**INTRODUCTION TO DATASCIENCE**

DATASCIENCE

Data Science is the process of extracting knowledge and insights from data by using scientific methods.

i.e. , Data Science is a simple process that will just extract the useful information from data. It is just going to process and analyze the entire data, and then it going to extract what is needed. Also using it to grow a business.

Data Science – skill set

* Statistics
* Programming Language
* Data extraction and processing
* Data wrangling and exploration
* Machine learning
* Big Data processing and framework

Some Examples of Data Science Job Roles

* Data Scientist
* Data Analyst
* Data Architect
* Data Engineer
* Statistician
* Data base Administrator
* Business Analyst
* Data and Analytics Manager

Data Life Cycle

* Business Requirements :

* Understand the problem
* Identify central objectives
* Identify variables that need to be predicted
* Data Acquisition :
* What data do I need for my projects?
* What are the data sources?
* How can I obtain the data/
* What is the most efficient way to store and access all of it?
* Data Processing:

Transform data into desired format.

Data Cleaning:-

* Missing values
* Corrupted data
* Remove unnecessary data
* Data exploration:
* Understand the patterns in the data
* Retrieve useful insight
* Form hypothesis
* Modelling:
* Determine optical data features for the machine learning model
* Create a model that predicts the target most accurately
* Evaluate and test the efficiency of the model
* Deployment :
* Check the deployment environment for dependency
* Deploy the model in a pre-production/test environment
* Monitor the performance