Data Science

An Introduction

Understand Terms…

- Data Science

- Analytics

- Artificial Intelligence

- Machine Learning

- Deep Learning

What is Data Science?

- Application of Scientific Methods like Statistical and Machine Learning in order to understand the phenomena to gain control on it

- It employs techniques from both the fields computer science and statistics

- Data science involves Machine Learning, Clustering, Visualization and many other things related to data
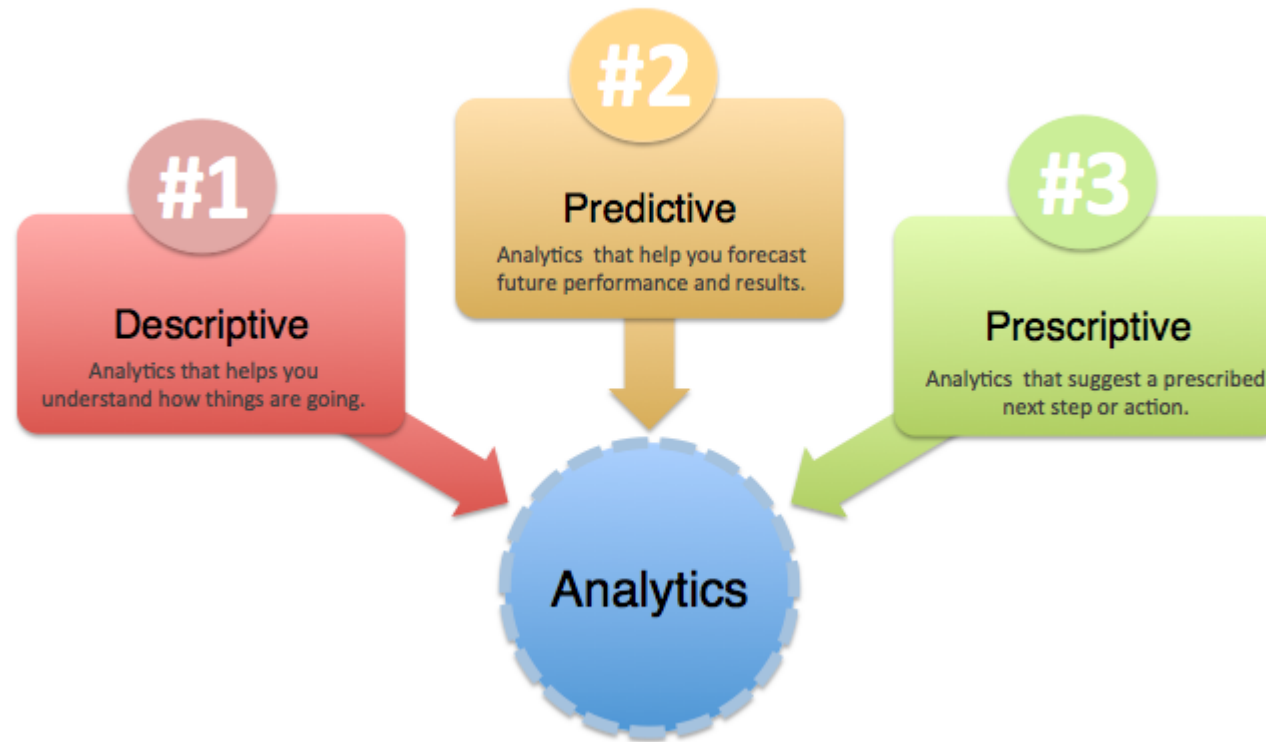
Data Science Composition
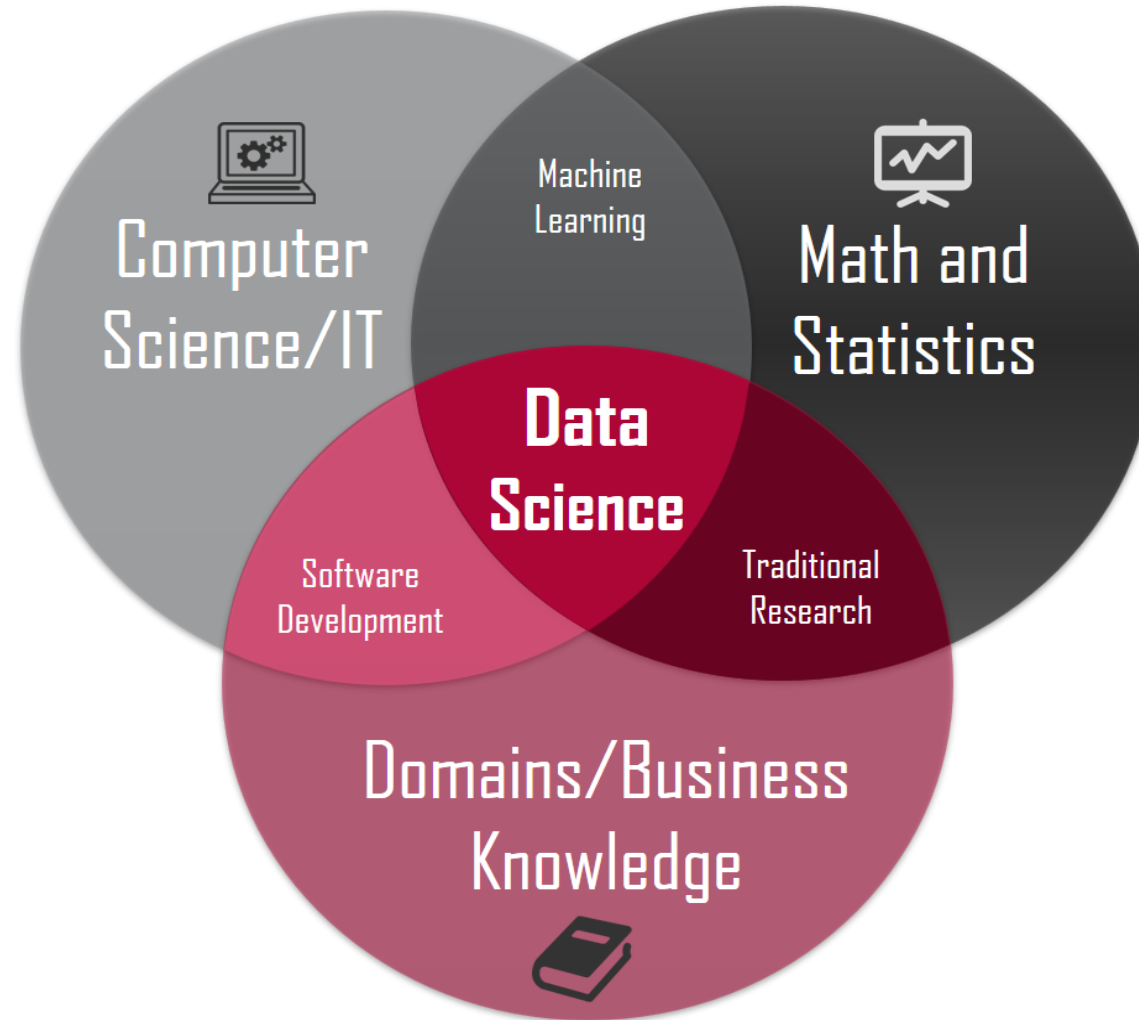
What is Analytics?

