

# BASIC OF DATASCIENCE

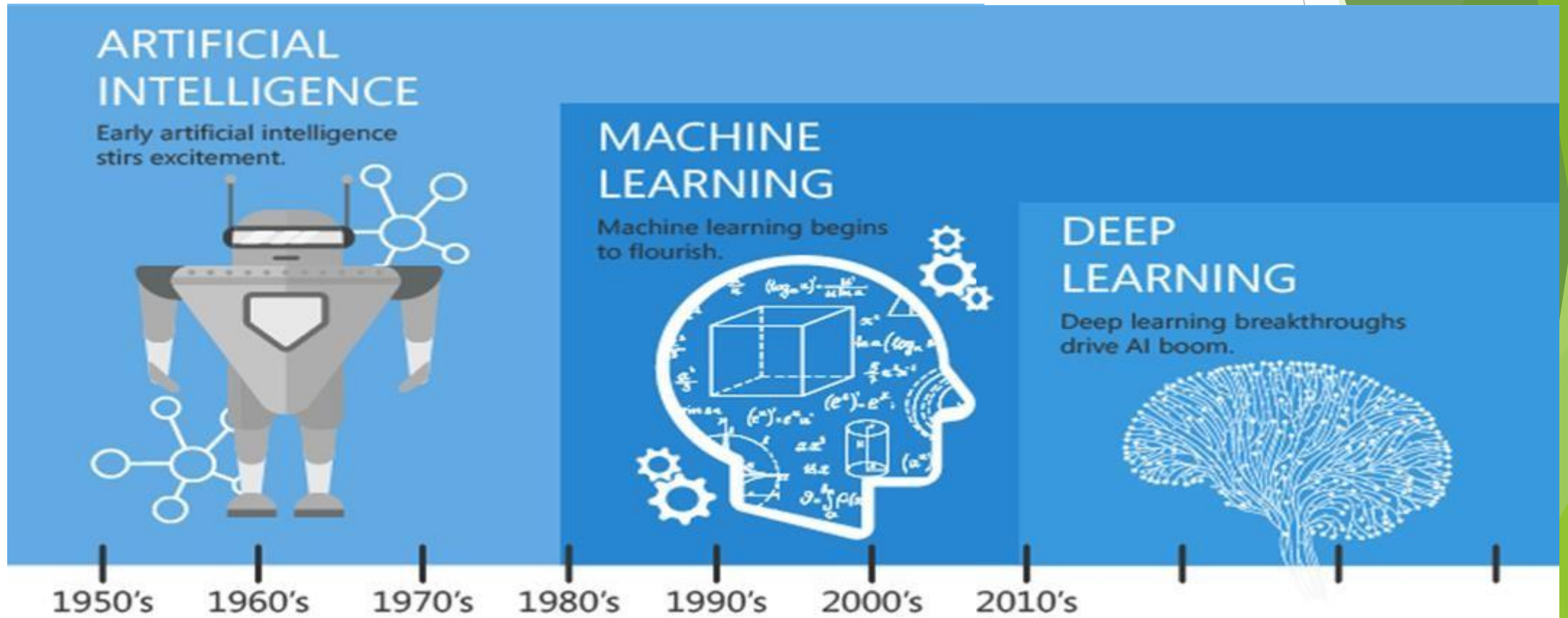
BY :- SHOBHIT TYAGI



# Content

- AI ,ML & DL
- What is Data Science
- Skills Required for Data Science
- What does Data Science enable
- Data Science Process
- What is Data & Type of Data
- Data Collection ,Data Preparation & Data Quality
- DataAnalytics
- Machine Learning and Types
- Machine Learning Use Cases

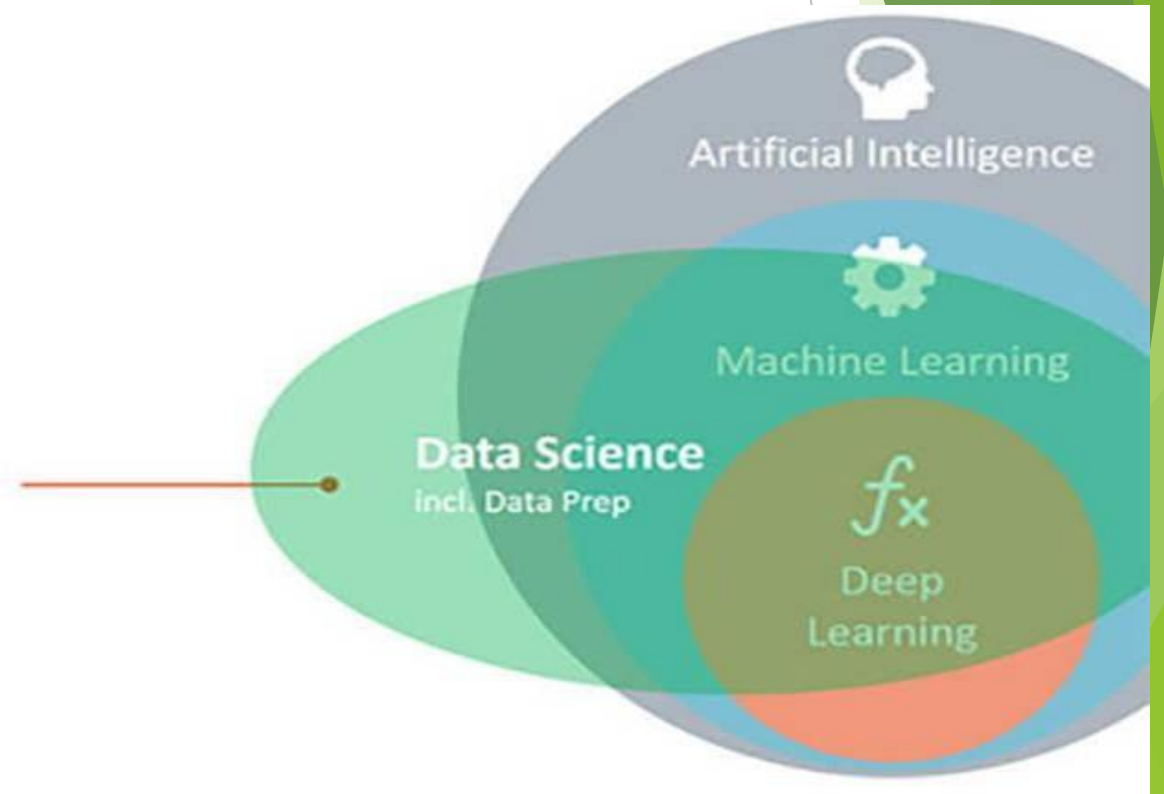
# Artificial Intelligence? Machine Learning? Deep Learning?



# Where does Data Science come?

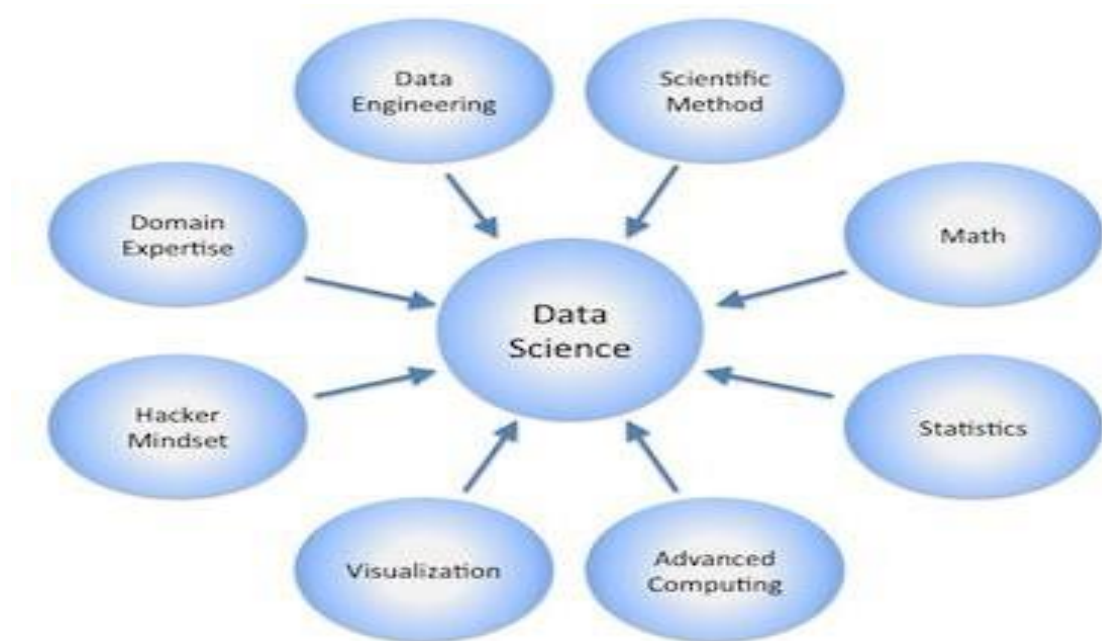
## Data Science

Covers the practical application of advanced analytics, statistics, machine learning, and the necessary data preparation in a business context.

