Machine Learning

Dr Muhammad Atif Tahir

Professor

School of Computer Science

National University of Computing & Emerging Sciences

Karachi Campus

Contents

- Introduction

- Applications

- Types of Machine Learning

- Conclusions

Introduction

- **Machine learning**: Field of study that gives computers ability to learn

Introduction

- Now what is meant by computers ability to learn
  - Tom Mitchell (1998): A computer program is said to *learn* from experience **E** with respect to some task **T** and some performance measure **P**, if its performance on **T**, as measured by P, improves with **E**

Example

- Task: Face Recognition
- Experience: Images or Examples
- Performance Measure: Accuracy

Training Data

??

Some Applications

- predict whether a patient, hospitalized due to heart attack, will have a second heart attack

- predict the price of a stock in 6 months from now

- face detection: find faces in images (or indicate if a face is present)

- spam filtering: identify email messages as spam or non-spam

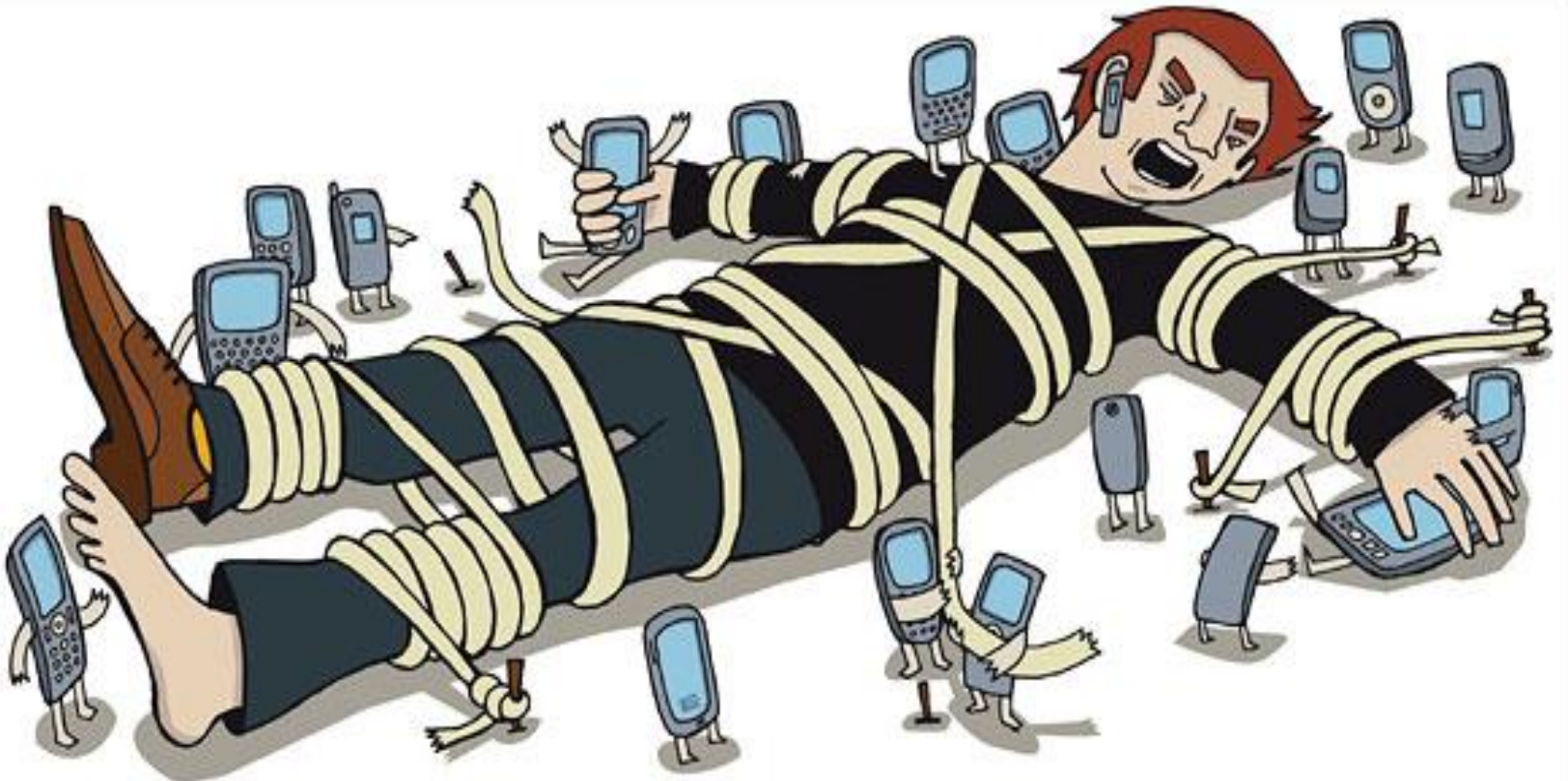- fraud-detection applications that seek patterns in jumbo-size data sets

