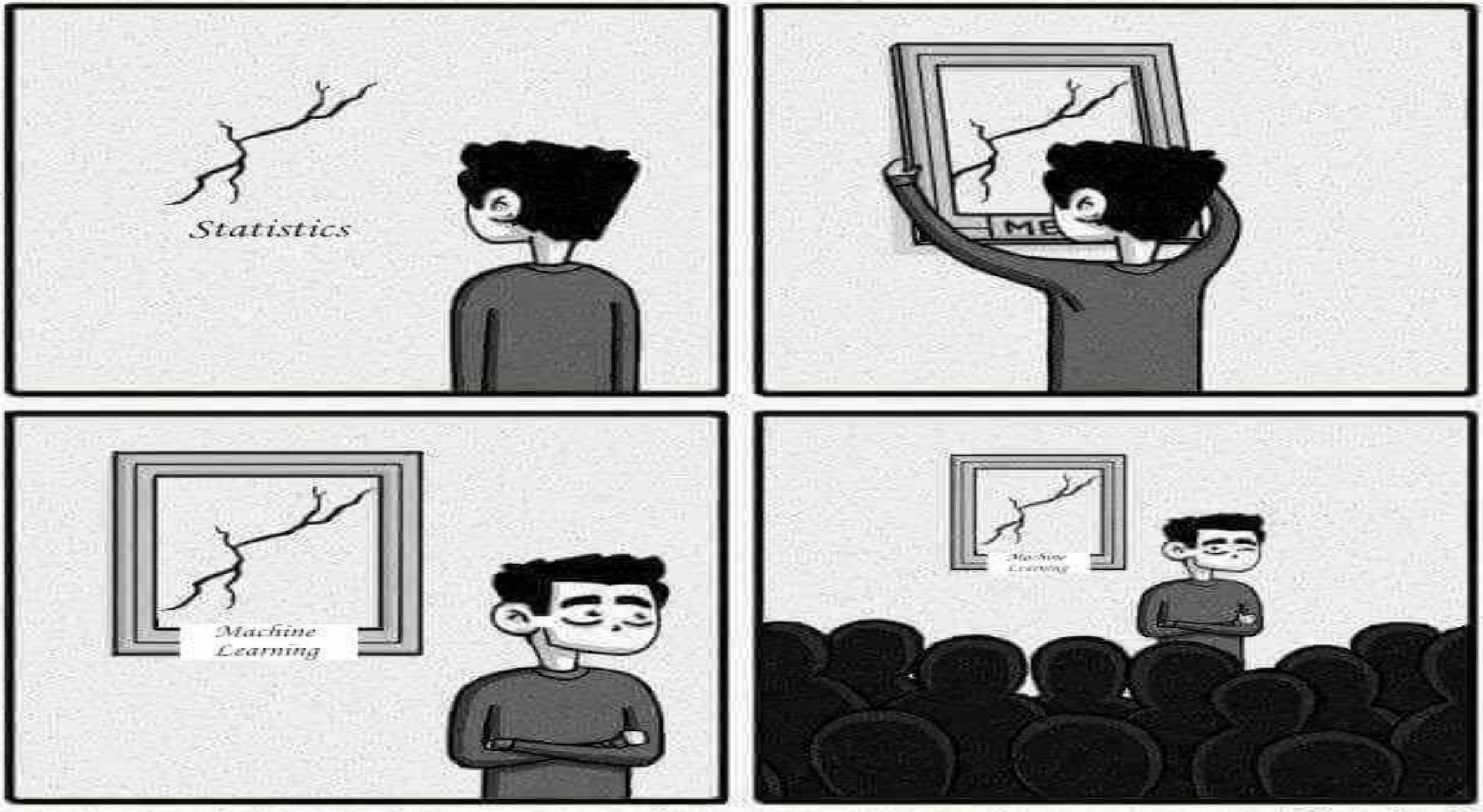
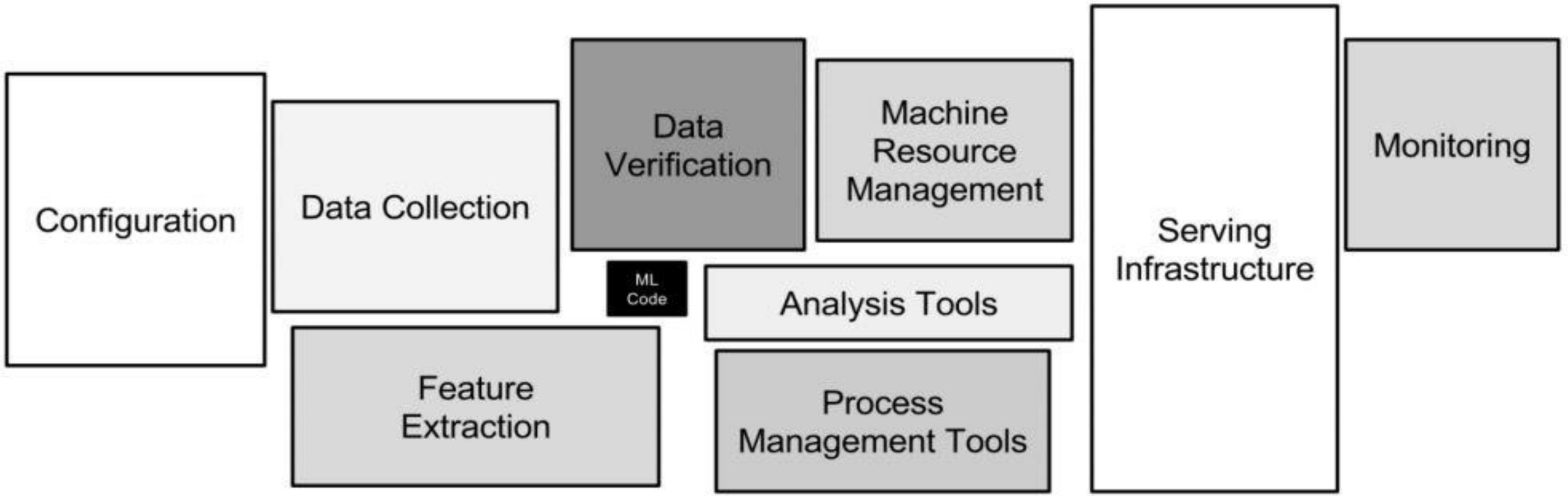

Statistics
Machine Learning
© sandserif

Hidden Technical Debt in Machine Learning Systems

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley,gholt,dgg,edavydov,toddphillips}@google.com
Google, Inc.



The Role of Statistics in data science
Regression Analysis

Agenda

Introduction – the context (30 -45 Mins)

- What is Analytics
- Need to learn Analytics

Overview of Analytics trends -Past/present/future

How it works

Industry Evolution

Why all of the sudden such a hype for data science - is it true

Applications in various industries

Manufacturing • Healthcare • FMCG

Applications in various functions

• HR • Supply chain • Marketing

The Process

Overview of the Process

Stages /Roles

How to Build career in data science

The Role of Statistics in Data Science

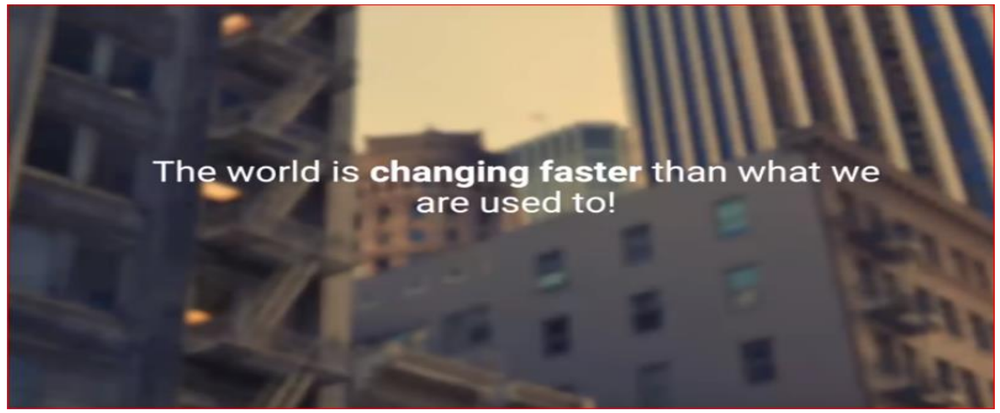
Analytics Continuum

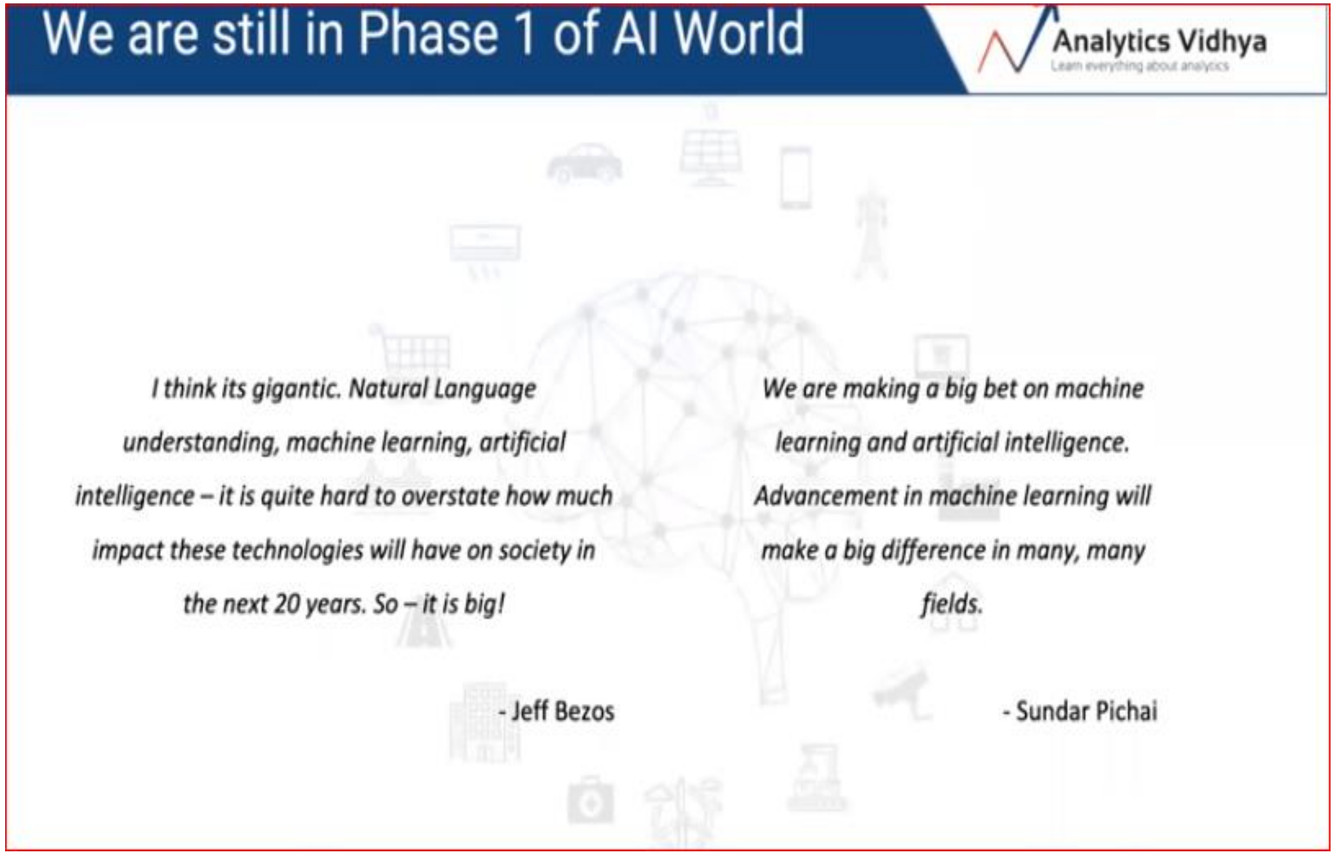
- Overview of Analytics Process
- Measurement Scales & Types of Data
- Distributions
- Understanding data using Descriptive statistics
- Understanding data using Viszvalization
- Inferential statistics
- Data Pre processing
- Modelling Concepts

- **Predictive Analytics - Regressions**
  - Multiple Regressions
  - Logistic Regression
  - Hands on using R

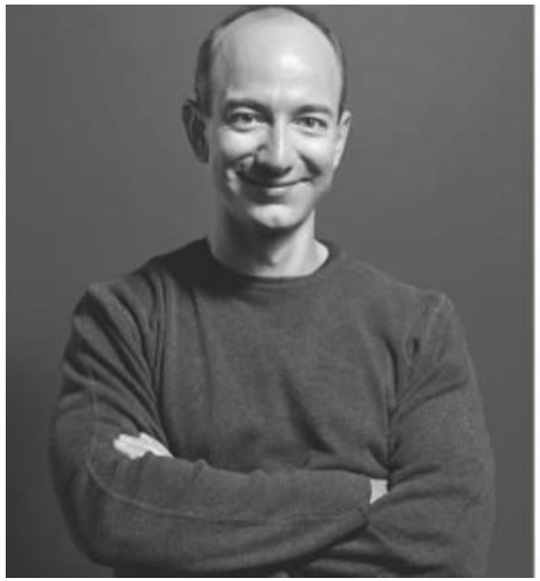
- **Q &A**

THE PRESENT WORLD STATUS





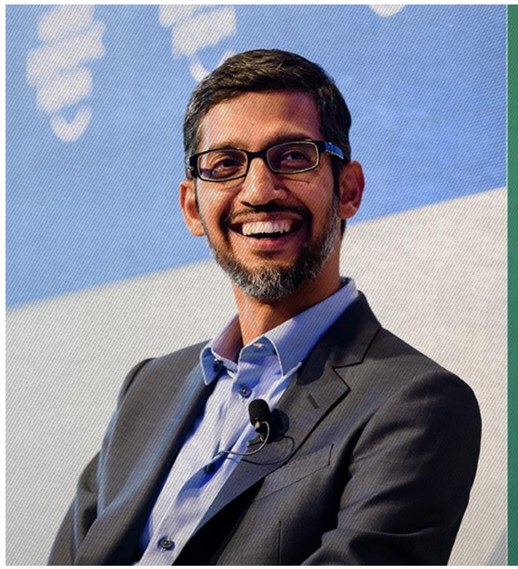
Opinions...



It is a renaissance, it is a golden age.

We are now solving problems with machine learning and artificial intelligence that were ... in the realm of science fiction for the last several decades.

