Introduction to Machine Learning

Gopi Krishna Nuti

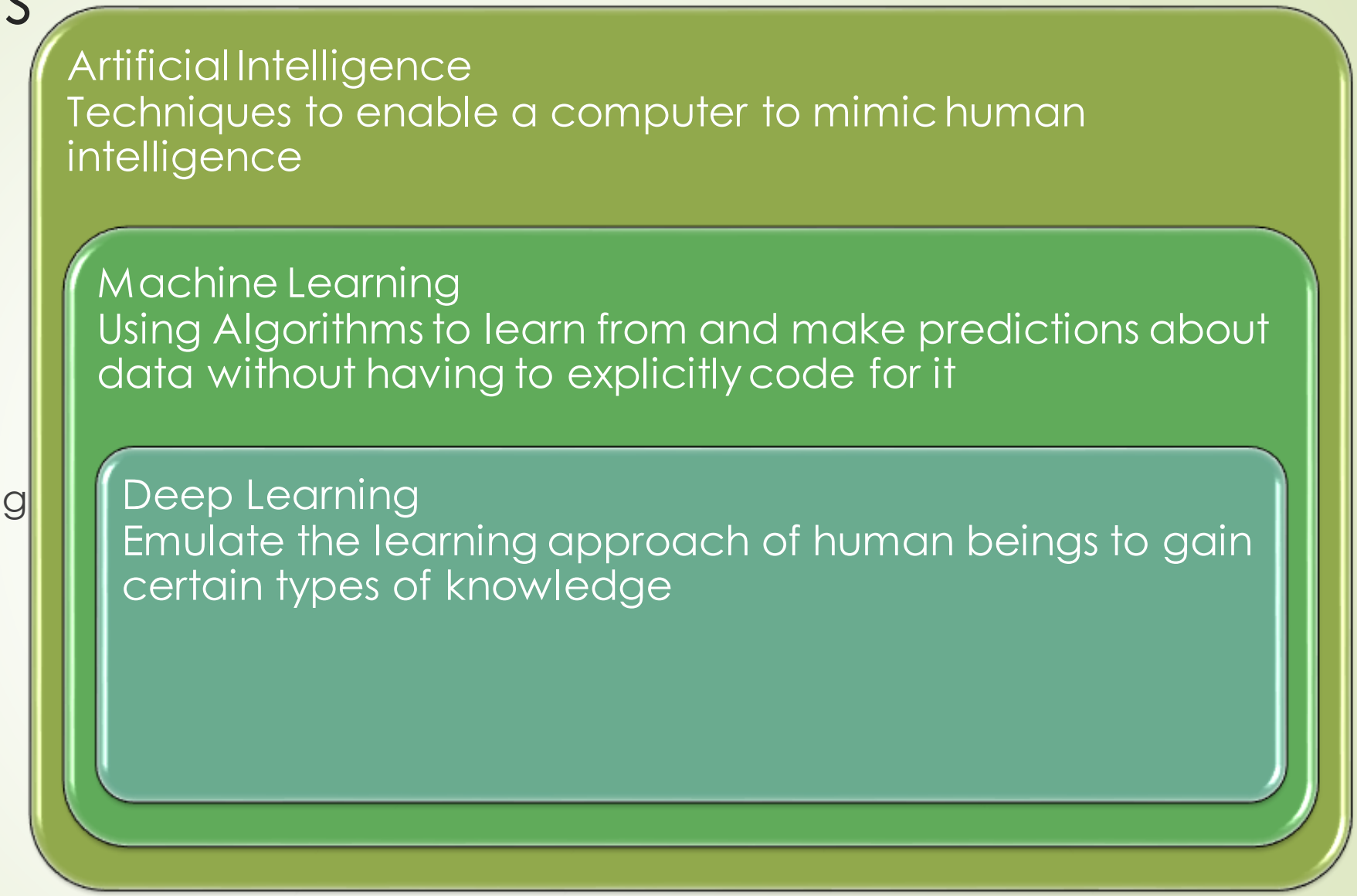
gopi.nuti@tcs.com

Machine Learning

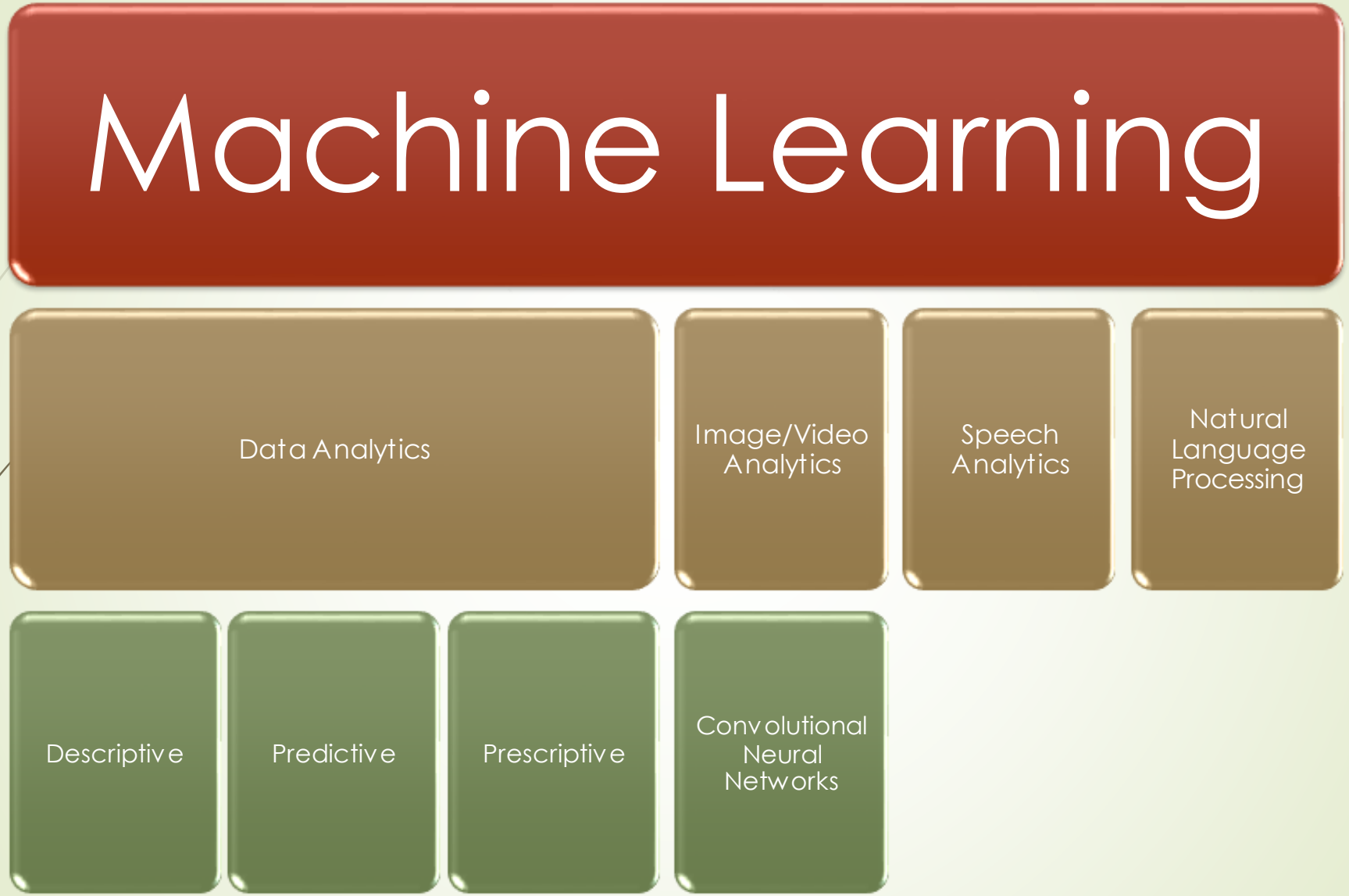
- Has been around since many decades with various names like Applied Statistics, Measurement Science etc.
- Why recent interest?
  - CPU capacity increased by orders of magnitude
  - Explosion of digital data
  - Cheaply available computing power

