Home

Chirag Goyal — **June 15, 2021**

Advanced   NLP   Project   Python   Text

This article was published as a part of the Data Science Blogathon

Introduction

Computers and Machines are great while working with tabular data or Spreadsheets. However, human beings generally communicate in words and sentences, not in the form of tables or spreadsheets, and most of the information that humans speak or write is present in an unstructured manner. So it is not very understandable for computers to interpret these languages.

Therefore, In natural language processing (NLP), our aim is to make the computer's unstructured text understandable and retrieve meaningful information from it.

Let's define Natural Language Processing (NLP) formally,

Natural language Processing (NLP) is a subfield of artificial intelligence, that involves the interactions between computers and humans.

So, In this article, we will discuss some of the basic concepts related to NLP. This article is part of a blog series on Natural Language Processing (NLP).

This is part-1 of the blog series on the Step by Step Guide to Natural Language Processing.

Important Note

After the completion of some topics, there is some practice (Test your Knowledge) questions given that you have to solve and give the answer in the comment box so that you can check your understanding of a particular topic.

Table of Contents

1. What is Natural Language Processing (NLP)?

2. Applications of Natural Language Processing

3. Understanding Natural Language Processing

4. Difference between Rule-based NLP and Statistical based NLP

5. Components of Natural Language Processing

6. Ambiguity and Uncertainty in Natural Language Processing

What is Natural Language Processing?

Natural Language Processing (NLP) is a subfield of Computer Science and Artificial Intelligence that deals with interactions between computers and human (natural) languages. This becomes crucial when we want to apply Machine Learning or Deep Learning Algorithms to a dataset that contains text and speech.

For Example, we can use NLP to create AI systems such as,

- Document Summarization,
- Machine Translation,
- Spam Detection,
- Named Entity Recognition,
- Question Answering,
- Autocomplete,
- Predictive Typing, etc.

In modern days, most of our smartphones have a speech recognition system. These smartphones use NLP to understand the natural language and give the response. Also, most of the peoples use laptops which operating system has built-in speech recognition.

Test Your Knowledge

Which of the below options is the field of Natural Language Processing?

- Computer Science
- Artificial Intelligence
- Computational linguistics
- All of the above

Applications of NLP

Some applications of Natural Langauge Processing are as follows:

Cortana

Image Source: Google Images

The Microsoft Operating system has a virtual assistant named [Cortana](#) that can recognize a natural voice. Its applications include

- Set up reminders,
- Open applications,
- Send emails to anyone,
- Play games for entertainment,
- Track flights and packages,
- Check the weather, etc.

If you want to read more about Cortana commands, refer to the link [here](#).
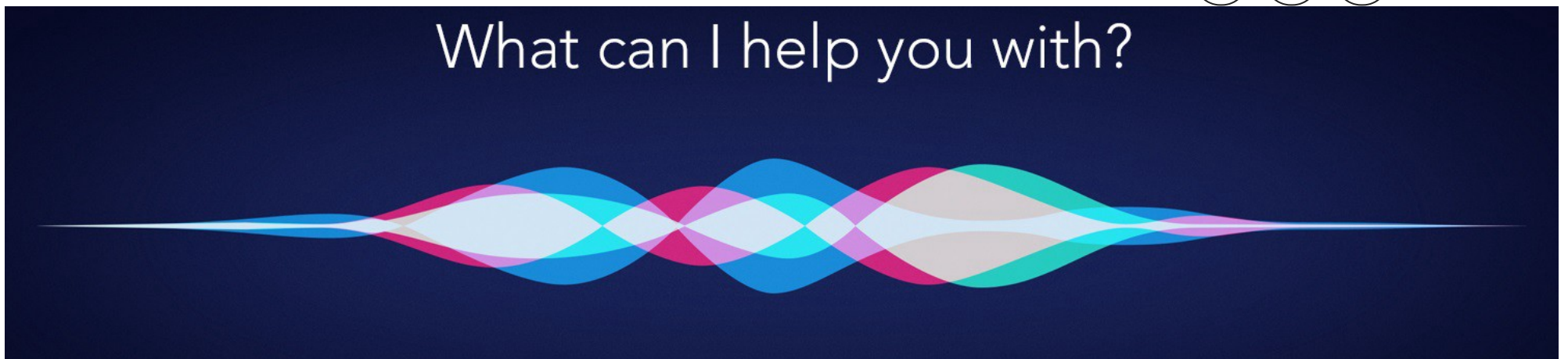
Siri

Image Source: Google Images

Siri is a virtual assistant made by Apple Inc.'s iOS, watchOS, macOS, HomePod, and tvOS operating systems. Again, with this you can do a lot of things with voice commands:

- Start a call with anyone
- Text someone
- Send an email
- Set a timer
- Take a picture
- Open an application
- Set an alarm
- Use navigation, etc.