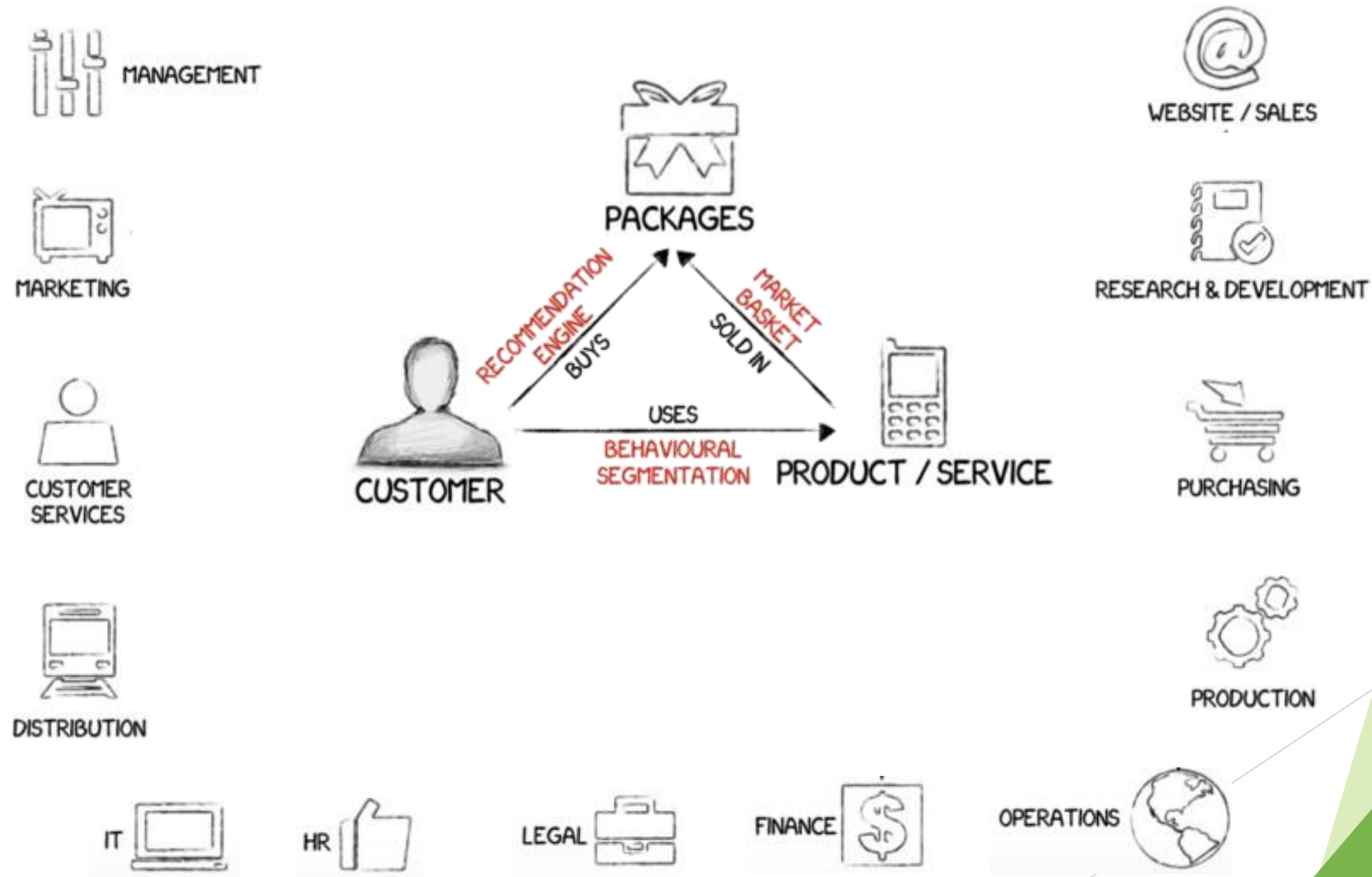


# What is Data Science?

- It is the study of the generalizable extraction of knowledge from data.



# What does Data Science enable?



# What skills are part of Data Science?

	Data Analyst	Machine Learning Engineer	Data Engineer	Data Scientist
Programming Tools	Very important	Very important	Very important	Very important
Data Visualization and Communication	Very important	Somewhat important	Somewhat important	Very important
Data Intuition	Somewhat important	Very important	Somewhat important	Very important
Statistics	Somewhat important	Very important	Somewhat important	Very important
Data Wrangling	Not that important	Not that important	Very important	Very important
Machine Learning	Not that important	Very important	Not that important	Very important
Software Engineering	Not that important	Somewhat important	Very important	Somewhat important
Multivariable Calculus and Linear Algebra	Not that important	Very important	Not that important	Somewhat important

