Data Science

Unit 1
Introduction to Business Analytics

A.Y. 2020-21 (Odd) Semester: 5

(Information Technology)

(Course Code: 3151608)

Vinita Shah

Asst. Professor, IT Dept., GCET

Business Analytics

Business analytics (BA) is the iterative, methodical exploration of an organization's data, with an emphasis on statistical analysis.

Business analytics is used by companies that are committed to making data-driven decisions.

Data-driven companies treat their data as a corporate asset and actively look for ways to turn it into a competitive advantage.

Business Analytics

Business analytics (BA) is the combination of skills, technologies and practices used to examine an organization's data and performance as a way to gain insights and make data-driven decisions in the future using statistical analysis.

The goal of BA is to narrow down which datasets are useful and which can increase revenue, productivity, and efficiency.

When used correctly, BA can be leveraged to accurately predict future events that are related to the actions of consumers, market trends, and also assist in creating more efficient processes that could lead to an increase in revenue.

Business Analytics

Analytics is the use of:

- data,
- information technology,
- statistical analysis,
- quantitative methods and
- mathematical or computer-based models

to help managers gain improved insight about their business operations and make better, fact based decisions.

Business Analytics Uses

Business analytics has many use cases like

- **Analyze data from a variety of sources. This could be anything from cloud applications to marketing automation tools and CRM software.**

- **Use advanced analytics and statistics to find patterns within datasets. These patterns can help you predict trends in the future and access new insights about the consumer and their behavior.**

- **Monitor KPIs and trends as they change in real-time. This makes it easy for businesses to not only have their data in one place but to also come to conclusions quickly and accurately.**

Why Business Analytics?

- Improves performance by giving your business a clear picture of what is and isn’t working

- Provides faster and more accurate decisions

- Minimizes risks as it helps a business make the right choices regarding consumer behavior, trends, and performance

- Inspires change and innovation by answering questions about the consumer

Business Analytics Example

Fast-food Restaurant –

Use business analytics to speed up the ordering process for your customers using the drive-thru.

When you use **BA to monitor the traffic** that the drive-thru receives, you'll be able to know your **peak hours and when to increase efficiency**.

Business Analytics Example

When you know the line is about to get long, you can move around your staff to get more employees working the drive-thru lane, or even have them recommend orders that can be completed quickly.

When lines are shorter, employees can recommend items with higher margins that are more expensive and take more time to create.

Business Analytics Example

The popular meal kit delivery service, Blue Apron, used business analytics to forecast demand for their orders and recipes.

Blue Apron Inc. is an American ingredient-and-recipe meal kit service.

The weekly boxes contain ingredients and also include suggested recipes that must be cooked by hand by the customer using the pre-ordered ingredients.

Business Analytics Example

Each week they sent its subscribers a mixed menu of meals for purchase and using predictive analytics, they were able to use various data insights to avoid product spoilage and fulfill orders.

Blue Apron looked at customer-related insights that consisted of historical data of how often a customer made specific orders. There was also recipe-related data that focused on a customer's preference for recipes in the past.

Finally, they looked at seasonal trends to see if there were purchasing patterns of higher or lower order rates for a specific time of year.

Blue Apron was able to better understand their customers, improve the user experience, predict shifting preferences, and even identify how tastes in meals change overtime.

Applications

Pricing

- setting prices for consumer and industrial goods, government contracts, and maintenance contracts

Customer segmentation

- identifying and targeting key customer groups in retail, insurance and credit card industries

Merchandising

- determining brands to buy, quantities and allocations

Location

- finding the best location for bank branches and ATMs, or where to service industrial equipment

Social Media

- understand trends and customer perceptions, assist marketing managers and product designers

Business analytics vs. Business Intelligence

BI also deals with historical data, but this data tends to be compiled from various places, like a company's CRM software, ERP systems, and marketing automation tools.

With both BA and BI, data is collected, sorted through, and displayed using data visualization software so that business executives can have a visual representation of any spikes or pain points that may be uncovered.

