AI Toolkit Scikit Learn, TensorFlow and Keras

We will be starting soon

Day 1

Welcome to Day 1
AI Toolkit

Please perform PRE-WORK

1. Access the virtual Lab using link https://html.inspiredvlabs.com Use the username TEKBD142-XX (replace XX with your number) and password

   TekBD142!23

https://tinyurl.com/bdtAIToolkit

We will be starting soon

Last Name	First Name	Login Id
AHMED	NAZEER	TEKBD142-01
AM	GANESH	TEKBD142-02
BISWAS	SOURAV	TEKBD142-03
CHACKO	THOMAS	TEKBD142-04
JAJOO	SANDESH	TEKBD142-05
MATTOX	CHRISTEL	TEKBD142-06
MAXSON	CRAIG	TEKBD142-07
MURPHY	MATTHEW	TEKBD142-08
PANDIAN	JEYARAJ	TEKBD142-09
PATURI	RAVIKIRAN	TEKBD142-10
RANI	AMITA	TEKBD142-11
SINGLA	SANJEEV	TEKBD142-12
VEERAMASU	BRAHMA RAO	TEKBD142-13

Logistics

- **Timing –** 8:30 AM – 4:30 PM CST
- **Lunch –** Approx. noon to 1 PM CST
- **Periodic Breaks –** At logical points

Virtual Class Tips

Audio

Please mute yourself if not speaking, unmute to ask questions

Notifications

✓ Give us green right check mark in participant panel when you are done

Notifications

🙂 Give a smiley when you are back from break

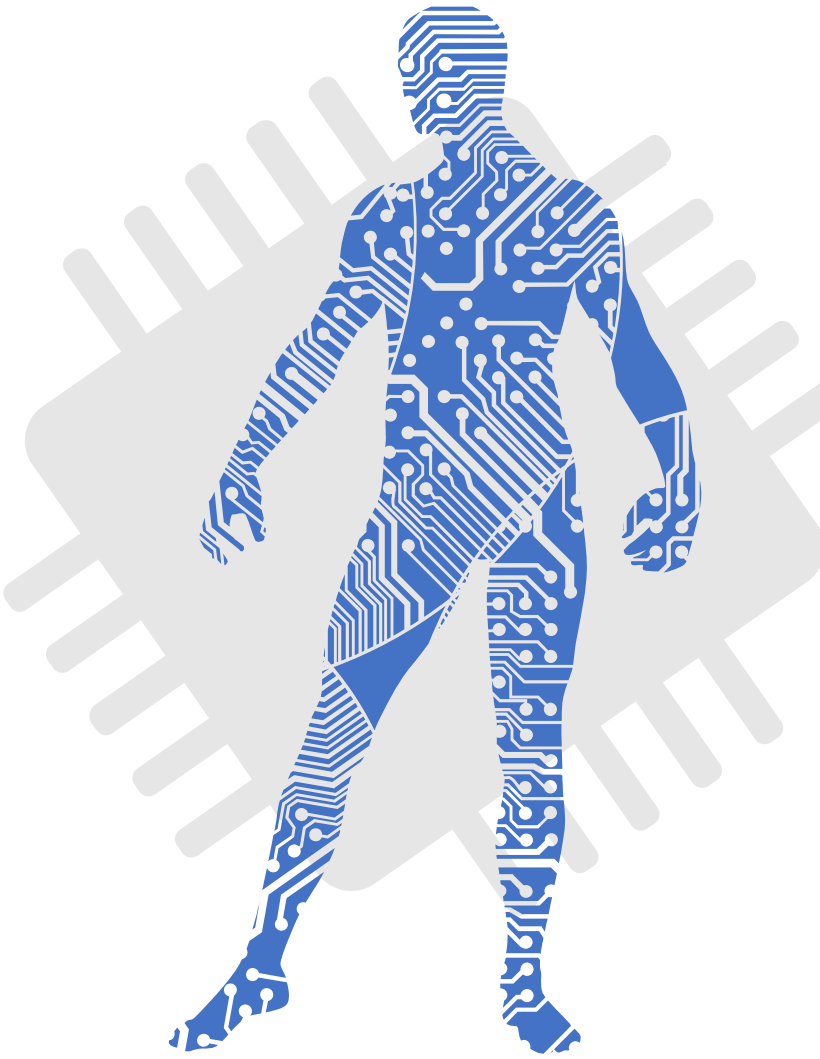
Notifications

 Please raise hand in participant panel or unmute and ask question

Let’s make it interactive !

Agenda – Day 1

1. Introductions
2. Machine Learning Introduction
3. Machine Learning Techniques
4. Machine Learning Development
5. Multiple Hands-on



Sanjay Oza

- 25+ years of experience in Software, Hardware and Data Engineering
- Multiple years of management experience in Networking and Insurance industries
- Provide training at multiple fortune 500 companies on wide variety of topics
- **Patents**
  - Detecting and mitigating a high-rate distributed denial of service - Approved
  - Individually assigned server alias address for contacting a server - Pending

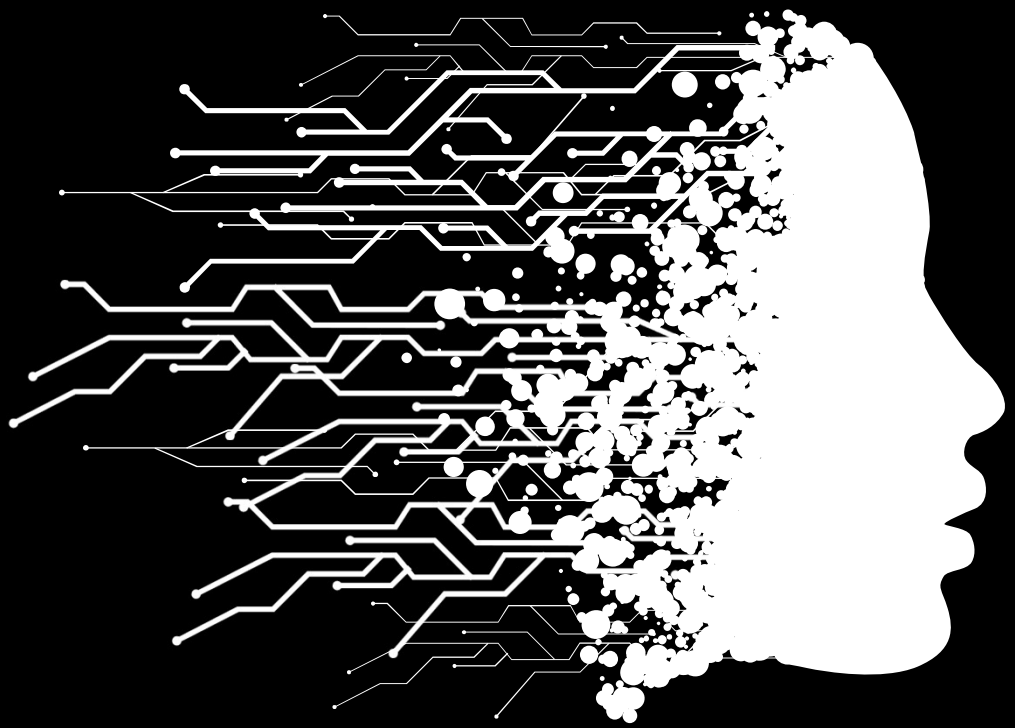
Your turn

Who? Name, Role & Group

What? Overall experience & Main Skill (Any Python, or ML Experience)

Why? Are you here (Key Motivations)

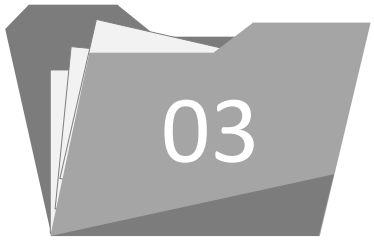
How? plan to use this (if already known)



Hands-on Labs

Labs Folder Structure

Assignment
Contains notebooks:
instructions for assignment



Solution
Contains solution notebooks

Code Samples
Contains code examples for
different topics.

Samples & Exercises

Folder	File Name	Dataset	Topic
CodeSamples	*Python Introduction*		Python Examples
CodeSamples	*Pandas Introduction*	*iris.csv*	Numpy and Pandas
CodeSamples	*Matplotlib Intro*	*iris.csv*	Matplotlib and Seaborn
CodeSamples	*Plotly Example*	*iris.csv*	Plotly - Interactive charts
CodeSamples	*Pandas-Profiling-Visualization*	*Telco-Customer-Churn.csv*	Autoviz, Pandas Profiling, Datasist
CodeSamples	*LinearRegression*	*50_Startups.csv*	Linear Regression
Assignment	*LinearRegression RealEstate*	*Boston dataset*	Implement Linear Regression
CodeSamples	*LogisticRegression*	*agent.csv*	Logistic Regression & Model Persistence
CodeSamples	*LoadModel-Predict*	*agent-new.csv*	Load pipeline, model, predict
CodeSamples	*PetFinder*	*pet-train.csv* *pet-test.csv*	Using wordcloud
CodeSamples	*Tensor Introduction*		TensorFlow (Tensors)
CodeSamples	*GradientDescent*		Gradient Descent
CodeSamples	*NeuralNetworks-Regression*	*auto-mpg (UCI)*	Neural Networks Regression
CodeSamples	*PredictPetFinder*	*pet-train.csv* *pet-test.csv*	Neural Networks Classification
CodeSamples	*ImageAugmentation*	*home-x.jpg*	Image Augmentation
CodeSamples	*NeuralNetworks*	*Keras MNIST Digits*	Neural Networks
CodeSamples	*CNN*	*Keras MNIST Digits*	Convolutional Neural Networks
Assignment	*CRM-CustomerChurn*	*CRM Dataset Shared.tsv* *CRM Churn Labels.tsv*	Classifier(s) and Neural Networks
Assignment	*CNN-Fashion*	*Keras MNIST Fashion*	Convolutional Neural Networks
CodeSamples	*SpacyExamples*		Using Spacy - text processing
CodeSamples	*SpacySentimentAnalysis*	*yelp_labelled.txt* *amazon_cells_labelled.txt* *imdb_labelled.txt*	Spacy -text processing & classification
CodeSamples	*TextClassification*	*IMDB Dataset.csv*	Recurrent Neural Networks

Lab Environment Verification

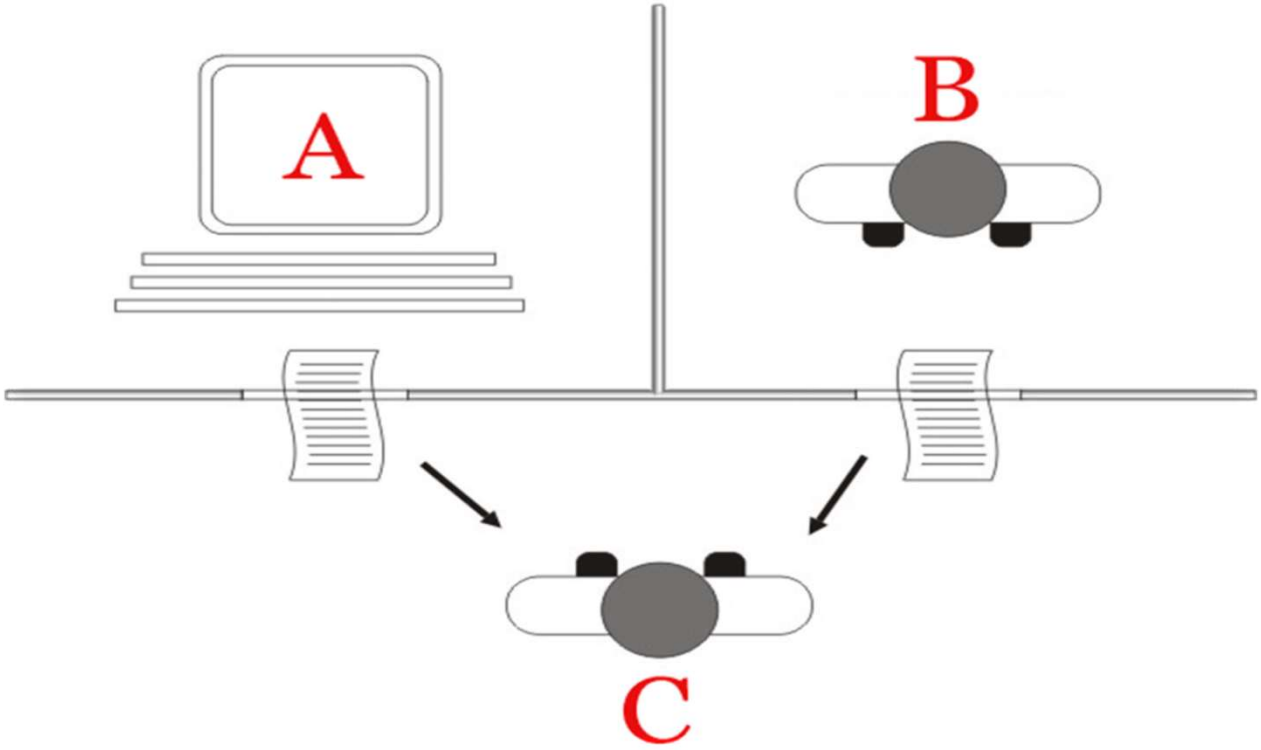
Notebooks

- Open notebook – "CodeSamples/Python Introduction"
- Demo of using Notebook
- Python refresher code – commonly used data structures, functions

When did Artificial Intelligence start?