● Not that important    ● Somewhat important    ● Very important

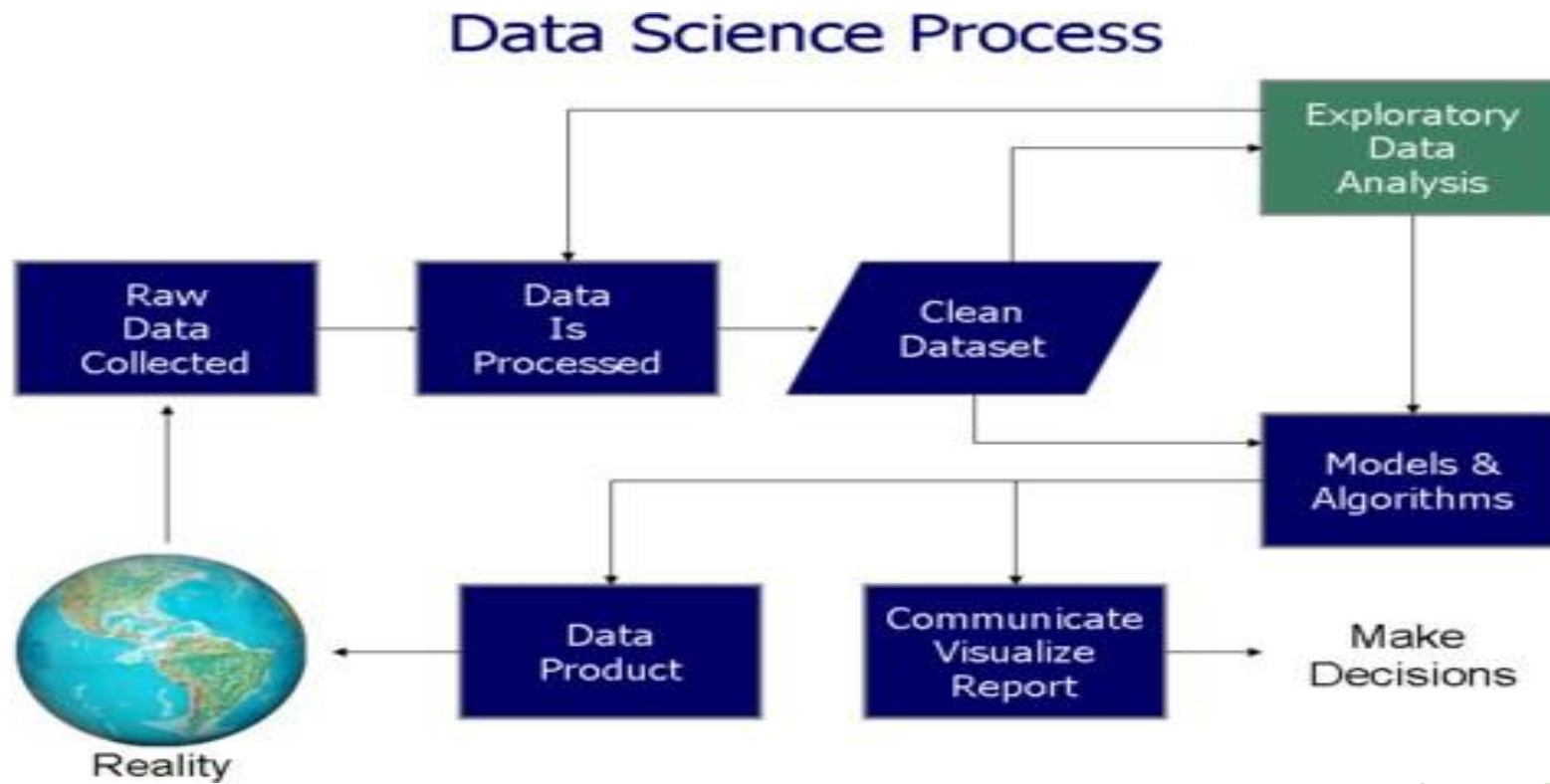


## Programming Skills

- SQL
- R or Python
- PySpark
- Tensorflow , Pytorch



# Data Science Process



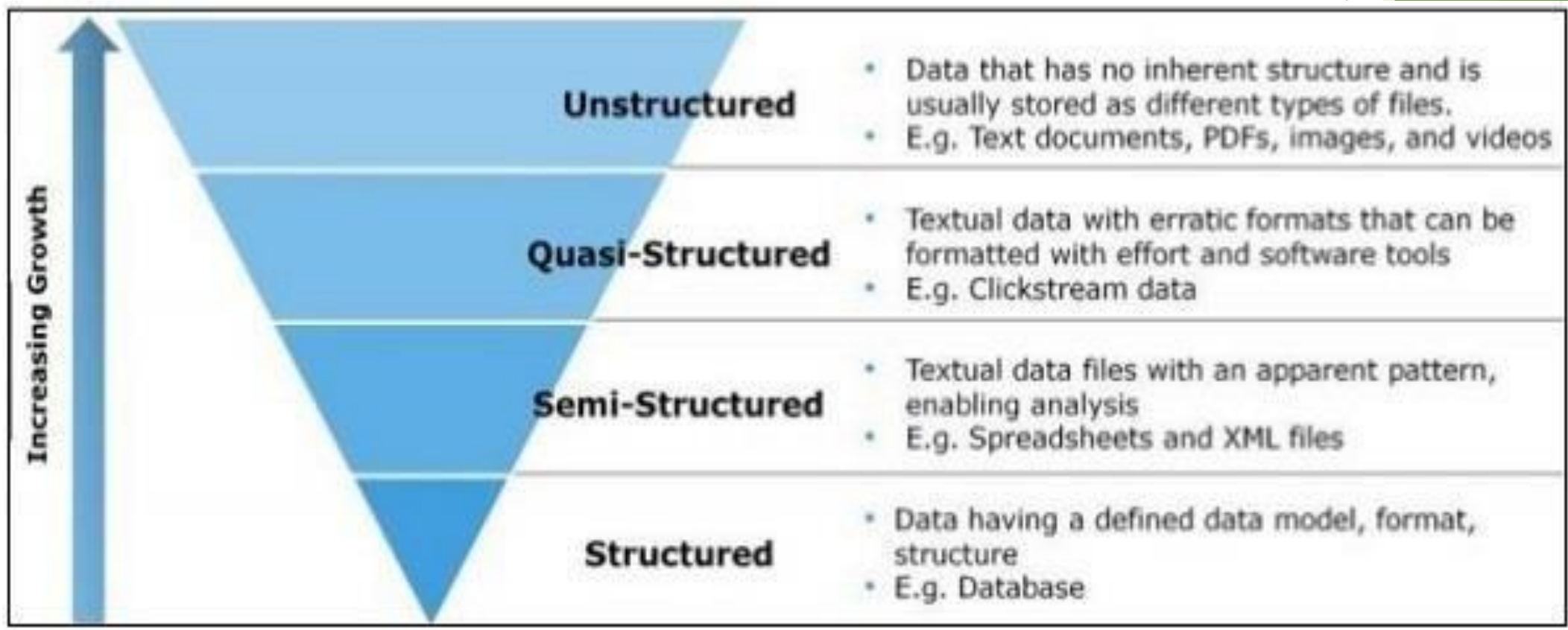


# What is Data?

---

- Data is information in raw or unorganized form (such as alphabets, numbers, or symbols)
- Raw Data also known as **primary data**, collected from a source.
- Example of sources:
  - Web data (Click-stream or click through data)
  - Engineering Info (Data from sensors, log reports, photos, video etc.)