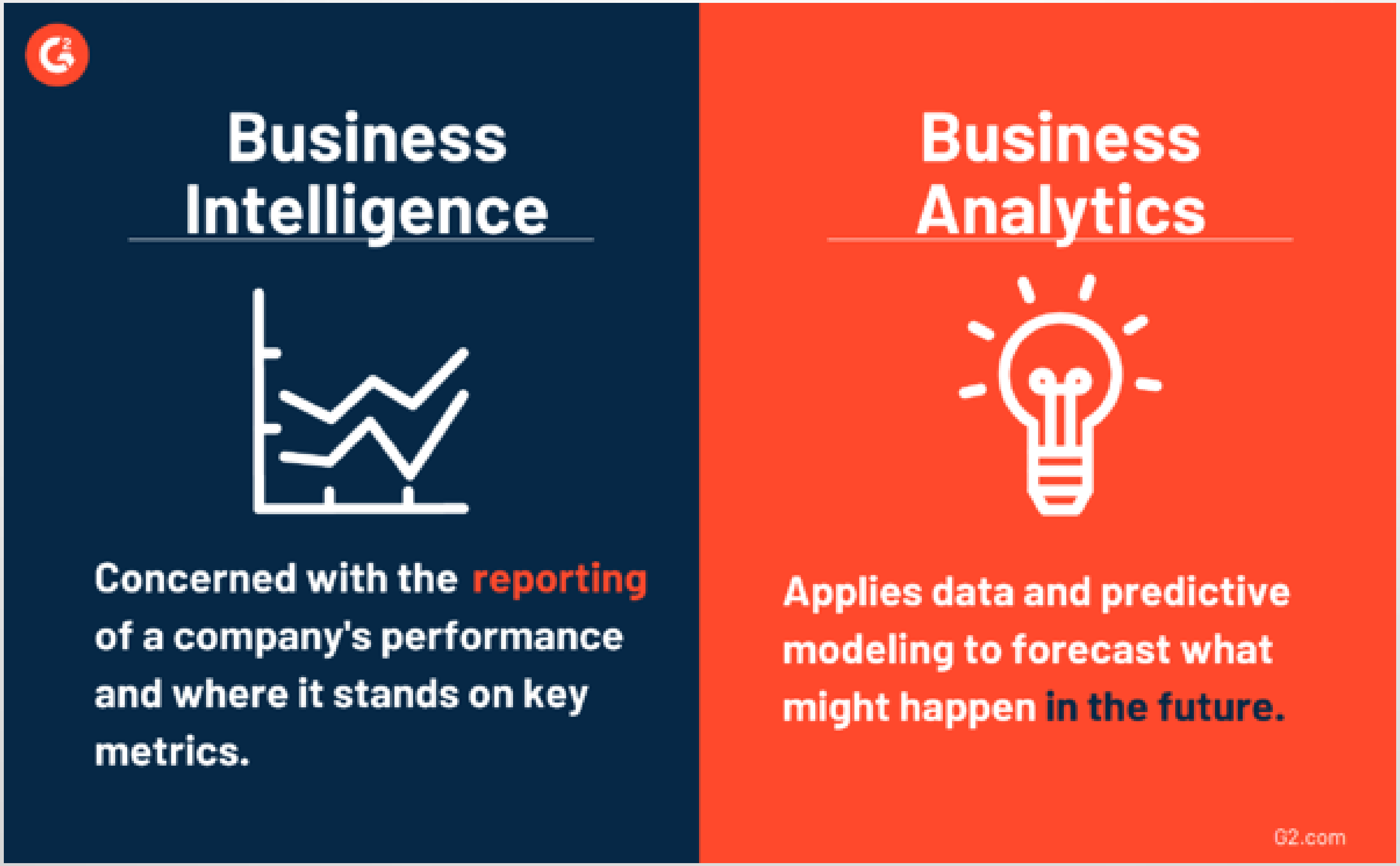
Business Analytics vs. business Intelligence

However, there is one main difference between the two:

Business Intelligence is more concerned with reporting a company's performance and where it stands on key metrics. It provides context to *what* happened in the past, *why* it may have happened, and *what* is happening now.

Business Analytics takes the context provided by business intelligence and applies statistical analysis, data mining, predictive modeling, and other techniques. These methods are more advanced, and they'll provide more context of what to expect in the future – also known as forecasting.