What is Analytics

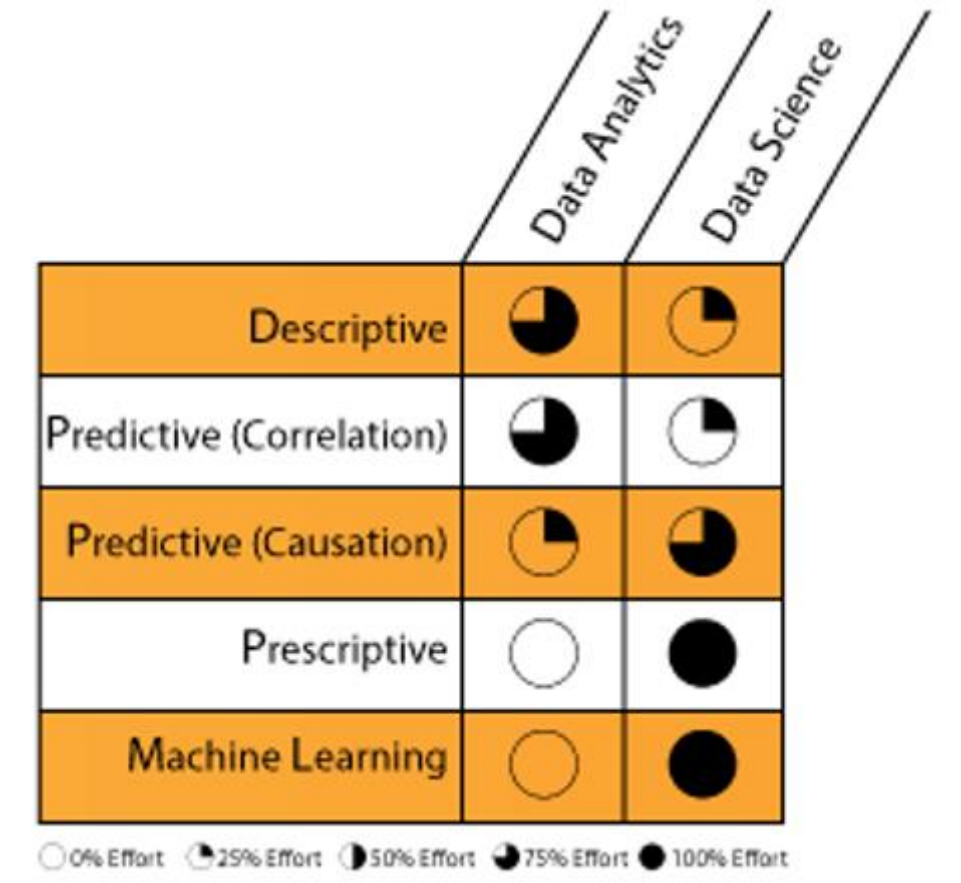
- Artificial Intelligence
- Machine Learning
- Data Science
- Image/Video Analytics
- Speech Analytics
- Natural Language Processing
- Statistics
- Big Data
- Big Data Analytics



Machine Learning



Data Analytics Vs Statistics (Data Science)