JEFF BEZOS
AMAZON

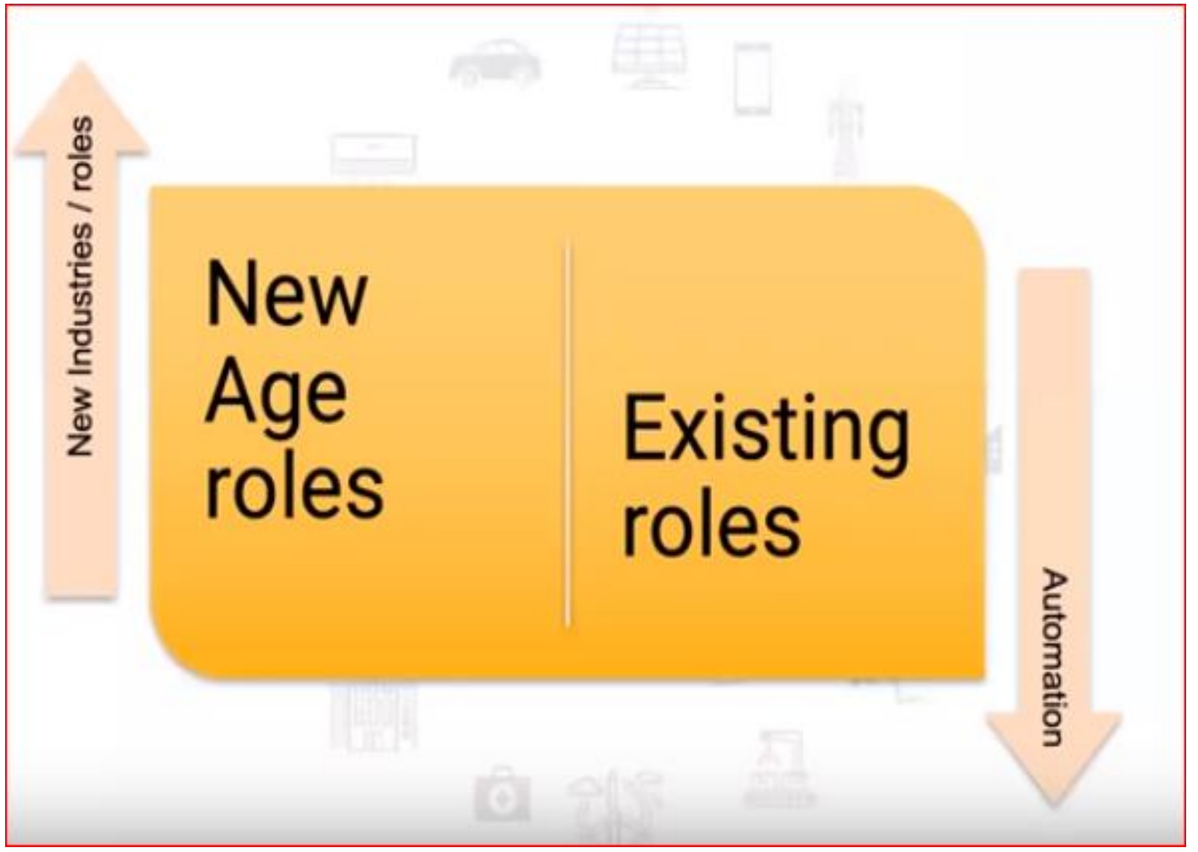


WORLD ECONOMIC FORUM

AI is probably the most important thing humanity has ever worked on.

Sundar Pichai
CEO of Google

ARE YOU READY FOR FUTURE... BECAUSE


New Industries / roles
New Age roles
Existing roles
Automation


Examples of New Roles being created
Analytics Vidhya
Learn everything about analytics

ARE YOU READY FOR THE FUTURE?

Around **2.7 million** Data Science and Analytics jobs will be created over the next three years.

International Data Corporation (IDC) predicts that worldwide revenues for big data and business analytics will reach more than **$210bn in 2020**.

PricewaterhouseCoopers (PwC) estimates that worldwide, AI will "increase global GDP by **$15.7 trillion, a full 14%, by 2030.**"

5mn jobs will be lost by 2020, according to World Economic Forum.

69% jobs in India are threatened by automation, according to World Bank.

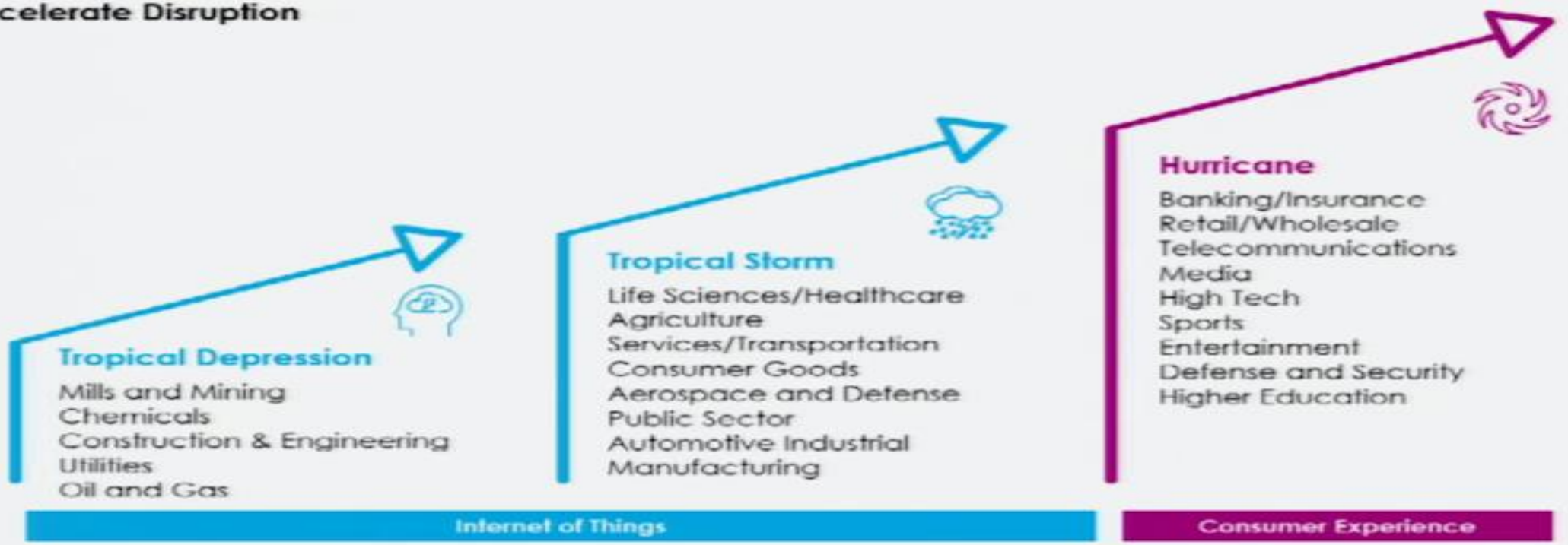
90% lawyers will become irrelevant, only specialists will remain.


85%
By 2020, customers will manage 85%
of their relationship with the enterprise
without interacting with a human
Source: Gartner


Transform
Customer Experience

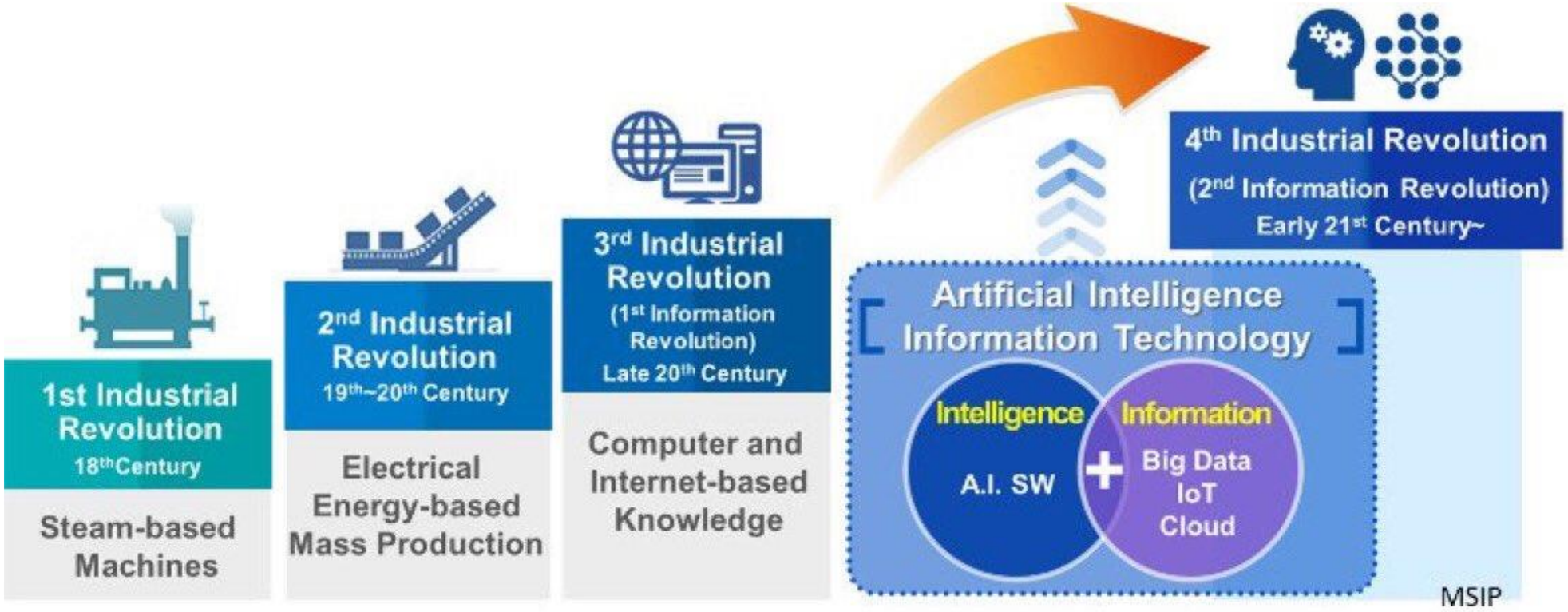
The Digital Storm Will Impact Every Industry

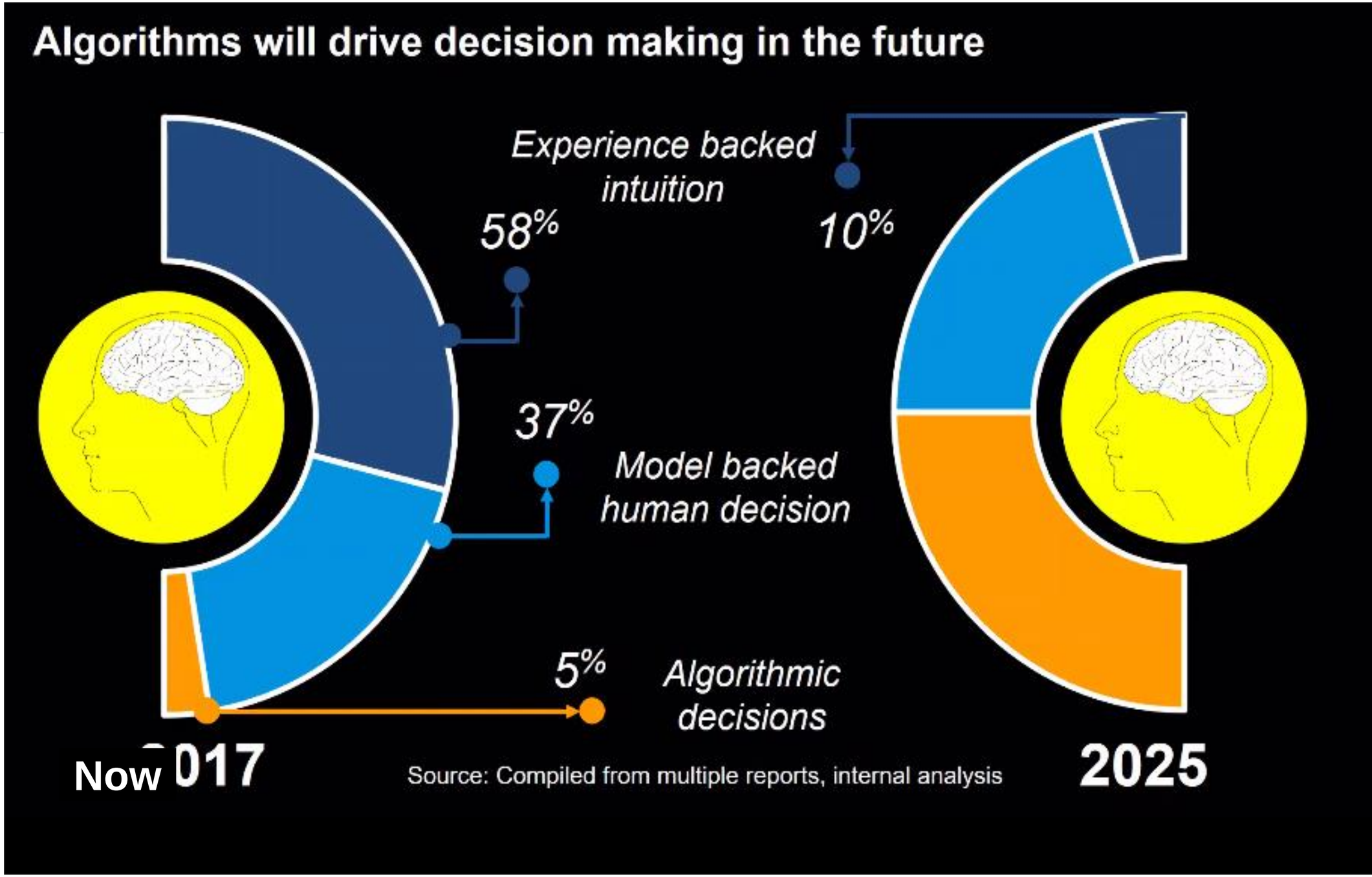
Accelerate Disruption



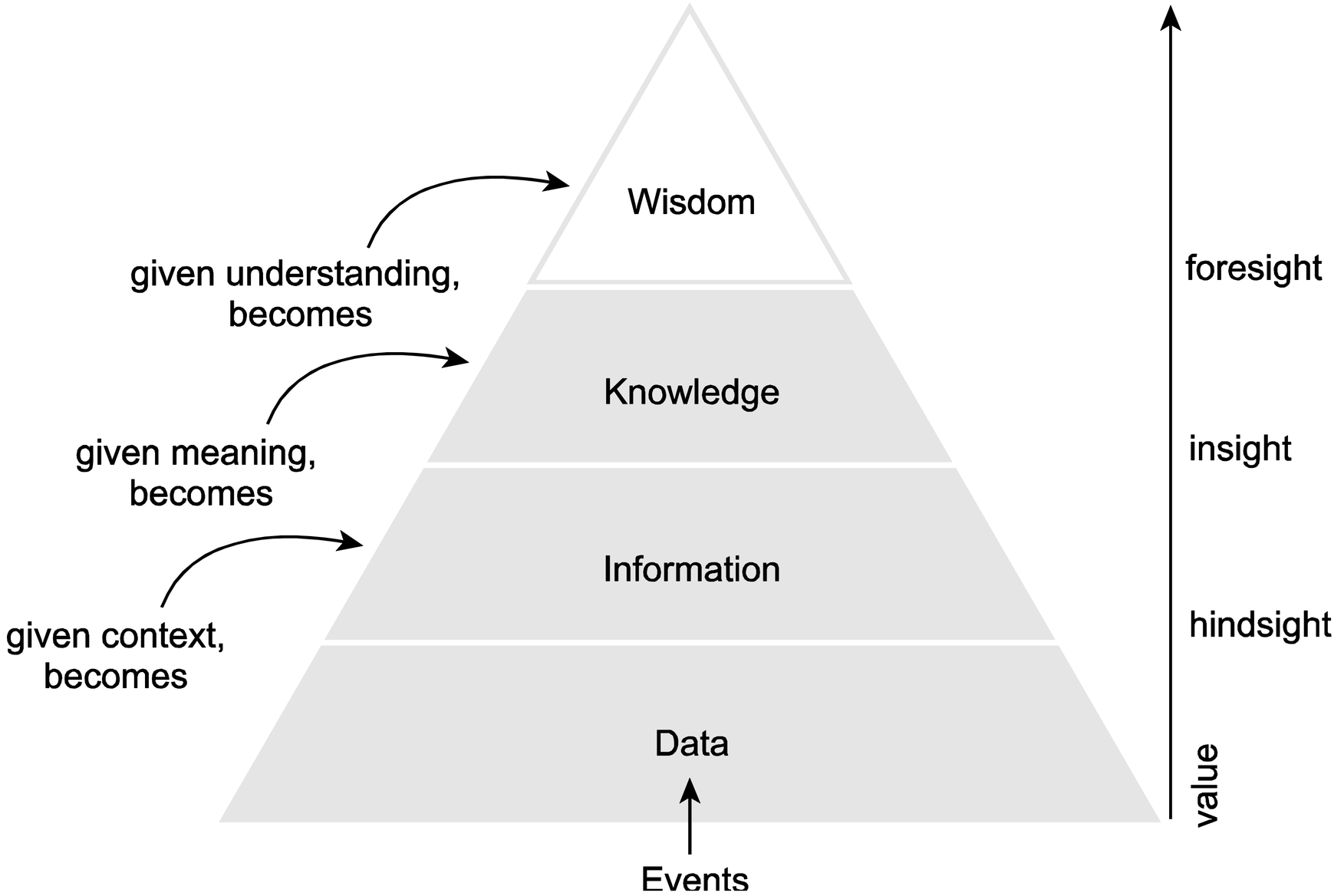
Source: SAP

Some history...