Business analytics vs. business Intelligence



Business analytics vs. Data Science

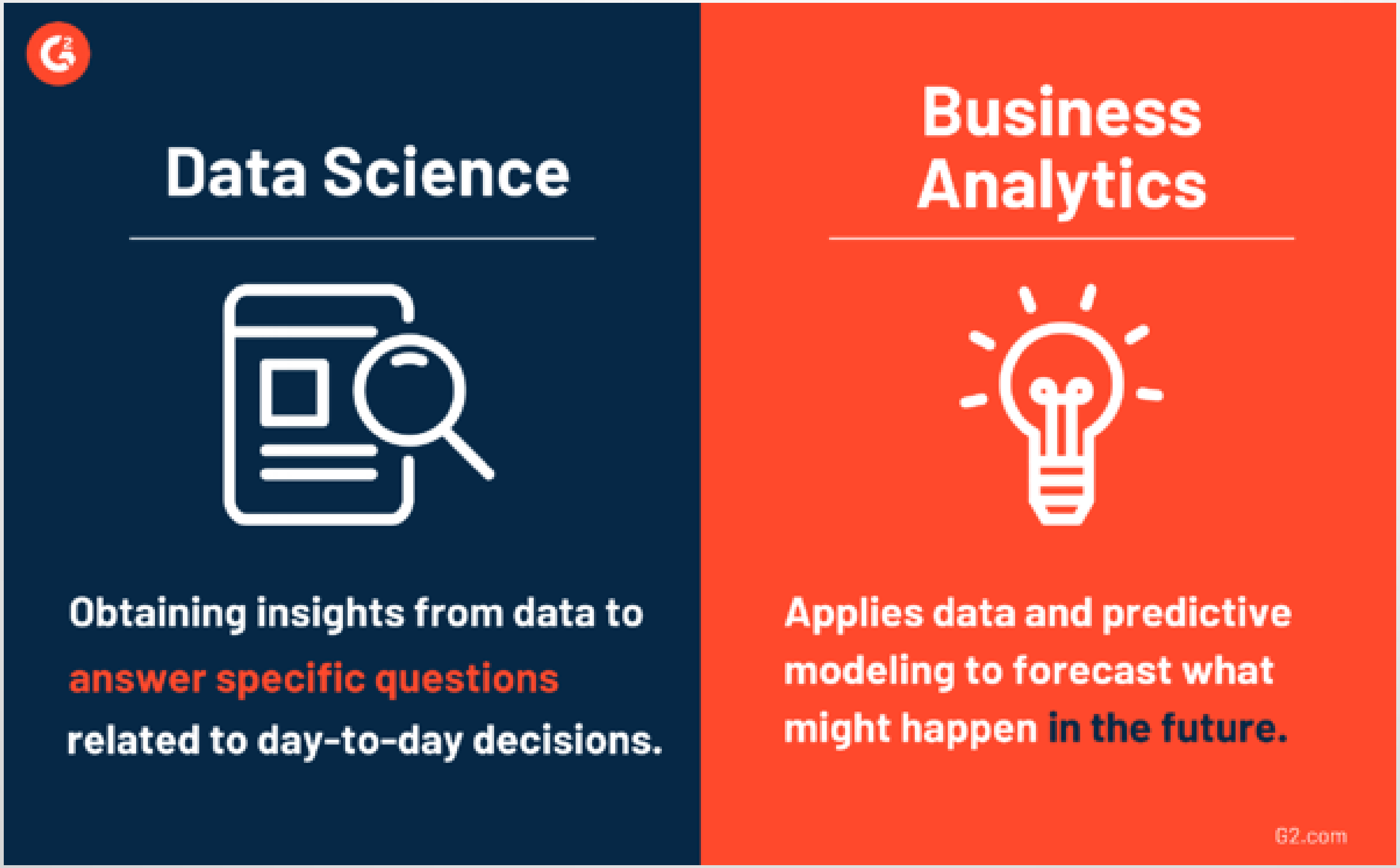
Both involve gathering data, modeling, and obtaining various insights.

The difference between the two stems from BA being specific to business-related problems, like cost and profit, and can predict what could happen in the future.

Data science is the larger or superset of the two, as its main focus is to answer questions related to customer preferences, seasonal factors, and geography within the business. It combines data with algorithm building and technology to answer these questions.

In short, data science is the science of studying data using statistics, algorithms, and technology. BA is the statistical study of business data.

Business analytics vs. Data Science



Types of Business Analytics

Descriptive analytics, which tracks key performance indicators (KPIs) to understand the present state of a business;

Predictive analytics, which analyzes trend data to assess the likelihood of future outcomes; and

Prescriptive analytics, which uses past performance to generate recommendations about how to handle similar situations in the future.

Types of Business Analytics

Descriptive analytics: the use of data to understand past and current business performance and make informed decisions

Predictive analytics: predict the future by examining historical data, detecting patterns or relationships in these data, and then extrapolating these relationships forward in time.

Prescriptive analytics: identify the best alternatives to minimize or maximize some objective

Example

Most department stores clear seasonal inventory by reducing prices.

Key question: When to reduce the price and by how much to maximize revenue?

Potential applications of analytics:
Descriptive analytics: **examine historical data for similar products (prices, units sold, advertising, …)**

Predictive analytics: **predict sales based on price**

Prescriptive analytics: **find the best sets of pricing and advertising to maximize sales revenue**

Descriptive vs Predictive vs Prescriptive Analytics

Descriptive Analytics is focused solely on historical data.

You can think of Predictive Analytics as then using this historical data to develop statistical models that will then forecast about future possibilities.

Prescriptive Analytics takes Predictive Analytics a step further and takes the possible forecasted outcomes and predicts consequences for these outcomes.

Descriptive vs Predictive vs Prescriptive Analytics

Descriptive Analytics

Descriptive analytics is a statistical method that is used to search and summarize historical data in order to identify patterns or meaning.

For example,
In an online learning course with a discussion board

Descriptive analytics could determine how many students participated in the discussion, or how many times a particular student posted in the discussion forum.

How does Descriptive Analytics Work?

Data aggregation and data mining are two techniques used in descriptive analytics to discover historical data.

Data is first gathered and sorted by data aggregation in order to make the datasets more manageable by analysts.

Data mining describes the next step of the analysis and involves a search of the data to identify patterns and meaning. Identified patterns are analyzed to discover the specific ways that learners interacted with the learning content and within the learning environment.

Descriptive Analytics - Examples

Many LMS platforms and learning systems offer descriptive analytical reporting with the aim of help businesses and institutions measure learner performance to ensure that training goals and targets are met.

The findings from descriptive analytics can quickly identify areas that require improvement - whether that be improving learner engagement or the effectiveness of course delivery.

Descriptive Analytics - Examples

Here are some examples of how descriptive analytics is being used in the field of learning analytics:

- **Tracking course enrollments, course compliance rates,**
- **Recording which learning resources are accessed and how often.**
- **Summarizing the number of times a learner posts in a discussion board.**
- **Tracking assignment and assessment grades.**
- **Comparing pre-test and post-test assessments.**
- **Analyzing course completion rates by learner or by course**
- **Identifying length of time that learners took to complete a course.**

Predictive Analytics

Predictive Analytics is a statistical method that utilizes algorithms and machine learning to identify trends in data and predict future behaviors.