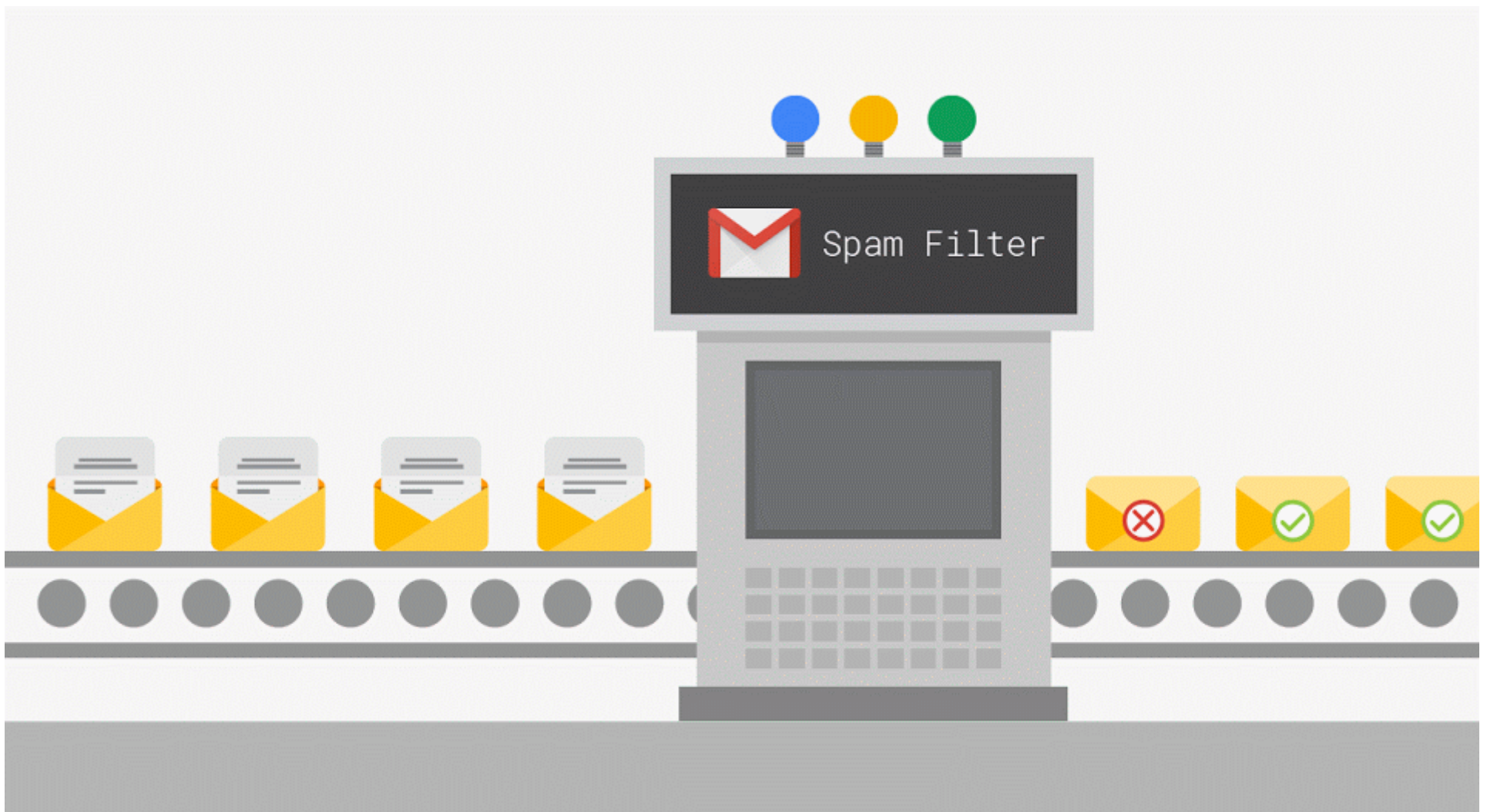[Here](#) is a complete list of all Siri commands.