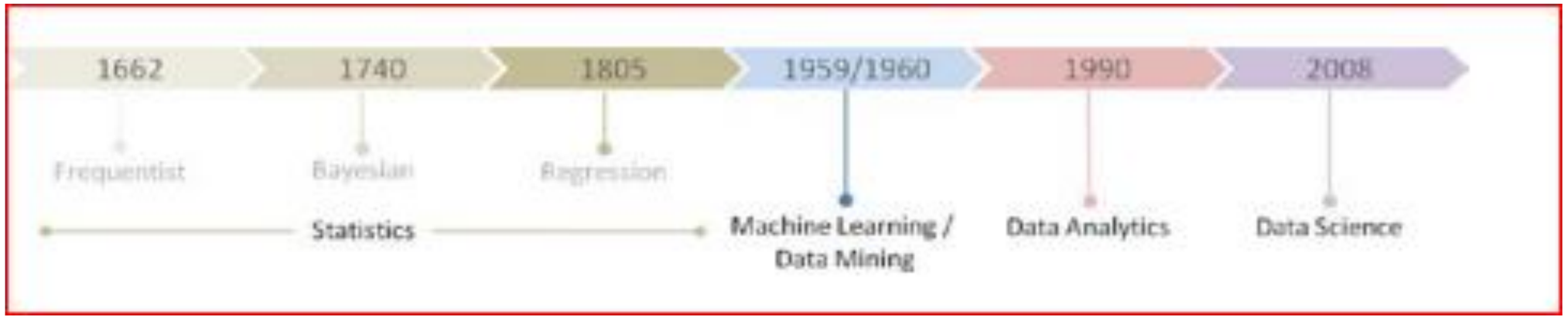



Algorithms will drive decision making in the future
Experience backed intuition
58%
10%
37%
Model backed human decision
5%
Algorithmic decisions
Now 2017
Source: Compiled from multiple reports, internal analysis
2025

What is data Analysis?


Wisdom
Knowledge
Information
Data
Events
given understanding, becomes
given meaning, becomes
given context, becomes
foresight
insight
hindsight
value

Overview of Analytics - Trends



- Past

  - Less volume of data and less information, mostly Numeric data (Traditional statistical models using customized softwares like Excel, SPSS, MatLab, MiniTab, SAS)
  - More of inferential models

Eight ways a Data Scientist can add value to any business

Courtesy: SimpliLearn

Present Trend of data Analysis (Data Science/Big data…etc)

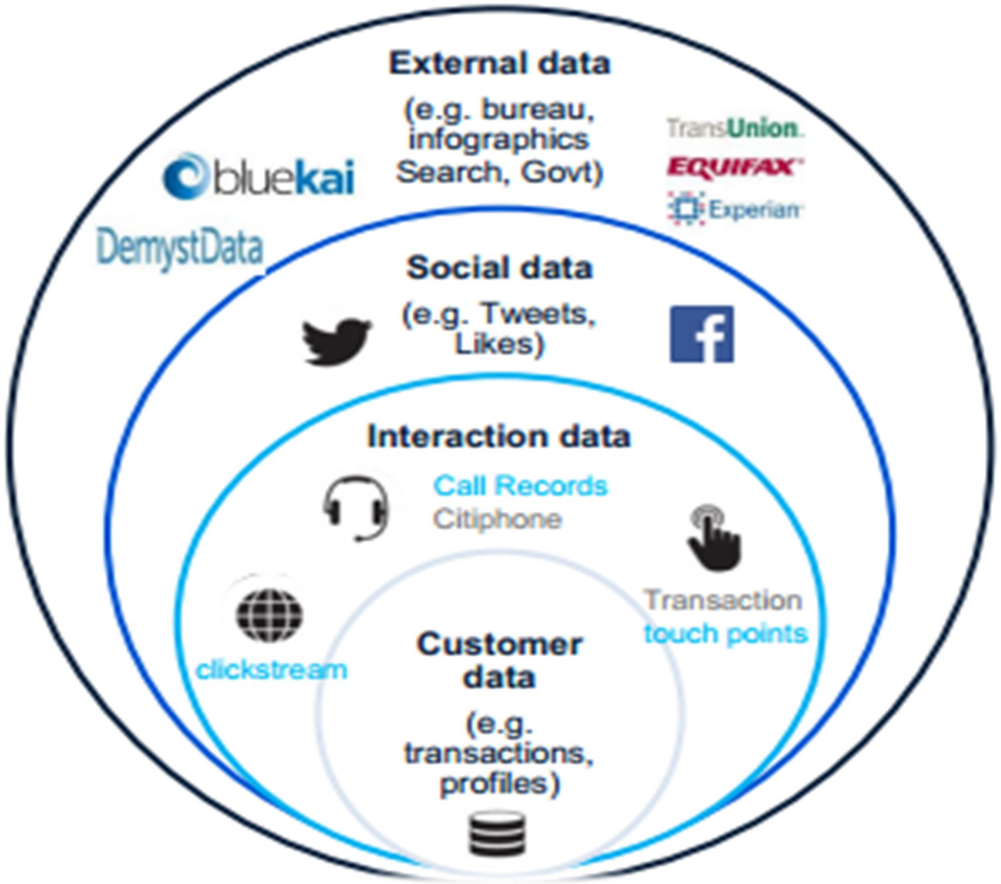
Present

- Huge data/Big data/ text Data /Image data /Audio data…
- Lot of information available and a Machine Learning era
- Modelling is seen more as a deployment model

Data Explosion

2.5 Exabyte
of data is being created in the world every day *

90%
of the data today has been created in the last 2 years*



50% decrease
In computing cost , every 18 months

Hadoop
reduces execution time

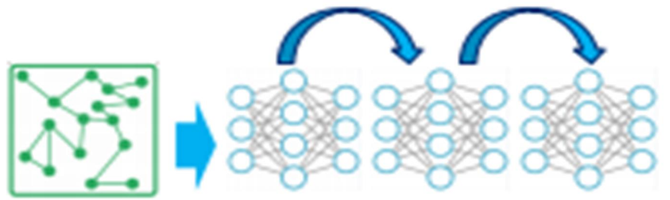
Adapt
to changes in data, automate learning

Faster results with distributed computing

>1000000 Faster
Enable quick builds and real time execution

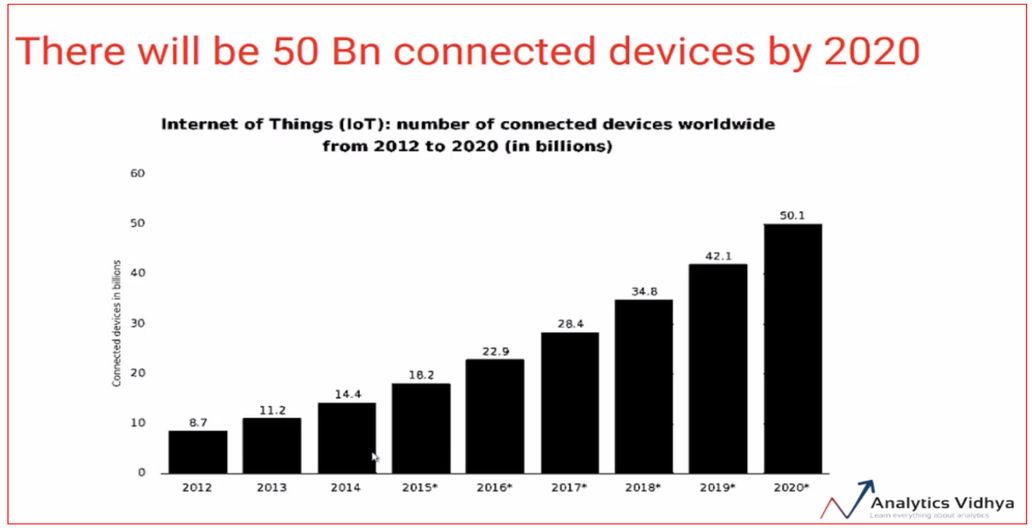
Billions of Computations in seconds

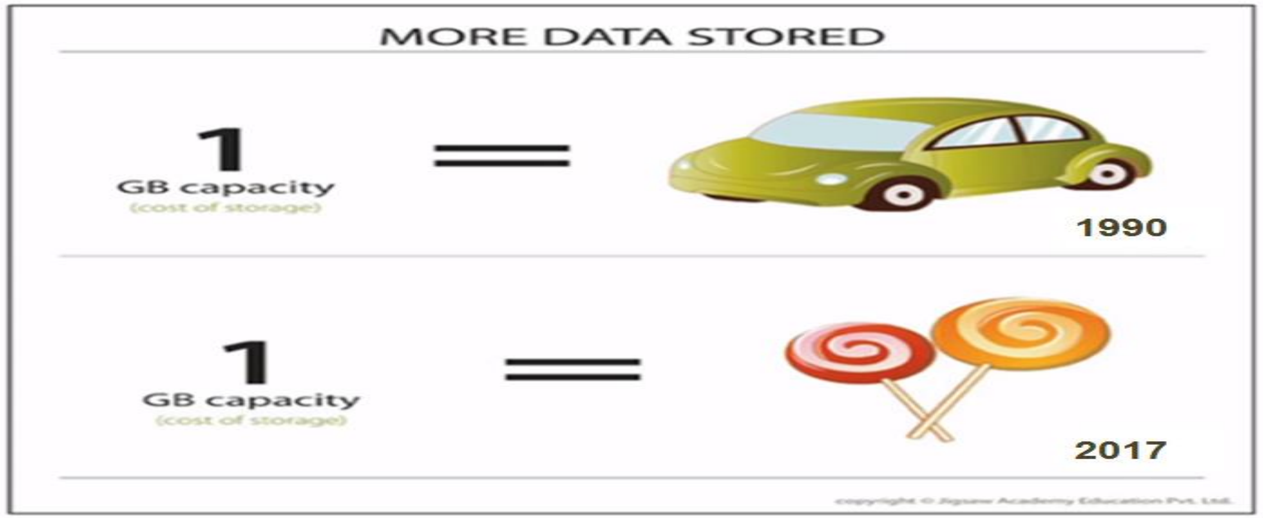
Learn Continuously with smarter algorithms



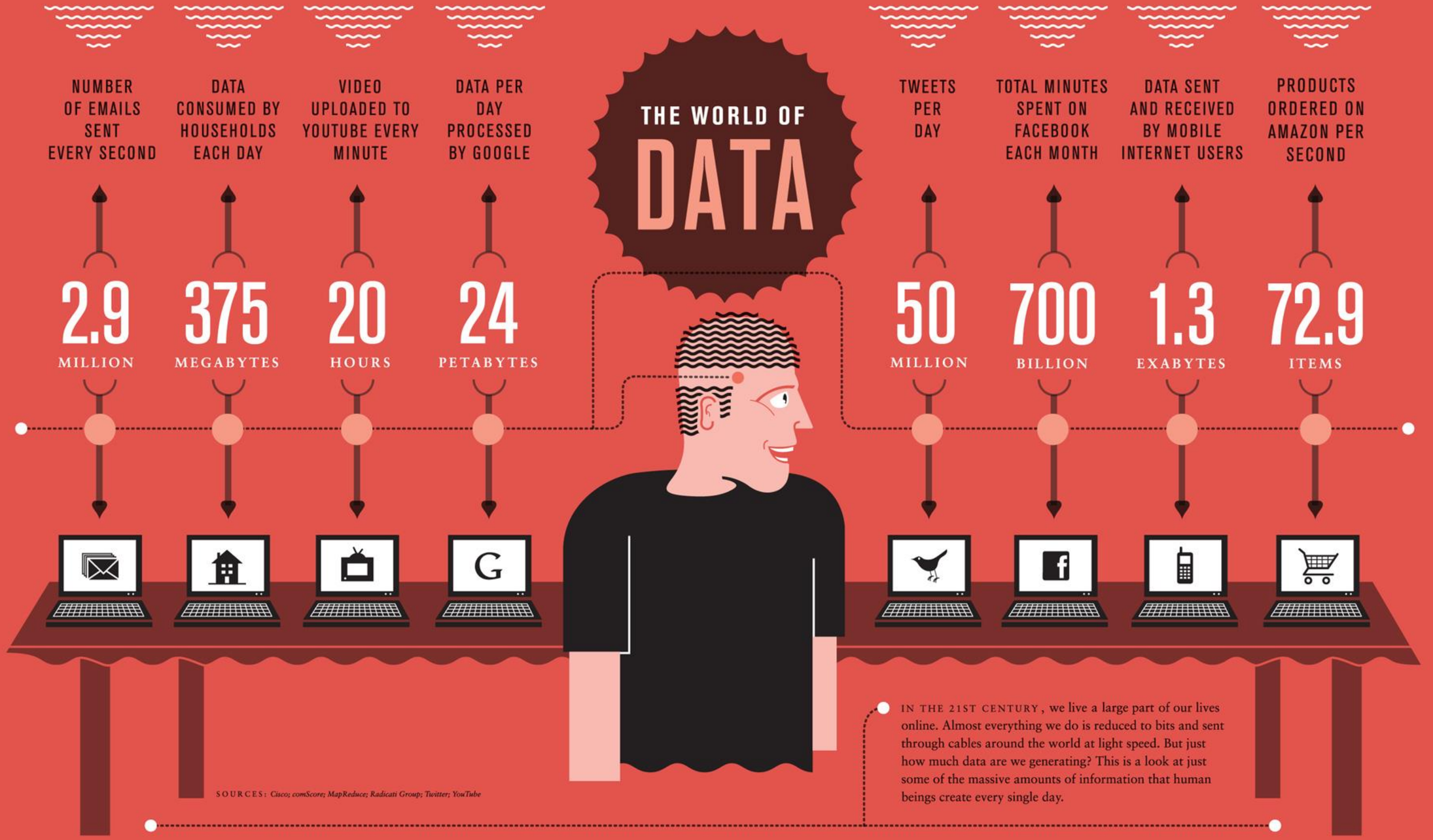
50% better
Adapts to new trends in consumer behavior

Present Trend of data Analysis Contd…





THE WORLD OF DATA



SOURCES: *Cisco; comScore; MapReduce; Radicati Group; Twitter; YouTube*

IN THE 21ST CENTURY, we live a large part of our lives online. Almost everything we do is reduced to bits and sent through cables around the world at light speed. But just how much data are we generating? This is a look at just some of the massive amounts of information that human beings create every single day.

A COLLABORATION BETWEEN GOOD AND OLIVER MUNDAY

IN PARTNERSHIP WITH IBM

Applications in various industries

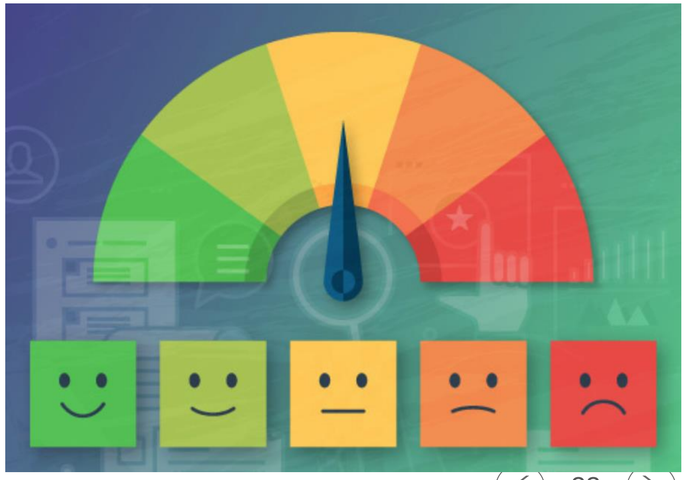
Retail

Recommendation engine

Market Basket Analysis

Inventory management

Customer sentiment analysis



How is the world changing?