Your thoughts?

Turing Test (1950)



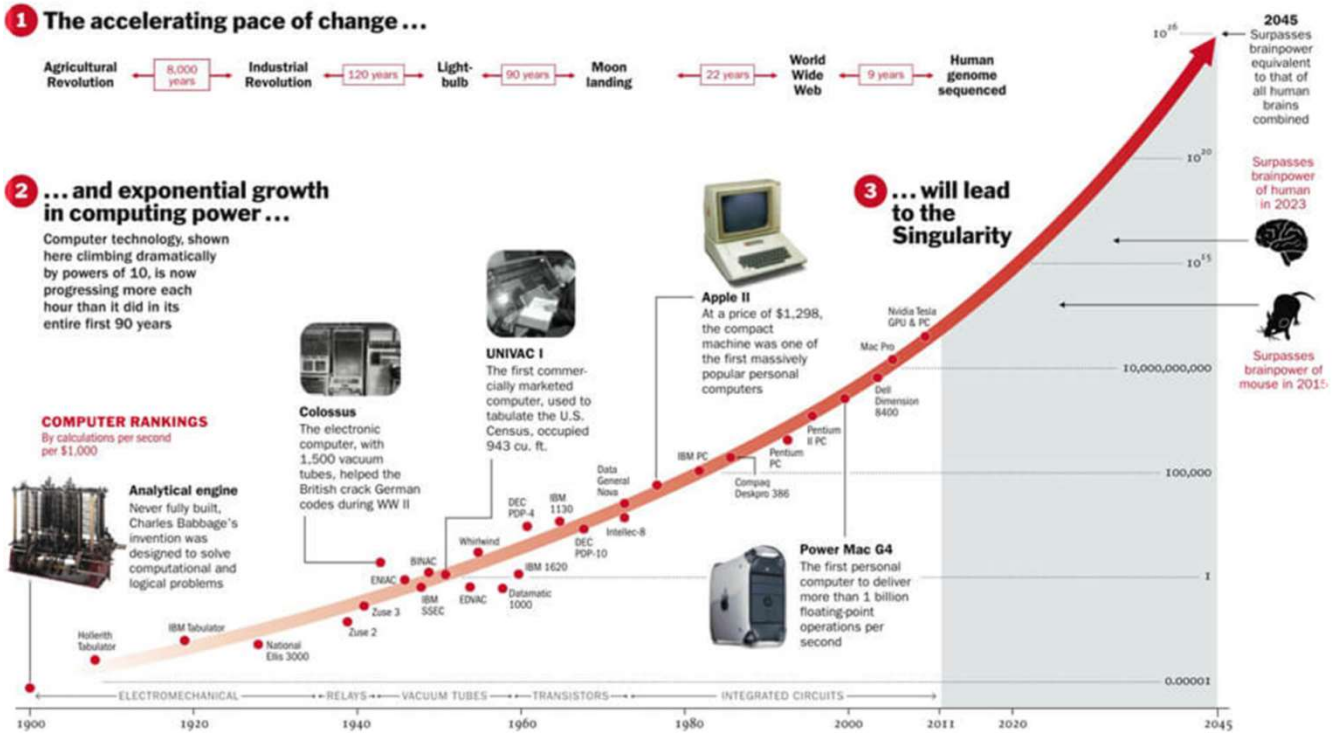
History of Artificial Intelligence

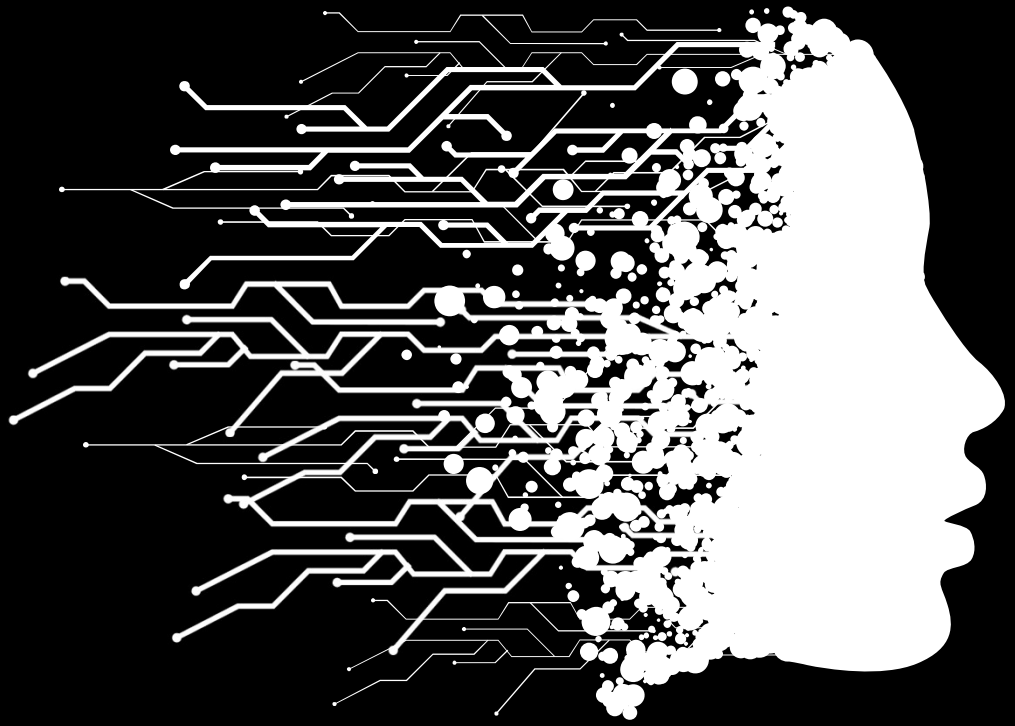
- 1950 – Turing Test
- 1951 – First Neural Network
- 1967 – “Nearest Neighbor” Algorithm is written
- 1974 – First AI winter
- 1979 – Stanford Cart
- 1997 – IBM Deep Blue beats Garry Kasparov
- 2006 - Geoffrey Hinton coins the term “deep learning”
- 2014 – Facebook develops Deep Face & Google Buys DeepMind
- 2015 – Amazon, Google & Microsoft ML offerings
- 2016 – AlphaGo beats Lee Sedol
- 2030 – Singularity?

Source https://www.bbc.com/timelines/zypd97h

www.bigdatatrunk.com

Singularity 2045?





Artificial Intelligence, Machine Learning, Deep Learning

Concepts & Terminology

Artificial Intelligence

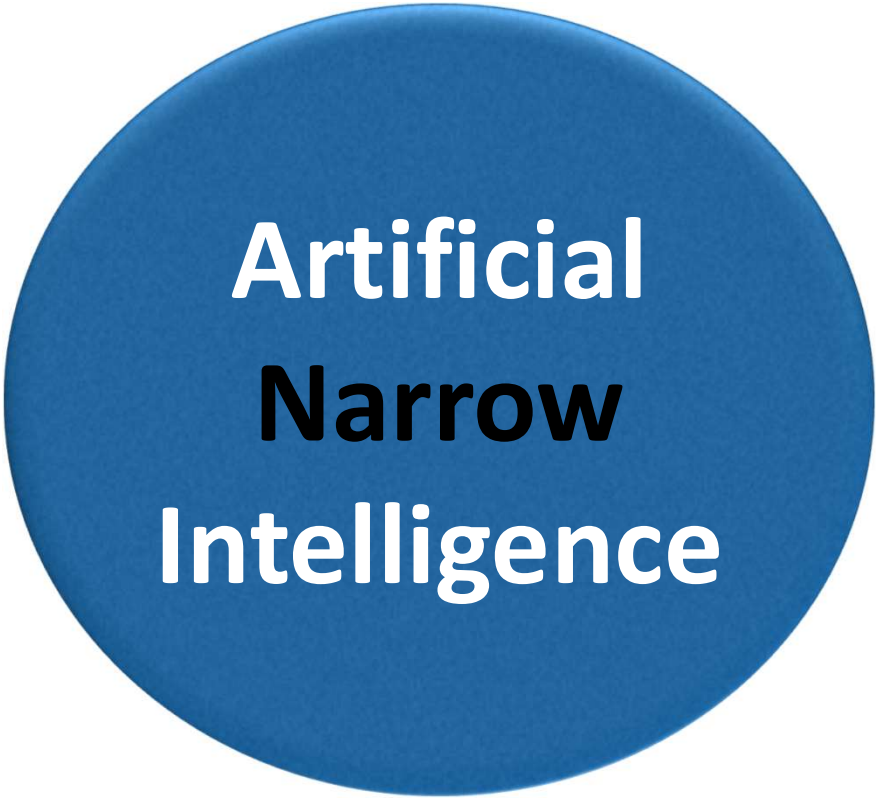
What is Artificial Intelligence?

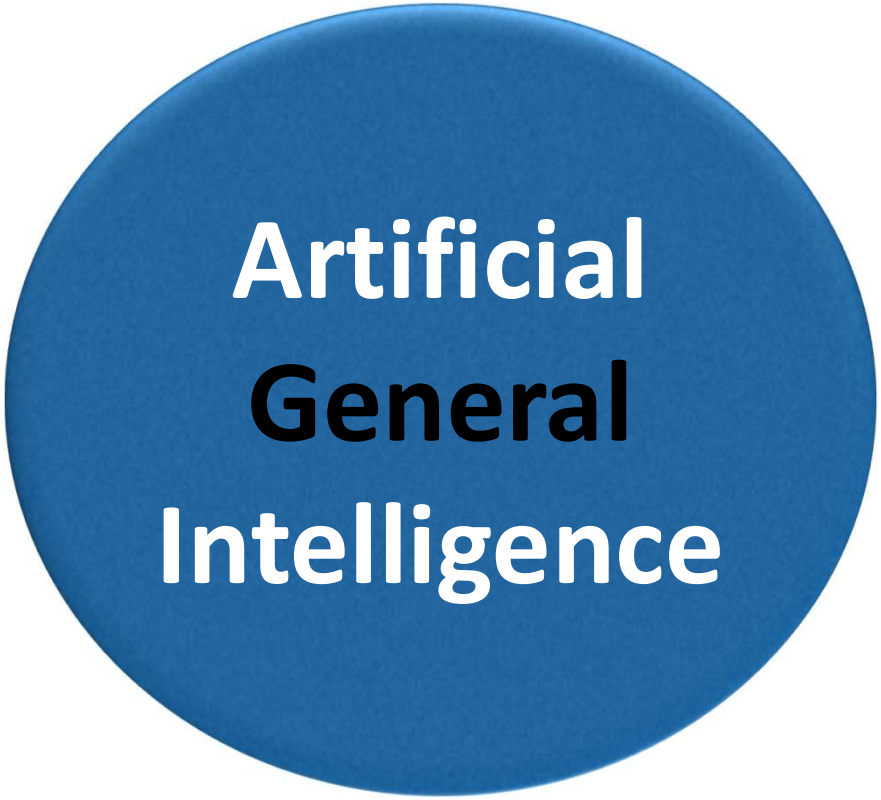
AI is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers "smart"

Applications include:

- Game playing
- Natural language processing
- Image Recognition

Artificial Intelligence - Types





Machine Learning

What is Machine Learning?

Humans learn from our past experiences

Machine follow instructions given by humans

What if we humans can train machines to learn from the historical data?

Instead of programming a computer you give a computer examples and it learns what you want

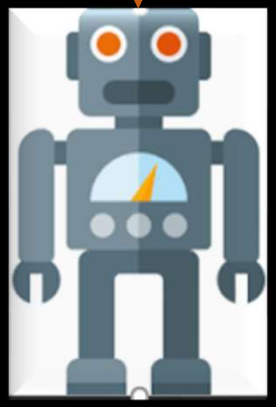
Machines Learn by Example

User ID	Gender	Age	Estimated Salary	Purchased
15624510	Male	19	19000	Yes
15810944	Male	35	20000	No
15668575	Female	26	43000	No
15804002	Male	19	76000	No
15728773	Male	27	58000	No
15598044	Female	27	84000	No
15694829	Female	32	150000	Yes
15600575	Male	25	33000	No