Gmail

Image Source: Google Images

Gmail is the famous email service developed by Google and is using spam detection to filter out some spam emails by doing text processing, in which it fetches texts from that particular mail that it tries to find as spam or not.

Test Your Knowledge

Which of the below are NLP use cases?

- Detecting objects from an Image
- Facial Recognition
- Speech Biometric
- Text Summarization

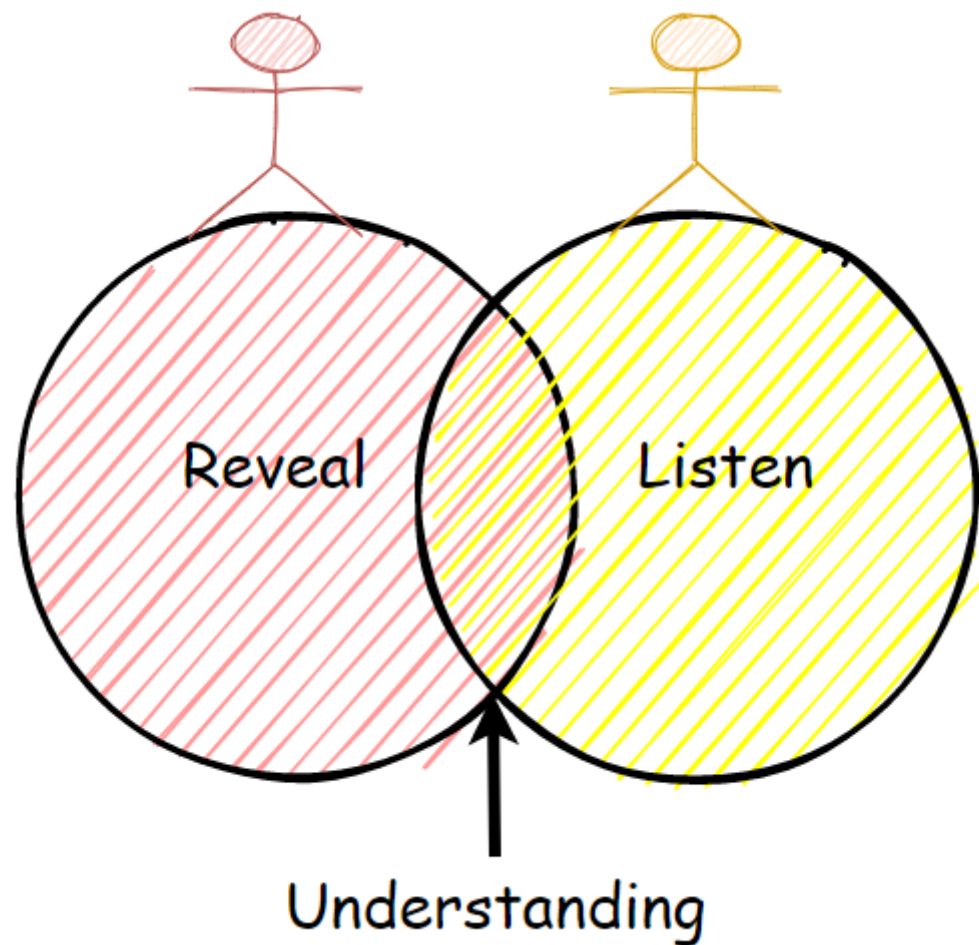# Understanding Natural Language Processing

Image Source: Google Images

We, as humans, it's not a very difficult task to perform natural language processing (NLP) but even then, we are not perfect. We often misunderstand one thing for another and often interpret the same sentences or words in a different manner.

For instance, consider the following sentences and try to understand its interpretation in many different ways:

Example 1

```
Sentence: I saw a student on a hill with a microscope.
```

These are various interpretations of the above sentence which are shown below:

- There is a student on the hill, and I watched him with my microscope.
- There is a student on the hill, and he has a microscope.
- I'm on a hill, and I saw a student using my microscope.
- I'm on a hill, and I saw a student who has a microscope.
- There is a student on a hill, and I saw him something with my microscope.

Example 2

```
Sentence: Can you help me with the can?
```

In the sentence above, we observed that there are two "can" words, but they have different meanings. Here.