Why Mine Data? Commercial Viewpoint

- Lots of data is being collected and warehoused
    - Web data, e-commerce
    - purchases at department/ grocery stores
    - Bank/Credit Card transactions

- Computers have become cheaper and more powerful

Data Everywhere

BIG DATA

Big Data is data that is too large, complex and dynamic for any conventional data tools to capture, store, manage and analyze.
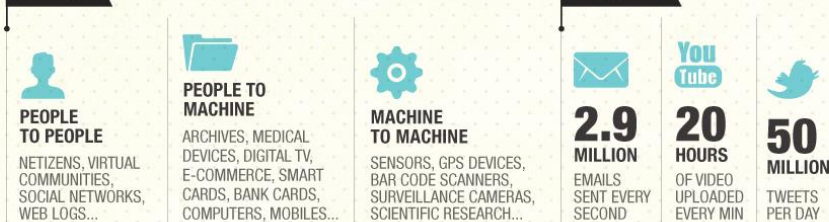
The right use of Big Data allows analysts to spot trends and gives niche insights that help create value and innovation much faster than conventional methods.

Visualization: Seventhinc

57.6% OF ORGANIZATIONS SURVEYED SAY THAT BIG DATA IS A CHALLENGE

72.7% CONSIDER DRIVING OPERATIONAL EFFICIENCIES TO BE THE BIGGEST BENEFIT OF A BIG DATA STRATEGY

50% SAY THAT BIG DATA HELPS IN BETTER MEETING CONSUMER DEMAND AND FACILITATING GROWTH

The "three V's", i.e the Volume, Variety and Velocity of the data coming in is what creates the challenge.

VOLUME

> **>3,500** NORTH AMERICA
> **>2,000** EUROPE
> **>250** CHINA
> **>400** JAPAN
> **>200** MIDDLE EAST
> **>50** INDIA
> **>50** LATIN AMERICA

Amount of Big Data stored across the world (in petabytes)

VARIETY

PEOPLE TO PEOPLE
NETIZENS, VIRTUAL COMMUNITIES, SOCIAL NETWORKS, WEB LOGS...

PEOPLE TO MACHINE
ARCHIVES, MEDICAL DEVICES, DIGITAL TV, E-COMMERCE, SMART CARDS, BANK CARDS, COMPUTERS, MOBILES...

MACHINE TO MACHINE
SENSORS, GPS DEVICES, BAR CODE SCANNERS, SURVEILLANCE CAMERAS, SCIENTIFIC RESEARCH...

VELOCITY

2.9 MILLION EMAILS SENT EVERY SECOND

You Tube **20 HOURS** OF VIDEO UPLOADED EVERY MIN

50 MILLION TWEETS PER DAY

VALUE

	PRODUCTIVITY INCREASE	SALES INCREASE
RETAIL	49%	$9.6B
CONSULTING	39%	$5.0B
AIR TRANSPORTATION	21%	$4.3B
CONSTRUCTION	20%	$4.2B
FOOD PRODUCTS	20%	$3.4B
STEEL	20%	$3.4B
AUTOMOBILE	19%	$2B
INDUSTRIAL INSTRUMENTS	18%	$1.2B
PUBLISHING	18%	$0.8B
TELECOMMUNICATIONS	17%	$0.4B

CASE STUDY - Healthcare

$300 billion is the potential annual value to Healthcare

$165B CLINICAL

$9B PUBLIC HEALTH

$108B R&D

$5B BUSINESS MODEL

$47B ACCOUNTS

TRANSPARENCY IN CLINICAL DATA AND CLINICAL DECISION SUPPORT

AGGREGATION OF PATIENT RECORDS, ONLINE PLATFORMS AND COMMUNITIES

PUBLIC HEALTH SURVEILLANCE AND RESPONSE SYSTEMS

RESEARCH AND DEVELOPMENT; PERSONALIZED MEDICINE; CLINICAL TRIAL DESIGN

ADVANCED FRAUD DETECTION; PERFORMANCE BASED DRUG PRICING

40% PROJECTED GROWTH IN GLOBAL DATA CREATED PER YEAR

1010

$$$

5% PROJECTED GROWTH IN GLOBAL IT SPENDING PER YEAR

The estimated size of the digital universe in 2011 was 1.8 zettabytes. It is predicted that between 2009 and 2020, this will grow 44 fold to 35 zettabytes per year. A well defined data management strategy is essential to successfully utilize Big Data.