Learns from the data

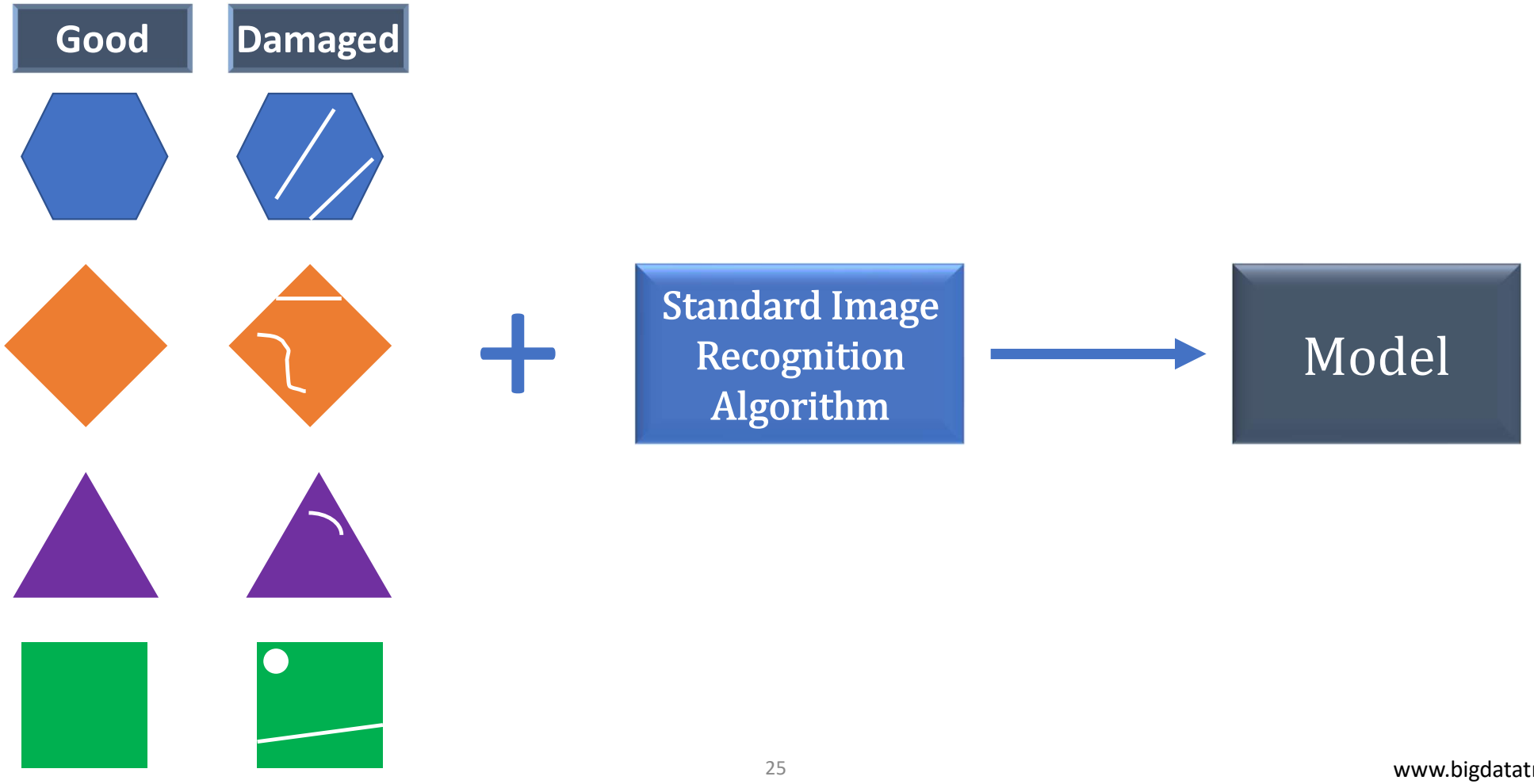
New Input Data

Gender, Age, Estimated Salary

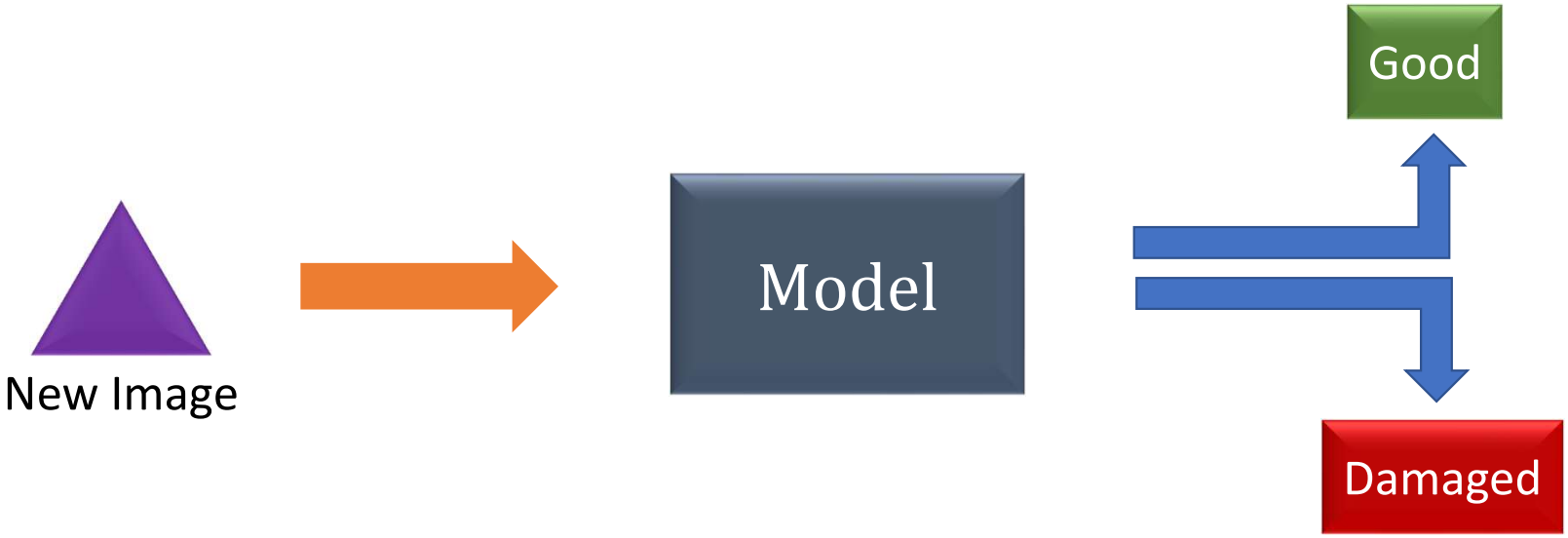


Purchase? Yes or No

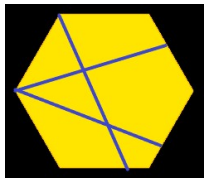
Machines Learn by Examples



Predict with Trained Model



Same Algorithm Different Trained Models



Flower Model (Trained)

≠

Parts Model (Trained)

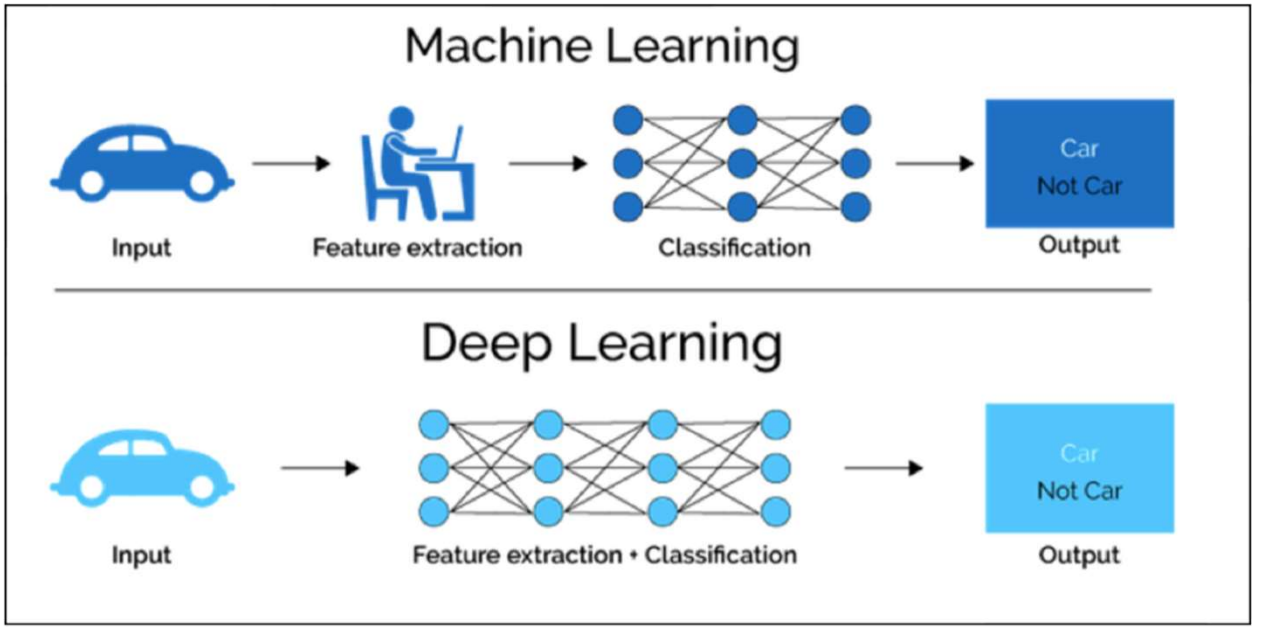
Iris Setosa

Damaged

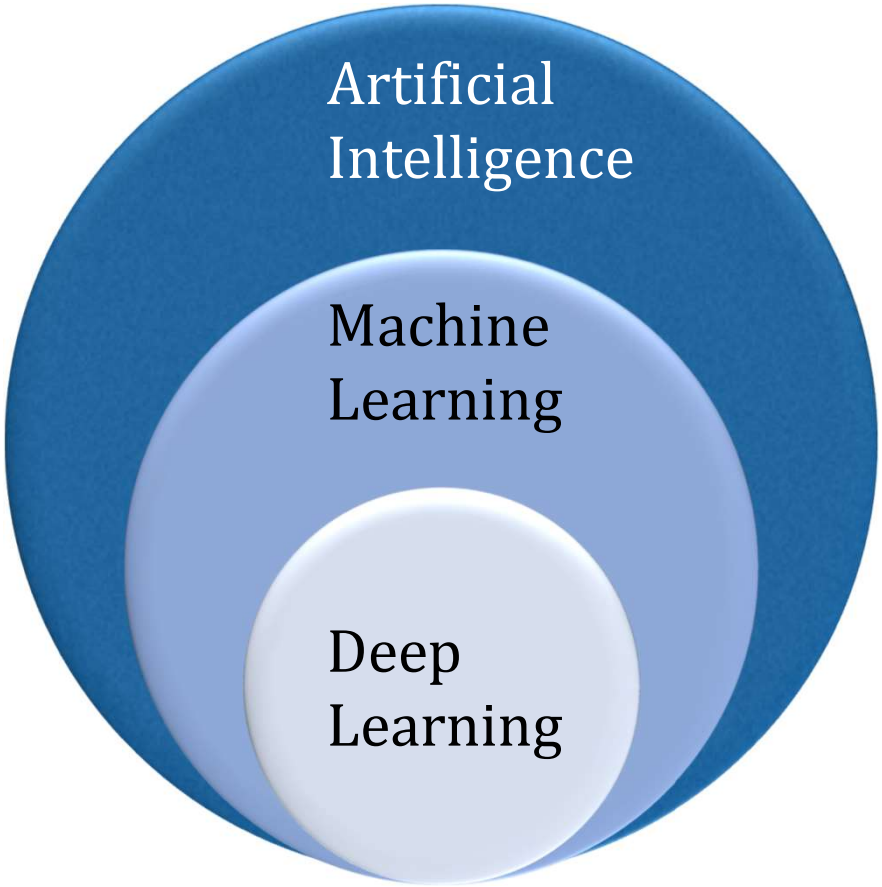
Deep Learning

What is Deep Learning?

Machine Learning v/s Deep Learning



Bringing Them Together



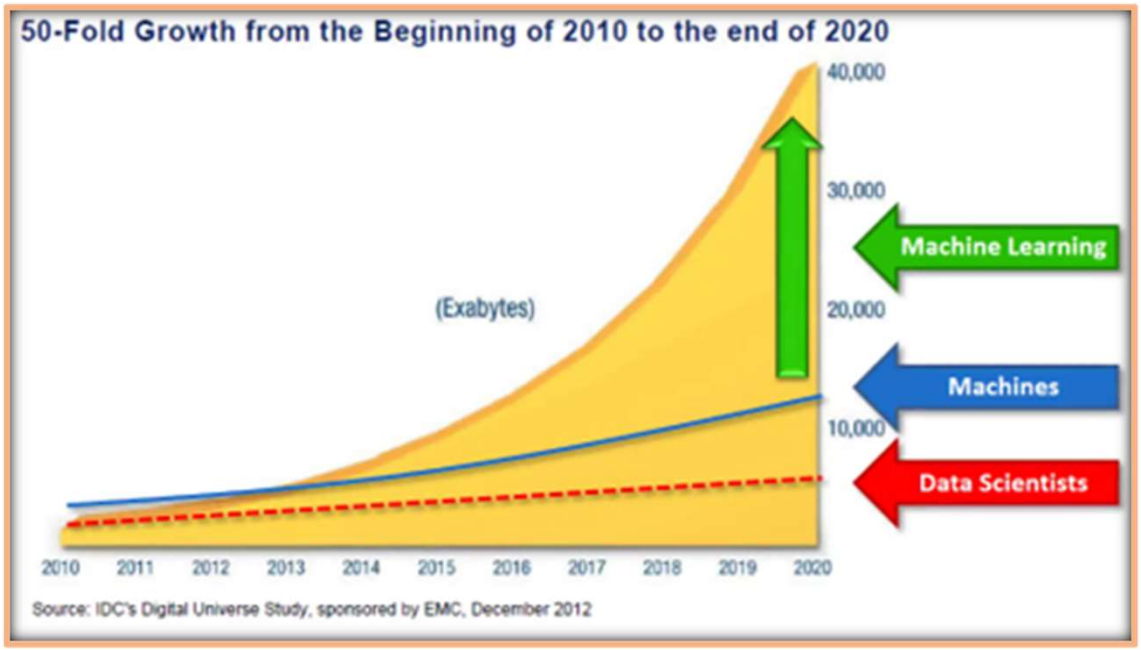
Intelligent Behavior

Learning from data to make predictions and gain insights

Learning using Artificial Neural Networks

Why is it popular in recent years?