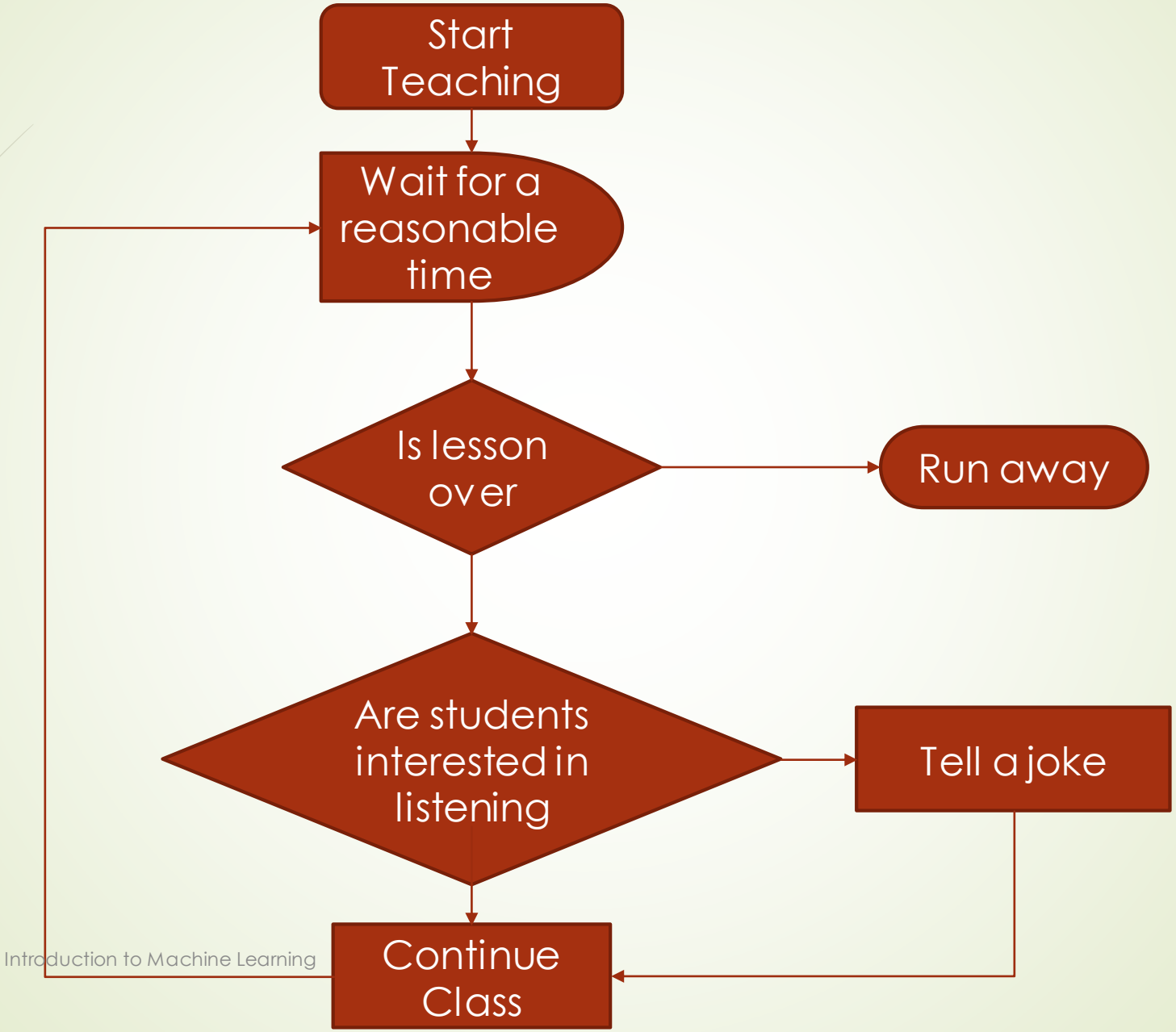
Data Science	Data Analytics
Mathematics of explaining population relationships based on samples.	• Extracting valuable information out of data • Predict values for new data
Scarcity of Data	Abundance of Data
Hypothesis comes first	Data comes first
Macro Decisioning	Micro Decisioning