Sources - ❶ Reaping the Rewards of Big Data - Wipro Report ❷ Big Data: The Next Frontier for Innovation, Competition and Productivity - McKinsey Global Institute Report ❸ comScore, Radicati Group ❹ Measuring the Business Impacts of Effective Data - study by University of Texas, Austin ❺ US Department of Labour.

DO BUSINESS BETTER

NYSE:WIT | OVER 130,000 EMPLOYEES | 54 COUNTRIES | CONSULTING | SYSTEM INTEGRATION | OUTSOURCING
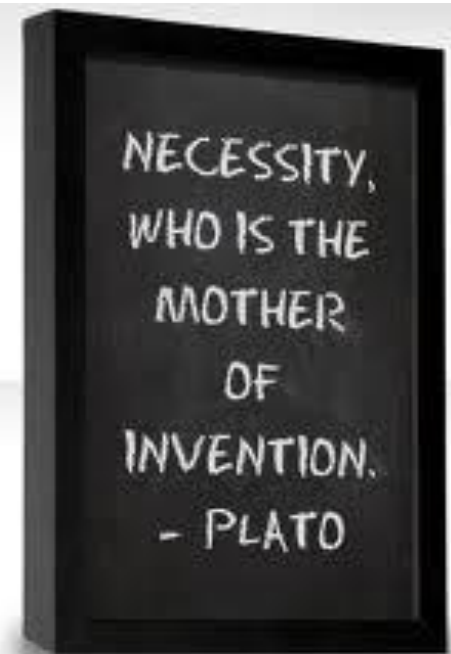
WIPRO Applying Thought
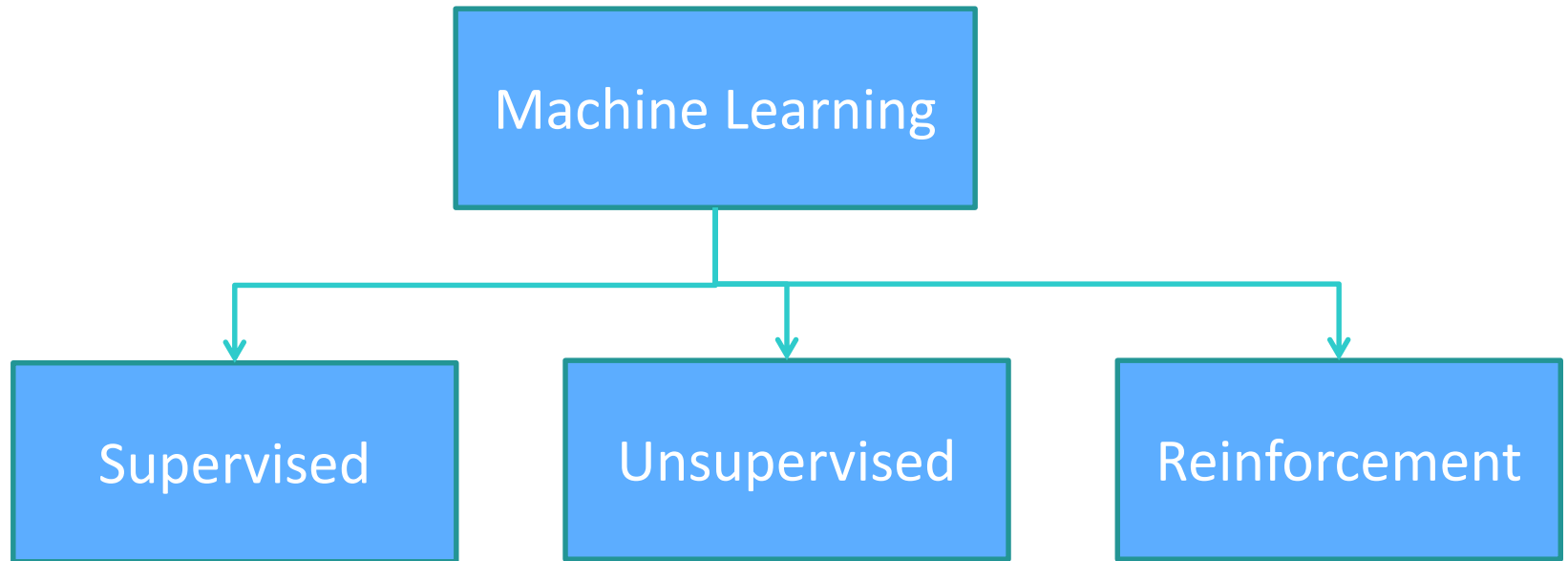
OK , there is too much Data, then WHAT?