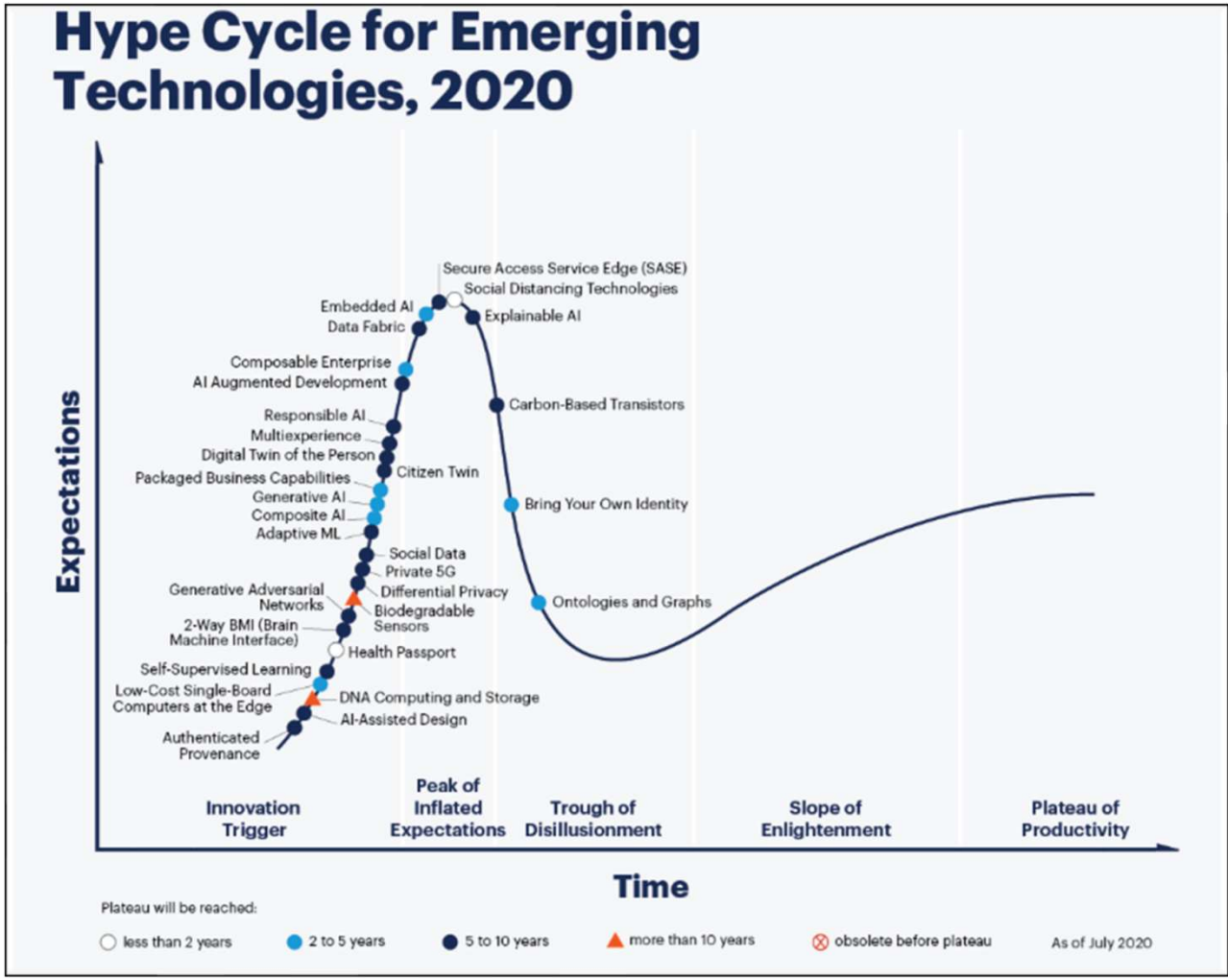
1 terabyte = 1,000 GB
1 petabyte = 1,000 terabytes
1 exabyte = 1,000 petabytes



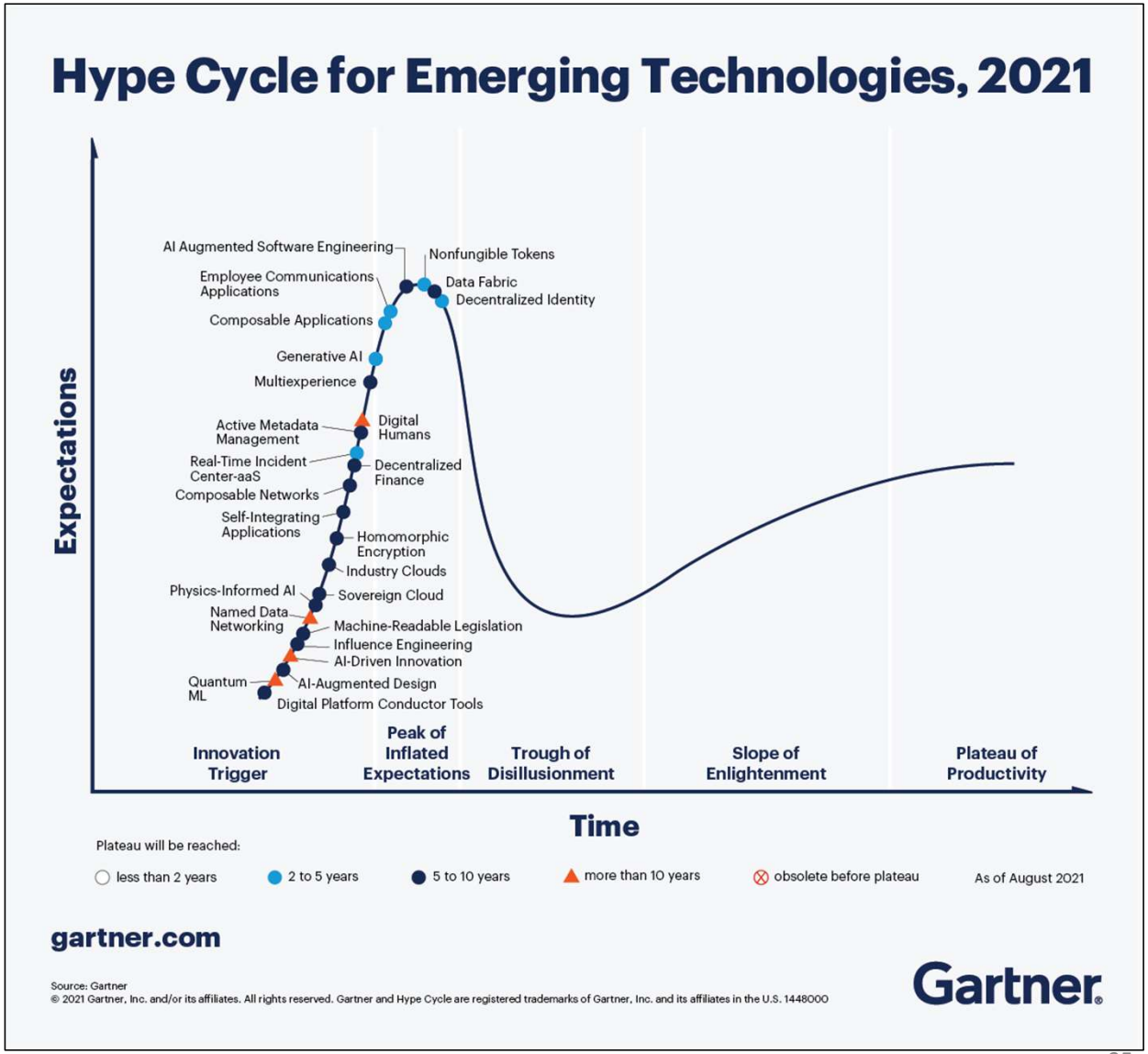
Industry Use Cases

Finance and risk	Sales and marketing	Customer and channel	Operations and workforce
$$$ Revenue Forecasting	Sales forecasting	User segmentation	Agent allocation
Portfolio optimization	Demand forecasting	Personalized offers	Warehouse efficiency
$$$ Investment modelling	Sales lead scoring	Product recommendation	Smart buildings
Fraud detection	Marketing mix optimization		Predictive maintenance
Risk management			Supply chain optimization

Gartner: 2020 Emerging Technologies



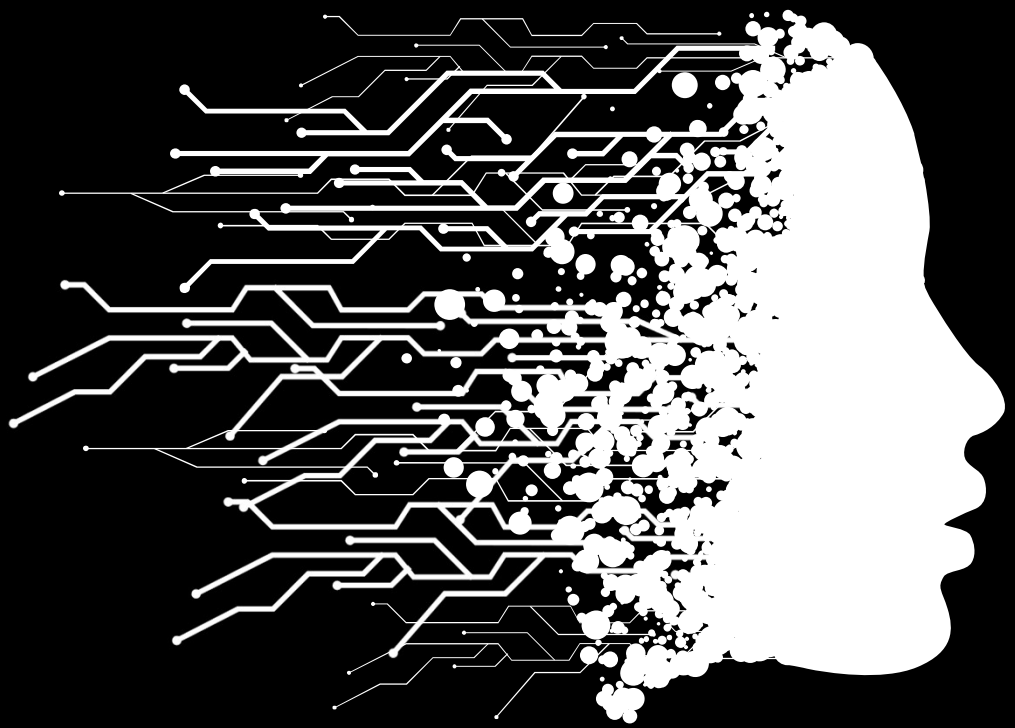
Gartner: 2021 Emerging Technologies



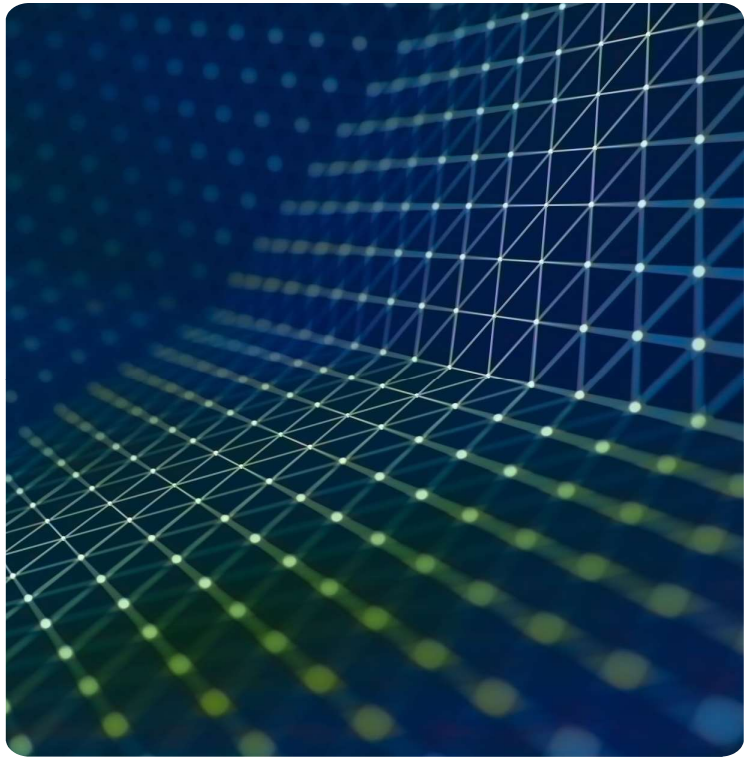
Hands-on

Pandas and Numpy

- Select “Jupyter Notebook” from start menu
- Open file Labs/CodeSamples folder ‘**Pandas Introduction**’ using Jupyter
- Numpy and Pandas examples