# Types of Data

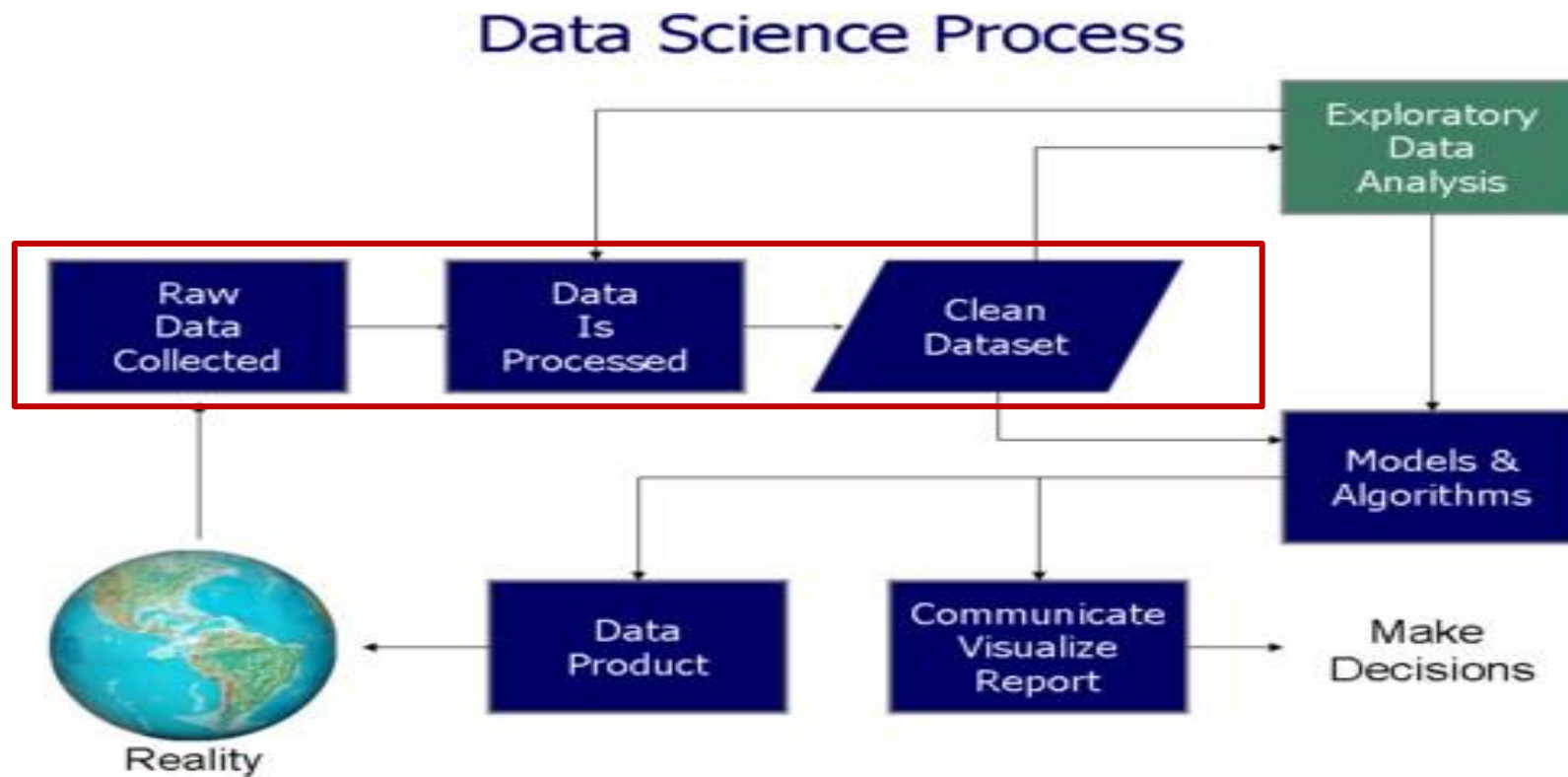


# Wake Up Call

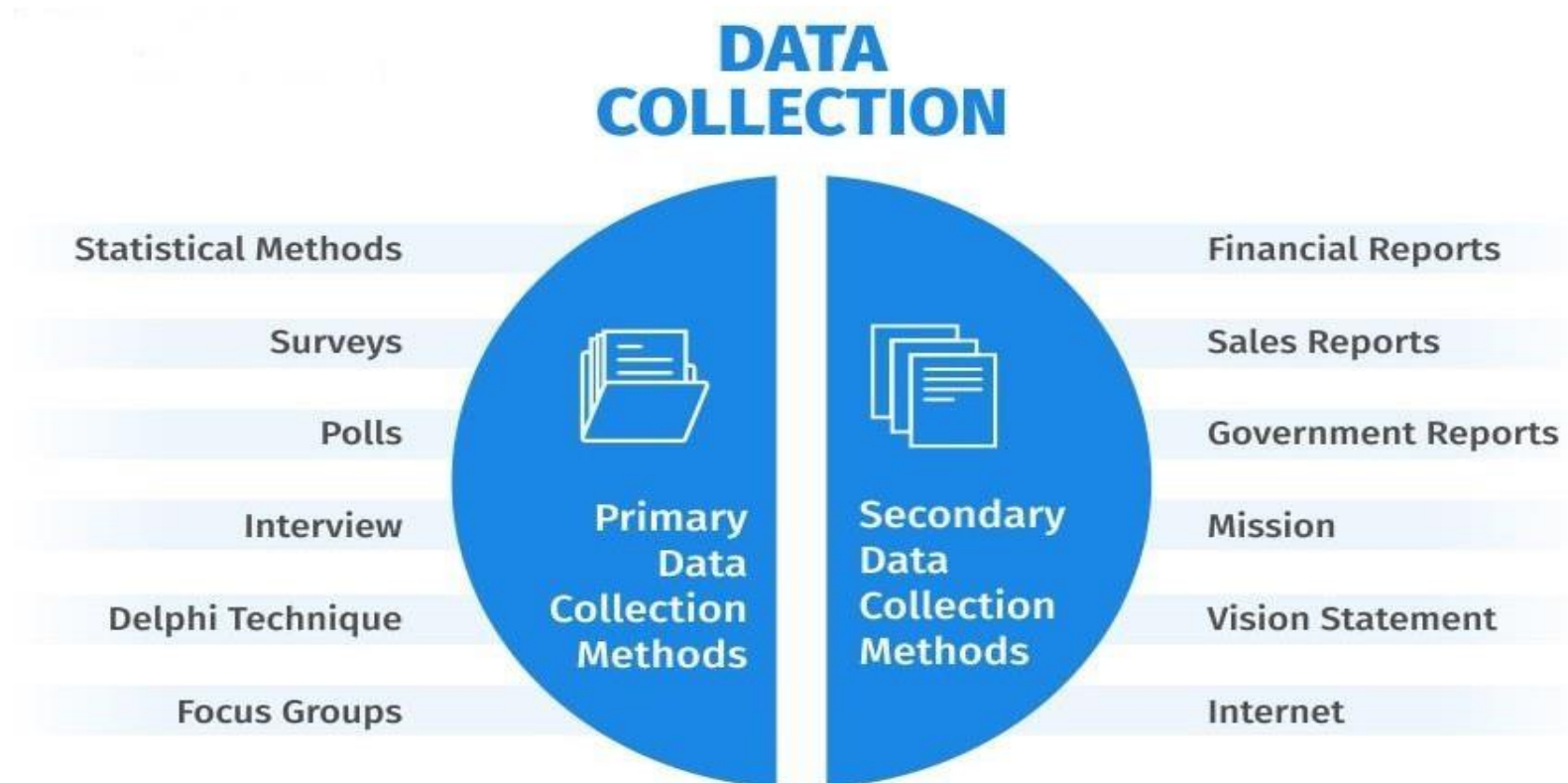


What do you think is the reason for growth of Data Science?