Analytics Vidhya
Learn everything about analytics

Swipe Phone and Enter in the store

↓

Pick up Items, will get added to cart

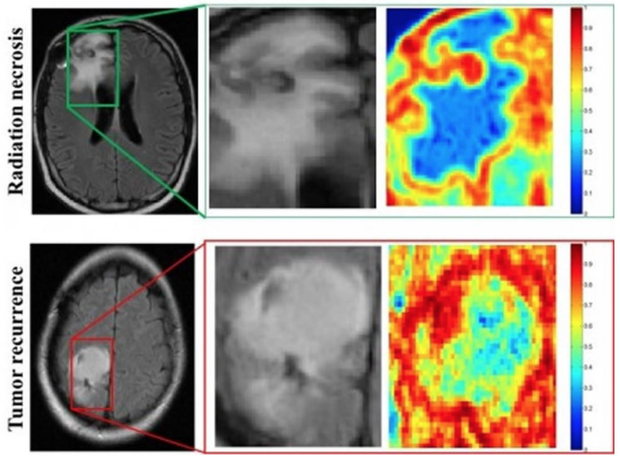
↓

Walk out of the store with items and get charged

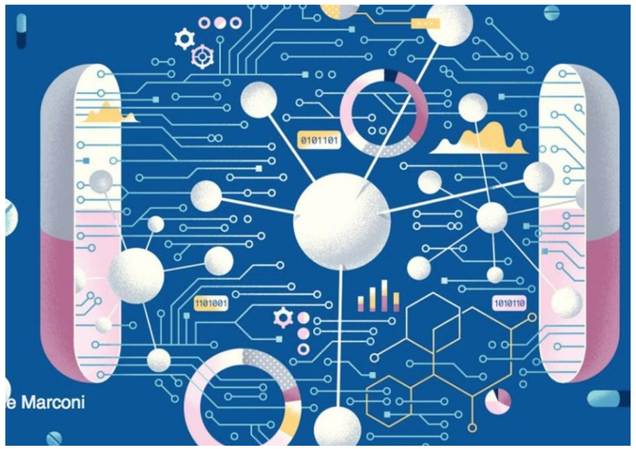
How is the world changing?

Healthcare

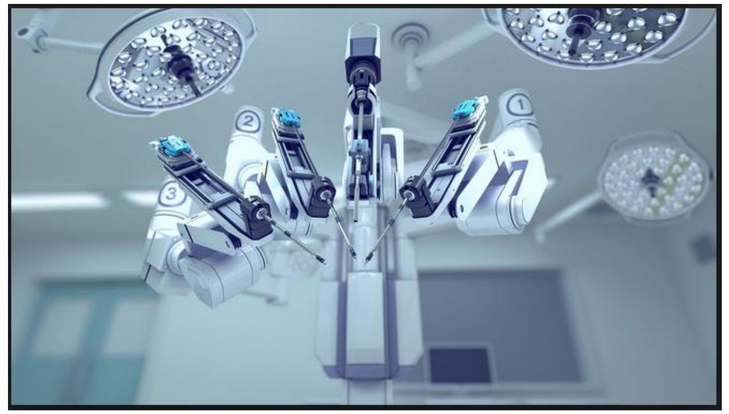
Medical Image Analysis



Drug Creation

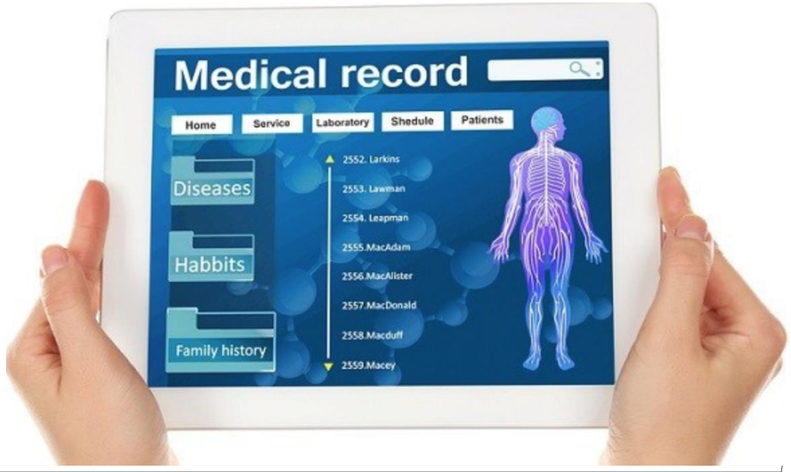


Assisted Robotic Surgeries



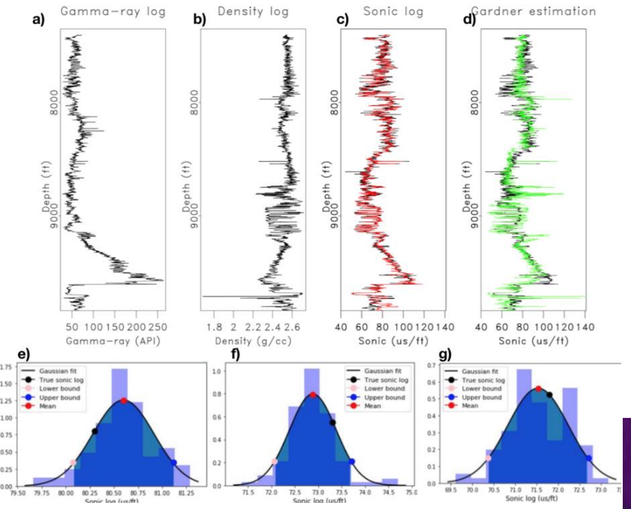
Virtual Nurses

Electronic Health record



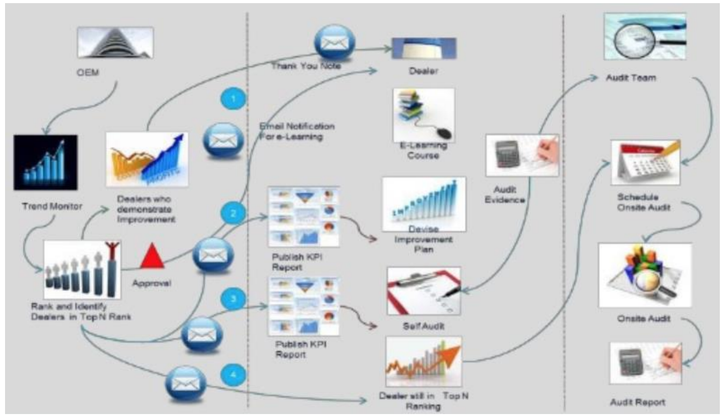
Manufacturing

Fault prediction and preventive maintenance



Computer vision applications

Warrant y

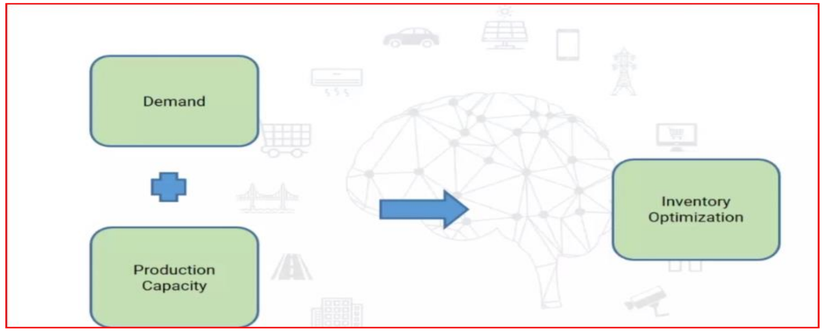


Robotizatio n

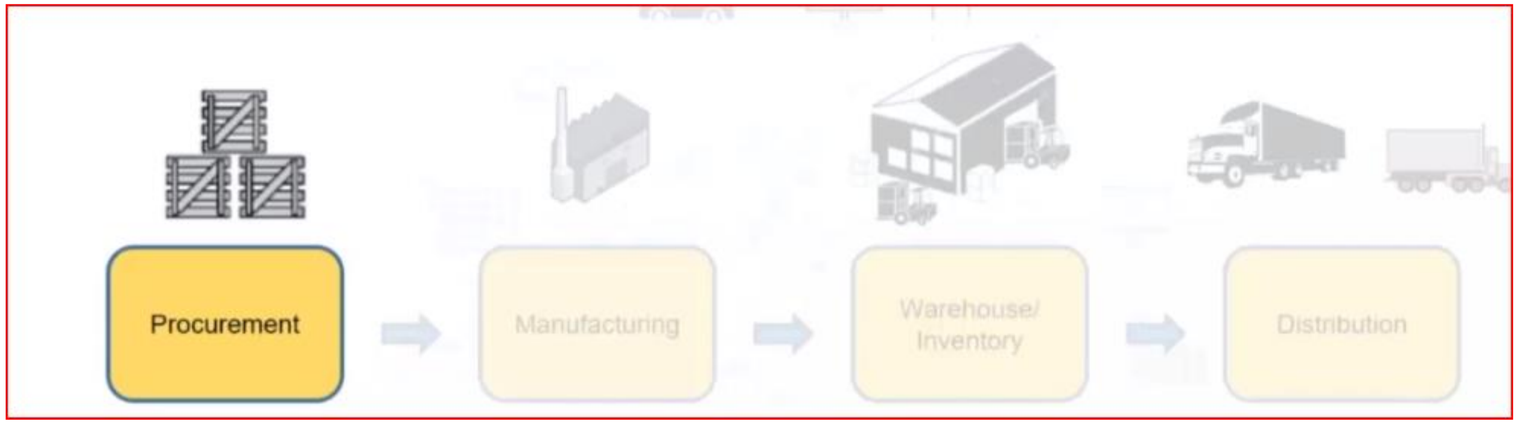
Supply Chain

Algorithms searches for multiple suppliers and fit with our requirement

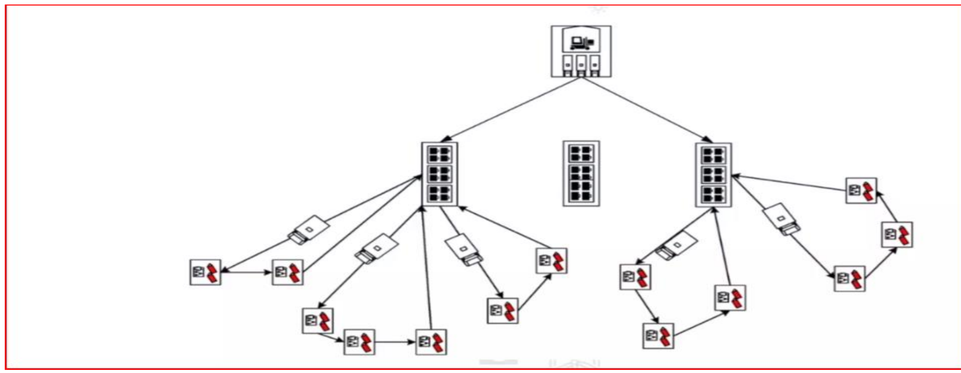
Inventory



Procurement



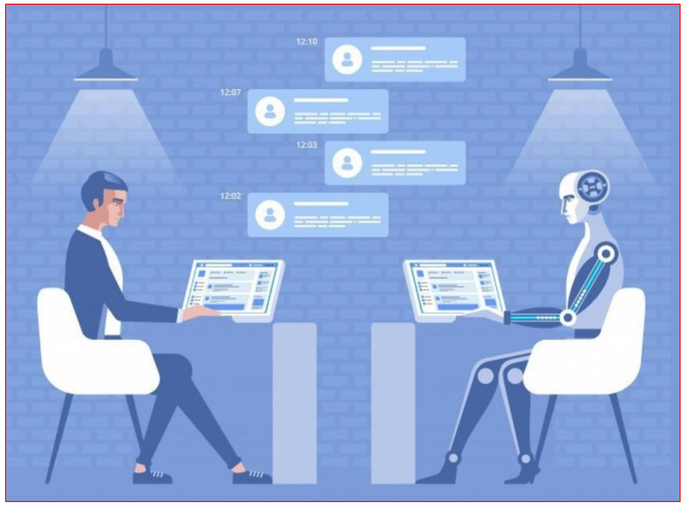
Route Optimization



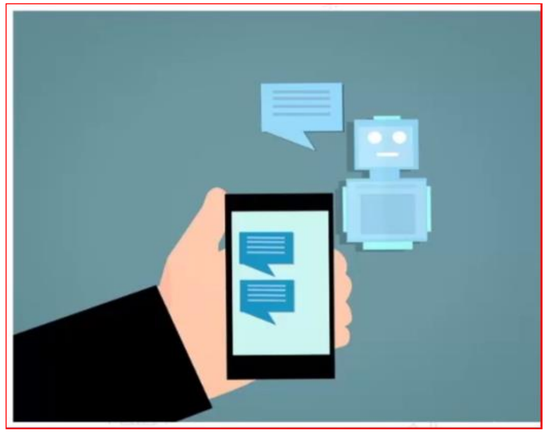
NLP Techniques uses to shortlist the candidates and does the scoring for candidates and then, selects

AI Companies sourcing right candidate

Recruitment process

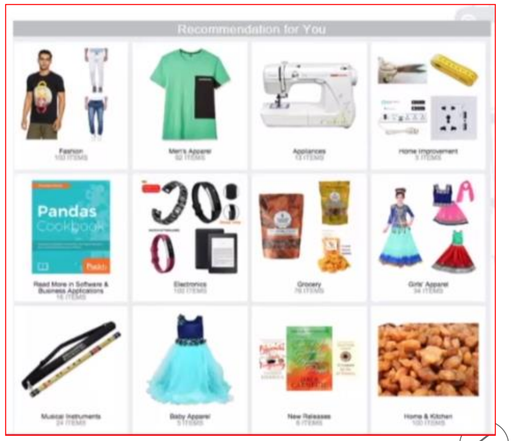


Chatbot helps in Onboarding

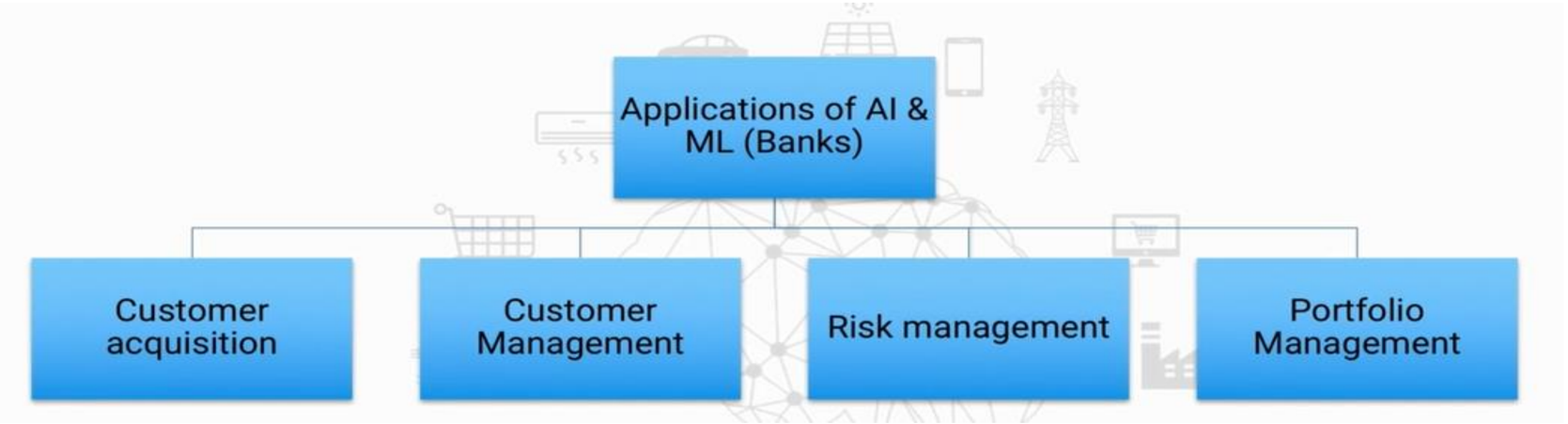


Predicting Attrition

Innovate employee learning experience



Banking

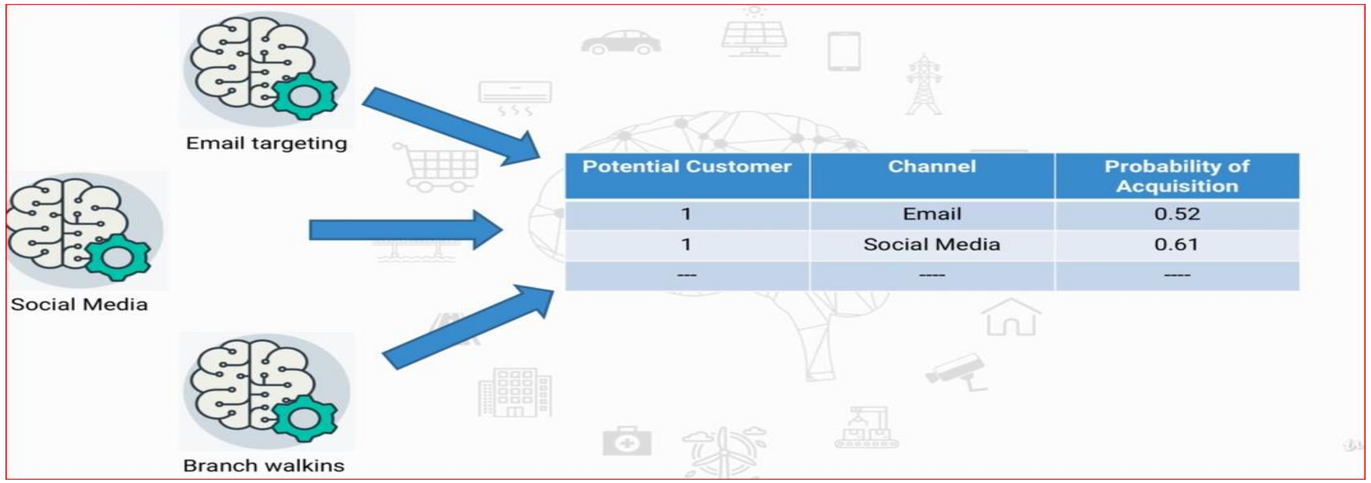


Banks use digital marketing and AI / ML algorithms to acquire new customers

They segment the customers into various groups and target them using E-mail / Social media campaigns