Predictive Analytics can take both past and current data and offer predictions of what could happen in the future.

This identification of possible risks or opportunities enables businesses to take actionable intervention in order to improve future learning initiatives.

Predictive Analytics

Predictive Analytics - Example

Training targets -
Some systems monitor and collect data on how employees interact within the learning environment, such as tracking how often courses or resources are accessed and whether they are completed.

Achievement level can also be analyzed, including assessment performance, length of time to complete training, and outstanding training requirements.

An analysis of these aggregated data patterns can reveal how employees may continue to perform in the future. This makes it easier to identify employees who are not on track to fulfilling ongoing training requirements.

Predictive Analytics - Example

Talent management –
Predictive reporting can also forecast how employees are developing in their role and within the company; this involves tracking and forecasting on individual employee learning paths, training, and up skilling activity.

This is important for Human Resources (HR) who may need to manage the talent pool for a large number of employees or training departments wanting to know what resources will be effective for individual skill development.

Predictive Analytics - Models

Some common basic models that are utilized at a broad level include:

Decision trees use branching to show possibilities stemming from each outcome or choice.

Regression techniques assist with understanding relationships between variables.

Neural networks utilize algorithms to figure out possible relationships within data sets.

Predictive Analytics - Models

Because predictive analytics goes beyond sorting and describing data, it relies heavily on complex models designed to make inferences about the data it encounters.

These models utilize algorithms and machine learning to analyze past and present data in order to provide future trends.

Each model differs depending on the specific needs of those employing predictive analytics.

Prescriptive Analytics

Prescriptive analytics is a statistical method used to generate recommendations and make decisions based on the computational findings of algorithmic models.

At present, prescriptive analytics is not widely used in the field of learning & development due to the complex requirements needed in the field of machine learning.

It can be found in adaptive learning and also within some learning experience platforms (LXP).

Difference between Prescriptive Analytics and Predictive Analytics

Prescriptive analytics is considered an extension of predictive analytics.

An insightful forecast from predictive analysis can be analyzed using specific models designed for prescriptive analysis in order to produce automated recommendations or solutions. Prescriptive analytics require complex algorithms in order to accomplish such machine-based decision-making.

With predictive analytics, it is understood that predictions may or may not happen.

Prescriptive Analytics - Example

The use of prescriptive analytics is growing and can already be found in some popular learning management systems (LMS) and learning technologies.

There are some tools that use prescriptive analytics to identify what content the learner has already learned so that new content not yet mastered is presented instead. This is an example of how prescriptive analytics is finding its way into adaptive learning.

Prescriptive Analytics - Example

Some LMS's enable administrators to define specific rules in order for automated feedback or actions to take place; for example, if an employee is struggling to complete a training course, the system may recommend they look at a different resource to obtain skills needed for the previous course.

Some LMS's are promising to reduce training time for employees by determining previous knowledge and proficiency baselines in order to recommend which training courses or resources are best suited for the learner.

Web Analytics

Web analytics is the collection, reporting, and analysis of website data.

The focus is on identifying measures based on your organizational and user goals and using the website data to determine the success or failure of those goals and to drive strategy and improve the user's experience.

Web Analytics

Web Analytics is the methodological study of online/offline patterns and trends.

It is a technique that you can employ to collect, measure, report, and analyze your website data.

Normally carried out to analyze the performance of a website and optimize its web usage.

We use web analytics to track key metrics and analyze visitors' activity and traffic flow. It is a tactical approach to collect data and generate reports.

Importance of Web Analytics

We need Web Analytics to assess the success rate of a website and its associated business.

Using Web Analytics, we can –

- **Assess web content problems so that they can be rectified**
- **Have a clear perspective of website trends**
- **Monitor web traffic and user flow**
- **Demonstrate goals acquisition**
- **Figure out potential keywords**
- **Identify segments for improvement**
- **Find out referring sources**