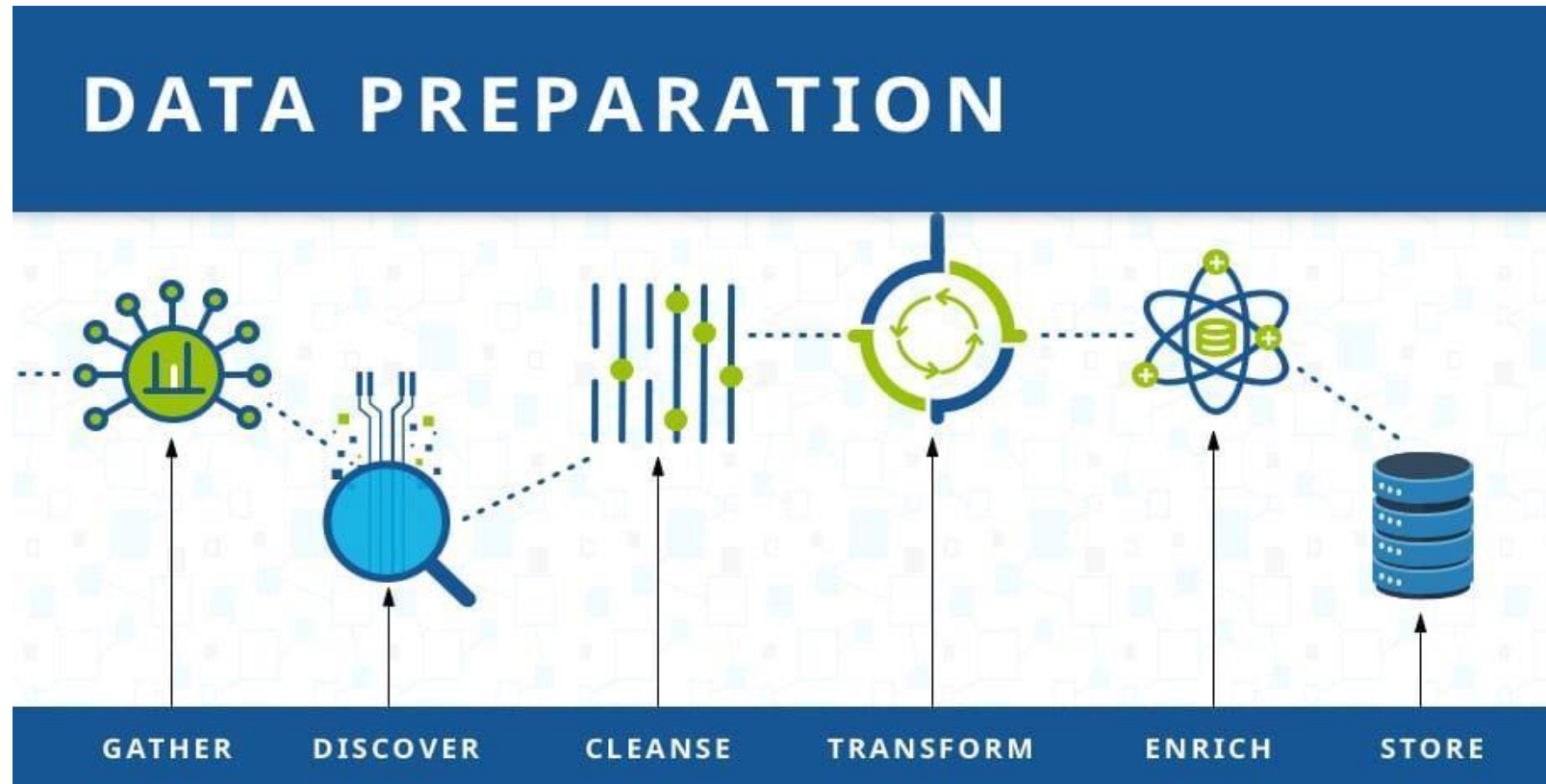
# Data Science Process



# Data Collection



# Data Preparation





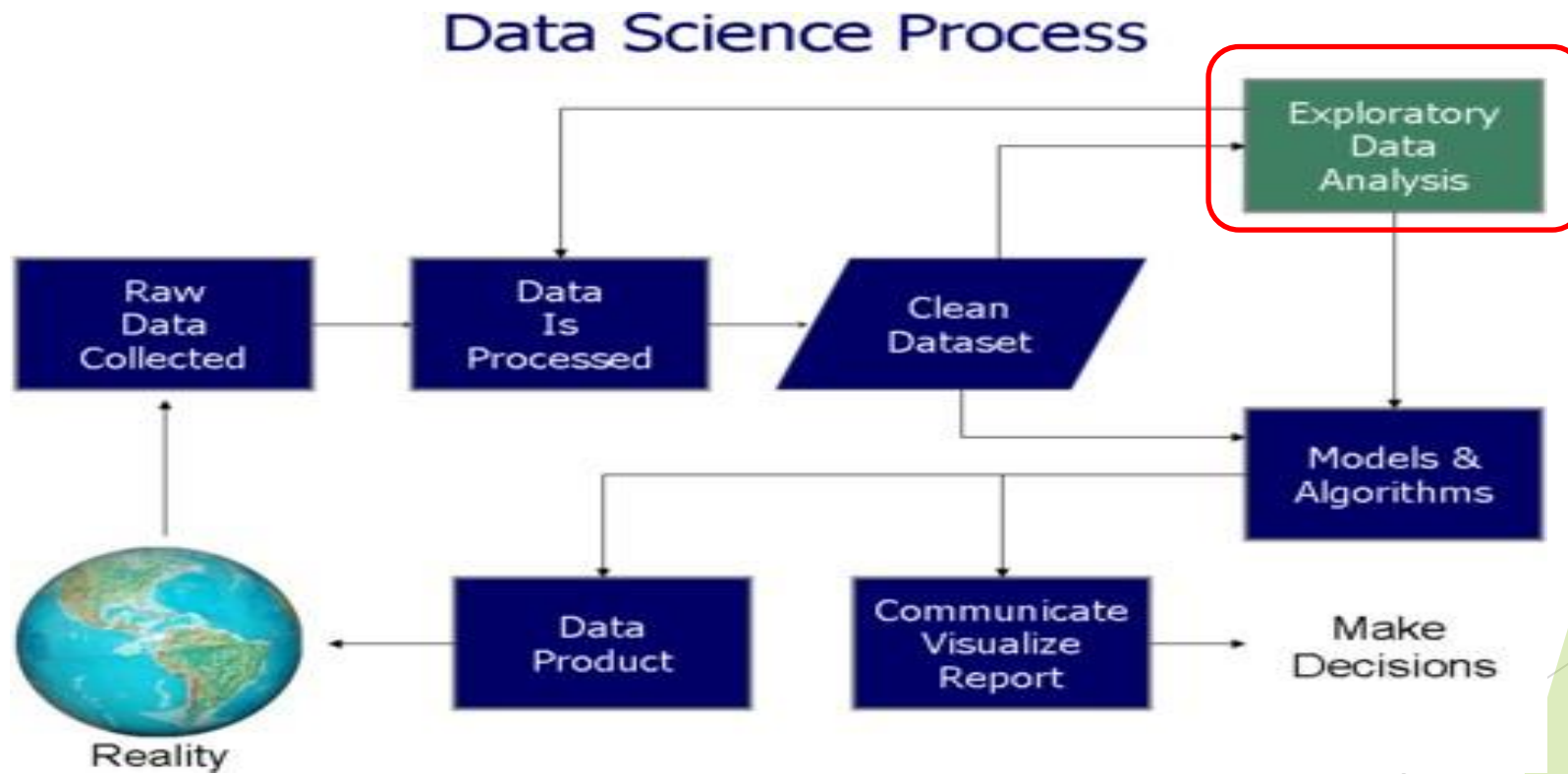
# Data Quality

**Figure 1: Data Quality Dimensions**





# Data Science Process



# Data Analytics

- It is the process of examining data sets in order to draw conclusions about the information they contain, increasingly with the aid of specialized systems and software.



At a high level, data analytics methodologies include **exploratory data analysis (EDA)**, which aims to find patterns and relationships in data

Its technologies and techniques are widely used in commercial industries to enable organizations to make more-informed business decisions and by scientists and researchers to verify or disprove scientific models, theories and hypotheses.



# Categorization of Analytics

#1

## Descriptive

Analytics that helps you understand how things are going.

#3

## Prescriptive

Analytics that suggest a prescribed next step or action.

#2

## Predictive

Analytics that help you forecast future performance and results.



## Quiz Time..

1. Weather Forecast is an example of:

- a. Predictive Analysis
- b. Descriptive Analysis
- c. Prescriptive Analysis

2. Reporting fashion trends is an example of:

- a. Predictive Analysis
- b. Descriptive Analysis
- c. Prescriptive Analysis

3. True or False

A Training Manager uses predictive analysis to discover that most learners without a particular skill will not complete the newly launched course.

Prescriptive analytics can be used to identify learners who don't have skills and can send recommend additional sources to acquire missing skills.

# Descriptive Analytics

#1

## Descriptive

Analytics that helps you understand how things are going.



90% of organizations use descriptive analytics.



Most of the social analytics are descriptive analytics.



Why is it happening?

DIAGNOSTICS



Analyses the data coming in real-time and historical data for insights on how to approach the future.

"The simplest class of analytics, one that allows you to condense big data into smaller, more useful nuggets of information." - Dr. Michael Wu

Business Intelligence and Data mining



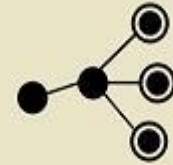
# Predictive Analytics

"Predictive analytics can only forecast what might happen in the future, because all predictive analytics are probabilistic in nature." - Dr. Michael Wu

Forecasting



Analytics is the next step of data reduction.



Predictive analytics provides answers to questions that cannot be answered by BI.



Predictive analytics can be further categorized as –



What will happen next if <condition>?  
Predictive Modelling



Why this actually happened?  
Root Cause Analysis



Identifying correlated data  
Data Mining



What if the existing trends continue?  
Forecasting



What could happen?  
Monte-Carlo Simulation



When should an action be invoked to correct a process.  
Pattern Identification and Alerts

#2

**Predictive**

Analytics that help you forecast future performance and results.

# Prescriptive Analytics



Used in producing the credit score which helps financial institutions decide the probability of a customer paying credit bills on time.



#3

## Prescriptive

Analytics that suggest a prescribed next step or action.