The first "can" word is used to form a question.

food or liquid, etc.

What Conclusions we can infer from the above two Examples?

From the above two examples, we can observe that language processing is not "deterministic" that is the same language has the same interpretations, and something suitable to one person might not be suitable to another person. Therefore, Natural Language Processing (NLP) has a non-deterministic approach.

In simple words, we can use Natural Language Processing to create a new intelligent or AI system that can understand in the same way as that of humans and interpret the language in different situations.

Difference between Rule-based NLP and Statistical NLP

Natural Language Processing is separated into two different approaches:

Rule-based Natural Language Processing

It uses common sense reasoning for processing tasks.

For Example,

- The freezing temperature can lead to death, or
- Hot coffee can burn people's skin,
- Some other common-sense reasoning tasks, etc.

However, these process can take more amount of time, and it requires manual effort.

Statistical Natural Language Processing

This type of NLP uses large amounts of data and aims to derive conclusions from it. To train NLP models, it uses machine learning algorithms. After completion of the training process on large amounts of data, the trained model will have positive outcomes with deduction.

Comparison (Pros and Cons)

Rule Based NLP	Statistical NLP
+ 1) Flexible	+ 1) Easy to scale
+ 2) Easy to debug	+ 2) Learn by itself
+ 3) Doesn't require much training	+ 3) Fast development
+ 4) Understanding language	+ 4) High coverage
+ 5) High precision	
- 1) Requires skilled developers	- 1) Requires large amount of data
- 2) Slow parsing	- 2) Difficult to debug
- 3) Moderate coverage	- 3) No understanding of context

Components of NLP

The two basic components in which NLP can be divided are as follows:

- Natural Language Understanding (NLU)
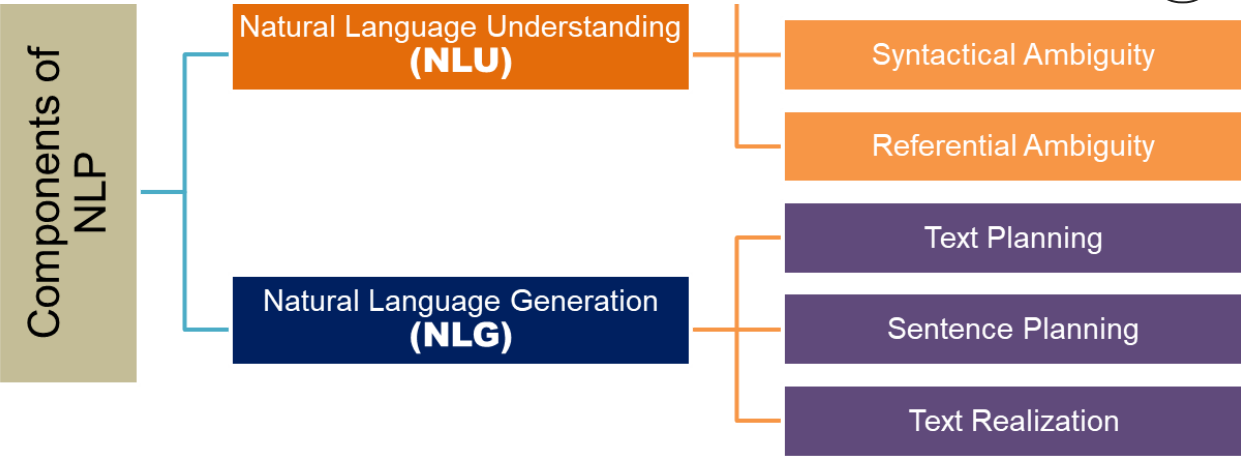- Natural Language Generation (NLG)

Image Source: Google Images

Natural Language Understanding (NLU)

NLU is naturally harder than NLG tasks. Let's discuss the challenges faced by a machine while it tries to understand the natural language.

While learning or trying to interpret a language, there are a lot of ambiguities.

```
Sentence: He is looking for a match.
```

Here, What do you understand by **"match"** – Partner or Cricket/Football Match.

Lexical Ambiguity can occur when a word carries a different sense, i.e. having more than one meaning, and the sentence in which that word is used can be interpreted differently based on its correct sense. To resolve these types of ambiguities to some extent, we can use parts-of-speech tagging techniques.

```
Sentence: The chicken is ready to eat.
```

Is the chicken ready to eat its food or the chicken is ready for someone else to it? You never know.