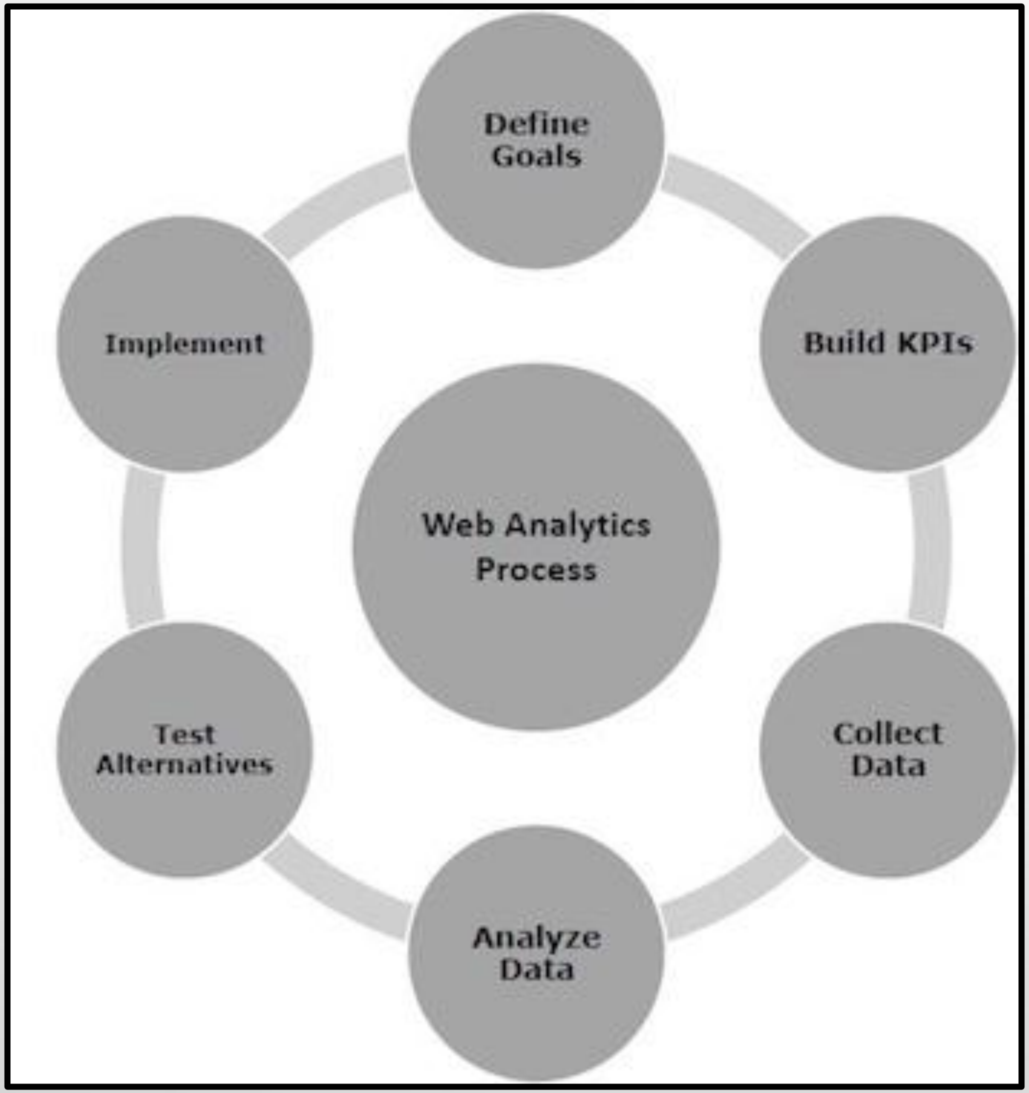
Web Analytics Process

The primary objective of carrying out Web Analytics is to optimize the website in order to provide better user experience. It provides a data-driven report to measure visitors' flow throughout the website.

Process of web analytics:

- **Set the business goals.**
- **To track the goal achievement, set the Key Performance Indicators (KPI).**
- **Collect correct and suitable data.**
- **To extract insights, Analyze data.**
- **Based on assumptions learned from the data analysis, Test alternatives.**
- **Based on either data analysis or website testing, Implement insights.**

Web Analytics Process



Web Analytics Tools

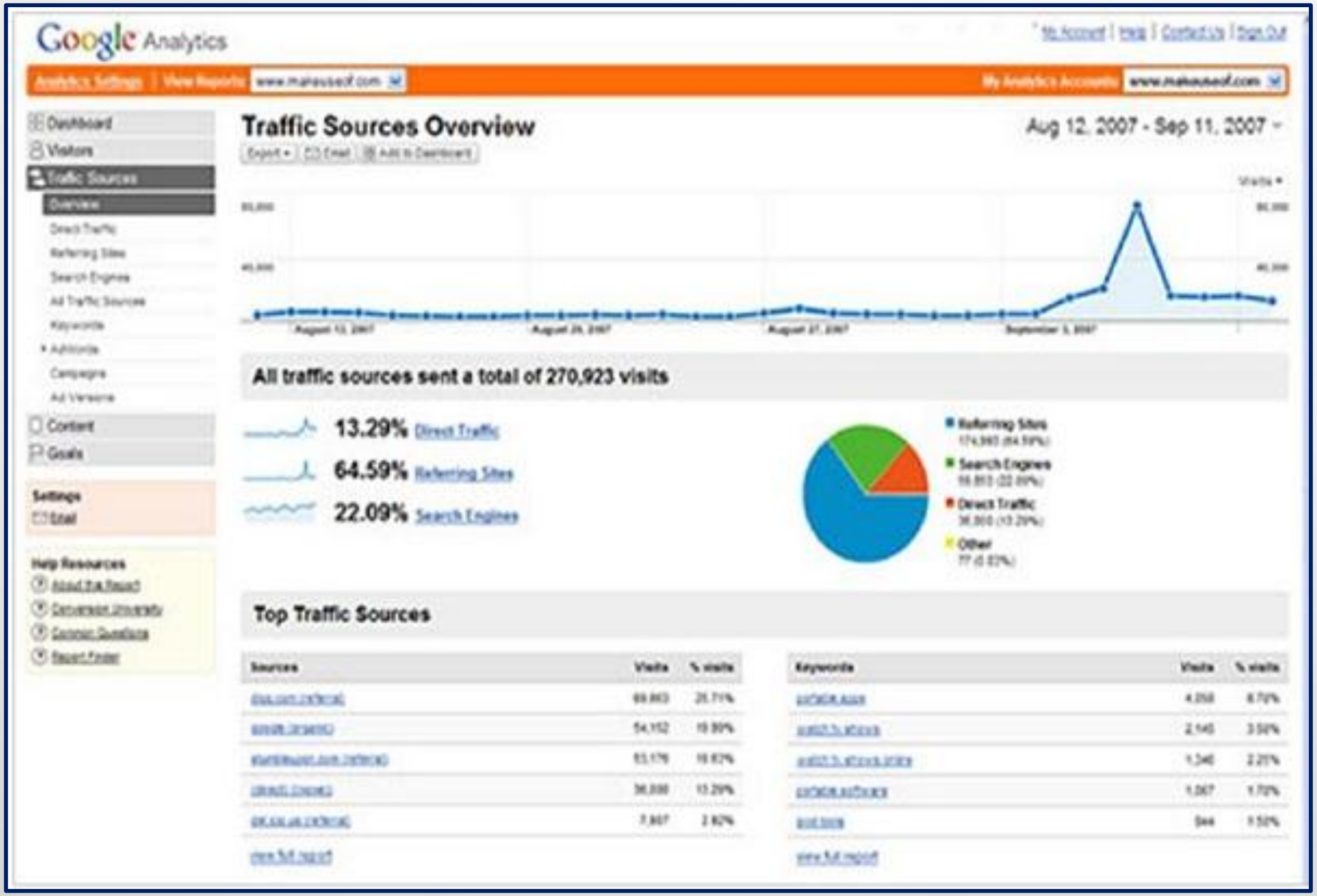
Analytics Tools offer an insight into the performance of your website, visitors' behavior, and data flow.