The segments will be derived using Income/Address…etc

Uses Geo Locations to open ATMS Branches using # of people, Life style, Other banks in that Locality



Potential Customer	Channel	Probability of Acquisition
1	Email	0.52
1	Social Media	0.61
---	----	----

RISK MANAGEMENT

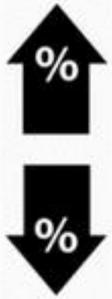
Social media based risk prediction

Analytics Vidhya
Learn everything about analytics

- Lenddo predicts risk based on social media behaviour

Lenddo™

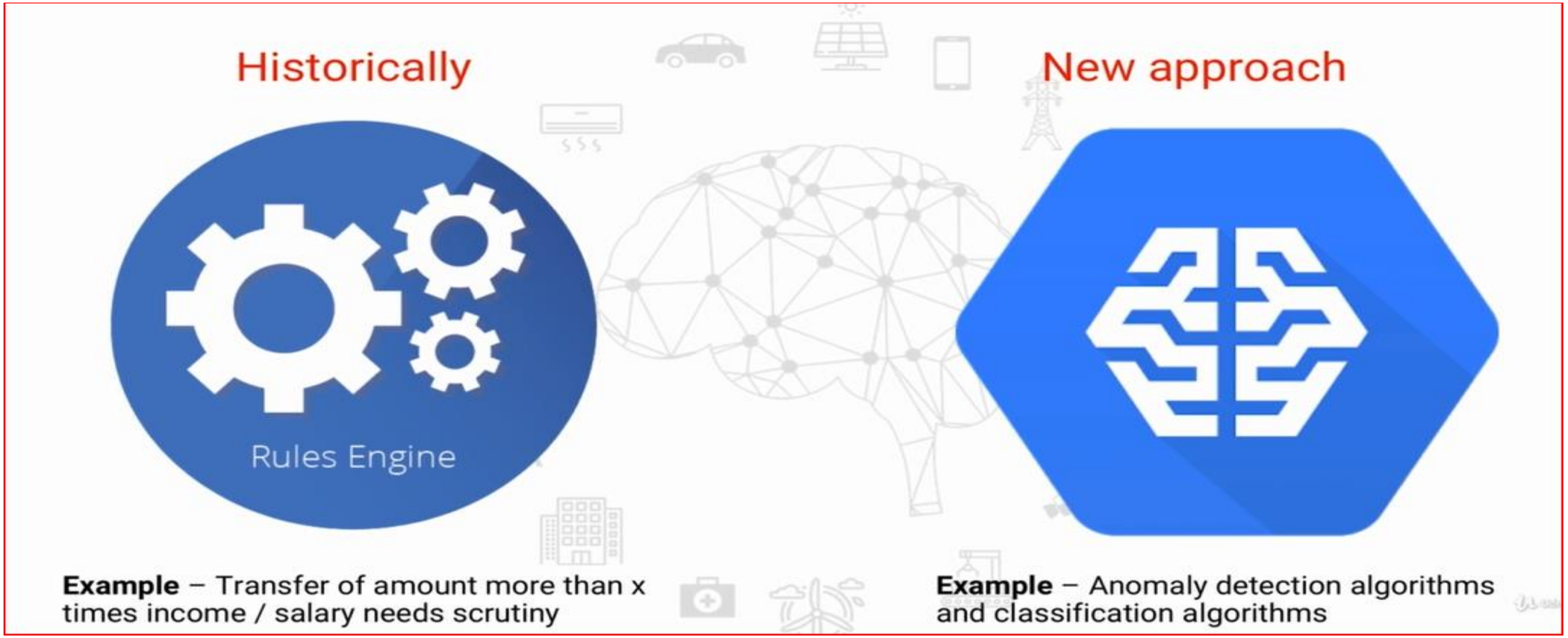
Better Customer Management



Banks uses the information like Customer purchases, Life style & Spending habits—Understand the customer needs/Risks and Wants and hence they do changes in interest rate, Credit limit and cross sell and up sell products For this Banks use lot of supervised Models

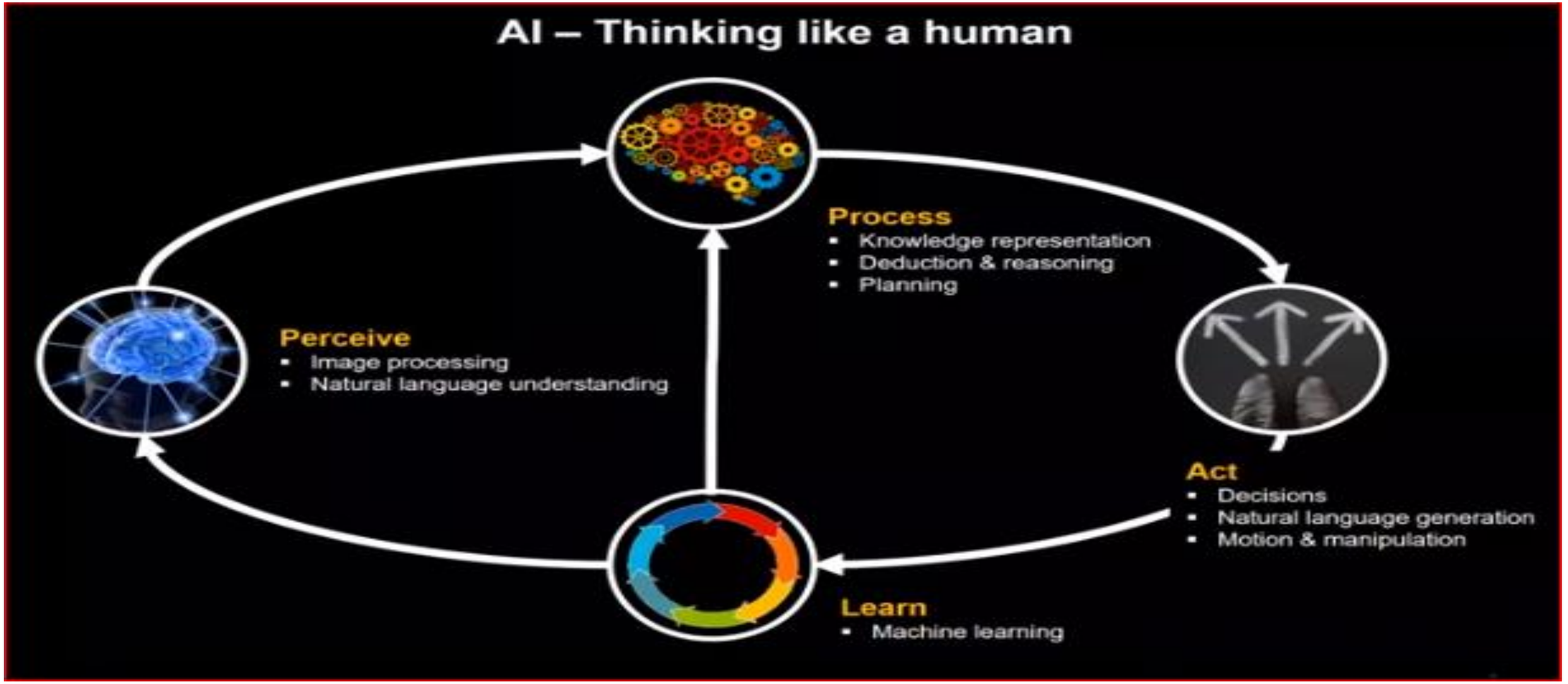
Anti Money Laundering

Needs to understand the history of transaction from the multiple accounts and multiple geo locations.. Hence need machine Learning algorithms solve to AML (Supervised/Anomaly detection using Logistic./Random forest)



HUMAN VS AI

Perceive-look/see/understand language/wor ld around us throuh 5 senses
Process- Understand /make sense of world/we can reason/ we can plan for future
Act- we can walk/ We can move/ we can make decisions
Learn- We learn from mistakes / learn from experiences

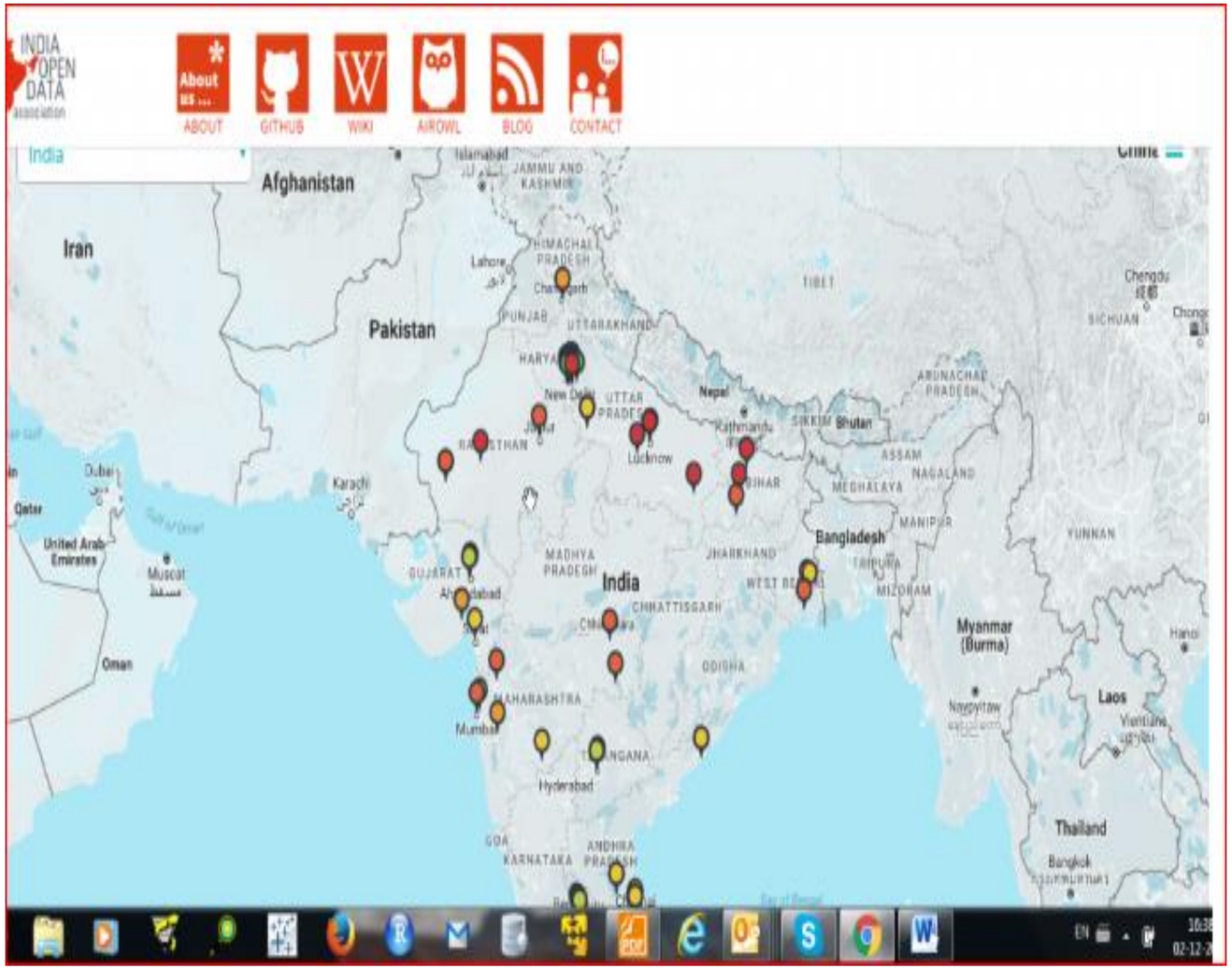


IOT APPLICATIOS

The Indian Open Data Association assembles low cost air quality monitoring systems of three different variants that record various parameters from dust particles to noise pollution levels and reveal real time quality of our air.

Compared to Central Pollution Control Board (CPCB) that also measures the air quality, and has 53 sensors installed, the number of sensors installed by Indian Open Data Association are 40. The existing sensors of CPCB collect real time data at sixty minutes intervals while the Open Data sensors collect data at five minute intervals

The Sensors captures the data like Nitrides(No2), CO and O3(Ozone), Co2,PMS(Particulate Matter is PM2.5, PM10..etc) and the data translates to server from there analytics platform gives the real time suggestions based on the data .



IoT Case study

We are all aware that machine data – from mobile apps, IoT devices etc are richer and more accurate. With the explosion of devices and data usage, we live in times of unprecedented data.

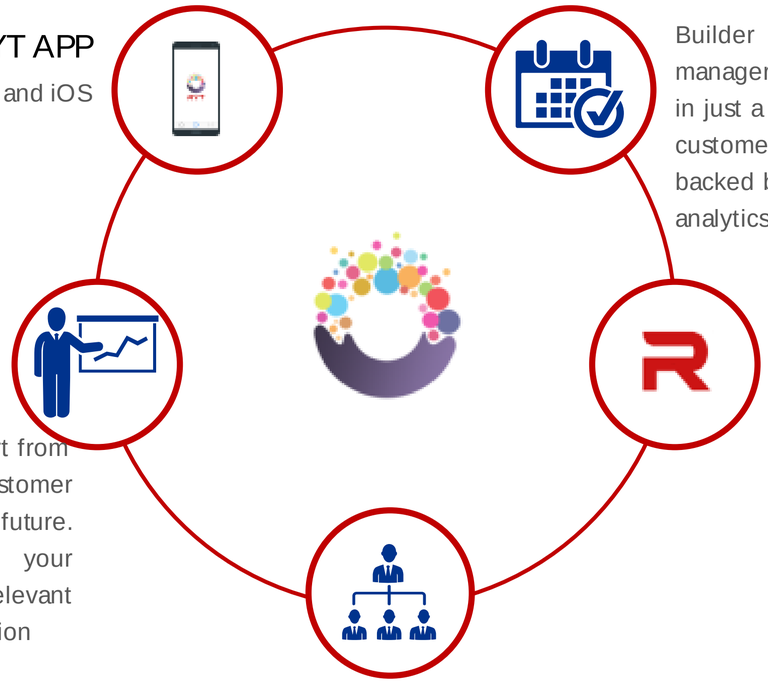
The challenge is in creating the right technology and analytics driven ecosystem to capture and utilize this information for business needs.