Data Exploration and Visualizations



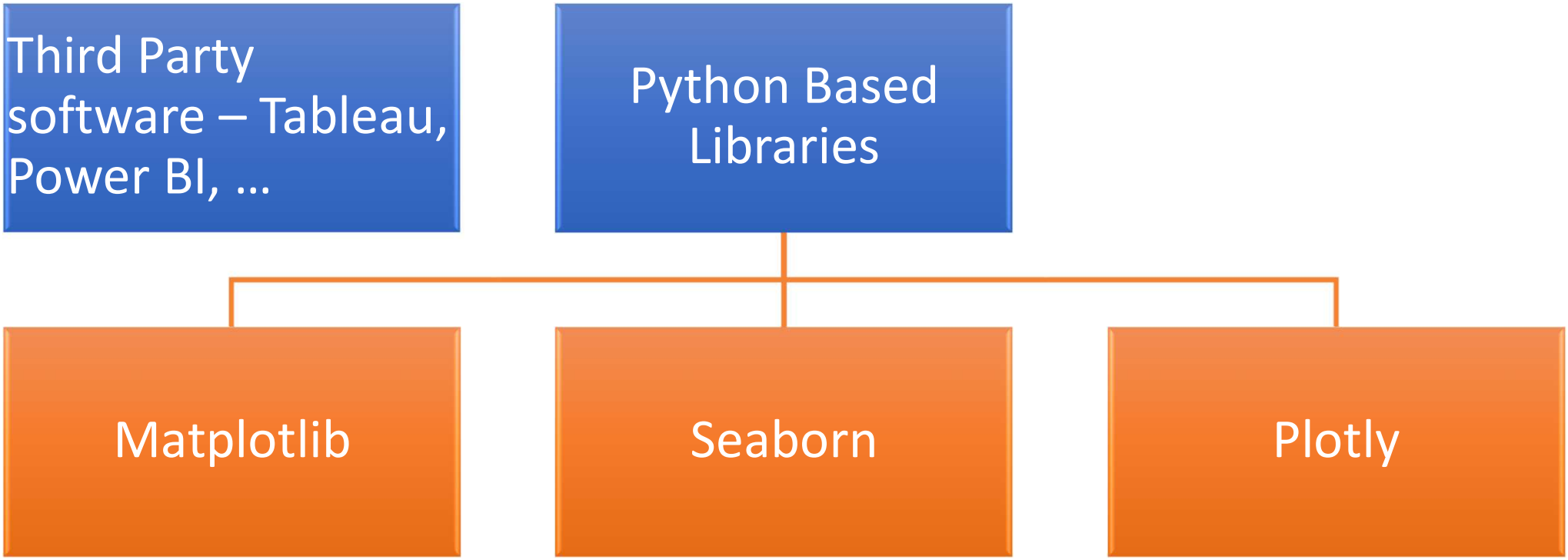
Data Visualization

Picture is worth 1000 words …

Visualizing the data is heavily used during the data exploration phase

There are multiple libraries that **help visualize data**

Data Visualization Tools



Matplotlib Examples

- Open file '**CodeSamples/Matplotlib Intro**' using Jupyter
- Creating line, scatter, histograms plots
- Use seaborn library
- Plot Time Series Data

Using Interactive Charts

- Open file '**CodeSamples/PlotlyExamples**' using Jupyter
- Using the Plotly library to create charts that we can zoom into, rotate it, etc.

Data Exploration

Data exploration is the initial step in data analysis

Want to explore any size of dataset to uncover patterns, characteristics

Objective is to get a broad view of the data

Methods include:

- Writing code and using utilities
- Performing data visualizations

Data Exploration

Data can be of various types

- Numbers – integers, float
- Textual

Software

- Tableau, Power BI, ….
- Python based libraries
  - Pandas, Numpy
  - Number of text processing utilities like wordcloud, spacy, ….

Tools to Profile and Visualize Data

- Look at couple of tools to Profile and Visualize data by writing minimal code
- **AutoViz**
  - Load a dataframe and the plot continuous variables in them
  - Plot with or without the target variable
- **Pandas Profiling**
  - Generates HTML report based on Pandas Dataframe
  - Report contains information about:
    - Overall summary (rows, columns, categorical columns, numerical columns, etc.)
    - Variables: summary for each feature including a chart
    - Correlations: feature correlations
    - Missing Values: Display count of missing values (np.nan, null)

Tools to Profile and Visualize Data

- **Datasist**
  - Open-source python library for doing data analysis, visualizations, modelling, etc.
  - Has features such as:
    - Visualization
    - Model – classifiers, regressors, compare models, feature importance, etc.
  - Will look at visualizations

Profiling and Visualizing

- Open file '**CodeSamples/Pandas-Profiling-Visualizations**' using Jupyter
- Examples of using
  - AutoViz
  - Pandas Profiling
  - Datasist

Machine Learning

Data

Data

Structured Data

User ID	Gender	Age	Estimated Salary	Purchased
15624510	Male	19	19000	Yes
15810944	Male	35	20000	No
15668575	Female	26	43000	No
15603246	Female	27	57000	Yes
15804002	Male	19	76000	No
15728773	Male	27	58000	No
15598044	Female	27	84000	No
15694829	Female	32	150000	Yes
15600575	Male	25	33000	No
15727311	Female	35	65000	No
15570769	Female	26	80000	No
15606274	Female	26	52000	No

Structured Data Types

Numerical

- Exact numbers as data points
- Two types
  - ***Continuous***
    - Can assume any value within a range
    - Height, weight, salary, etc. e.g. 6.3, 40.32, 32.5
  - ***Discrete***
    - Data has distinct values
    - Number of students, units sold, etc. e.g. 15, 30, 10

Categorical

- Qualitative data that has no inherent mathematical meaning
- Can assign numbers to categories in order to represent them, however the numbers have no mathematical meaning
- 1 for red, 2 for blue, 3 for green

Ordinal

- A mixture of numerical and categorical data
- Categorical data has the mathematical meaning
- Movie ratings in scale 1 to 5
  - Ratings can be 1, 2, 3, 4 or 5
  - 1 means worst rating, 5 being the best

Unstructured Data

```
[ [216, 203, 125,  10,  84, 241, 149, 159, 212, 118, 135, 158,  11,  91,  36, 177, 176, 253, 132, 210, 159,  20, 153, 131, 132,  55, 16, 132],
  [184,  34,  95, 225,  60, 218,  49, 193,  93, 119,  68, 133, 195, 104, 248,  18,  18, 136,  90,  71,  81,  41, 233,  53,  46,  87, 86, 243],
  [ 85,  61, 220, 170, 206,  34, 141,  97,  66, 217, 124, 143, 241, 205,  76, 123,  66,  72, 231, 116, 244,  74, 155, 144,  47, 230, 171, 165],
  [156,  87, 181,  90, 160,   2, 184, 112, 108,  62, 223, 153,  93, 244,  83, 187,  83,  18, 134,  28, 121, 244, 202, 176, 228, 233, 76,  13],
  [ 76, 238, 128, 183, 119, 130,  34,  12, 112, 254,  90, 167,  64,  89, 170, 221, 196,  69,  82,  11,  65,  86, 254, 111, 134,   0, 148, 246],
  [105, 178, 254,  31,  32, 133,  57,  40,   6,  85, 115,  56, 132,  64,  35, 119, 158, 182, 106,  77,  84, 106, 164, 230,  54,  42, 55, 130],
  [ 25,  86, 222,  59, 242, 111,  59, 183, 236, 214, 251,   7, 142,  90, 179,  80, 163, 159,  26, 143, 108, 109, 229, 223, 220, 196, 21,  18],
  [ 21,  42, 109, 188,  91,  93, 246, 236, 125,  48, 151,  12, 178,  26, 118, 135,  77,  84, 179, 208, 114, 224,  99, 246,  68,  21, 69,  39],
  [253,  66,  78,  55,  39, 107, 248,  90, 124, 107,  51,  92, 150, 234,  91, 177, 146,  80,   8, 179, 148, 229, 233,  59, 164, 199, 252,  43],
  [ 79,  60,   5,  70,  37, 218,  19,   9,  90,  74, 198, 129,  61, 160, 206,  11,  37, 171,  44, 241, 228, 190, 232,  99,   7, 100, 83, 225],
  [211,  38,  52, 167, 206, 139, 215, 209, 202, 102, 122,  77,  86, 117, 134,  22, 176,  94,  22, 201,   6,  73, 156, 226,  36,   0, 50, 119],
  [159,  24, 197, 215,  16, 243, 177,  13, 108, 211,   6,  97,  75, 214, 121,  92, 154, 109, 213, 163, 123,  20, 190, 174,  89,   6, 136, 164],
  [183, 136, 245, 175, 233,  62, 141, 117, 150,  74, 182, 175,  36, 230,  93, 109, 212,  43,  10,  75, 234, 124,  70, 244, 161,  76, 241, 223],
  [150,   7, 184,  20, 133,  22, 112, 212,  48,  30, 156, 113, 127, 207, 219, 173, 223, 127, 202, 172,  39,  98, 134, 124, 130,  34, 210, 101],
  [101,  77,  87,  37, 152, 112,  34, 106,  30,  23,  79, 214, 245, 152, 129, 243, 109, 213, 170, 190, 220,  25,  76, 205, 135, 227, 225, 165],
  [108, 184, 172, 121,   8,  83, 106, 116, 236,  55,  73, 204,  50,  40, 124, 153, 225, 157,  13,  28, 105,  62, 242, 214,  56, 159, 137,  67],
  [ 14,  75,  26,  47,  74, 205,  46, 219,  27,  18,  79,  28,  49, 224,  85, 214, 180, 106, 183,  87,  18,  64,   7,  61, 125,  87,  38,  98],
  [122, 146,   4,  72, 150, 249,  77,  90,   6, 132, 134, 151, 164,  29,  94, 188, 251, 177,   0, 206, 193, 182, 231,  43,  32,  32,  80, 147],
  [ 26,  39,  76,  12,  35,  81, 103, 233, 204, 136,  82,  28,   5,  68, 229, 197,  52, 215, 224, 117, 101,   4, 154,   4, 205,  50, 251, 114],
  [ 68, 176,  23, 246,  11,  57,  62,  25,  38,  17, 136, 106, 113, 140, 254,  43, 231, 150,  12, 114,  77,   8, 214, 187,  92,  66, 195,  70],
  [ 20, 241, 148, 151,  37,   4,  14, 231, 225,  53, 232, 240, 223,  59, 234, 134, 247, 242, 212,  63, 201,  38,  63, 200, 128, 139, 167, 173],
  [ 80, 244,  33, 111, 143, 127, 168, 237, 189,  63, 126, 181,  92,  91,  14, 211,  21,  26, 253, 109, 174, 100, 138, 138, 221, 204, 29, 230],
  [ 81, 174, 217,  93,  65, 134,   7,  36, 176, 122, 226,  23, 223,  28, 202,   5,  54, 205, 169,  14,  88, 178,  84, 198,  96, 201, 230, 193],
  [215, 168, 125,  92,  70, 151, 183, 210,  36,  32,  19,  51,  42,  64,  19, 146, 183, 246,   0, 184, 236,   7, 226, 118, 113, 241,  85,  89],
  [ 31, 158, 210,  16, 199,  58, 224,   7, 203,  86, 103,  45,  28,  54,  92, 204, 243, 117,  75, 208, 248, 223,  87, 250,  14,  43, 102,  66],
  [ 13, 236, 138,  67, 236, 109, 113,  46, 115,  19, 214, 154, 199, 248,  55, 172, 214, 249, 125, 154, 139, 141, 188,  78, 107, 200, 196,  16],
  [ 65, 150, 158, 254, 114, 177, 120,  15,  65,  58,  79, 171, 118,  32, 250,  81,  27,  85, 128, 146, 144, 234, 139,  26,   6,  68, 133, 205],
  [123,  68, 216,  34, 139,  34,  34, 175, 213,  72,  76,  19,  32, 138, 132, 111, 242, 249, 177,  89,  61,  72, 252,  79,  20, 171, 174, 177] ]
```