Image courtesy of Datascientistinsights.com

Machine Learning Vs Classical Programming

- Why is machine learning disruptive?
- Lets look at an example.
  - The business is an educational institution.
  - KPI being measured is interest level of students attending the class in a 1 hour duration.
  - Goal : KPI should always be above a certain threshold.

Classical Programming



Classical Programming

- Starts with business rules or Business Requirements Spec.
- Business rules
  - Extracted by Business Analyst or Product Owner
  - Coded by developers
  - Tested by the testing team with simulated test cases.

Machine Learning

- Starts with historically available data
- Business rules
  - Extracted by the machine by looking at the data.
- Testing
  - Has to be done on real world data.
  - Simulation is of little value

Classical Programming Vs Machine Learning

S. No	Duration Since class started	Current Interest Level	Action Taken	Resultant Interest level
1	0	High	Continue Class	High
2	30	Low	Punish the class	Low
3	30	Low	Tell a joke	High
4	30	Low	Continue class	Low
5	15	High	Continue Class	High
6	5	Low	Punish the class	High
N	60	Low	Stop class	High

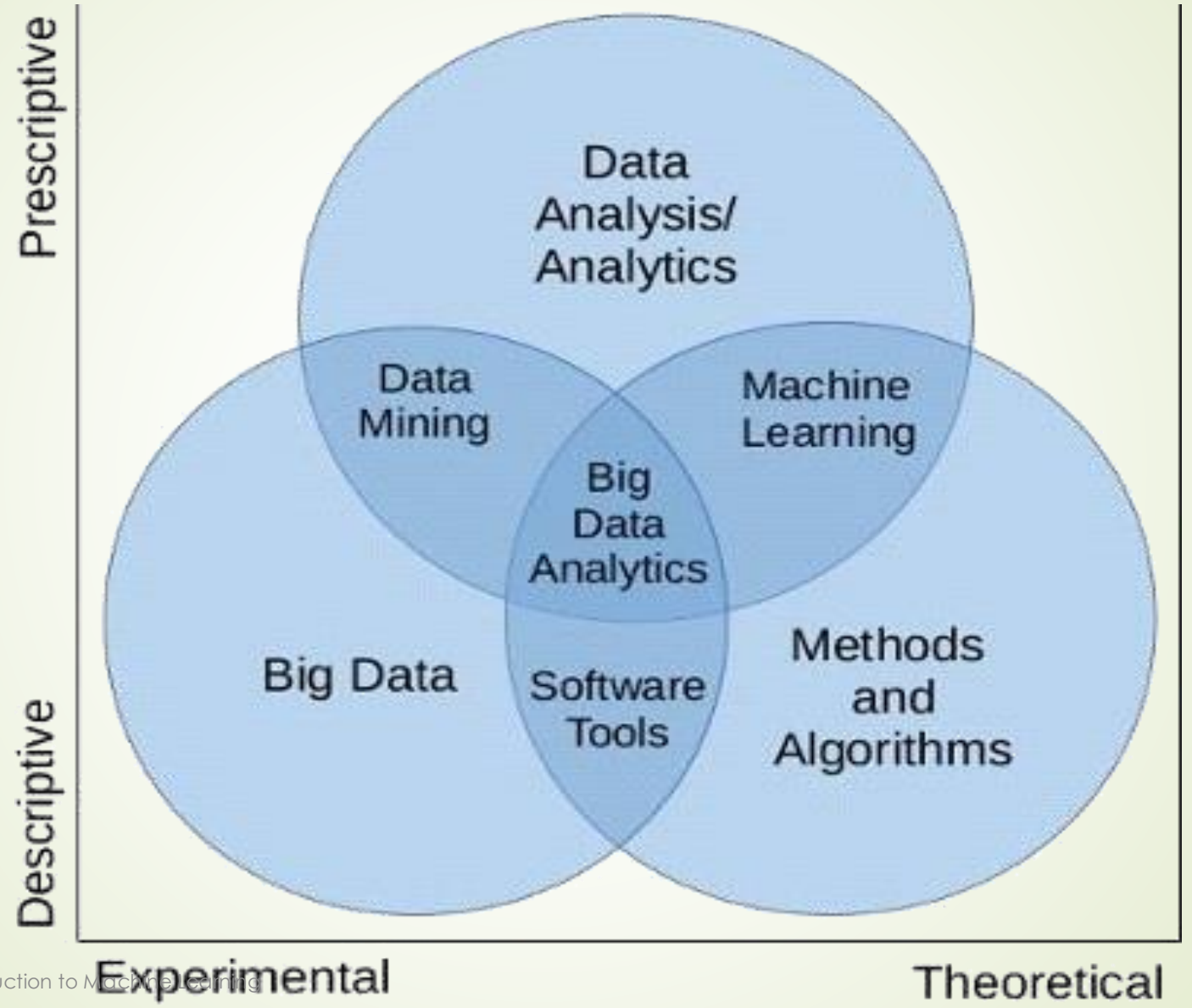
- "Answer" is "Resultant Interest Level".
- "Rules" is "Action Taken" i.e. what we are modelling for.
- Machine analyses this data and extracts the rules
- When in future, we query for a scenario like below, the appropriate "Action Taken" will be identified by the machine
  Duration since class started = 38
  Current Interest Level = Medium
  Resultant Interest Level = High
  What action should be taken?

Classical Programming Vs Machine Learning

Classical Programming

Machine Learning

Big Data Vs Analytics



Roles & Activities

- Data Scientist
- Data Engineer
- DBA
- Performance Engineer
- Hardware Developer