The RYT platform: A cloud-based intelligence platform enabling businesses to deliver context-based real-time targeting (personalization) to customers, leading to enriched user engagement, and sustainably higher revenues.

RYT platform demonstrates JSM's ability to conceptualize/ architect and execute:

a) Complex technology platform
b) Conduct deep and actionable analytics
c) Close the loop on providing business benefits

COMPONENTS: RYT ENGAGEMENT PLATFORM



RYT APP

Android and iOS

RYT BUILDER

Builder tool allows the business managers to create live campaigns in just a few seconds to target the customers in real-time. The tool is backed by deep machine learning analytics.

RYT INTEL : Deep learning Analytics

Context is king and apart from location deeper customer understanding is the future. Segregate and target your customers, create relevant campaigns, communication

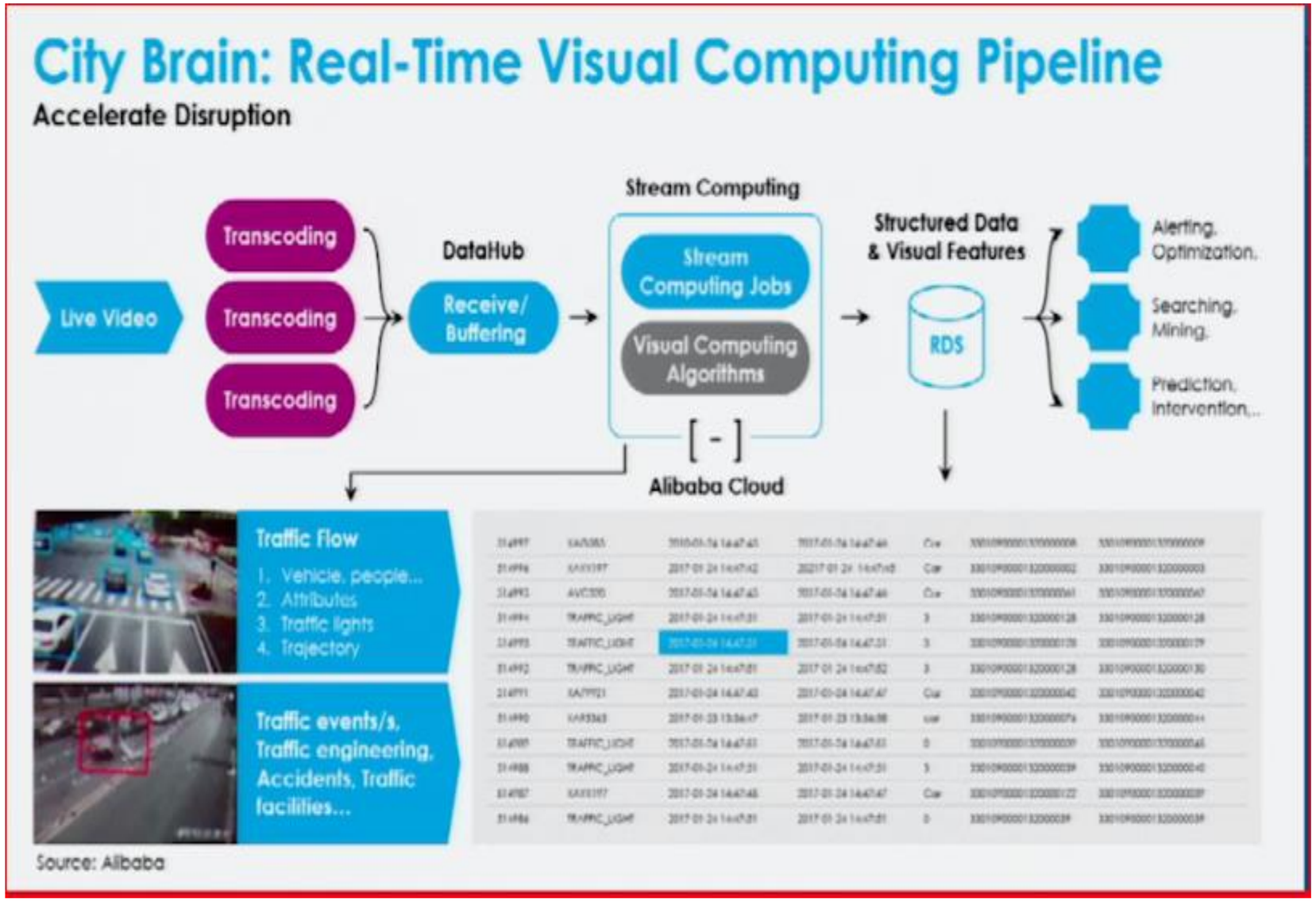
RYT ENGAGE

Recognize and engage customers in real-time, in-premise. Make the front line staff 10X more effective by helping them in delivering assisted selling and customer experience.

RYT MOMENTS

Experiences are made up of many unique moments. RYT allows businesses to create and customize your own Moment Builder.

Smart City



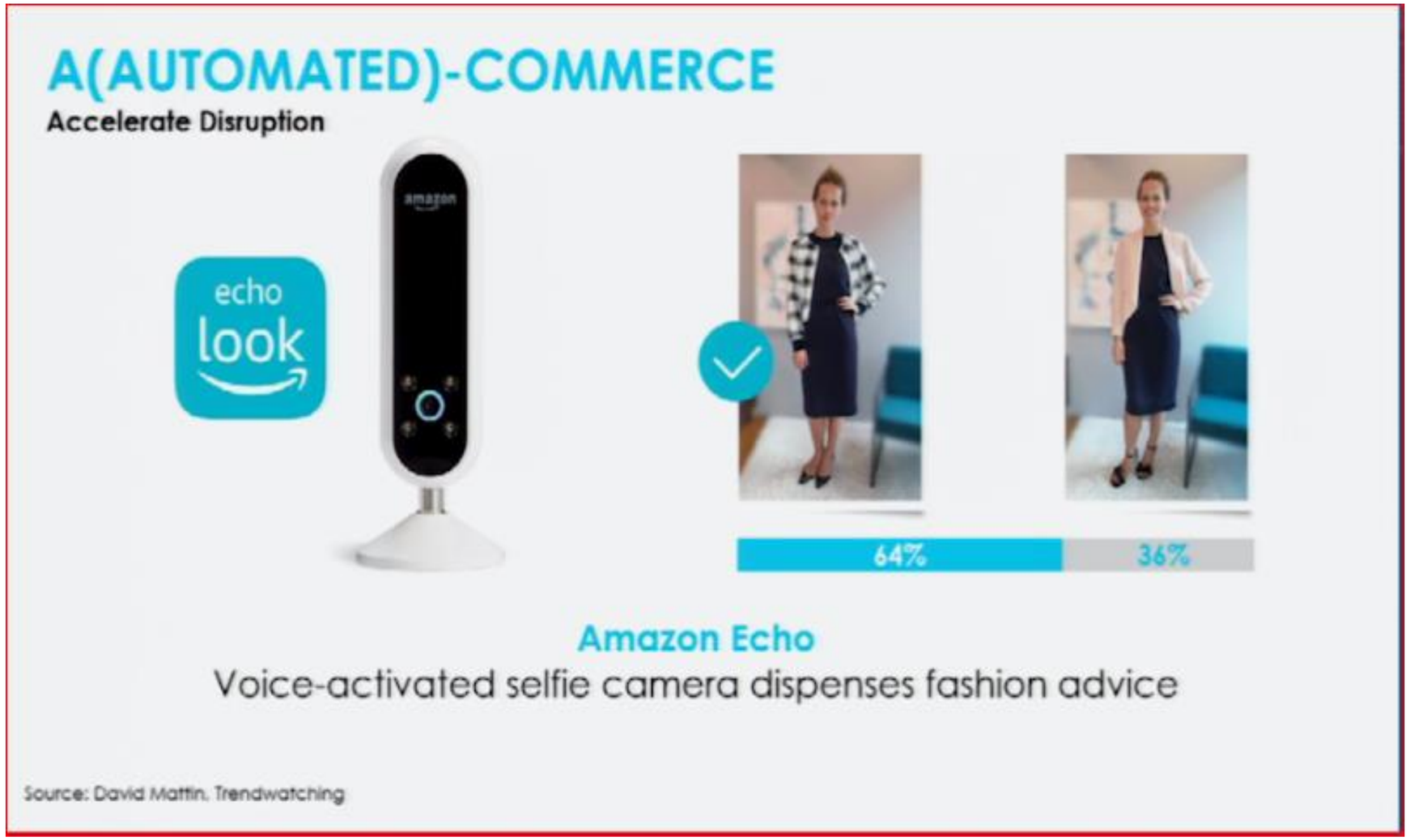
Track everything through Cameras

Track traffic floating like how many Cars/ vechicles passing

With this tracking data and pipilines they analyse predict how to re route the traffic in order to reach the Ambulance and police

Reduced congestion by 15%

E Commerce Turns To A Commerce

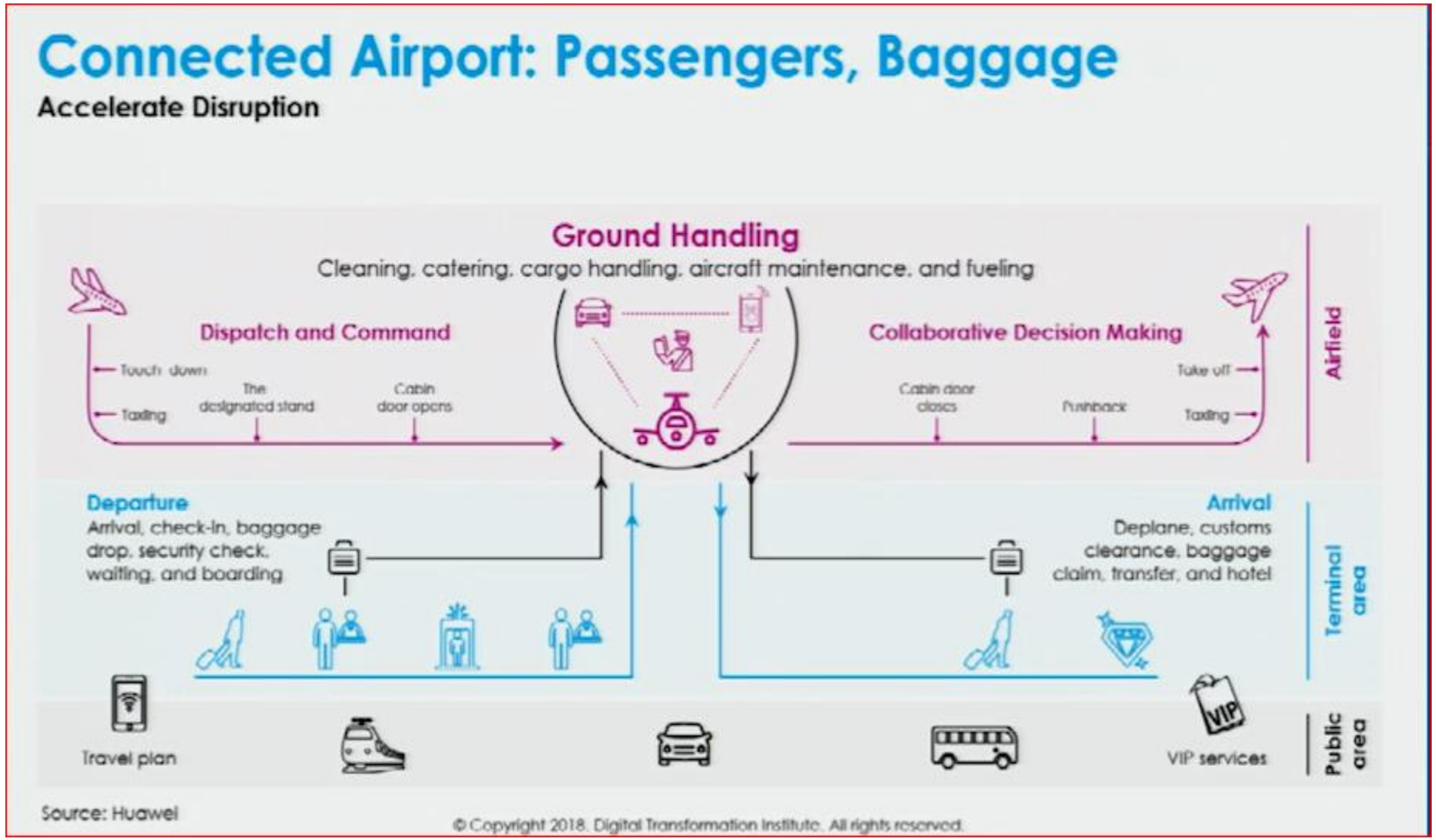


With this Product we can try out Various clothes and it automatically

Suggests what fit best for you based on Previous decisions and based on what all the people do

And based on Trend ..etc

FACE RECOGNITION

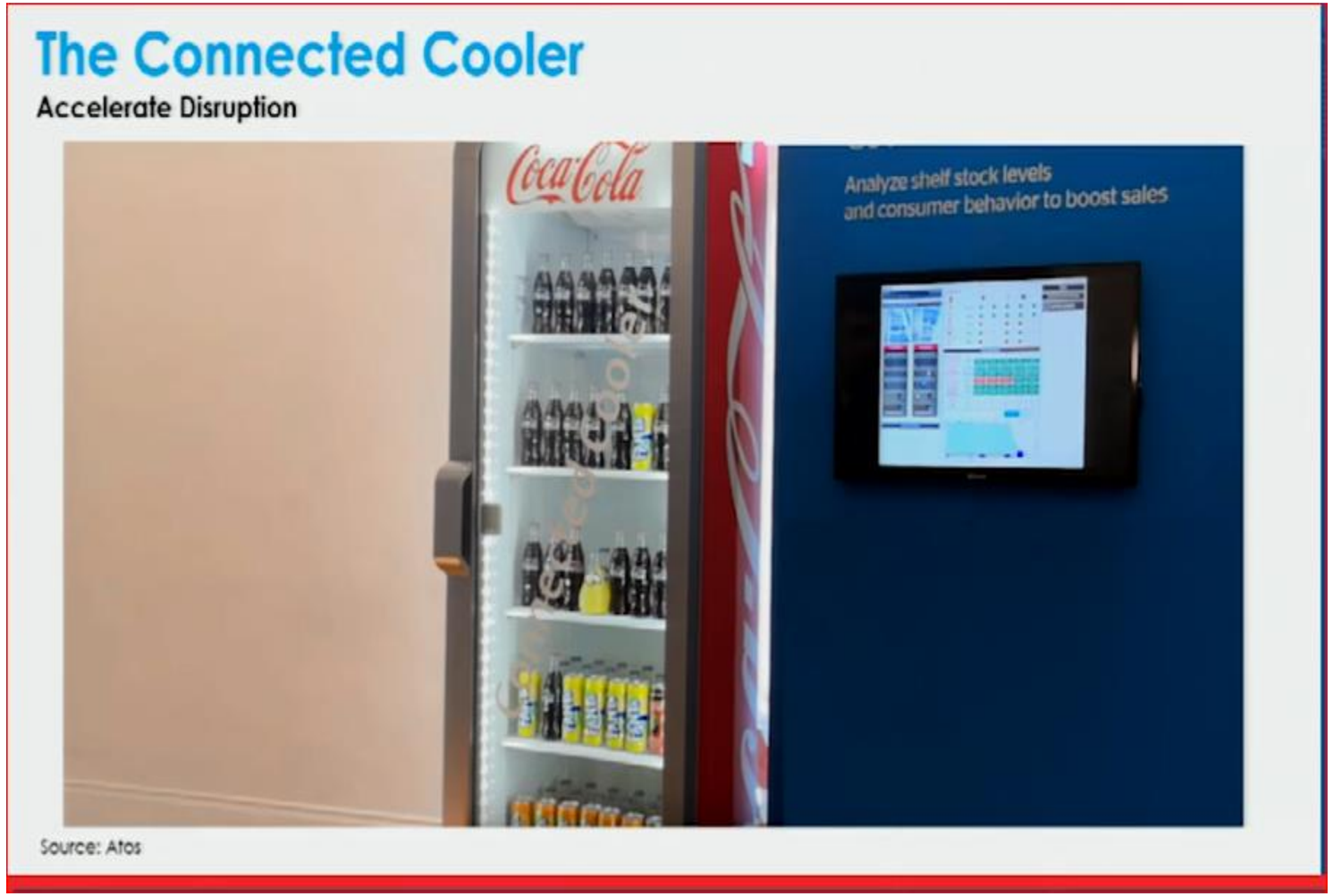


Airports are started using Face recognition for customs and security (No need to check the ticket and passport) we just need to walkthrough