Best Data Qualities

Clean

Coverage

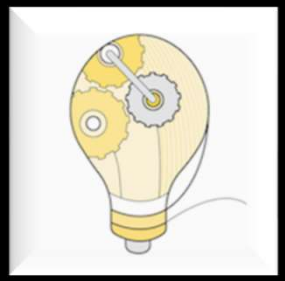
Complete

Otherwise, Garbage in Garbage out

Even for Images

Machine Learning Offerings

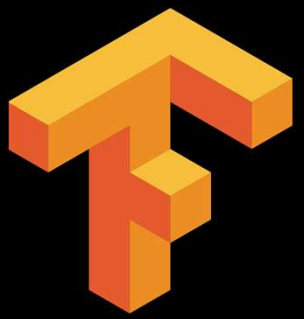
Cloud Based Offerings



Programming Languages

Python Based Libraries

Machine Learning

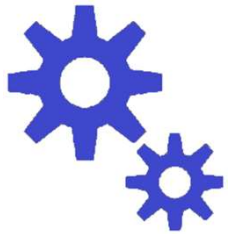


Deep Learning

Traditional Software Development v/s Machine Learning

Traditional Software Development

Define an algorithm/logic to solve problem



Implement the algorithm/logic in code

Implemented algorithm

Apply

Input Parameters

Implemented algorithm

Results

Traditional Software Development

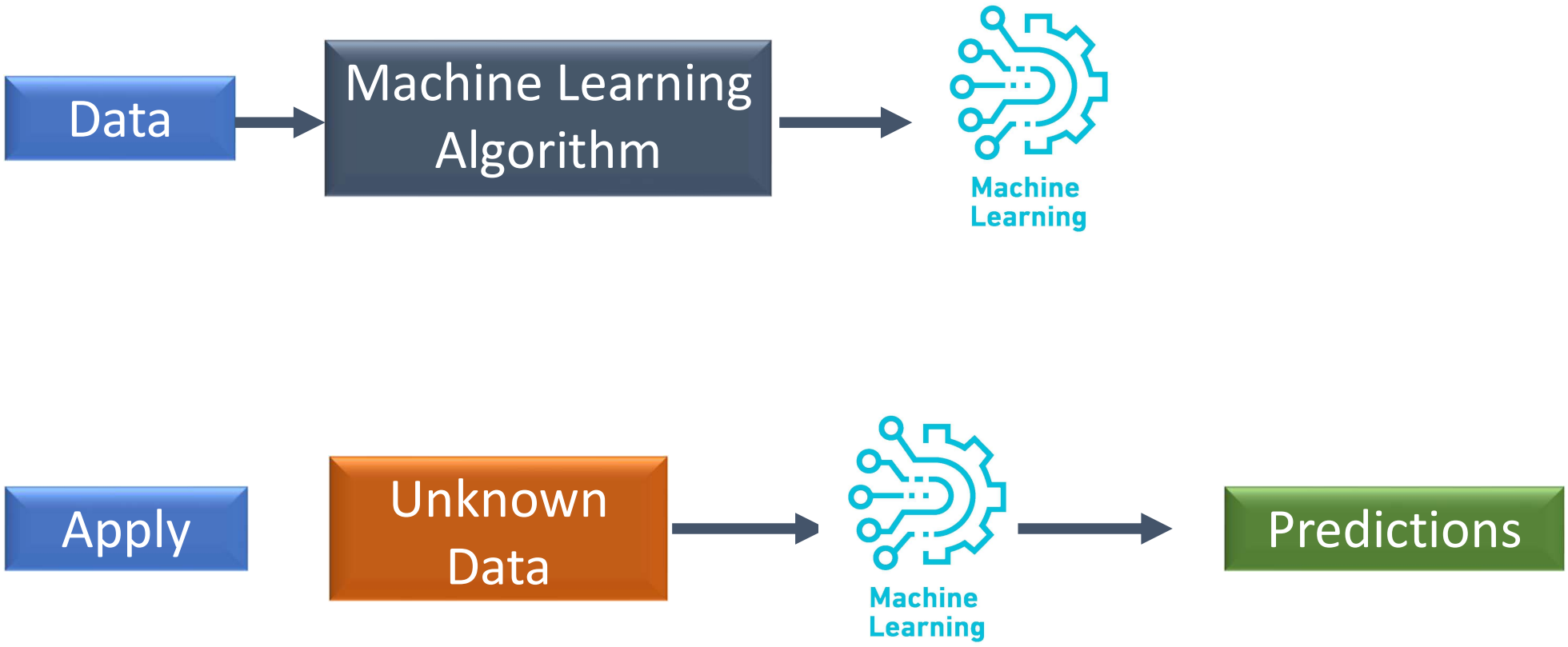
Convert inches to centimeters (cms)

Input: **inches**

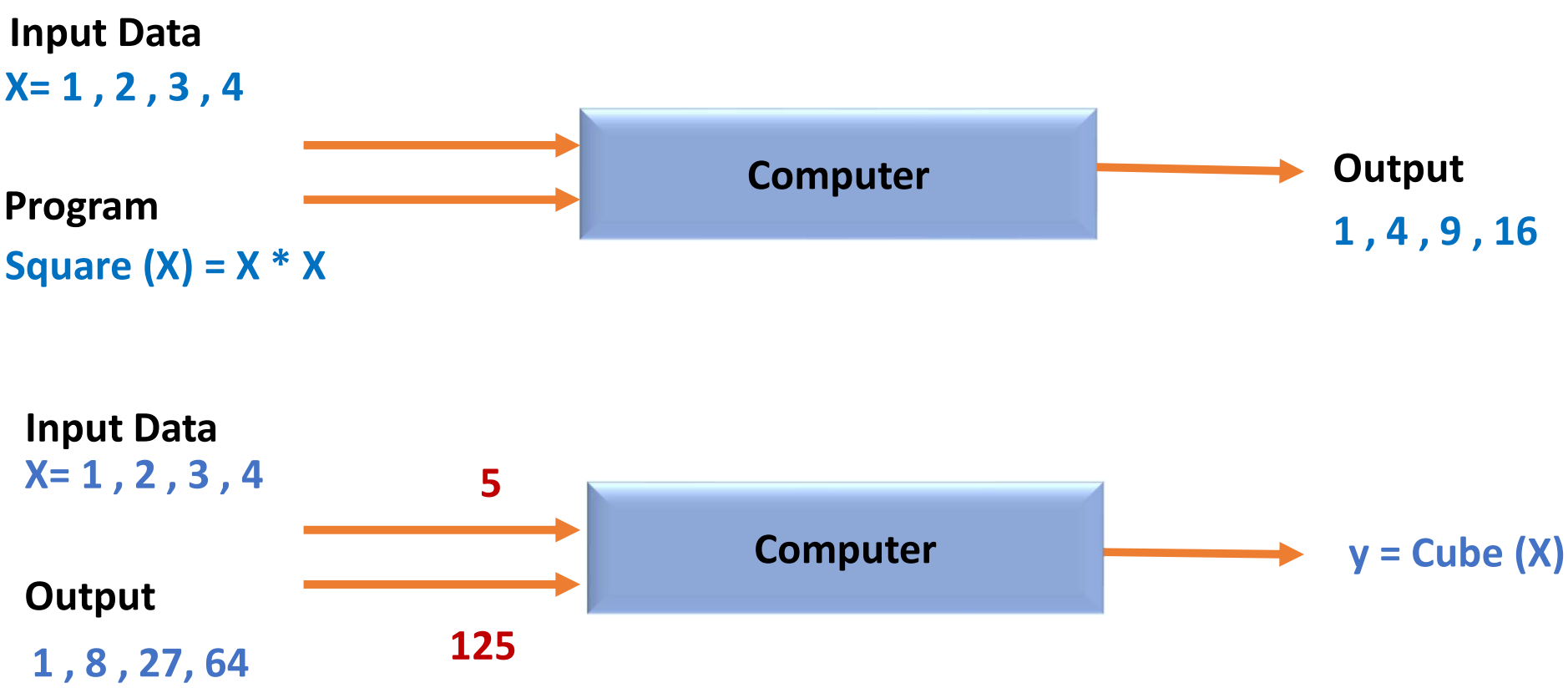
Relationship: **cms** = **inches * 2.54**

