. Here, you will learn about working with text data in general, and the different aspects of text data and text preprocessing.

Learning objective:

- Different text feature extraction techniques
- Basic text pre-processing techniques
- Advanced text pre-processing techniques

Module folder: Step 2 – Handling text data

Step 3 – Learning Regular Expressions

Ask any NLP expert about text data cleaning and you are bound to hear Regular Expressions as the first thing they say. Yes, it is that important! This technique is old but gold! Therefore, here you will learn about cleaning your text data using Regular Expressions. A ton of practical examples will be covered here that will make you highly proficient with their usage.

Learning objective:

- Understanding Regular Expressions
- Match specific patterns
- Searching for pattern within text
- Finding patterns from entire text
- Splitting text on specific pattern
- Substituting a text pattern with new sub-string
- Real world applications of Regular Expressions

Module folder: Step 3 – Regular Expressions

Step 4 – Text Preprocessing

Real-world text data is very unstructured. There is a lot of unnecessary information. If you feed all of this to your machine learning model, you will never achieve good results. Therefore, in this step you will learn about the key text preprocessing steps in NLP using different libraries like NLTK, SpaCy, Gensim, and more!

Learning objective:

- Tokenization
- Stop words removal
- Stemming
- Lemmatization

Module folder: Step 4 – Text Preprocessing

Step 5 – Projects

No learning is complete without a practical project. But here at Analytics Vidhya we go the extra mile, therefore we bring to you two projects on NLP! These will really solidify your understanding of concepts covered in this starter kit, and you will have something to add to your resume!

Learning objective:

- Twitter Sentiment Analysis – Free course
- Multi-class classification – Predicting Movie Genres

Module folder: Step 5 – Projects