They also detect where are more crowd and they will direct you to shift resources to improve the whole flow in the airport

Preventing safety and improves better customer experience (By connecting everything with webcams
(language, bus ..etc)

IOT- SURVEILLANCE

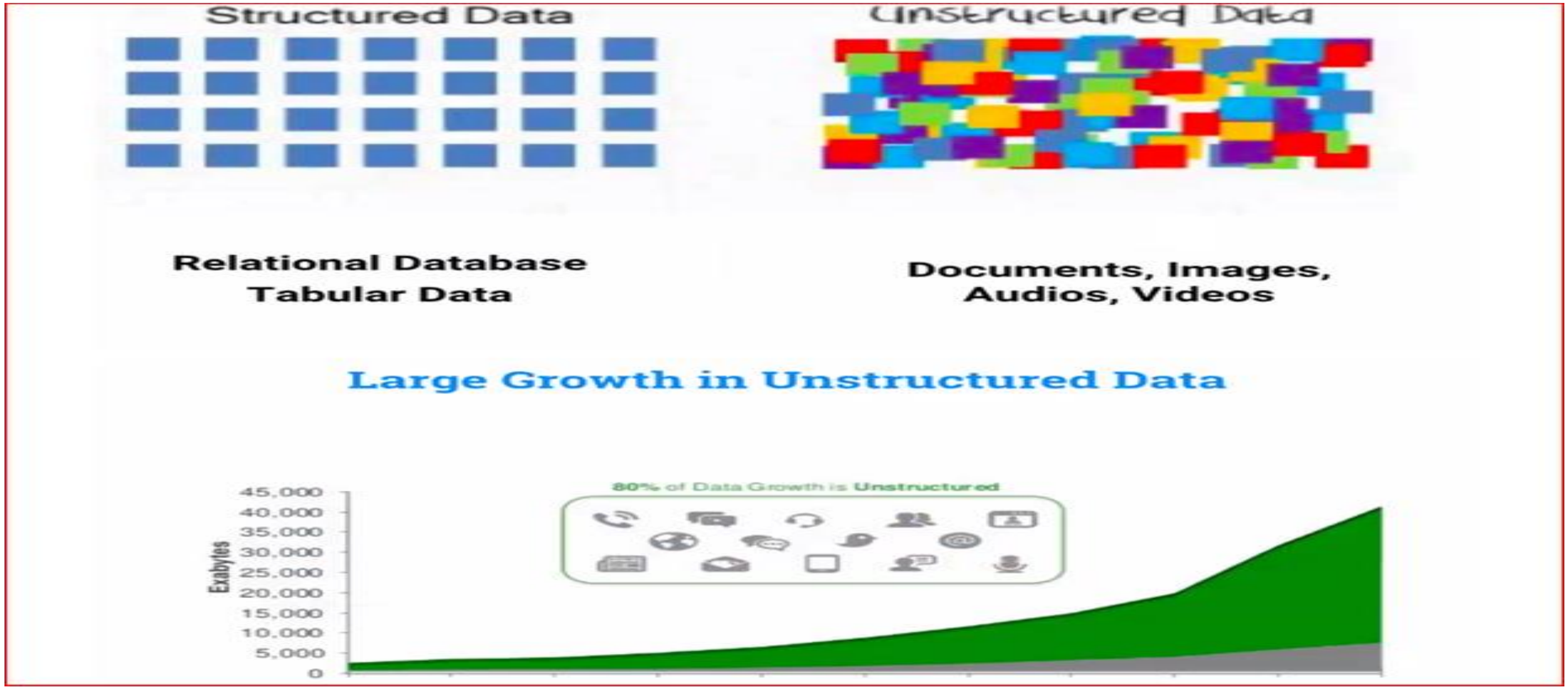


By Connecting Vending Machine through IOT We can Monitor Surveillance of brand

Also Monitor Stock of the brand

(By Taking Picture of it when over it opens and use that information for the Analysis

GROWTH OF UNSTRUCTURED DATA



CHATBOTS IN FUTURE

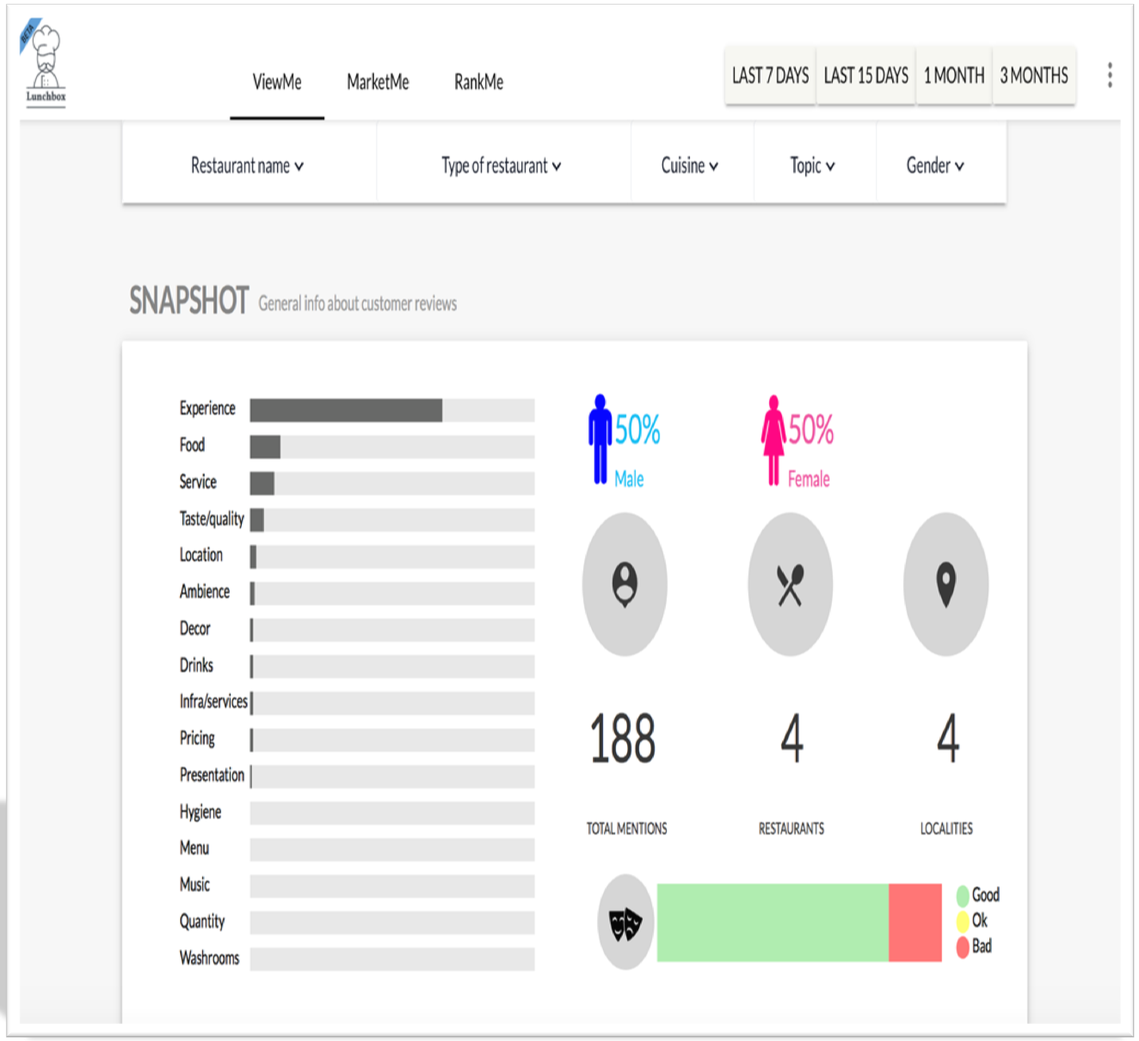
80% of typical customer questions can be handled by AI chatbots

$8 billion per year can potentially be saved by investing in AI chatbots

90% success rate with chatbot interactions expected by 2022

MINING CUSTOMER REVIEWS

CASE STUDY



Challenge

Offering critical insights into brand perception and competitive scenarios to experience brands wanting to manage and learn from data on review sites that is viewed by millions.

Solution

We scrapped over 10 million customer reviews for more than 100000 restaurants across India.

We focused on the following questions for analysing the reviews:

- What are the keywords and topics of discussion across the comments?
- What elements of the restaurant would they want improved – service, staff behaviour, ambience etc.?
- What is the overall mood of the customers visiting the restaurant?

Using advanced techniques of text analytics - machine learning for topics classification, sentiment analysis, and more, we developed a module where restaurant collected reviews could also be analysed and compared for taking better decisions.

WHAT A DATA SCIENTIST NEEDS TODAY.

Computing Knowledge

Statistical Skills

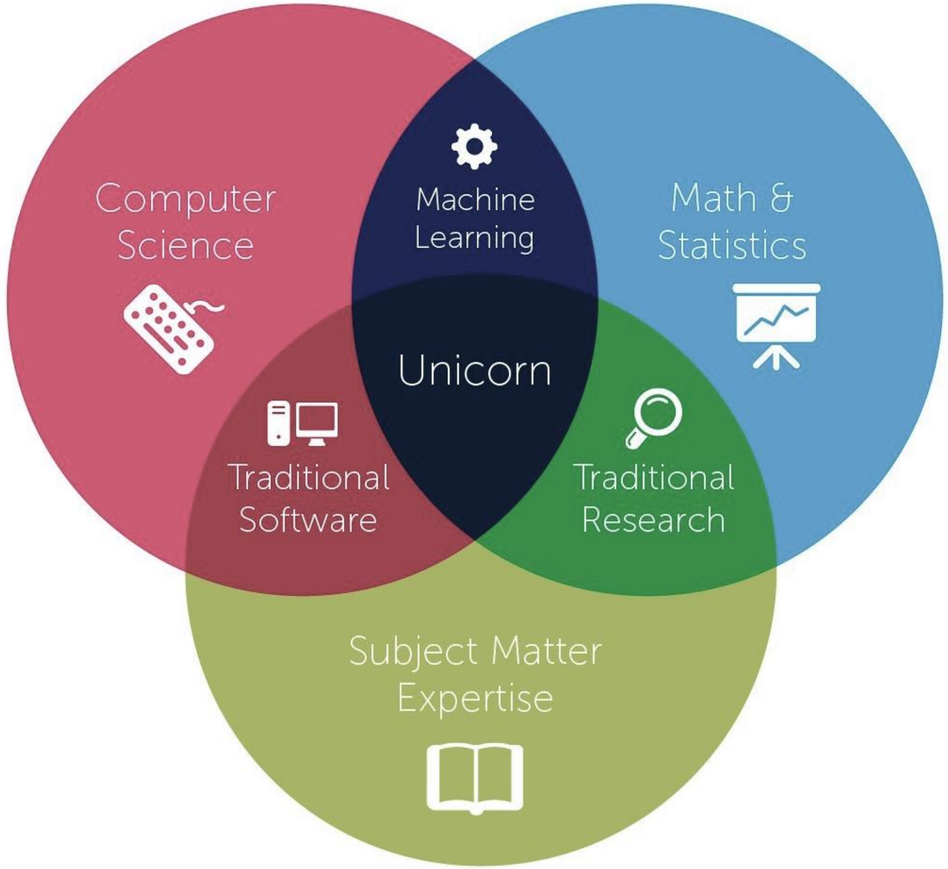
Domain Knowledge

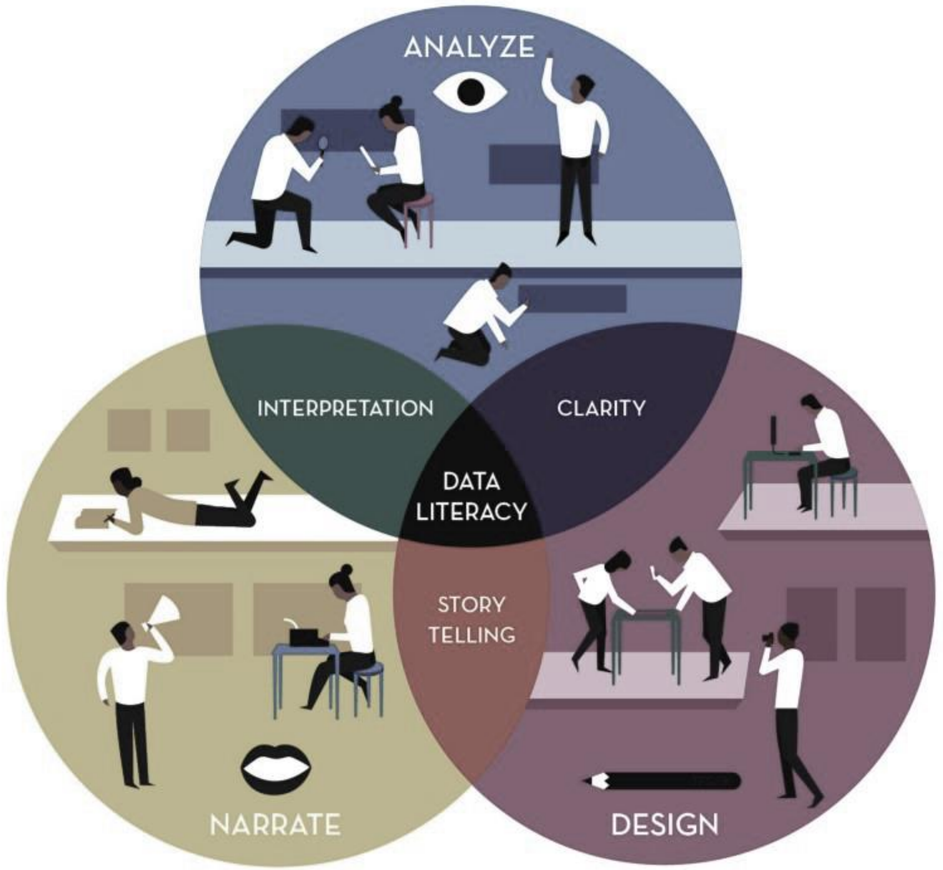
Story-telling and visualization skills

TalentForge
The Finishing Touch

just smart mandate

Role of a Data Scientist





Broadly Analytics

- Decision = Data + Statistics

Analytics – Requires two major Skills

- Architectural
  - Need to understand the problem and need to prepare the complete cycle for the solution in terms of solution as well as interpretation

- And, Executional
  - Able to run various Softwares - Data Bases / ETL

Data Engineer vs. Data Scientist

Source : cognitiveclass.ai

Data Engineer

Data Engineers are the data professionals who prepare the “big data” infrastructure to be analyzed by Data Scientists. **They are software engineers who design, build, integrate data from various resources, and manage big data.**

Because Data Engineers focus more on the design and architecture, they are typically not expected to know any machine learning or analytics for big data.
Skills: Hadoop, MapReduce, Hive, Pig, Data streaming, NoSQL, SQL, programming.

Data Scientist

A data scientist is the alchemist of the 21st century: someone who can turn raw data into purified insights. **Data scientists apply statistics, machine learning and analytic approaches to solve critical business problems**. In addition to data analytical skills, **Data Scientists are expected to have strong programming skills, an ability to design new algorithms, handle big data, with some expertise in the domain knowledge.**

it is essential to know computer science fundamentals and programming, including experience with languages and database (big/small) technologies.
Skills: Python, R, Hadoop, machine learning, deep learning, and Mathamatics & statistics.

source analyticsbuddhu
via @mikequindazzi

DATA SCIENTIST

WHO AM I?