Syntactical Ambiguity occurs when we observed that there can be more than one meaning in a sequence of words. It is also known as Grammatical ambiguity.

```
Sentence: Chirag met Kshitiz and Dinesh. They went to a restaurant.
```

Here, they refer to Kshitiz and Dinesh or all.

Referential Ambiguity: It is very often in a text that it mentions an entity (something/someone), and then refers to it again, possibly in a different sentence, with the help of another word. So, these different pronouns can cause ambiguity when it is not clear which noun it is referring to.

Natural Language Generation (NLG)

It is defined as the process of generating or extracting some meaningful phrases and sentences in the form of natural language with the help of some internal representation.

This component involves the three basic steps:

- **Text planning:** It involves the retrieving of the relevant information from the knowledge base.
- **Sentence planning:** It involves processes such as choosing required words, forming meaningful phrases, setting the tone of the sentence.
- **Text Realization:** It involves the mapping of sentence plans into sentence structure.

Test Your Knowledge

<u>**Question-1:**</u> NLP is divided into two subfields:

- algorithmic and heuristic
- time and motion
- understanding and generation

<u>Question-2:</u> Which of the following is used to mapping sentence plans into sentence structure?

- Text planning
- Sentence planning
- Text Realization
- All of the Above

Ambiguity and Uncertainty in NLP

In natural language processing, Ambiguity can be referred to as the ability to be understood in more than one way. In simple terms, we can understand ambiguity as to the capability of being understood in more than one way. Natural language is very ambiguous.

NLP has the following five types of ambiguities:

Lexical Ambiguity

Lexical ambiguity is the ambiguity that involves the ambiguity of a single word.

For Example, Let's consider the following Sentences:

```
She won two silver medals
She made a silver speech
His worries had silvered his hair
```

In the above sentences how we treat the word silver- as a noun, an adjective, or a verb.

Syntactic Ambiguity

Syntactic ambiguity occurs when a sentence is parsed in different ways.

For Example, Let's have a sentence

```
Sentence: The man saw the girl with the microscope
```

This sentence is ambiguous as:

whether the man saw the girl carrying a microscope or he saw her through his microscope.

Semantic Ambiguity

This type of ambiguities occurs when the meaning of the words themselves can be misinterpreted. In simple words, semantic ambiguity occurs when a sentence contains an ambiguous word or phrase.

For Example, Let's have a sentence

```
Sentence: The bus hit the pole while it was moving
```

The above sentence is having semantic ambiguity because this sentence can have two interpretations

- "The bus while moving hit the pole"
- "The bus hit the pole while the pole was moving".

Anaphoric Ambiguity

Anaphora means when the same beginning of a sentence is repeated several times and Anaphoric ambiguity occurs due to the use of anaphora entities in discourse.

```
Sentence: The dog ran up the hill. It was very steep. It soon got tired.
```

Here, the anaphoric reference of "it" in two situations causes ambiguity.

Pragmatic Ambiguity

These types of ambiguities occur when the context of a phrase gives it multiple interpretations. In simple words, we can say that these ambiguities arise when the statement is not specific.

For Example, Let's have a sentence

```
Sentence: I like you too
```

that can have multiple interpretations as :

- I like you (just like you like me)
- I like you (just like someone else does).

This ends our Part-1 of the Blog Series on Natural Language Processing!

End Notes

Thanks for reading!

If you liked this and want to know more, go visit my other articles on Data Science and Machine Learning by clicking on the **Link**

Please feel free to contact me on **Linkedin**, **Email**.

Something not mentioned or want to share your thoughts? Feel free to comment below And I'll get back to you.

<u>About the Author</u>

Chirag Goyal

Currently, I am pursuing my Bachelor of Technology (B.Tech) in Computer Science and Engineering from the **Indian Institute of Technology Jodhpur(IITJ).** I am very enthusiastic about Machine learning, Deep Learning, and Artificial Intelligence.

The media shown in this article are not owned by Analytics Vidhya and are used at the Author's discretion.