Machine Learning Pipeline

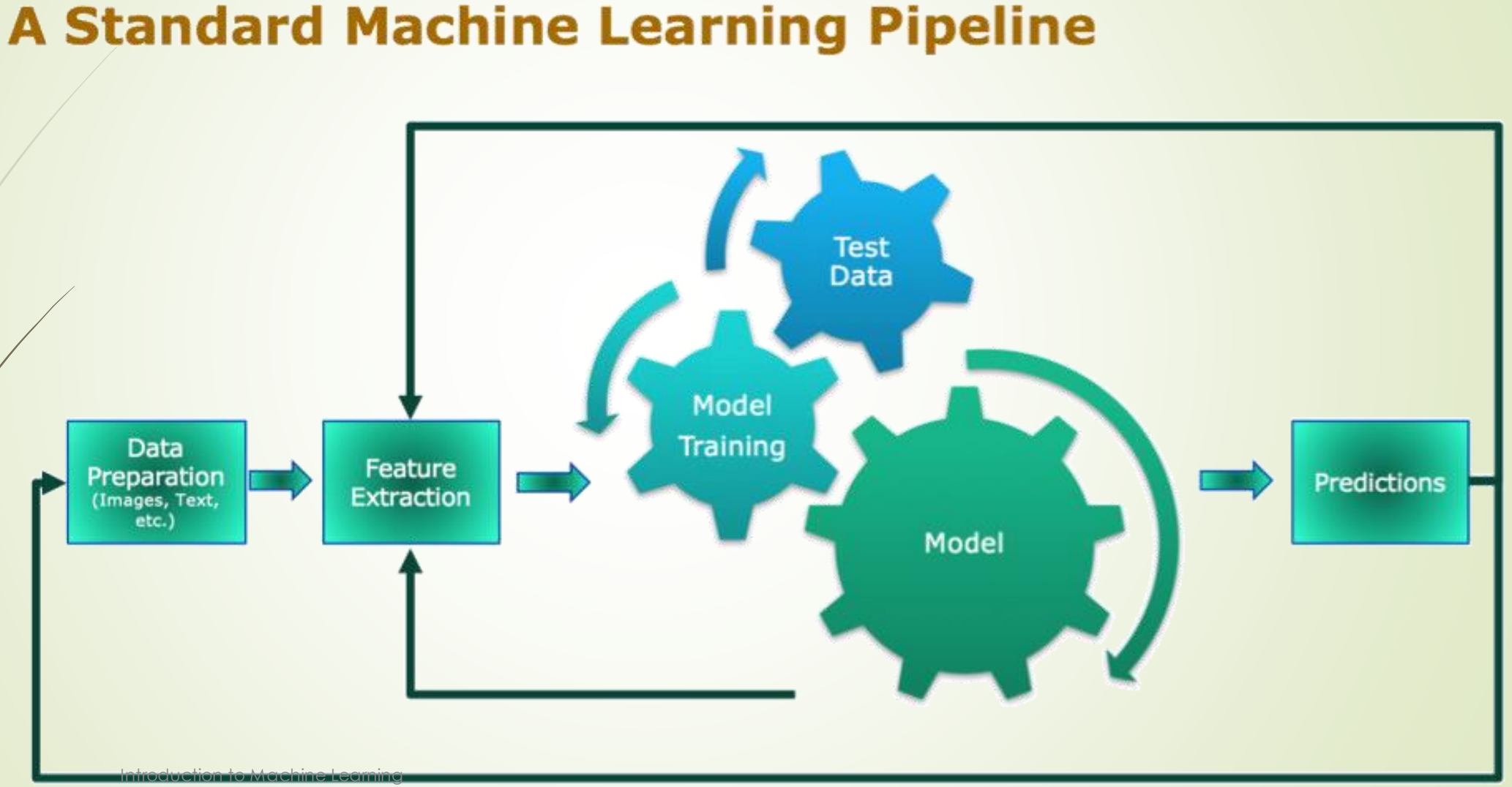


Image courtesy Western Digital

Recent Advances

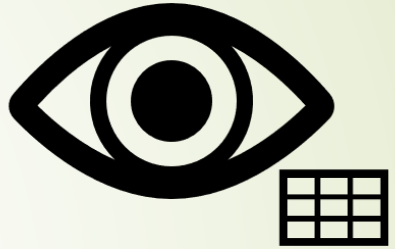
- Video Analytics
  - Convolutional Neural Networks
  - Generative Adverserial Networks
- Speech Analytics
  - Recurrent Neural Networks
  - Long Short Term Memory
  - Connectionist Temporal Classification
- Text Analytics/Natural Language Processing
- Re-inforecement Learning

Data Analytics

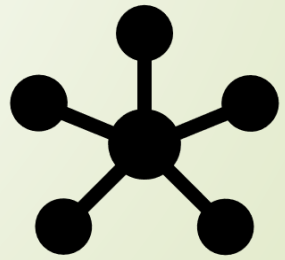
- Information Types
  - Structured Data : Databases etc.
  - Semi structured Data: XML files, JSON files
  - Unstructured Data: Images, Videos, Sound

Data Analytics – Machine Learning Types

- Supervised Learning



- Unsupervised Learning



Learning modes

Supervised Learning

- Predict outcome for new data
- Regression
  - Linear Regression, Polynomial Regression, Support Vector Regression, Decision Tree, Random Forest,
- Classification
  - Logistic Regression, k-Nearest Neighbours, Support Vector Machines, Naïve Bayes, Decision Tree, Random Forest

Unsupervised Learning (Descriptive)

- Essentially descriptive.
- Clustering
  - k-Means, Hierarchical Clustering
- Associate Rule Learning
  - Market Basket Analysis

Levels of Data

Nominal

- Algebraic operations are not possible

Ordinal

- Logical operations are possible but not mathematical operations . Ex: Account Number

Interval

- Addition/Subtraction is possible but not multiplication/division
- Interval between two continuous elements is always same and meaningful
- Zero is arbitrary
- Ex: Temperature

Ratio:

- Zero makes sense and negative values are not possible
- Mean, Median, Mode etc can be calculated
- Account Balance

Regression Vs Classification

	Regression	Classification
Dependent variable	Continuous	Categorical
Purpose	Predict output value using training data	Group the output to a class
Output level	Ratio or Interval	Ordinal or Nominal

Clustering

Group similar data together

Behavioural segmentation:	Segment by purchase history, activities on application, website, or platform
	Define personas based on interests
	Create profiles based on activity monitoring
Inventory categorization	Group inventory by sales activity, manufacturing metrics
Sorting sensor measurements:	Detect activity types in motion sensors
	Group images
	Separate audio
	Identify groups in health monitoring
Anomaly Detection	Fraud Analytics, Strange behaviour in Bigdata databases
	Security systems

How does Prediction work?