- Analytics is the discovery, interpretation, and communication of meaningful patterns in data.

- Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming and operations research to quantify performance

Types of Analytics

#1

Descriptive

Analytics that helps you understand how things are going.

#2

Predictive

Analytics that help you forecast future performance and results.

#3

Prescriptive

Analytics that suggest a prescribed next step or action.

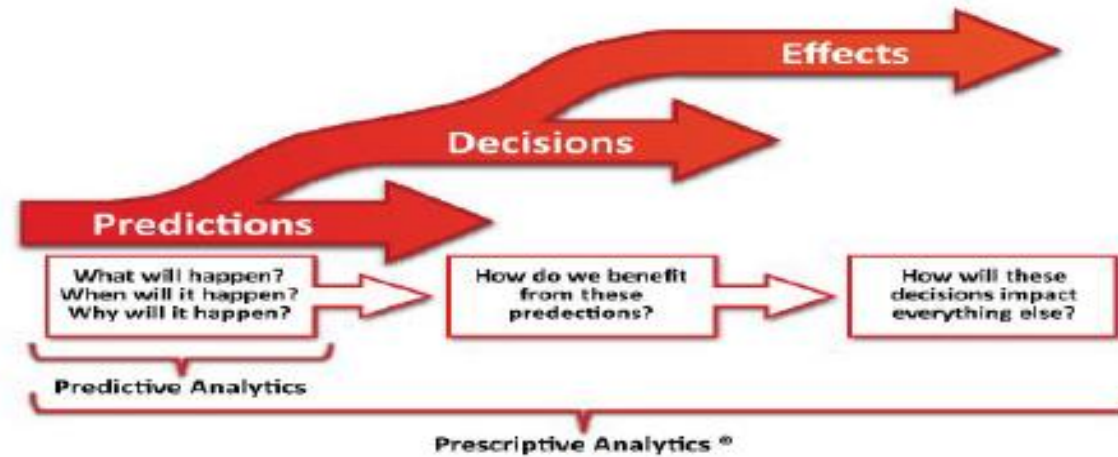**Analytics**

Descriptive Analytics

- Gain insight from historical data with reporting, scorecards, clustering etc.
- Can involve data visualization for knowing the basic characteristics of the data
- Descriptive analytics answers the questions what happened and why did it happen.
- Implementations : Business Intelligence, Visualizations
- Software: Informatica, Business Objects, TIBCO Spotfire, Tableau etc.

Predictive Analytics
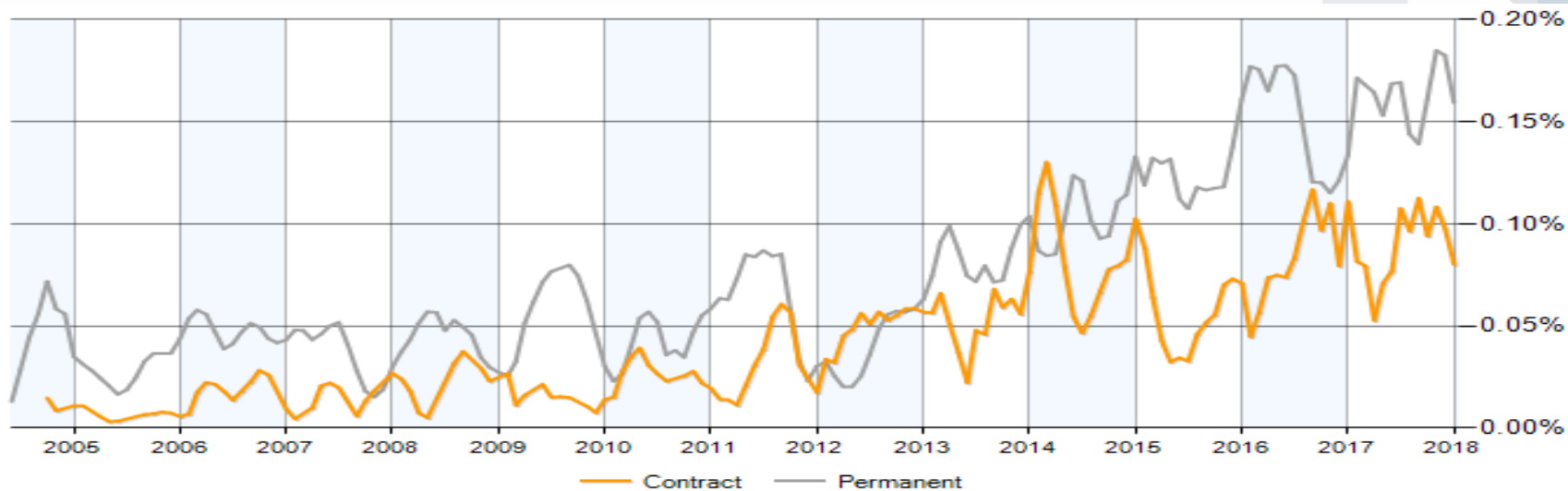
- Involves statistical and machine learning techniques