Prescriptive analytics advises on possible outcomes and results in actions

Simulation and Optimization



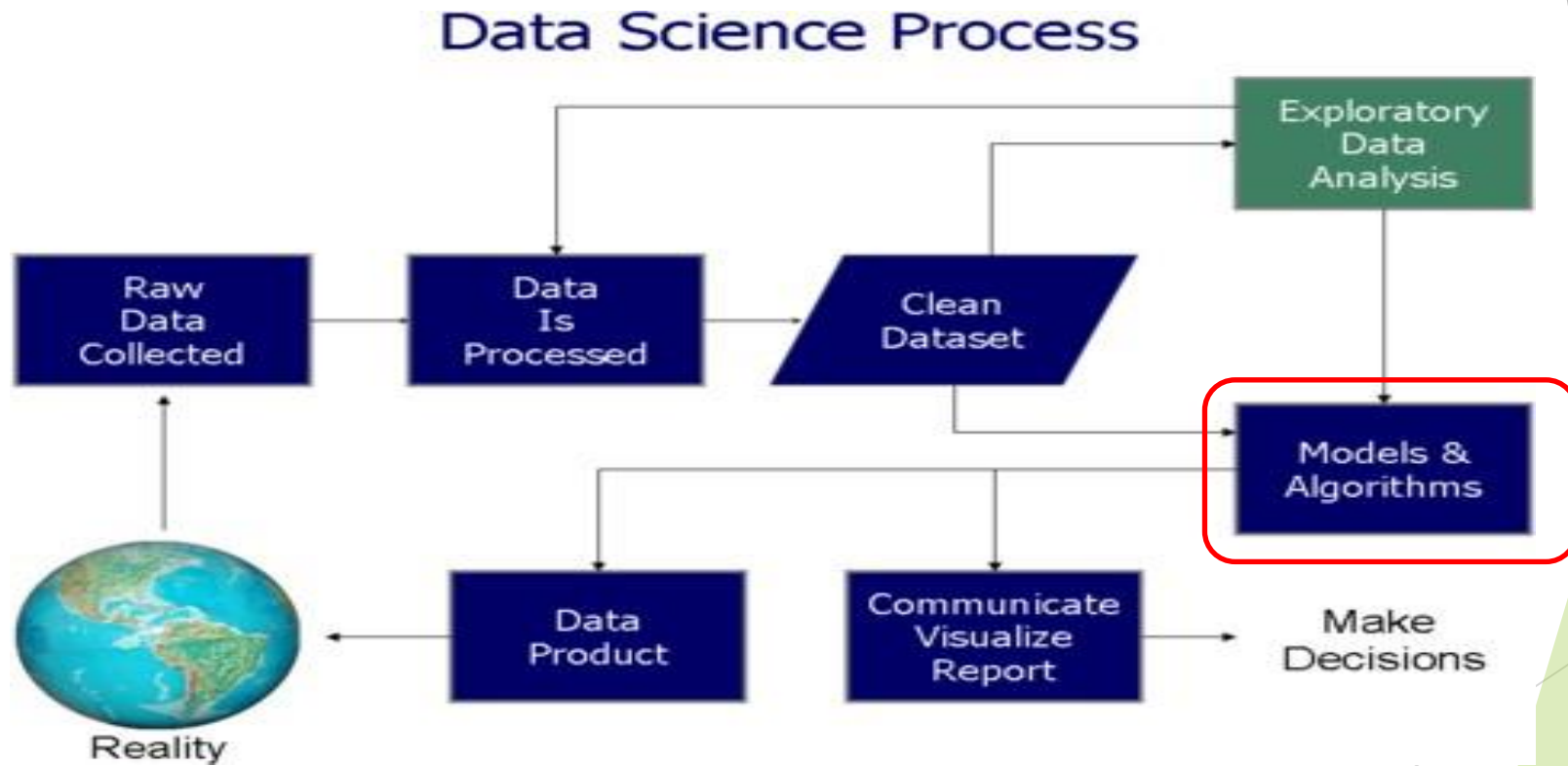
Optimization that helps achieve the best outcomes.



Stochastic optimization that helps understand how to achieve the best outcome and identify data uncertainties to make better decisions.



# Data Science Process



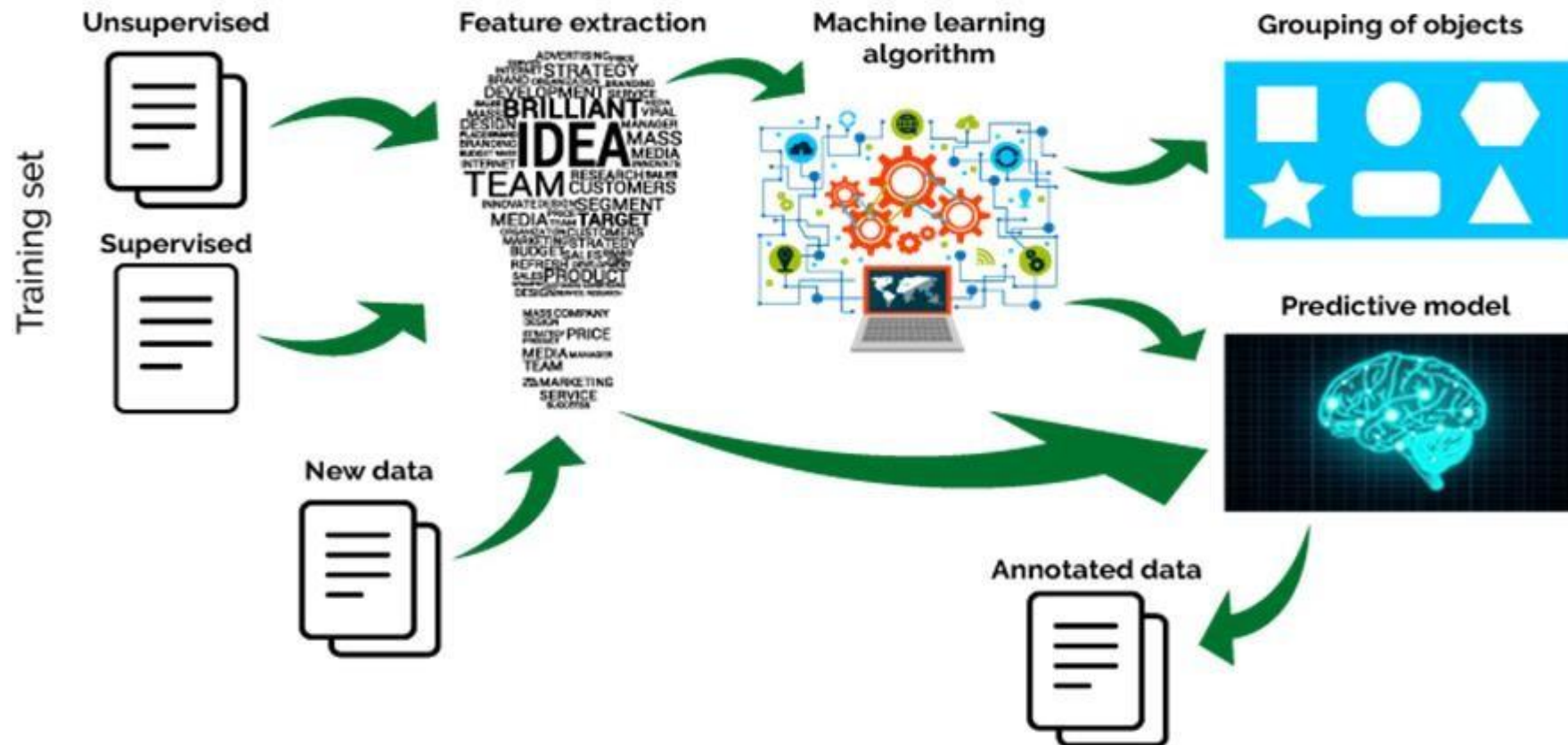
# Machine Learning

- It is a subfield of Artificial Intelligence
- Deals with “construction and study of systems that can learn from data”
- It can be seen as building blocks to make computer learn to behave more intelligently
- It is a theoretical concept. There are various techniques with various implementations.