I am a part analyst & part artist. I use my analytical and technical abilities to extract meaning / insights from massive data sets.

WHAT DO I DO?

1. I cleanse existing raw data & build models to predict future data.
2. I go beyond merely collecting and reporting data, to look at data from multiple angles & give meaning to it.
3. I identify the correct business problem(s) & offer solutions (via visualizations, reports or blogs) by best applying the data.

WHAT DO I RELY ON?

1. Analytics
2. Predictive Models
3. Statistical Analysis & Modeling
4. Data Mining
5. Sentiment Analysis
6. What-if Analysis

THE PROCESS I FOLLOW

- Define Problem
- Structure Data
- Use Programming Language

WHAT DO I EARN?

After oil & gas geologists, mine is the 2nd highest paid job in the world!

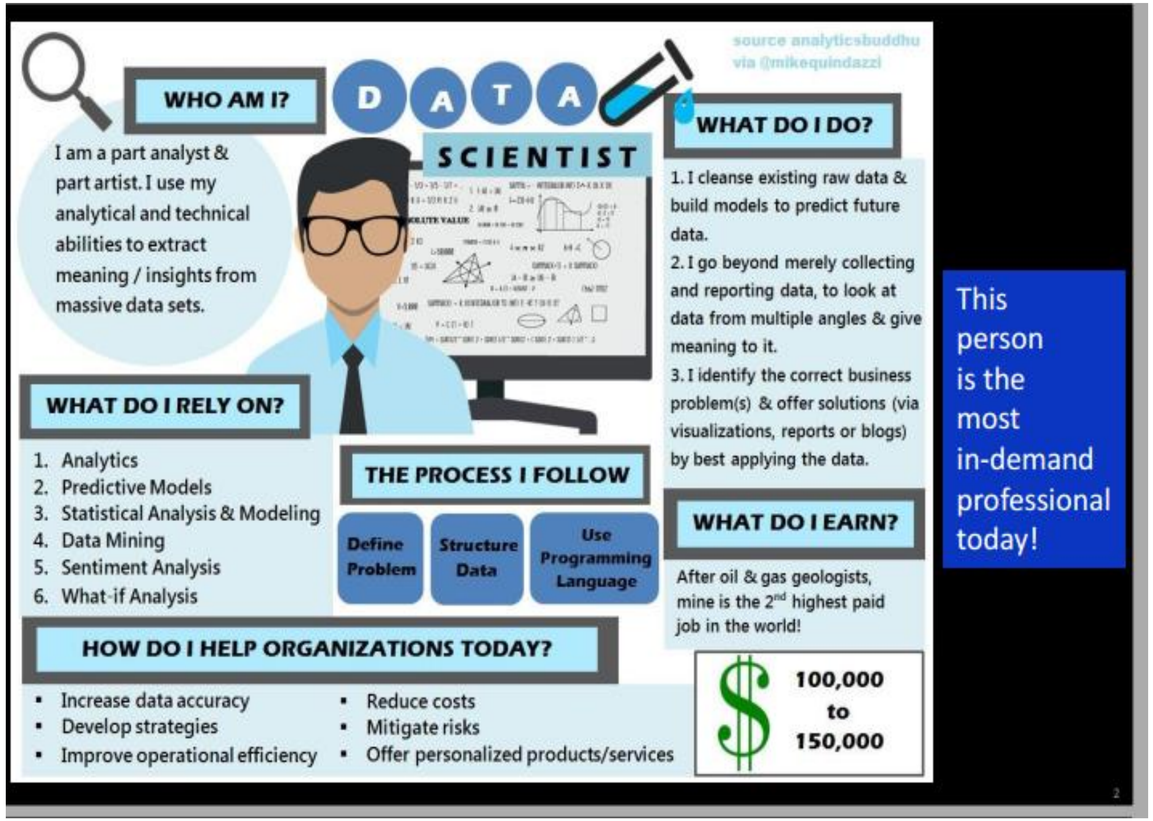
100,000 to 150,000

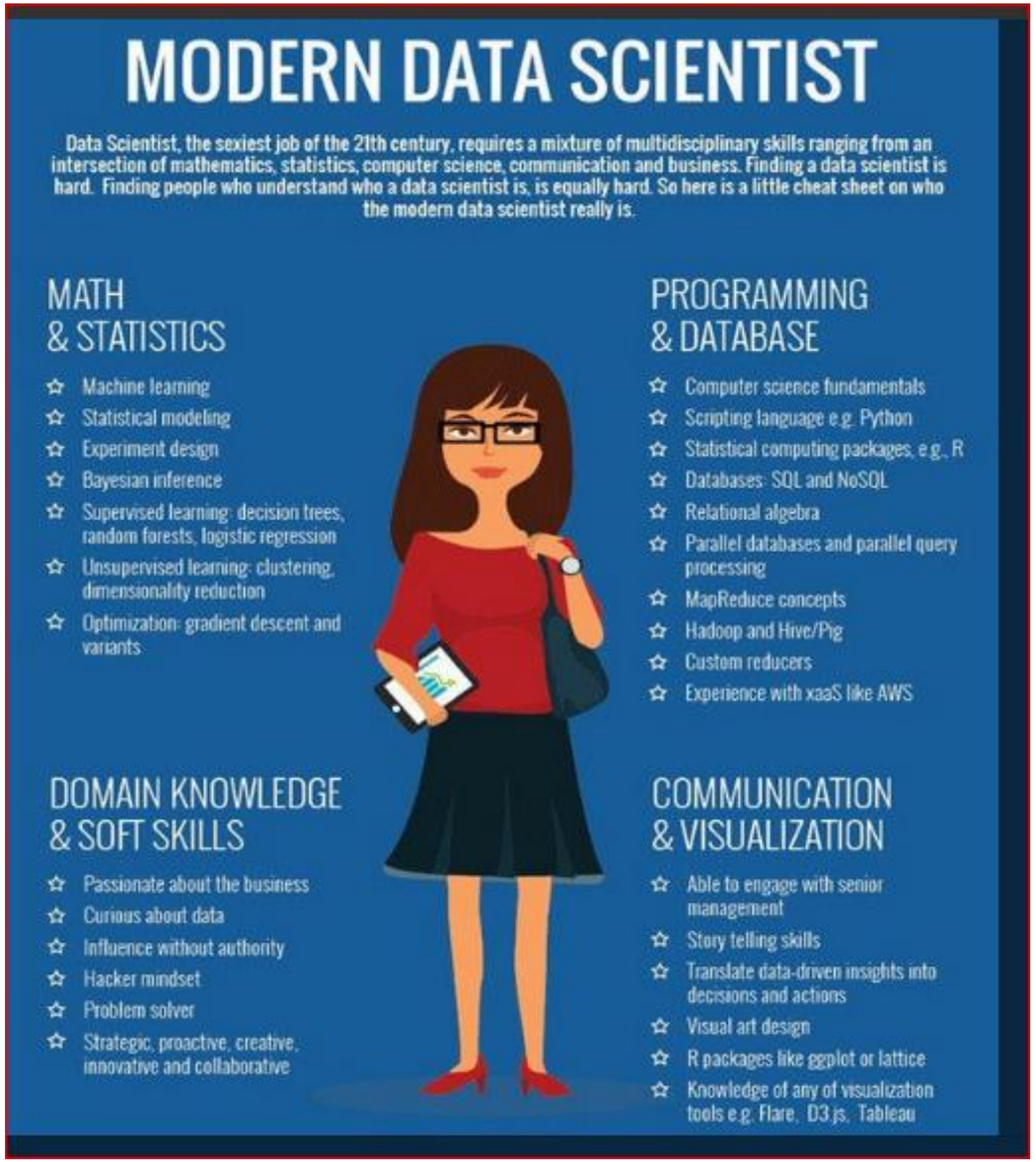
HOW DO I HELP ORGANIZATIONS TODAY?

- Increase data accuracy
- Develop strategies
- Improve operational efficiency
- Reduce costs
- Mitigate risks
- Offer personalized products/services

This person is the most in-demand professional today!

Understanding Roles





PROFESSIONAL OPPORTUNITIES

The most common job titles in Big Data Analytics are:

Data Scientist

Analyzes the data from various angles, determines what it means, and recommends ways to apply the data

Data Analyst

Specializes in collecting, organizing, and analyzing data from various resources

Data Engineer

Creates code to designs, manages, and interpret large datasets to achieve business goals

WHAT IS NEXT, AS A CAREER?

India has second most analytics jobs in the world – 90,000 vacancies.

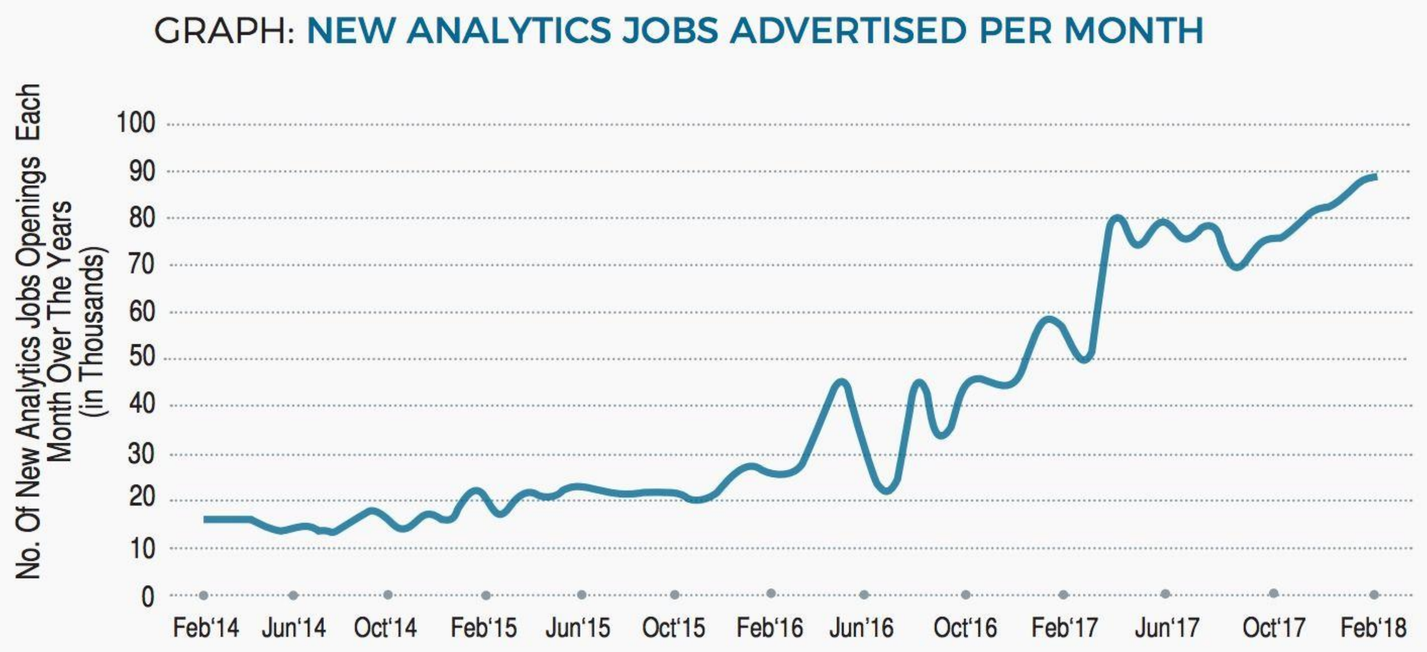
The analytics/data science/Big Data industry in the country is currently estimated to be $2.03 billion annually in revenues.

The industry is expected to almost double by 2020 with a sizable portion of around 24 per cent being attributed to Big Data.

Besides the large corporations, mid-size organisations employ 33 per cent of all analytics professionals in India.

Start-ups employ 27 per cent of analytics professionals.

It is expected that revenue from data science and AI, from both IT and non-IT industries, would be around 16 billion U.S. dollars, providing jobs to nearly 150,000 professionals in the next seven to



Data science and analytics job openings in India are growing significantly. (Edvancer)

WHO ARE HIRING?

Top 10 companies with most vacancies this year.

J.P.Morgan

AIG

accenture

EY Building a better working world

Microsoft

Adobe

wipro

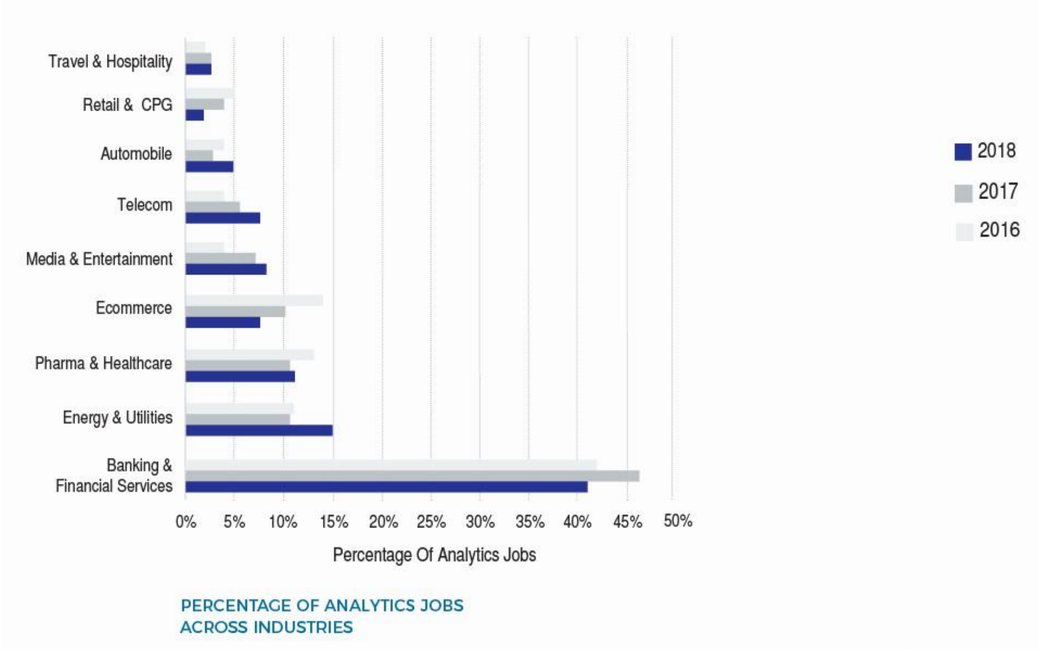
Flipkart

Deloitte.

vodafone

57%

jump in the open job requirements, compared to the same time in 2017



Courtesy: Analytics India Magazine

TalentForge The Finishing Touch

just smart mandate

Systems

Volume

Refers to the scale and amount of data


Microsoft Excel


Microsoft Access


SQL

Variety

Structured data

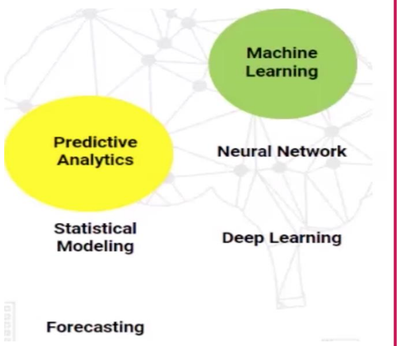
Unstructured data

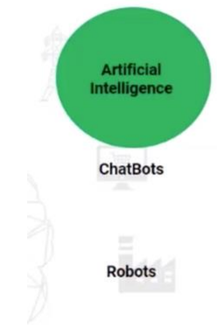
Velocity

Refers to the speed at which data is captured

Tools


Detective Analysis
Data Visualization
MIS
Business Intelligence
Reporting
Dashboard


Machine Learning
Predictive Analytics
Neural Network
Statistical Modeling
Deep Learning
Forecasting


Artificial Intelligence
ChatBots
Robots

How to Build Career in Data Science

- Understanding the Business Need
- Tools / Technologies required
- Community
  - Data science central/AV
  - Kdnuggets
  - Intact social Media
  - You tube / On line Free Courseras
  - Work shops

  - And Finally The Logic here is the Tip
    - I Hear and I forgot
    - I See and I remember
    - I do and I understand