- Can we use this data to find things that are not apparent but can be very useful to us?

We are drowning in data, but starving for knowledge!

Knowledge Discovery is needed to make sense and use of data.

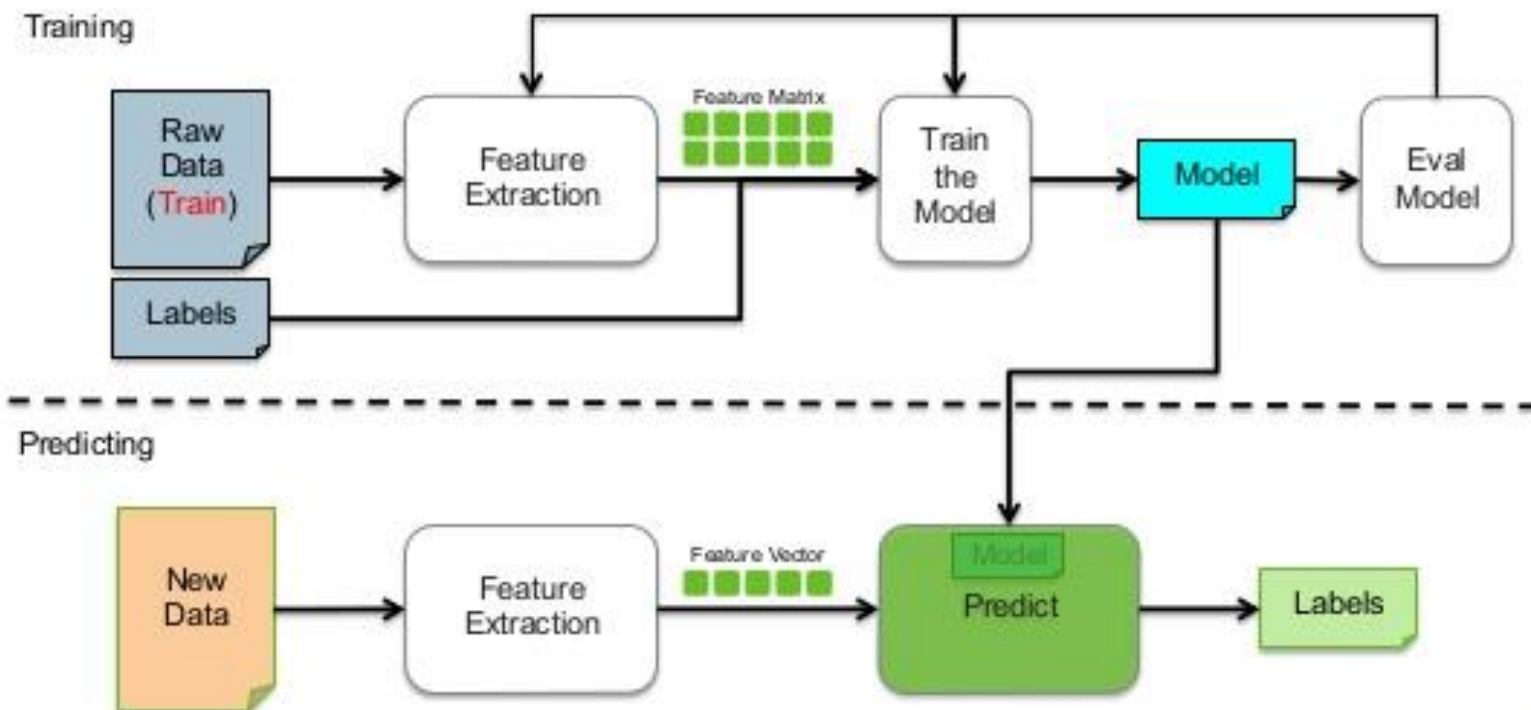
NECESSITY, WHO IS THE MOTHER OF INVENTION. - PLATO