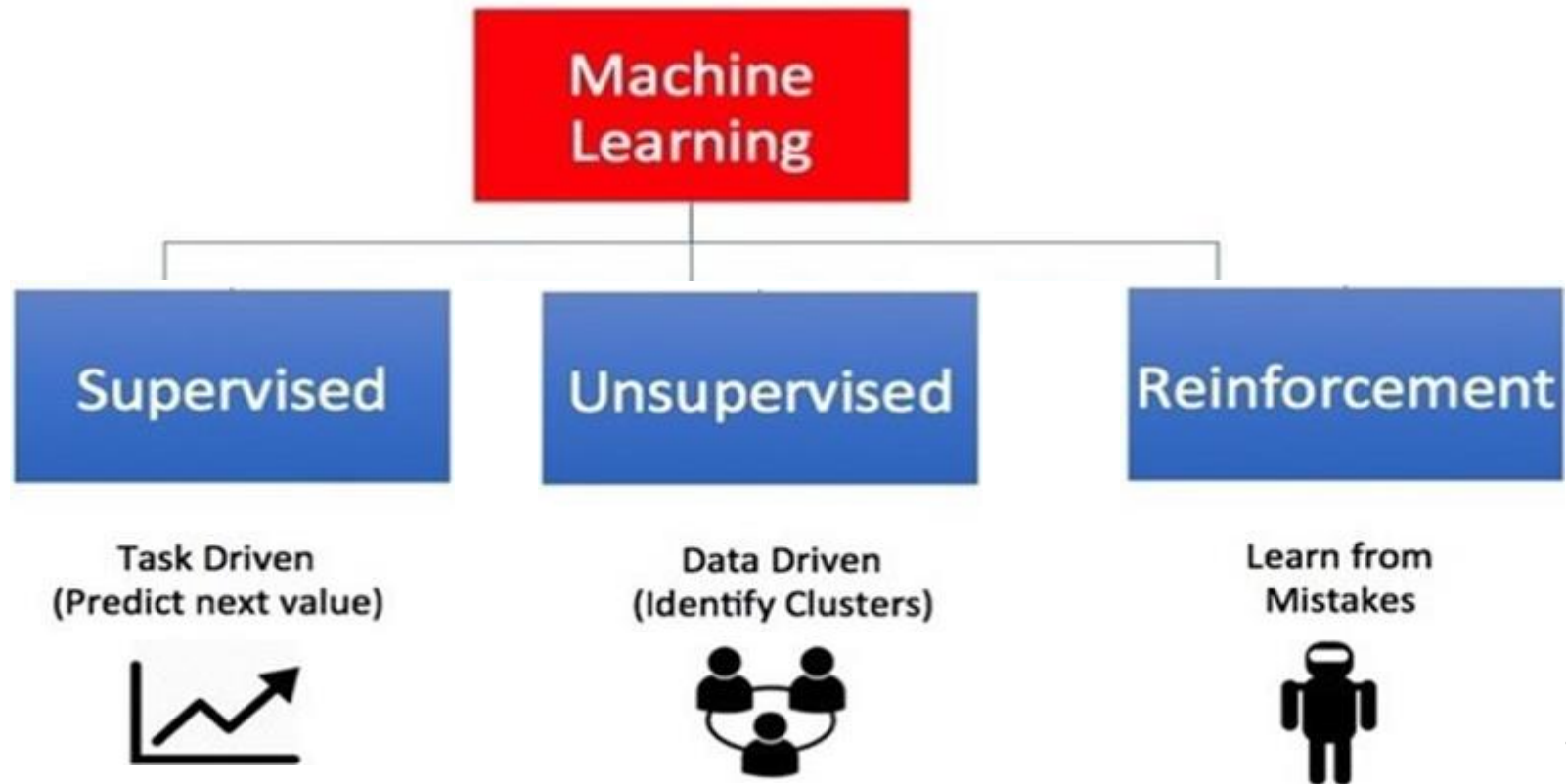
# Machine Learning vs Rule Based

Approach	Rule-based	ML-based
Sub-Goal	To recognize parent-child relations based on event attributes.	
Operation	Rule generation, rule-based identification	Feature extraction, train & test
Model	Empirically derived rules	Automatic generated
Labeled Data	No	Yes (for training)
Manual Effort	Needed for generating rules	Minimal
Flexibility	Low (cannot recognizes beyond rules)	High (can adapt to subtle cases)

# ML Process Flow



# Type of Machine Learning



# Let's dig deep into it Learning

What do you mean by

## Apple

Learning



Apple



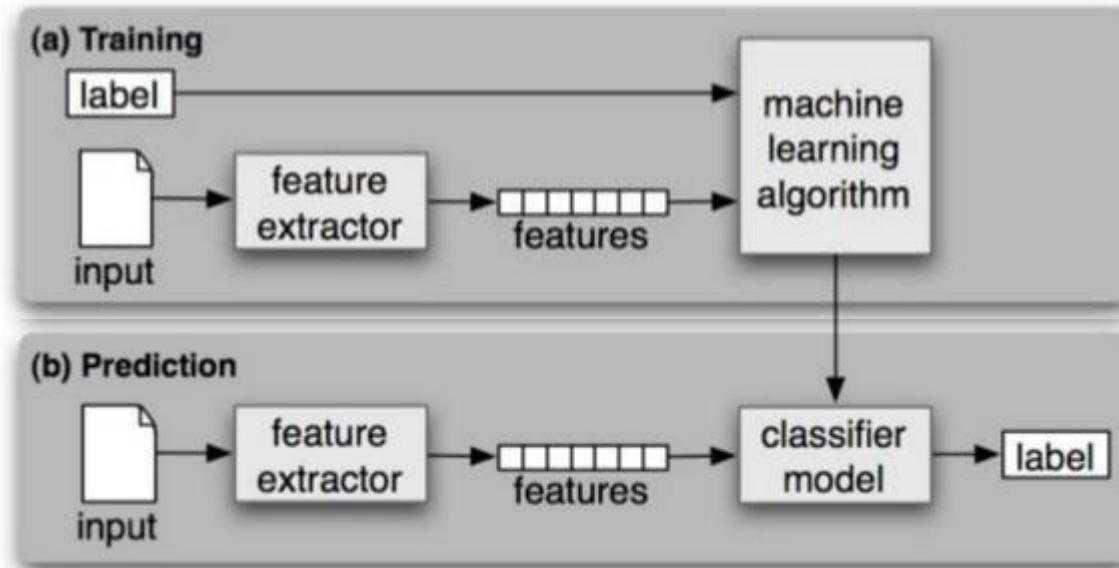
Features

1: Color: **Red**  
2: Type: Fruits  
3: Shape  
etc

1: Color: **SkyBlue**  
2: Type: Logo  
3: Shape  
etc

1: Color: **Yellow**  
2: Type: Fruits  
3: Shape  
etc

# Workflow



Example of a supervised machine learning workflow.



# Machine Learning Algorithms

- Linear Regression
- Logistic Regression
- Decision Tree
- SVM
- Naive Bayes
- KNN
- K-Means
- Random Forest
- Dimensionality Reduction Algorithms
- Gradient Boost & Adaboost

# Real Life Use Cases - Spam mail detection

it's not (Snapshot of Spam folder)