Consider an example for Linear Regression

- Let us imagine a world where Sir Issac Newton was not born.
  So, we don't know that F = m.a
- We have been asked by the client to create a model for calculating Force. He thinks Force is related to mass and height from sea level.
- Historic data looks like below

S. No	Mass	Height	Force observed

- Our linear regression equation can be mathematically explained as

$$F = \beta 0 + \beta 1.m + \beta 2.h$$

- B0, β, β2 are calculated using historic data

Prediction

- Is it perfect? Obviously not.
- However this approach is a mathematically verifiable estimate with a calculable error.
- Multiple such approaches a.k.a models are possible.
- What does a Data Scientist do?
  Explore the different models and identify the one with least error.
- What does a Data Engineer do?
  Explore the different datasets available, does ETL and provides it to Data Scientist.

Terminology and Metrics

- Force – Supervised Variable or Dependent Variable
- Mass, Height – Independent variable

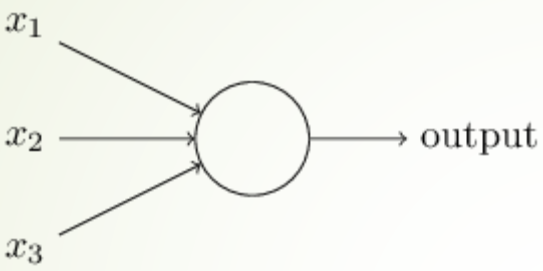
- Metrics for Prediction
  - R^2, RMSE – Root Mean Square Error, MAE – Mean Absolute Error,
- Metrics for Classification
  - True Positive, False Positive, True Negative, False Negative
  - Accuracy, Precision, Recall,
  - Specificity, Sensitivity, ROC, AUC, Gini Index

Deep Learning

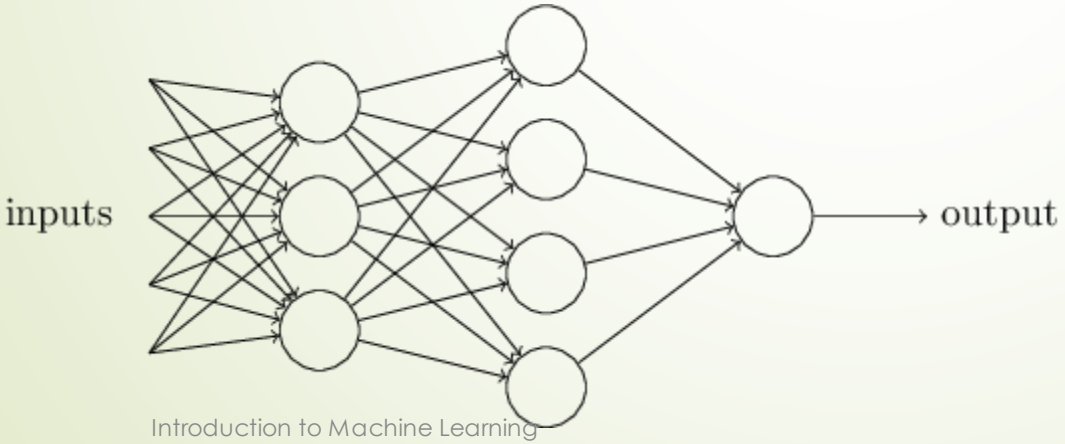
- An approach of Machine Learning where the learning approach simulates a human brain
- Advantages
  - Generalize very well. Can identify the hidden patterns even in complicated data sets.
  - Generally out perform the statistics based approaches
  - With enough training data, can represent any function. NAND Gate representation.
  - Universal Approximation Theory
  - In words of Elon Musk, "It's quite simple, really".
- Disadvantages
  - Data Hungry
  - Computation Intensive
  - By and large, a black box. Difficult to explain why a particular answer was arrived at for the given data because of in-built Stochasticity and mathematical complexity.