Types of Machine Learning

Supervised Learning

- Supervised Learning
  - Learn from Supervised Training Data
  - For example; spam filtering where Large number of email messages labelled as either
    - spam
    - non-spam
  - New email message will then be classified as spam or non-spam

Supervised Learning Workflow
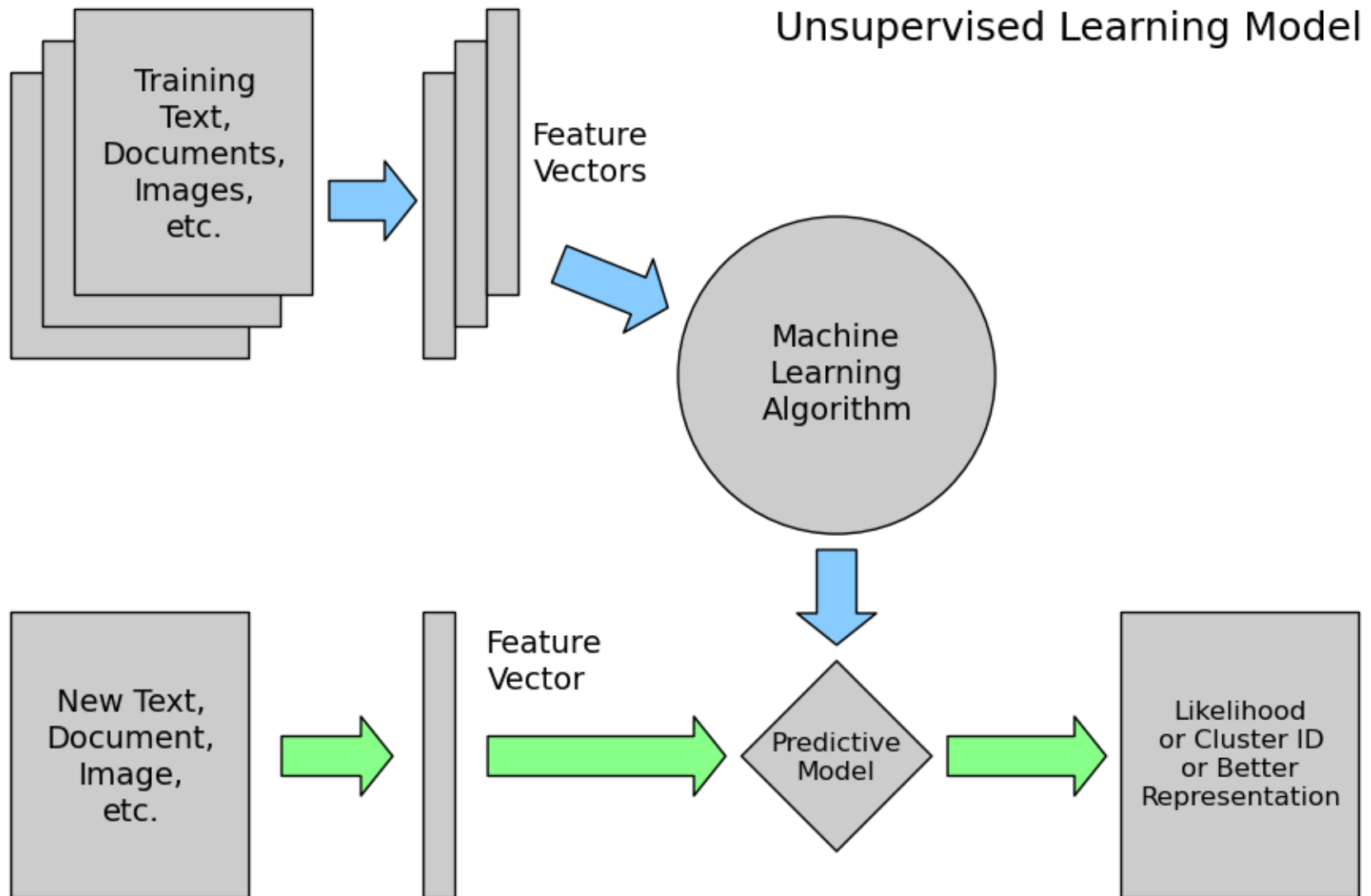
Supervised Learning