These tools are inexpensive and easy to use. Sometimes, they are even free.

Google Analytics

Google Analytics is a freemium analytic tool that provides a detailed statistics of the web traffic. It is used by more than 60% of website owners.

Google Analytics



Google Analytics

Google analytics helps you to track and measure visitors, traffic sources, goals, conversion and other metrics.

It basically generates reports on –

Audience Analysis
Acquisition Analysis
Behavior Analysis
Conversion Analysis

Social Media Analytics

Social Media is a platform that lets us participate in social networking.

We can share our posts on various social media platforms to improve business visibility.

Today it is the best source for news updates, marketing, education, and entertainment.

Social Media Analytics

Social Media Analytics (SMA) refers to the approach of collecting data from social media sites and blogs and evaluating that data to make business decisions.

This process goes beyond the usual monitoring or a basic analysis of retweets or likes to develop an in-depth idea of the social consumer.

Social Media Analytics

Process of gathering and analyzing data from social networks such as Facebook, Instagram, LinkedIn and Twitter.

It is commonly used by marketers to track online conversations about products and companies.

One author defined it as "the art and science of extracting valuable hidden insights from vast amounts of semi-structured and unstructured social media data to enable informed and insightful decision making."

Social Media Analytics

Social media analytics is data offered by most social media sites that gives insight into what people are responding to and engaging with on your social channels.

You can use this data to measure the growth and effectiveness of your social channels, usually to improve brand awareness, profits, and return on investment (ROI).

Analytics also can help you understand what works for your competitors and their audiences.

Social Media Analytics

Each social platform has its' own analytics or insights tool:

- **Twitter uses Twitter Analytics**
- **Facebook offers in-depth analytics on the Insights tab of Facebook pages**
- **Instagram uses the Facebook Insights platform**
- **LinkedIn offers basic, free data on your company page and full analytics software with a premium account**
- **YouTube uses the YouTube analytics dashboard**

Necessity

Data is a set of values of qualitative or quantitative variables. The individual pieces of data contain information.

Data could be measured, collected and analyzed. It could be visualized as well using images, graphs, pie charts or other analysis tools.

The existence of digital data is growing at a rapid speed. According to few surveys, by the year 2020, about 1.7 billion of new information will be developed every second for all the human inhabitants on this planet.

Since the amount of data is increasing, various ways have been introduced to handle and process the data. Among which, Data Science and Machine Learning are considered are some of the ways to process data.

Types / Categories of Data

- Structured
- Unstructured
- Natural language
- Machine-generated
- Graph-based
- Audio, video, and images

Structured Data

Structured data is data that depends on a data model and resides in a fixed field within a record.
It is often easy to store structured data in tables within databases or Excel file.
SQL, or Structured Query Language, is the preferred way to manage and query data that resides in databases.

You may also come across structured data that might give you a hard time storing it in a traditional relational database.

Hierarchical data such as a family tree is one such example.

The world isn't made up of structured data, though; it's imposed upon it by humans and machines. More often, data comes unstructured.