Getting started with NLP using NLTK Library   Understanding Natural Language Processing -A Beginner's Guide   How to Use Google's NLP API to Analyze and Produce Better Content

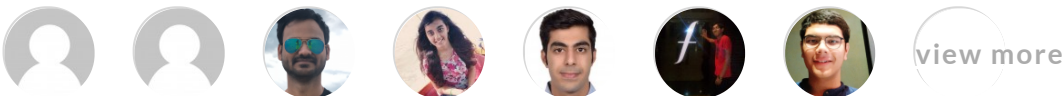blogathon   Natural Language Processing Python

About the Author

Chirag Goyal

Our Top Authors

view more

Download

Analytics Vidhya App for the Latest blog/Article

Previous Post

Fairness and Ethics in Artificial Intelligence!

Next Post

Part 2: Step by Step Guide to NLP – Knowledge Required to Learn NLP

Leave a Reply

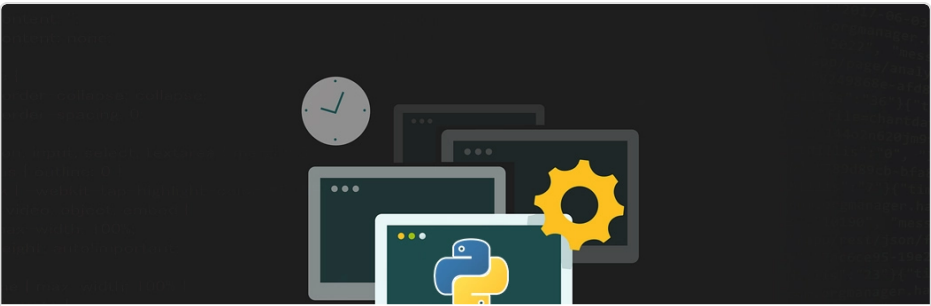**Your email address will not be published. Required fields are marked ***

Comment

Name*

Email*

Website

☐ Notify me of follow-up comments by email.
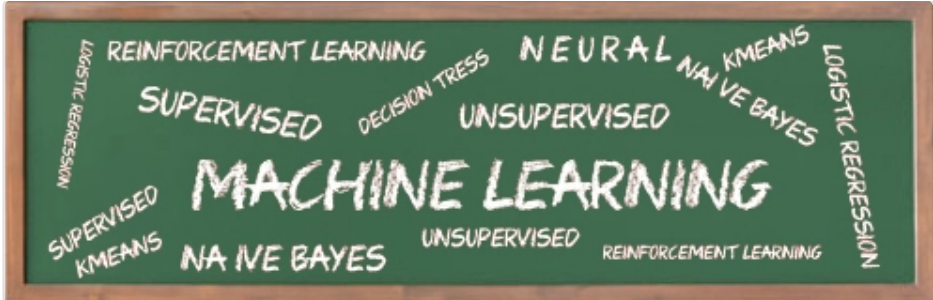
☐ Notify me of new posts by email.

Submit

Top Resources

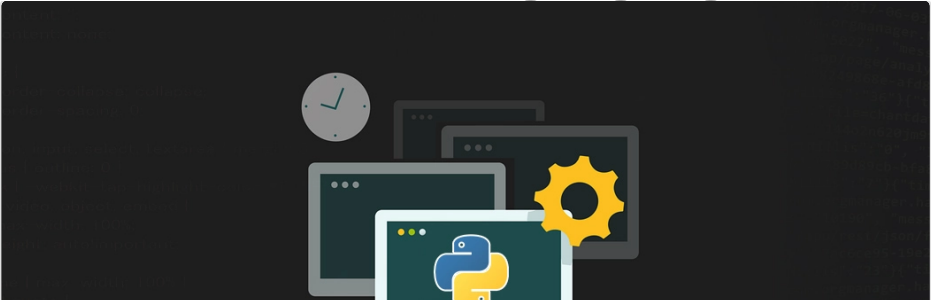40 Questions to test a Data Scientist on Clustering Techniques..

Python Tutorial: Working with CSV file for Data Science

[Commonly used Machine Learning Algorithms (with Python and R Codes)](#)

Sunil Ray - SEP 09, 2017

[Basic Concepts of Object-Oriented Programming in Python](#)

Himanshi Singh - SEP 01, 2020

Analytics Vidhya

About Us

Our Team

Careers

Contact us

Companies

Post Jobs

Trainings

Hiring Hackathons

Advertising

Download App

Data Scientists

Blog

Hackathon

Discussions

Apply Jobs

Visit us

Privacy Policy   Terms of Use   Refund Policy