**Introduction To NLP**

Writng a basc Sentenc is easy Rigt? I apologize for the spelling errors I made, but this is all because the ‘autocorrect' was off 😠. Ever consider the possibility that you're speaking with a bot? over the phone, or on Telegram and Discord? If you're wondering why you weren't aware of anything Heh, heh 😁.

When it comes to checking claims of plagiarism or copyright, things have gotten considerably simpler. Visit any of the sites and use the software to quickly verify it. There are 6500 different languages spoken worldwide, so what would you do if you were a writer who wanted your work to be translated into another language?

Have you ever wondered how many people remark on different social media and e-commerce websites? Wouldn't it be better if someone or something could analyze all of these comments and give the producers a sense of the users' attitudes so that they could better understand their customers?

Let's examine this from a societal standpoint. Everyone today utilizes social media, correct? According to a survey, several effects could occur there, such as detecting nasty remarks, suicidal behavior, harsh conversations, and even crime 😮. If and only if the machine could grasp our language, everything could be kept an eye on and prevented. Are you able to come up with a solution to the aforementioned issues?

You made a good approximation. Natural Language Processing, a subfield of artificial intelligence, enables machines to comprehend "Language" and do all the tasks listed above in a "Human-Like" fashion. Well, sometimes machines are more articulate than we are 😅.

As the population grows and social media becomes more and more popular, zettabyte-scale amounts of data (also known as "Big Data") are produced (21 zeroes). As a result, it has become vital to utilize NLP to evaluate this data to better understand human behavior because it is largely present in an unstructured form (>80%). This can also assist to save a lot of time...

Some more applications of NLP :

1. Sentiment Analysis
2. Chatbot
3. Speech Recognition (Siri, Alexa)
4. Machine Translation
5. Spell Checking
6. Keyword Searching
7. Information extraction
8. Advertisement matching

After learning about the uses and significance of NLP in the modern marketplace, you must be concerned that learning this Technology will be very challenging. But don't worry, we'll make studying more enjoyable and uncomplicated in this "Practical Handbook on NLP." where you'll feel secure enough to create models of the real world at the end of it. So, let’s keep reading...

NLP is divided into 7-10 steps just like how a child would learn a language for example a paragraph.

1. Segmentation - Breaking the big paragraph into segments. (Like each with full stops)
2. Tokenizing - Take those sentences and break them into words.
3. Stop Words - Removing the non-important words like - ‘and’ ‘the’
4. Stemming - verb forms, Like we should tell the child that walks, walking, and walking all have similarities like (walk + ).
5. Lemmatization - Then understanding the base words like Am, Is, are
6. Speech Tagging - Understanding Nouns, Verbs, etc.
7. Named Entity Tagging - Also the child should know the important names of capital, animals, and state right?
8. Classification - As a Machine is still only a machine but to make it is intelligent enough to make it understand the language.

**NATURAL LANGUAGE with Python and NLTK - Tokenizing Words and Sentences**

Is there a toolkit that can teach us all the basics so we can do all these above-mentioned steps?

The answer is yes !!! .

NLTK - Natural Language Toolkit ([Documentation](https://www.nltk.org/)) is a python toolkit for Natural Language Processing. It includes everything that you need to know to get started in NLP and we will perform the steps mentioned above with it in ([Google Colab](https://colab.research.google.com/drive/1n0QaAGofS9HQEFe_VA__gGB7cv8_N2Rw?usp=sharing)).

Things to be remembered in the Google Colab →

1. Segmentation in the Colab file - Think it as the my\_text string is the segment of a bigger paragraph like a part of my whole portfolio
2. Tokens - These are the broken down words or sentences from a big paragraph into simpler language-assignable elements or comprehensible components.
3. The importance of Tokens - Tokenization because this is also used while any programming language gets executed.