Structured Data

1	Indicator ID	Dimension List	Timeframe	Numeric Value	Missing Value Flag	Confidence Inte
2	214390830	Total (Age-adjusted)	2008	74.6%		73.8%
3	214390833	Aged 18-44 years	2008	59.4%		58.0%
4	214390831	Aged 18-24 years	2008	37.4%		34.6%
5	214390832	Aged 25-44 years	2008	66.9%		65.5%
6	214390836	Aged 45-64 years	2008	88.6%		87.7%
7	214390834	Aged 45-54 years	2008	86.3%		85.1%
8	214390835	Aged 55-64 years	2008	91.5%		90.4%
9	214390840	Aged 65 years and over	2008	94.6%		93.8%
10	214390837	Aged 65-74 years	2008	93.6%		92.4%
11	214390838	Aged 75-84 years	2008	95.6%		94.4%
12	214390839	Aged 85 years and over	2008	96.0%		94.0%
13	214390841	Male (Age-adjusted)	2008	72.2%		71.1%
14	214390842	Female (Age-adjusted)	2008	76.8%		75.9%
15	214390843	White only (Age-adjusted)	2008	73.8%		72.9%
16	214390844	Black or African American only (Age-adjusted)	2008	77.0%		75.0%
17	214390845	American Indian or Alaska Native only (Age-adjusted)	2008	66.5%		57.1%
18	214390846	Asian only (Age-adjusted)	2008	80.5%		77.7%
19	214390847	Native Hawaiian or Other Pacific Islander only (Age-adjusted)	2008	DSU		
20	214390848	2 or more races (Age-adjusted)	2008	75.6%		69.6%

Unstructured Data

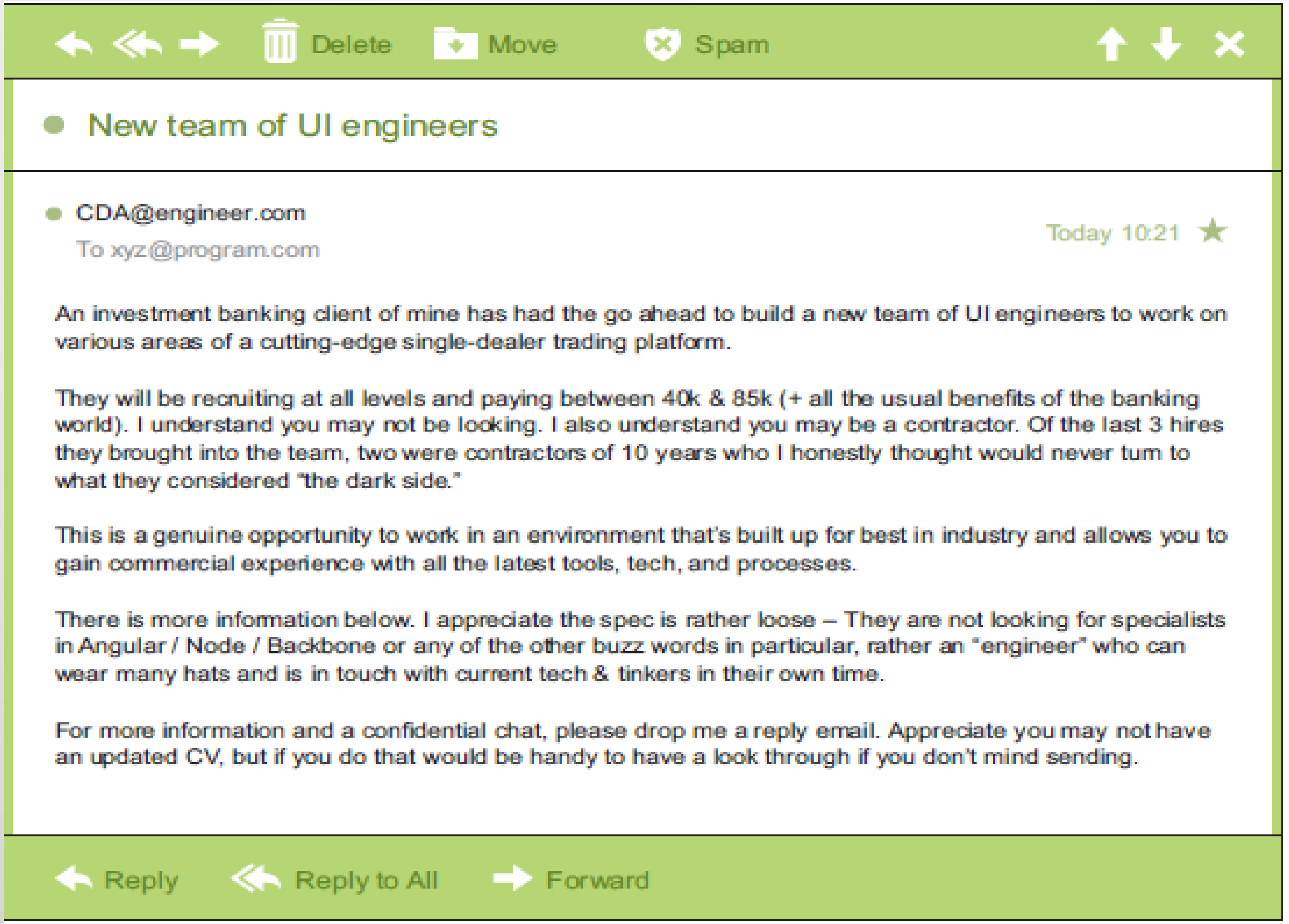
Unstructured data is data that isn’t easy to fit into a data model because the content is context-specific or varying.

