

Data Analytics

What is Analytics ?

Analytics is the extensive use of data, statistical and quantitative analysis, exploratory, predictive models, and fact based management to drive decisions and actions.”

Analytics can be defined as “the analysis of data to draw hidden insights to aid decision making”.

..... and many more !!!

Frequently used terms

Data Analytics	Data Analysis	Big Data	Data Types
Data Warehouse	Data Mining	Data Cleansing	Data Definition
Data Manipulation	Data Transformation	Data Wrangling	Databases
Data Sources	Data Forms	Raw and Processed Data	
Data Collection	Statistics	Statistical measures	Mathematics Linear Algebra
Artificial Intelligence	Normalization	R / Python	Hadoop
Text Analytics	Algorithms	Predictions	Patterns
Supervised learning	Unsupervised learning	Clustering	etc....

Analysis vs Analytics

Analysis	Analytics
Uses data to perform functions to standardize processes	Uses data to find insights, predict, forecast and optimize models
Tools like SWOT (Strength, Weakness, Opportunities, Threats) Analysis are associated with business process	Tools used are Decisive(visual analytics), Descriptive(insight from historical data), Predictive(Machine Learning) and Prescriptive analytics(Optimisation and simulation)
Eg: <ul style="list-style-type: none">• Creating a business case• Risk Assessment• Requirement Analysis• Business Processes• Documentation requirements	Eg: <ul style="list-style-type: none">• Dashboard for tracking• Data Warehouse for tracking historical data• Finding patterns using data mining• Predictions

Definitions

Statistics is just about the numbers, and quantifying the data. There are many tools for finding relevant properties of the data but this is pretty close to pure mathematics.

Data Mining is about using statistics as well as other programming methods to find patterns hidden in the data so that you can *explain* some phenomenon. Data Mining builds intuition about what is really happening in some data and is still little more towards math than programming, but uses both.

Machine Learning uses **Data Mining** techniques and other learning algorithms to build models of what is happening behind some data so that it can *predict* future outcomes. Math is the basis for many of the algorithms, but this is more towards programming.

Artificial Intelligence uses models built by **Machine Learning** and other ways to *reason* about the world and give rise to intelligent *behavior* whether this is playing a game or driving a robot/car. Artificial Intelligence has some goal to achieve by predicting how actions will affect the model of the world and chooses the actions that will best achieve that goal. Very programming based.

In short

Statistics *quantifies* numbers

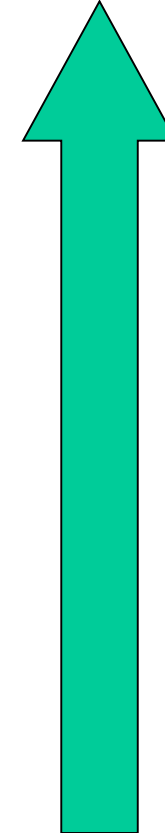
Data Mining *explains* patterns

Machine Learning *predicts* with models

Artificial Intelligence *behaves* and *reasons*

Types of Analytics

	<i>Types of report, analytics and query</i>	<i>Focus</i>
Analytics	Optimization	What's the best that can happen ?
	Prediction	What will happen next ?
	Forecasting	What if this trend continues ?
	Statistical Analysis	Why is this happening ?
Query and Reports	Alerts	What actions are needed ?
	Drilldown reports	Where is the problem ?
	Ad-hoc reports	How many, how often ?
	Standard Reports	What happened ?



Data Science

- Art of transforming *hypotheses* and *data* into actionable predictions
- **For example**, we can use models and data to
 - ✓ Predict who will win an election
 - ✓ What products will sell well together (Apriori / Market-Basket analysis)
 - ✓ Who is likely to default on loans
 - ✓ Which advertisements will be clicked on
 - ✓ etc.

- **Tools used (but not restricted to)**

Empirical Sciences

Statistics

Business Intelligence

Databases

Data Warehousing

Visualization

Expert Systems

Analytics

Machine Learning

Big Data

Data Mining

Reporting

- **Central goal of Data Science**

To deploying effective decision-making models to a production environment

Data Science

These systems share a lot of features:

- Amazon's product recommendation systems
- Google's advertisement valuation systems
- LinkedIn's contact recommendation system
- Twitter's trending topics
- Walmart's consumer demand projection systems

Built on a large dataset

Most of the systems are live or online

Allowed to make mistakes

Not concerned with any cause

Machine Learning

- The ability to write a mathematical function that will read an input and produce output
- We provide the function – machine does not pick its own function
- ML considerations
 - Training data (lots of it)
 - Model
 - Cost function (eg: Ordinary Least Squares)
 - Optimisation (eg: Gradient descent)

Why is learning possible?

- **Generalisation is possible**
eg: if dataset contains travel time between places A and B, function would not generalise if we predict travel distance between A and C
- **IID (independent and identical distribution) of data**
That's why gradient descent needn't go through the entire dataset, since data is similar

... Eventually data will surpass in oil and water in importance

Thank You !!!