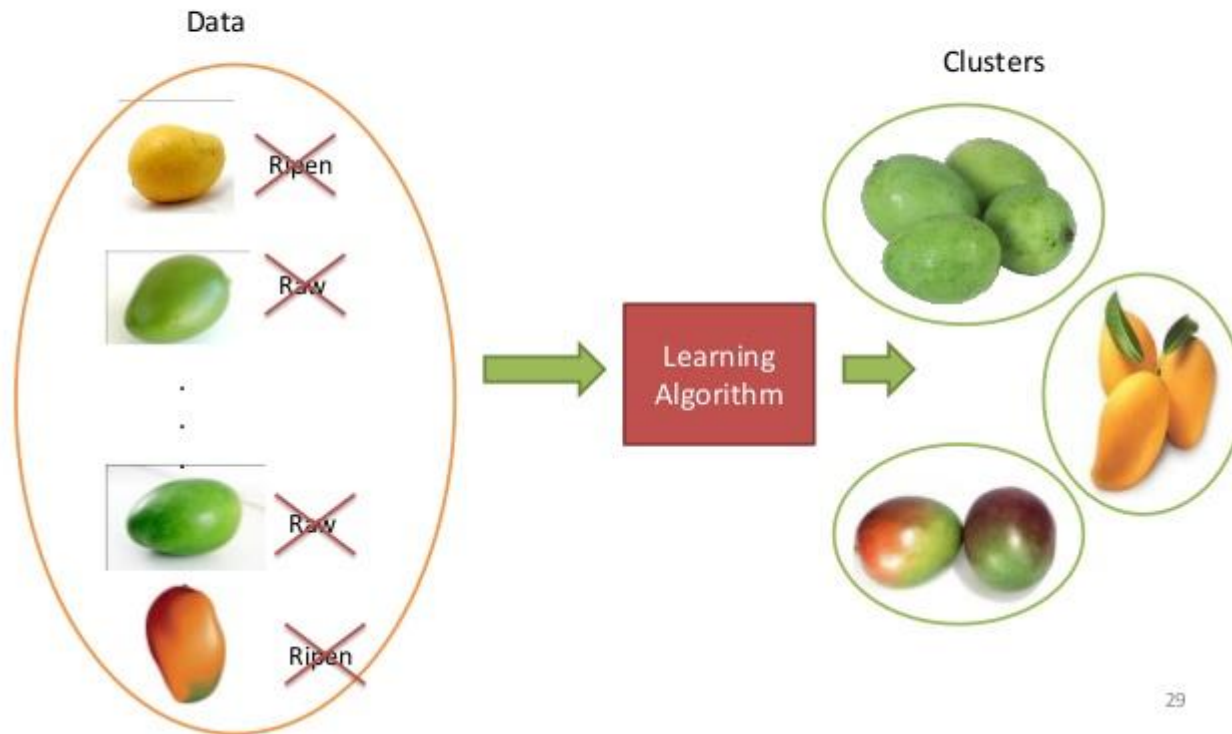

Training Images

Test Image

Unsupervised Learning

- The correct classes of training data are not known

- Applications
  - Fraud Detection: Identify groups of motor insurance policy holders with a high average claim cost
  - Social Networks: Recognize communities within large groups of people

Unsupervised Learning

Unsupervised Learning Model

Training Text, Documents, Images, etc.

Feature Vectors

Machine Learning Algorithm

New Text, Document, Image, etc.

Feature Vector

Predictive Model

Likelihood or Cluster ID or Better Representation

Unsupervised Learning

Unsupervised Learning

Objective is simply to divide above Images into N groups
Here ideally N = 2

Reinforcement Learning

- Allows the machine to learn its behaviour based on feedback from the environment

- Applications: Card Games, Chess etc

Tools

- Java (WEKA)
- Matlab
- C++
- Phyton
- Apache Spark (For Clouds)
- R

Final Remarks

- 12 IT skills that employers can't say no to (ComputerWorld)
  - Number one is Machine Learning
    - "There are lots of applications that have big, big, big data sizes, which creates a fundamental problem of how you organize the data and present it to users"

Questions!