Data Science

Data Science is an inter-disciplinary field which consists of:

- **Mathematics**
- **Statistics**
- **Computer Science**
- **Information Science**

that employs scientific methods and processes to extract knowledge and insights from the data either structured or unstructured.
The professionals who practice data science is called as "Data Scientist". Dj Patil along with Jeff Hammerbacher claims that they have first coined the term **Data Scientist** in the year 2008 to describe their jobs at **LinkedIn** and **Facebook** respectively.

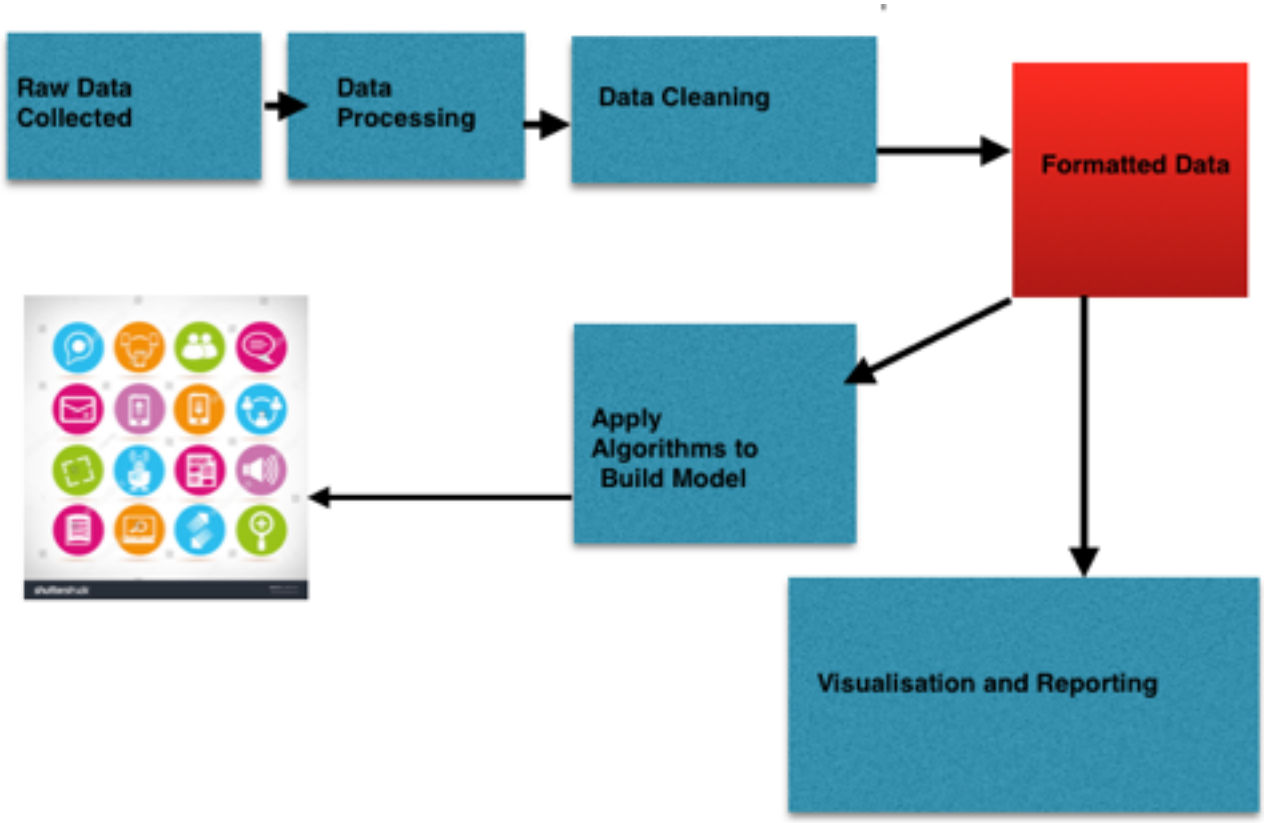
The term Data Science and Data Scientist becomes the buzz word or hashtag when, Harvard Business Review wrote an article titled:

"**Data Scientist: The Sexiest Job of the 21st Century**"

Since then, the demand for the Data science professionals are increasing day by day, so is their pay scale and challenges.

In this session we will be learning the best practices and method employed in the Data Science domain to get the best possible results and insights from the data.

The Data Science flow chart:



The Fundamentals Of Machine Learning

Machine Learning is the study and design of software system that learns from past experiences to make future decisions.

Popular computer scientist Tom Mitchell defines Machine Learning as-

"**A program can be said to learn from the experience E with respect to some class of task T and performance measure P, if its performance at task in T, as measured by P, improves with the experience E** "

There are two types of machine learning task:

1. Supervised Learning:

In supervised learning problem a program learns to predict an output, from an already labelled input and output.

2. Unsupervised Learning:

In unsupervised Learning, the program attempts to discover patterns in the data.

The supervised learning task is again categorised in to:

Regression Problem: In regression problem the program predicts the value of continuous response variables. Example: price of house, sales of new product, revenue etc.

Classification Problem: In classification problem, the program predicts the probable class, category or labels, in much boarder sense the program tries to predict the discrete values for the explanatory variables. Example: Gender, rise/ fall of stock market price, whether the given article is spam or ham etc.