Output: **cms**

Machine Learning Development

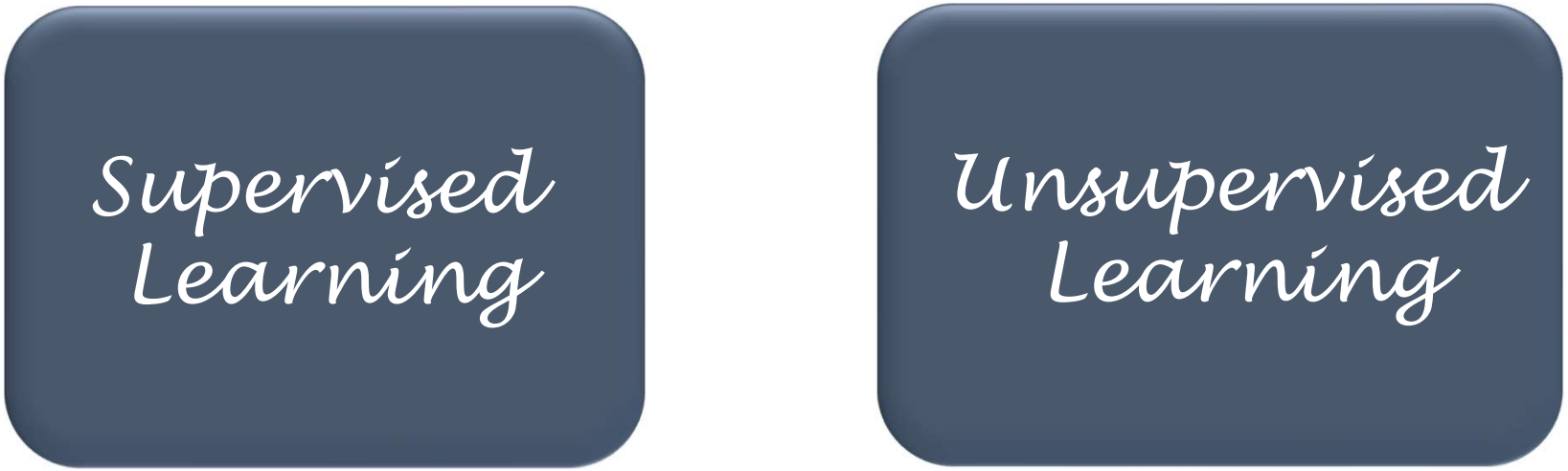


Programming v/s Machine Learning



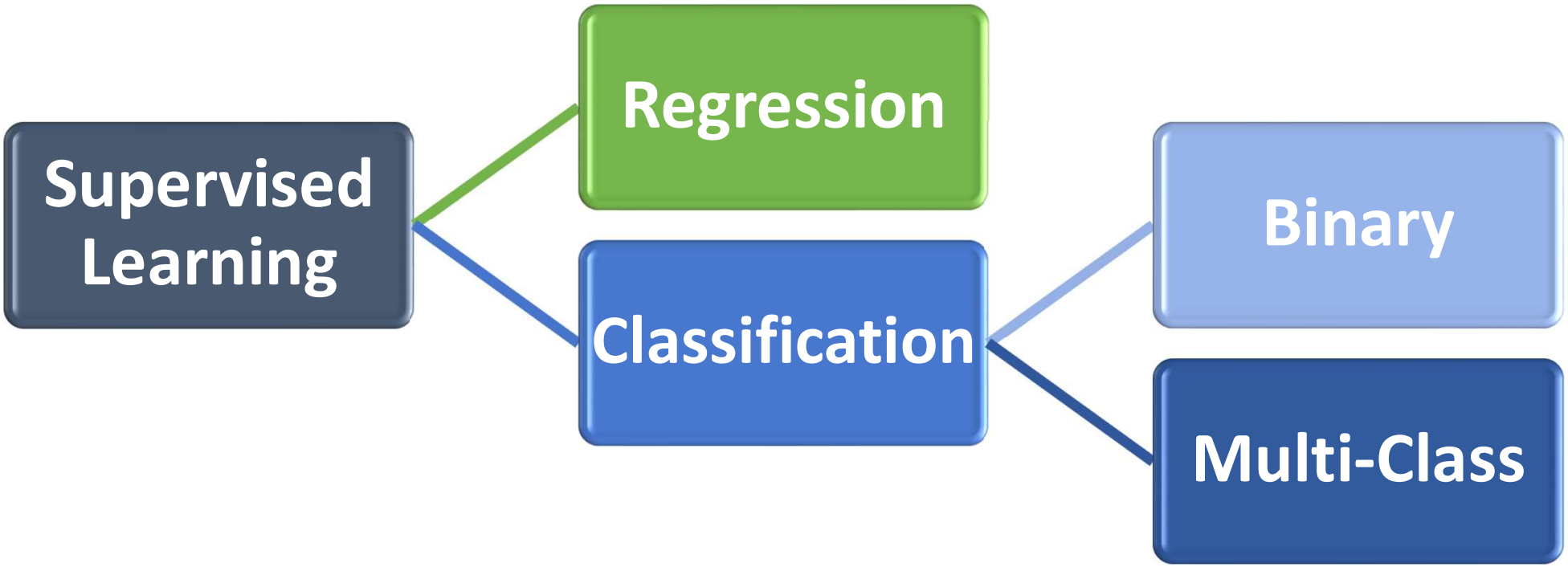
Machine Learning Techniques

Machine Learning Techniques



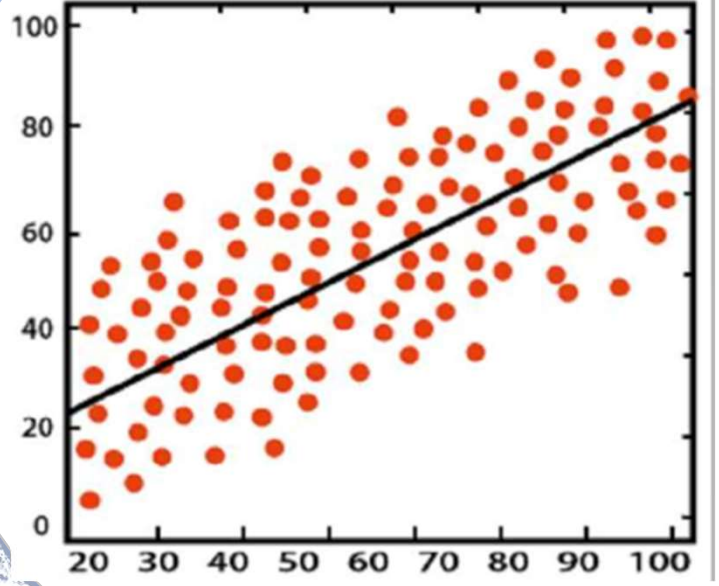
Supervised Learning

Supervised Learning – Predict a Label



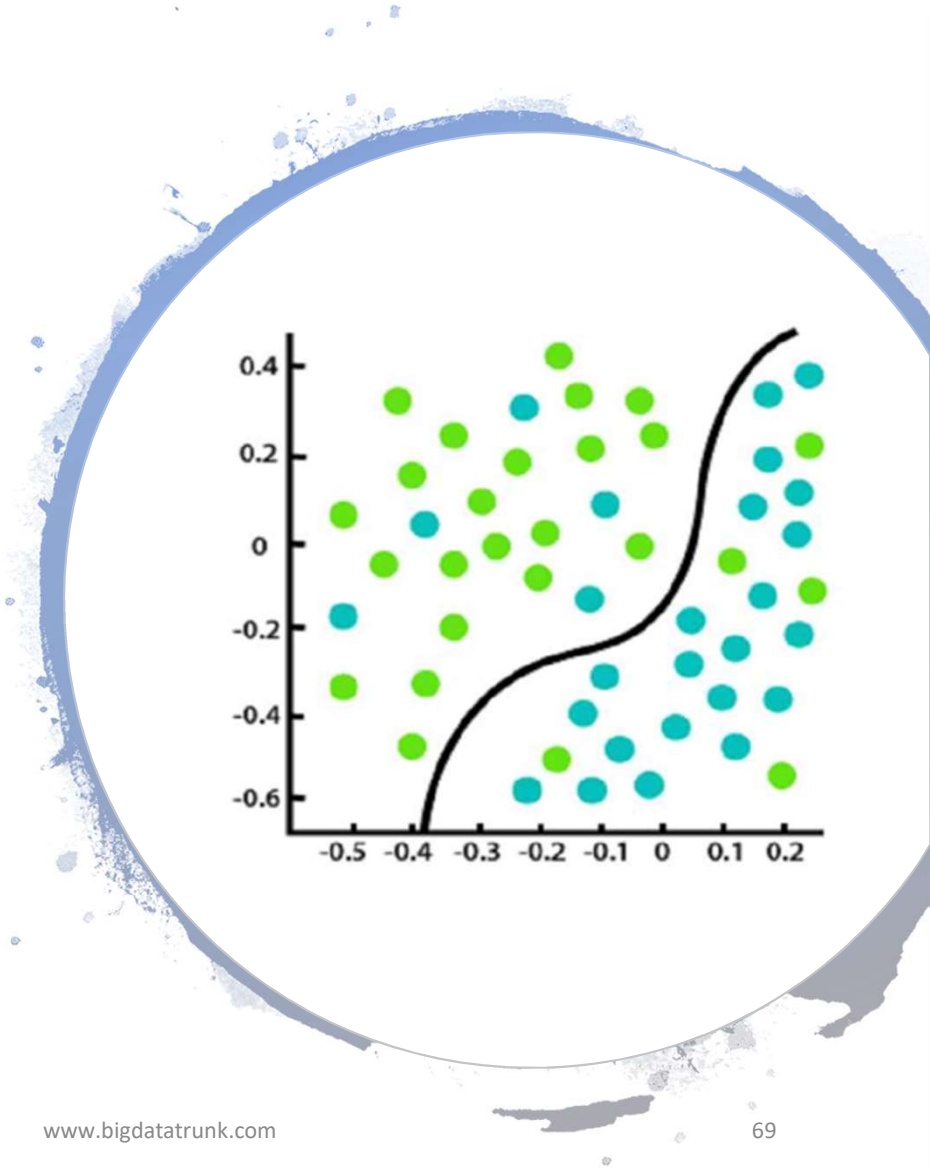
Regression

- **Use cases**
  - Forecasting Sales
  - Stock, Real Estate Prices
  - Auto/Home Insurance premiums
  - Customer support – number of issues over a time period, etc.
- **Algorithms**
  - Linear Regression
  - Decision Tree
  - Random Forest

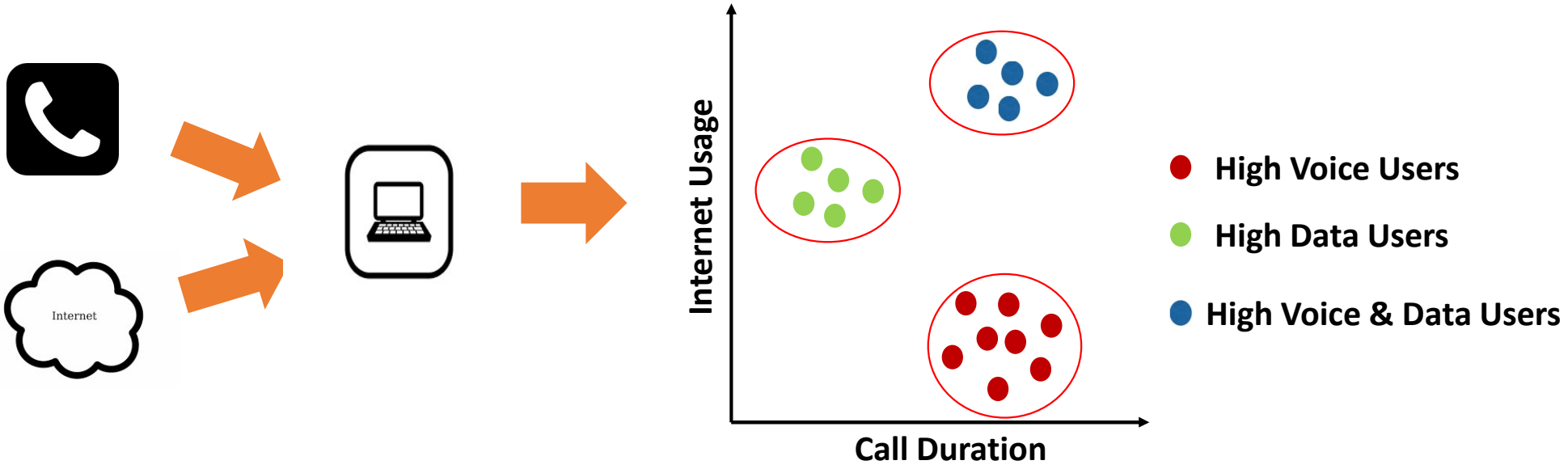


Classification

- **Use cases**
  - Patient is likely to have diabetes or not
  - Transaction is fraudulent or not
  - Which price quotes are likely to turn into orders
- **Machine Learning Algorithms**
  - Decision Tree
  - Logistic Regression
  - Random Forest
  - Support Vector Machine
  - Many more ….



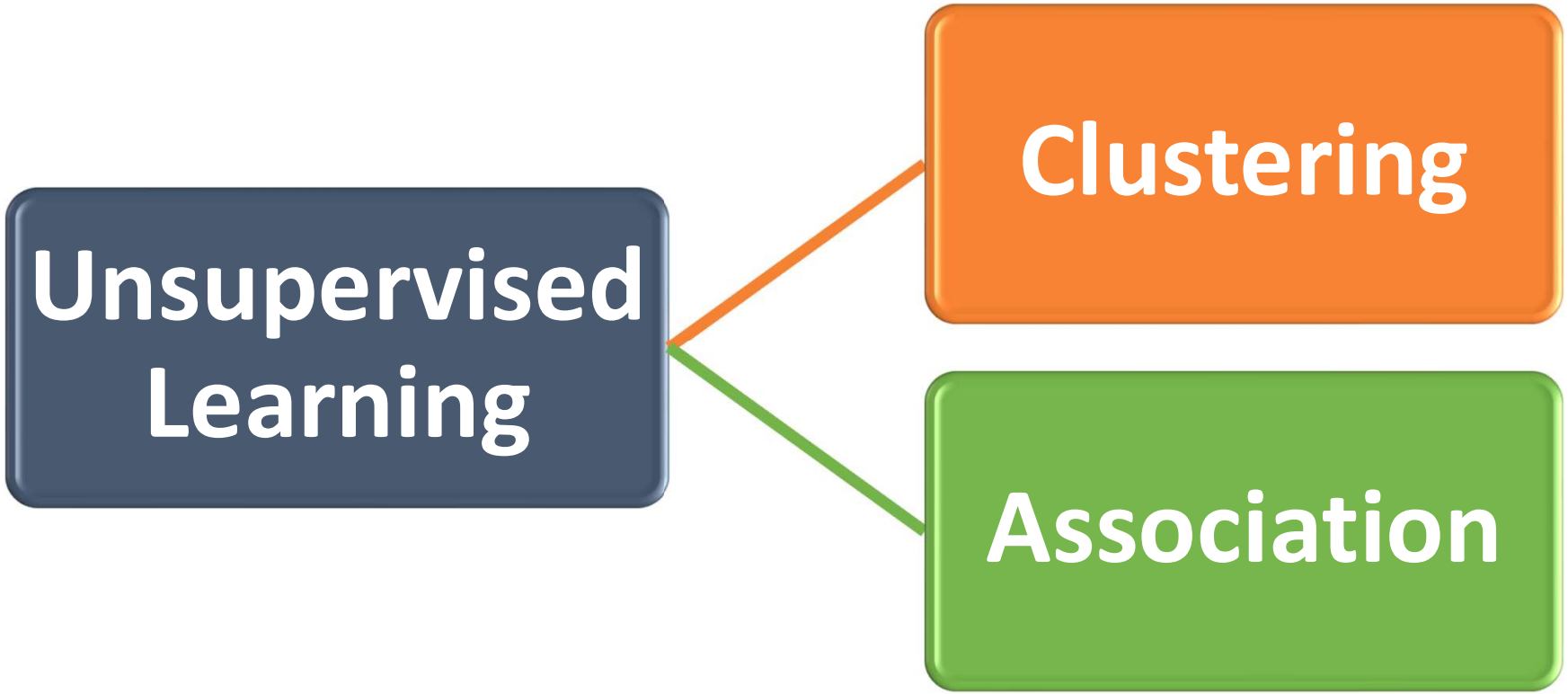
Unsupervised Learning



Unsupervised Learning – Creating grouping of data

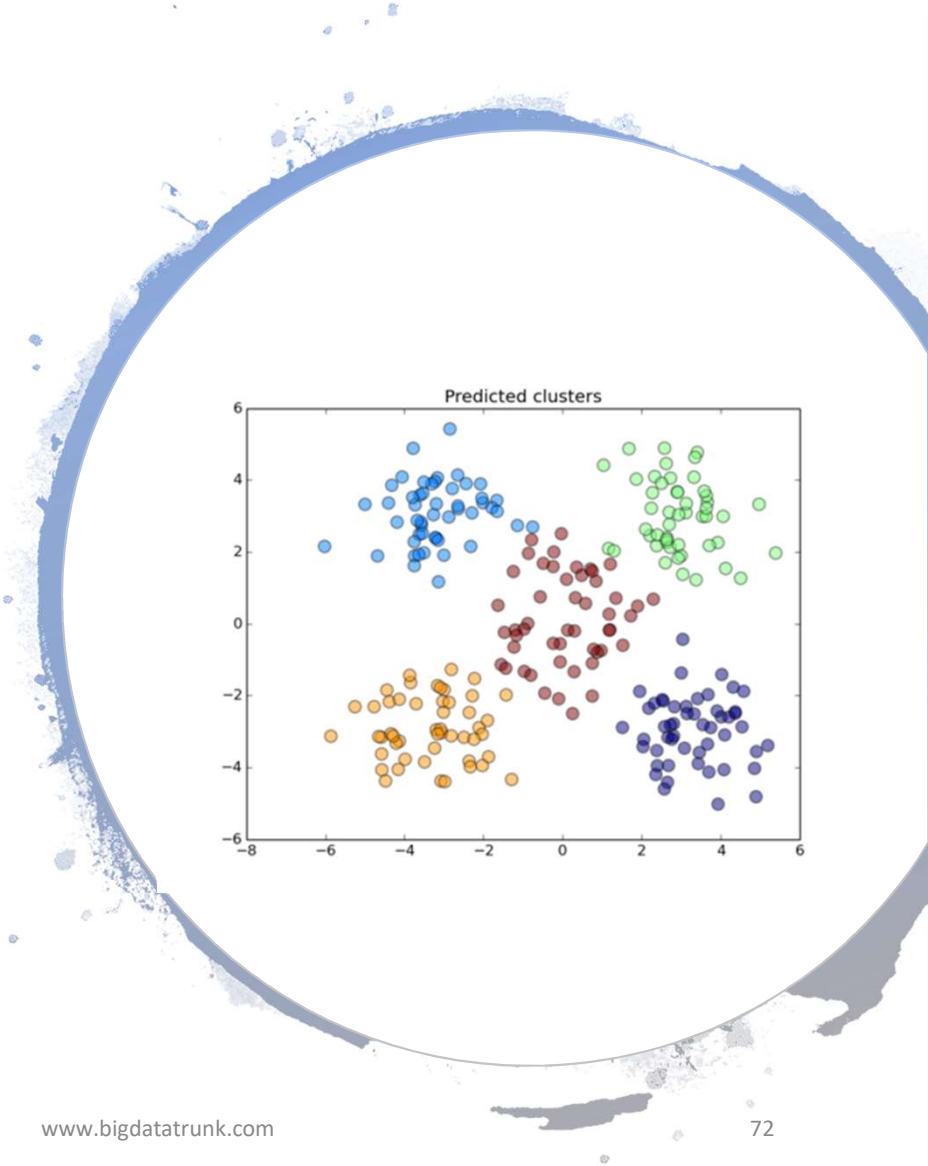
Not predicting a label

Unsupervised Learning- Types

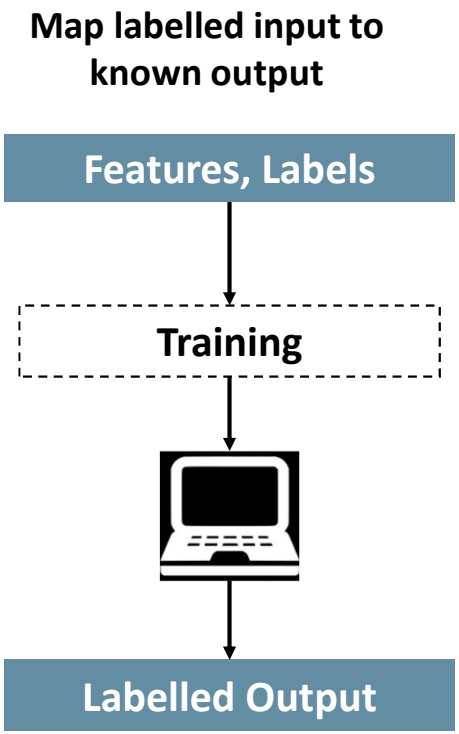


Clustering

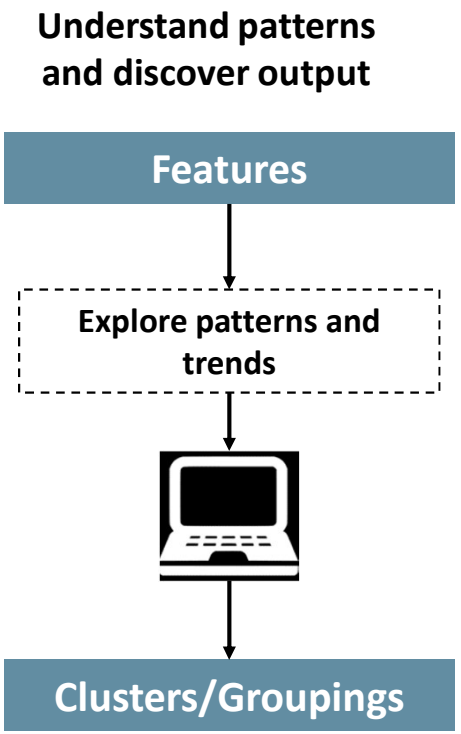
- **Use cases**
  - Market segmentation
  - Pattern recognition
  - High Data, High Voice, High Data and Voice users – telecom company
- **Algorithms**
  - K-Means
  - Hierarchical Clustering