- Analyzing the historical patterns in the data and predicting the future patterns

- Predictive analytics answers the question what will happen

- Implementation: Machine Learning, Deep Learning

- Software: R, Python, Libraries like TensorFlow, h2o.ai etc.

Prescriptive Analytics

- Prescriptive analytics goes beyond predicting future outcomes by also suggesting actions to benefit from the predictions and showing the implications of each decision option.

- Implementation: Optimization Techniques like Linear programming Problems, Non-linear programming Problems, Genetic Algorithm etc.

What is AI?

- AI or artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems.

- These processes include learning (the acquisition of information and rules for using the information), reasoning (using the rules to reach approximate or definite conclusions), and self-correction.

- AI is a discipline just like Physics.

Role of Machine Learning (ML)

- Machine Learning is a tool set for implementing AI, today
- AI elements which don't include ML are expert systems
- ML Algorithms are driven by mathematical concepts
- ML Algorithms analyse the patterns in the captured data and can be used to build a predictive model on the existing phenomena in business
- Broadly, there are three types of ML Algorithms
  - Supervised Learning Algorithms
  - Unsupervised Learning Algorithms
  - Re-inforcement Learning Algorithms

Supervised Learning

- Supervised learning algorithms are those used in classification and prediction.

- We must have data available in which the value of the outcome of interest (e.g., purchase or no purchase) is known.

- The objective is to predict the values of the outcome of interest

Models for Supervised Learning

- We identify strong links between variables of a data table (columns).

- Such a link may translate into an expression between one variable y (the so-called "dependent" or "response" variable) and a group of other variables $\{x_i\}$ (the so-called "independent variables" or "predictors" or "features") :

    $y = f(x_1, x_2, ..., x_n) +$ Small random noise

Types in Supervised Learning

- When the response variable is numerical, predictive modeling is called Regression.

- When the response variable is categorical (nominal / ordinal), predictive modeling is called Classification.

Examples

- **Regression Case**: Sales are influenced by the variables like advertisement expenses, manpower deployed for sales, cost of products, number of dealers etc. Hence we see here

  Sales = function (Adv. Exp , Manpower , Cost , Dealers , …)

- **Classification Case**: The customer may purchase a particular product based on some conditions like his need, his age, his income, his place of residence etc. Hence we see here

  Prob(Customer Purchases) = function(Age, Income, Residence,…)

Algorithms of Supervised Learning

- Naïve Bayes

- K-NN

- Decision Trees

- Regression Models

- Neural Nets

- Support Vector Machines

Unsupervised Learning

- Unsupervised learning algorithms are those used where there is no outcome variable to predict or classify.

- Association rules, data reduction methods, and clustering techniques are all unsupervised learning methods.

Examples

- Customer Segmentation like RFM (Recency, Frequency, Monetary)
- Market Basket Analysis
- Product Grouping

Algorithms of Unsupervised Learning

- Clustering Techniques
    - Hierarchical
    - K-means
- Principal Component Analysis
- t-SNE
- Association Rules
- Auto-encoders

Re-inforcement Learning

- In this type, there is an agent which/who receives information from the environment and learns to choose actions based on rewards or punishment received

- Examples include:
  - Self-driving cars
  - Robotics

- Algorithms:
  - Upper Confidence Bound
  - Thomson Sampling