Neural Networks

- Artificial Neural Network
  - Perceptron

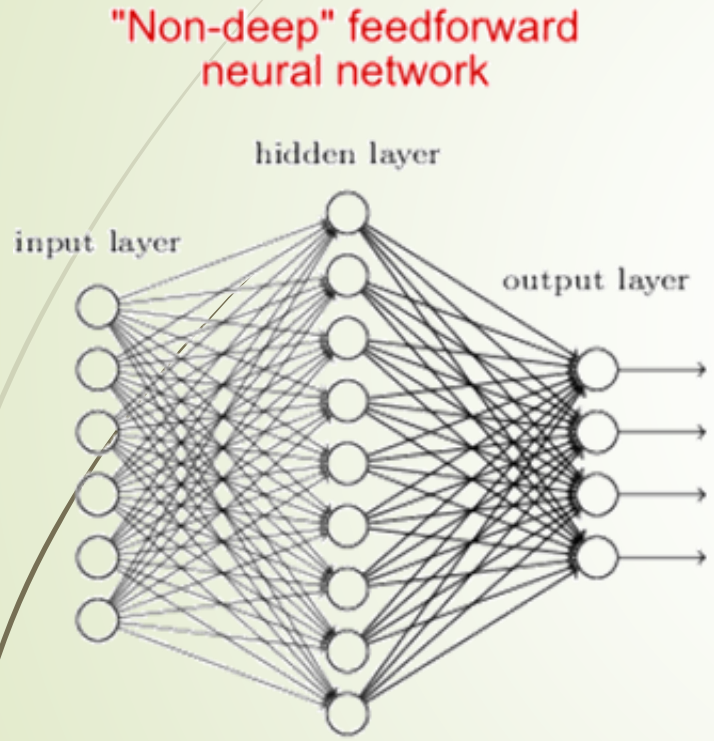


output= 0 if $\sum w_j x_j \leq$ threshold
1 if $\sum w_j x_j >$ threshold



Neural Networks

- Deep Network vs Shallow Network



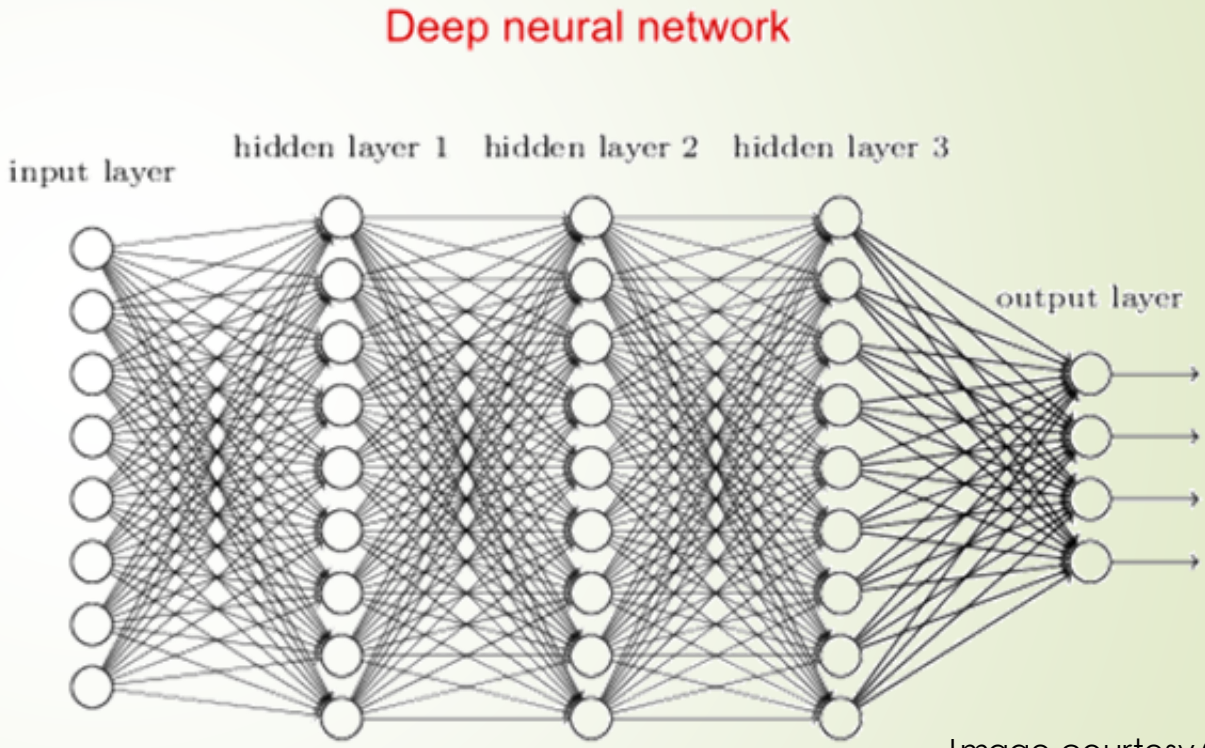


Image courtesy Quora

Neural Networks

- Activation Functions
- Gradient Descent & Loss

Important Concepts

- Feature Engineering
- Dimensionality Reduction
- Principal Component Analysis
- Training Data, Validation Data, Testing Data
- Outliers and Missing value treatment
- Overfit & Underfit
- Precision & Recall
- Feature Scaling
  - Manhattan Distance, Mahalanobis Distance, Euclidean Distance