Learning Approaches



Supervised



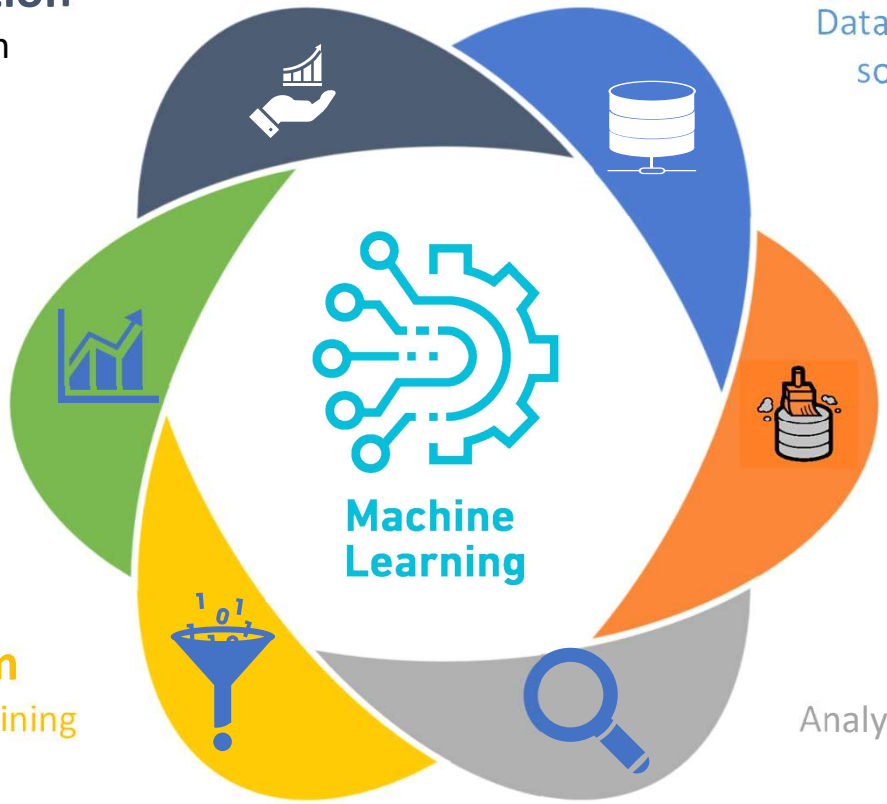
Unsupervised

Model Development

Define Problem

Model in production
Deploy the model in production

Collect Data
Data can come from various sources – social media, databases, etc.



Evaluate Model
How does the model perform with unseen data?

Prepare Data
Clean data and prepare for machine learning

Train Algorithm
Train algorithm with training dataset

Analyze Data
Analyze data, perform feature selection.

Data Split (Randomly)

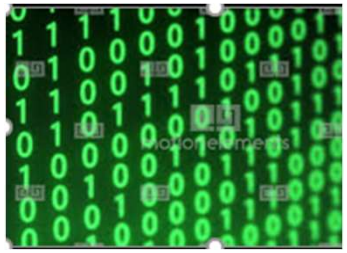
Training Data

70 %

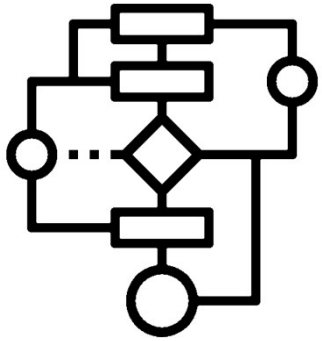
Test Data

30 %

Training Model



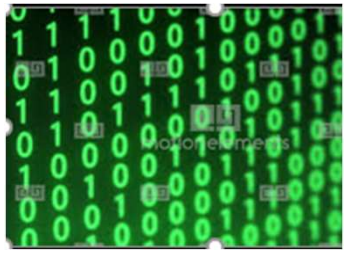
Training Data



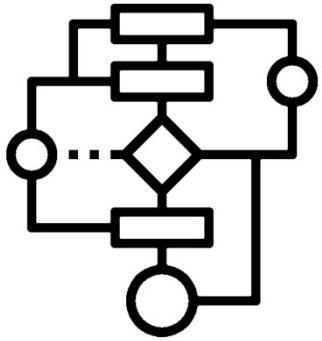
Algorithm

Prediction	Label
0	1
1	1
0	0
0	1
1	0
0	0

Training Model



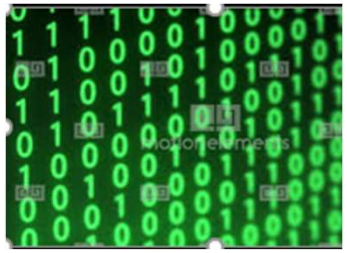
Training Data



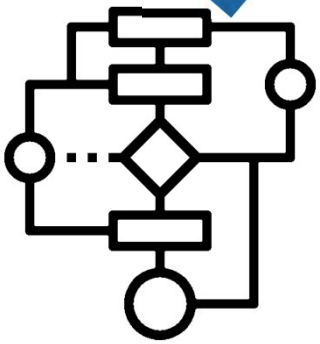
Initial Model

Prediction	Label
0	1
1	1
0	0
0	1
1	0
0	0

Training Model



Training Data

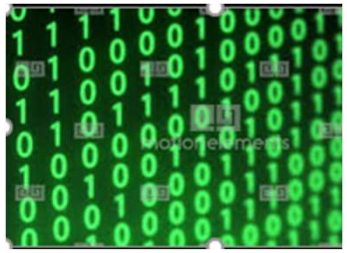


Initial Model

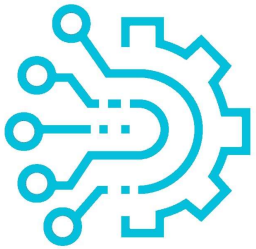
Prediction	Label
0	1
1	1
0	0
0	1
1	0
0	0

Adjust Parameters

Training Model



Training Data

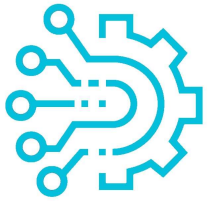


Model Ready

Prediction	Label
1	1
1	1
0	0
1	1
0	0
0	0

Model Evaluation

Test Data



Model

Prediction	Label
0	1
1	1
0	0
1	1
1	0
0	0

Machine Learning - Code

Training

algorithm.fit(training dataset)

Evaluate with Test Data

model.predict(test dataset)

Evaluation the Model

model.score(test dataset, predictions)

Algorithms

Machine Learning - Predictions

$y = f(X)$

$\hat{y} = \hat{f}(X)$ Prediction

$\hat{y} = 22.5$ Regression – continuous value

$\hat{y} = 0.89$ Classification – probability

$\hat{y} = 1$ Classification – prediction

Algorithms

Linear Regression

- Predicting real numbers – home prices, expected sales, etc.
- Drawing a best fit line to describe the relationship between the features

Logistic Regression

- Used in classification – buy a product or not
- Predicts probability of an event occurring

Support Vector Machines

- Can be used for either classification or regression – but mostly in classification
- Algorithm looks at data from non probabilistic view & uses linear separation of data

K-Nearest Neighbors

- Assumes that similar things exist in close proximity to each other
- Used for both classification and regression

Decision Trees

- Builds binary tree based on the data points
- Used for both classification and regression – mostly for classification

Algorithms

Random Forest

- Uses multiple decision trees to make predictions
- Majority prediction wins, can be used for both classification and regression

K-Means Clustering

- Partitions the data into K clusters (or grouping)
- Have to define how many clusters are required

Hierarchical Clustering

- Builds hierarchy of clusters by grouping data points (builds something called Dendrograms)
- Two types
  - Agglomerative – bottom up approach, every point starts as its own cluster and pairs of points are merged as one moves up the hierarchy
  - Divisive – top down approach, all points start as one cluster and they are split recursively as one moves down the hierarchy

Regression

Multiple Linear Regression - Multivariate

$$y = b_0 + b_1 x_1 + b_2 x_2 \ \ldots\ldots + b_n x_n$$

- *y* is the dependent variable
- x_i are the independent variables

Multiple Linear Regression

R&D Spend	Administration	Marketing Spend	State	Profit
165349.2	136897.8	471784.1	0	192261.83
162597.7	151377.59	443898.53	1	191792.06
153441.51	101145.55	407934.54	0	191050.39
144372.41	118671.85	383199.62	0	182901.99
142107.34	91391.77	366168.42	0	166187.94
131876.9	99814.71	362861.36	0	156991.12
134615.46	147198.87	127716.82	1	156122.51
130298.13	145530.06	323876.68	0	155752.6

$$profit = 1.63 + 7.91x_1 + 3.8x_2 + 6.88x_3 - 0.12x_4$$

x_1 = R&D Spend

x_2 = Administration

x_3 = Marketing Spending

x_4 = State (0 = California, 1 = New York)

Model Metrics

Model Evaluations

Loss Function

- A method to evaluate how well our algorithm is modeling our dataset
- If predictions are way off then the loss function will have higher value
- Absolute value of (prediction – actual value)

Loss Function – Mean Squared Error

Mean Square Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y)^2$$

Root Mean Square Error (RMSE)

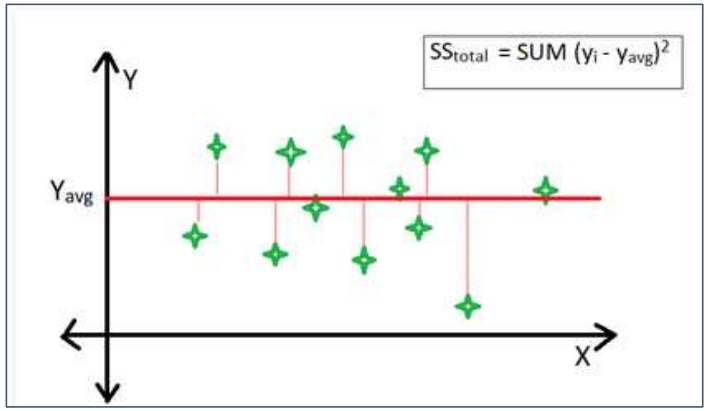
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y)^2}$$

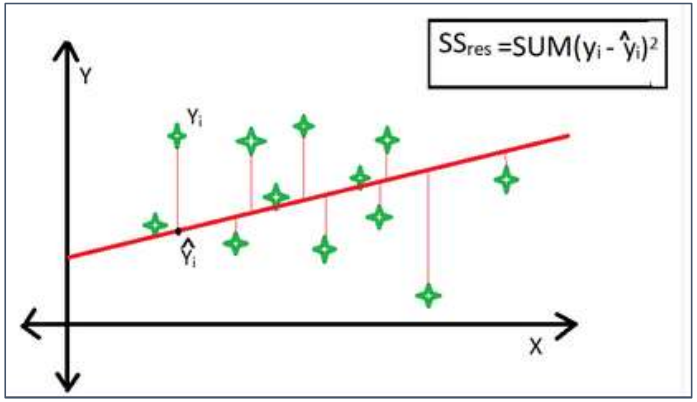
R-Squared (R^2)

Statistical measure of how close the data are to the fitted line

Usually, values are between 0.0 and 1.0

Higher the R^2 value, the better the model fits the data





$$R^2 = 1 - \frac{SSres}{SStot}$$

Linear Regression Demo

- Open file ‘CodeSamples/LinearRegression’ using Jupyter
- Start up dataset with 50 samples.
- Want to build a Linear Regression Model to predict a new start up will be profitable

Assignment – Linear Regression

- Open file '**Assignment/LinearRegression RealEstate**' using Jupyter
- Implementation requirements are defined in the notebook in red