Clip slide

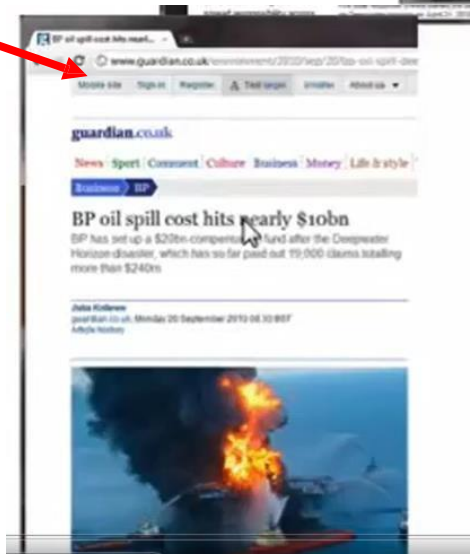
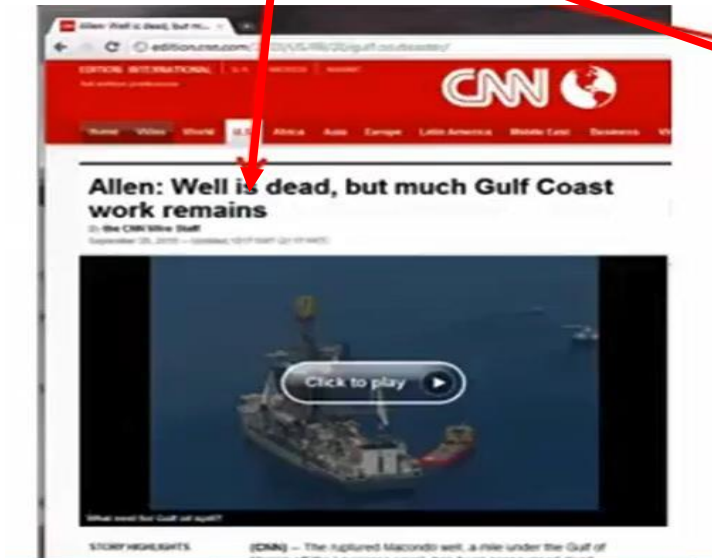
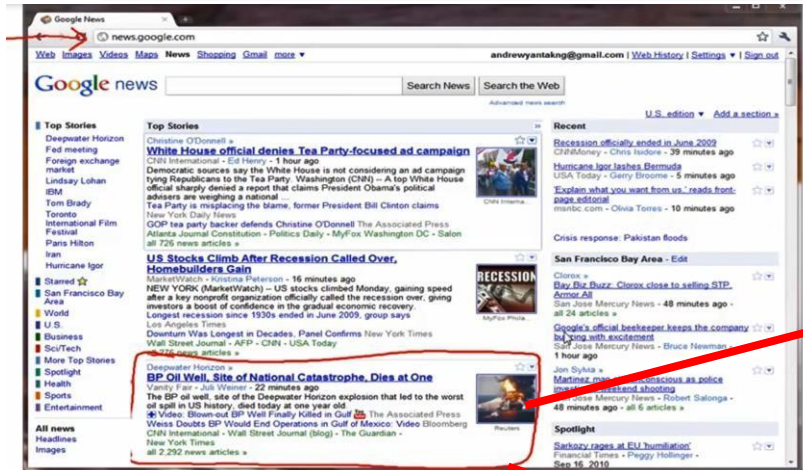
Delete all spam messages now (messages that have been in Spam more than 30 days will be automatically deleted)

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	HDFC Bank	LOAN upto Rs 25 lac - Disbursal in 2 days - open in fresh tab - If you do not want to receive any more newsletters, please click here	9:40 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	iEntry	Welcome iEntry Member - Ultimate Guide To Assessing	9:23 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	New Zealand Jobs.067L	Come to New Zealand to find a great job and settle here [Search for all Jobs from diff... - Search for all Jobs from different kinds of industries Find a Job in Enchanting	8:18 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CarSizzler	Assured Free Luxurious Ride worth Rs.300 with Uber Cabs - Home Home Buy New Car Buy New Car Sell Car Sell Car Tech Tips Tip & Tale Facebook 41727 others	6:05 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Supermarket Promotion	Enjoy Rs.1700 voucher valid at any supermarket! - If you are unable to view this mailer Click here HOW TO CONTACT US? BY EMAIL: support@savefdeals.in	4:51 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Entireweb Newsletter	Hire an SEO the Right Way - 6 Tips You Must Remember for Life - Unsubscribe me View web version Become a fan on Facebook Follow us on Twitter September 5th, 2014	1:24 pm
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Mix Buys	A policy that understands your family's medical need - open in fresh tab - If you do not want to receive any more newsletters, please click here	11:08 am
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Scoop.it	Your Scoop.it Daily Summary - How to Maximize Your LinkedIn Publishing Exposure   SME a... - Scoop.it Facebook Twitter G+ H... - Scoop.it	9:30 am
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	standard charterer Bank	Instant approval on your Credit Card	7:27 am
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CAR TRADE	Sell your car at no cost at all - If you are having trouble viewing this email, view web version   View this message in your mobile	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Uday	VPS Web Hosting Services Provider - Dear Sir, I am Uday Sharma, Business development executive. We are providing quality VPS hosting for	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Mark Regan, SPN	How to Find Your Most Valuable Keywords [Free Guide] - This is a SiteProNews/ExactSeek Webmaster Exclusive Mailing! To drop your subscription, use the link	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	HDFC Bank	LOAN upto Rs 25 lac - Disbursal in 2 days - open in fresh tab You have received this mailer from Shop@Best on behalf of HDFC Bank because you	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CAR TRADE	Sell your car at no cost at all - open in fresh tab - If you do not want to receive any further newsletters, please click here	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ICICI Bank	Home Loan Interest Rate starting from 10.15%. Get Instant Approval! - open in fresh tab - If you do not want to receive any more newsletters, please Click Here	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	calculateyourwealth	It's good when your bank helps you manage your wealth and fulfill your ambitions - Calculate Now Dreams you wish to realize in your lifetime require enough wealth. C	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Angel Broking	Get Low Brokerage - Free Demat & Trading Account - open in fresh tab - If you do not want to receive any further newsletters, please click here	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Bankbazaar	7 Minute Instant Online Approval for your PERSONAL LOAN - Now get instant online Personal Loan approval in 7 minutes by BankBazaar.com from leading Banks in	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Jayde	Welcome To The Jayde Newsletter! Welcome To The Jayde Newsletter! Welcome To The Jayde Newsletter! Welcome To The Jayde Newsletter! Welcome To The Jayde Newsletter!	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ineedhits no-reply	[ineedhits] Your ineedhits Account and Password - ACCOUNT CREATION Account ID : A1588368 Dear Raj, Welcome to ineedhits. You have successfully created your account. Your account details are as follows:	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Rakha	Monthly Apps for Your Business - While we look at the span of last 20 years, we could broadly look at two distinct eras - Life in	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	SlideShare Newsletter	Top Tips From the World Champions of PowerPoint - View online version Remember to display images Meet the PowerPoint World Champs Top Tips From the	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Dilshad Pathan	Feeling Hesitate to Discuss personal Health Queries - My Life Care Follow Us on facebook twitter linkedin Google+ Feeling Hesitate to Discuss personal	Sep-4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Vaishu	TAKE YOUR PICK. Register in SimplyMarry - TAKE YOUR PICK. Register in SimplyMarry - Regards Vaishu	Sep-4

Not a Spam

Not a Spam

# Real Life Use Cases - Google News

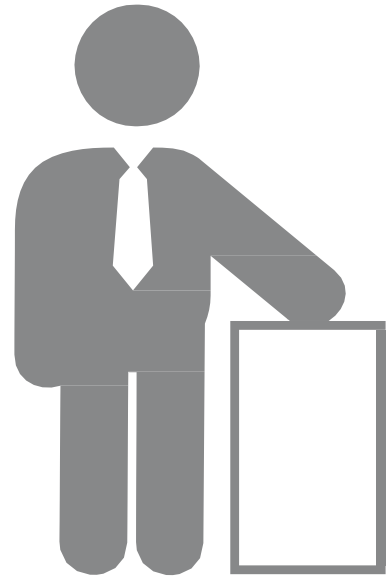


# Real Life Use Cases

- Clustering :Amazon Recommendations
- Image Search (Similarity)
- Text Summarization :Google News
- Fraud Detection :Credit card providers
- Sentiment Analysis
- Face Detection

# References

- <https://www.usoft.com/blog/difference-between-ai-ml-and-deep-learning>
- <https://www.edureka.co/blog/what-is-data-science/>
- <https://www.kdnuggets.com/2016/03/data-science-process.html>
- <https://www.newgenapps.com/blog/descriptive-vs-predictive-vs-prescriptive-analytics>
- <https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-eed6024fdb08>



Thank you.  
Questions?