One example of unstructured data is your regular email.

Although email contains structured elements such as the sender, title, and body text, it’s a challenge to find the number of people who have written an email complaint about a specific employee because so many ways exist to refer to a person.

A human-written email, as shown in figure, is also a perfect example of natural language data.

Unstructured Data



Delete Move Spam

New team of UI engineers

CDA@engineer.com
To xyz@program.com

Today 10:21

An investment banking client of mine has had the go ahead to build a new team of UI engineers to work on various areas of a cutting-edge single-dealer trading platform.

They will be recruiting at all levels and paying between 40k & 85k (+ all the usual benefits of the banking world). I understand you may not be looking. I also understand you may be a contractor. Of the last 3 hires they brought into the team, two were contractors of 10 years who I honestly thought would never turn to what they considered "the dark side."

This is a genuine opportunity to work in an environment that's built up for best in industry and allows you to gain commercial experience with all the latest tools, tech, and processes.

There is more information below. I appreciate the spec is rather loose – They are not looking for specialists in Angular / Node / Backbone or any of the other buzz words in particular, rather an "engineer" who can wear many hats and is in touch with current tech & tinkers in their own time.

For more information and a confidential chat, please drop me a reply email. Appreciate you may not have an updated CV, but if you do that would be handy to have a look through if you don't mind sending.

Reply Reply to All Forward

Natural Language

Natural language is a special type of unstructured data; it's challenging to process because it requires knowledge of specific data science techniques and linguistics.

The natural language processing community has had success in entity recognition, topic recognition, summarization, text completion, and sentiment analysis, but models trained in one domain don't generalize well to other domains.

Even state-of-the-art techniques aren't able to decipher the meaning of every piece of text.

Natural Language

This shouldn't be a surprise though: humans struggle with natural language as well. It's ambiguous by nature. The concept of meaning itself is questionable here.

Have two people listen to the same conversation. Will they get the same meaning? The meaning of the same words can vary when coming from someone upset or joyous.

Machine Generated Data

Machine-generated data is information that's automatically created by a computer, process, application, or other machine without human intervention.

Machine-generated data is becoming a major data resource and will continue to do so.

The analysis of machine data relies on highly scalable tools, due to its high volume and speed.

Examples of machine data are web server logs, call detail records, network event logs, and telemetry.

Machine Generated Data

```
CSIPERF:TXCOMMIT;313236
2014-11-28 11:36:13, Info                  CSI    00000153 Creating NT transaction (seq
69), objectname [6]"(null)"
2014-11-28 11:36:13, Info                  CSI    00000154 Created NT transaction (seq 69)
result 0x00000000, handle @0x4e54
2014-11-28 11:36:13, Info                  CSI    00000155@2014/11/28:10:36:13.471
Beginning NT transaction commit...
2014-11-28 11:36:13, Info                  CSI    00000156@2014/11/28:10:36:13.705 CSI perf
trace:
CSIPERF:TXCOMMIT;273983
2014-11-28 11:36:13, Info                  CSI    00000157 Creating NT transaction (seq
70), objectname [6]"(null)"
2014-11-28 11:36:13, Info                  CSI    00000158 Created NT transaction (seq 70)
result 0x00000000, handle @0x4e5c
2014-11-28 11:36:13, Info                  CSI    00000159@2014/11/28:10:36:13.764
Beginning NT transaction commit...
2014-11-28 11:36:14, Info                  CSI    0000015a@2014/11/28:10:36:14.094 CSI perf
trace:
CSIPERF:TXCOMMIT;386259
2014-11-28 11:36:14, Info                  CSI    0000015b Creating NT transaction (seq
71), objectname [6]"(null)"
2014-11-28 11:36:14, Info                  CSI    0000015c Created NT transaction (seq 71)
result 0x00000000, handle @0x4e5c
2014-11-28 11:36:14, Info                  CSI    0000015d@2014/11/28:10:36:14.106
Beginning NT transaction commit...
2014-11-28 11:36:14, Info                  CSI    0000015e@2014/11/28:10:36:14.428 CSI perf
trace:
CSIPERF:TXCOMMIT;375581
```

Graph-based or Network data

Graph or network data is, in short, data that focuses on the relationship or adjacency of objects.

The graph structures use nodes, edges, and properties to represent and store graphical data.

Graph-based data is a natural way to represent social networks, and its structure allows you to calculate specific metrics such as the influence of a person and the shortest path between two people.

Graph-based or Network data

Examples of graph-based data can be found on many social media websites.

For instance, on LinkedIn you can see who you know at which company.

Your follower list on Twitter is another example of graph-based data.

The power and sophistication comes from multiple, overlapping graphs of the same nodes.

Graph-based or Network data

For example,

Imagine the connecting edges here to show “friends” on Facebook.

Imagine another graph with the same people which connects business colleagues via LinkedIn.

Imagine a third graph based on movie interests on Netflix. Overlapping the three different-looking graphs makes more interesting questions possible.

Graph-based or Network data

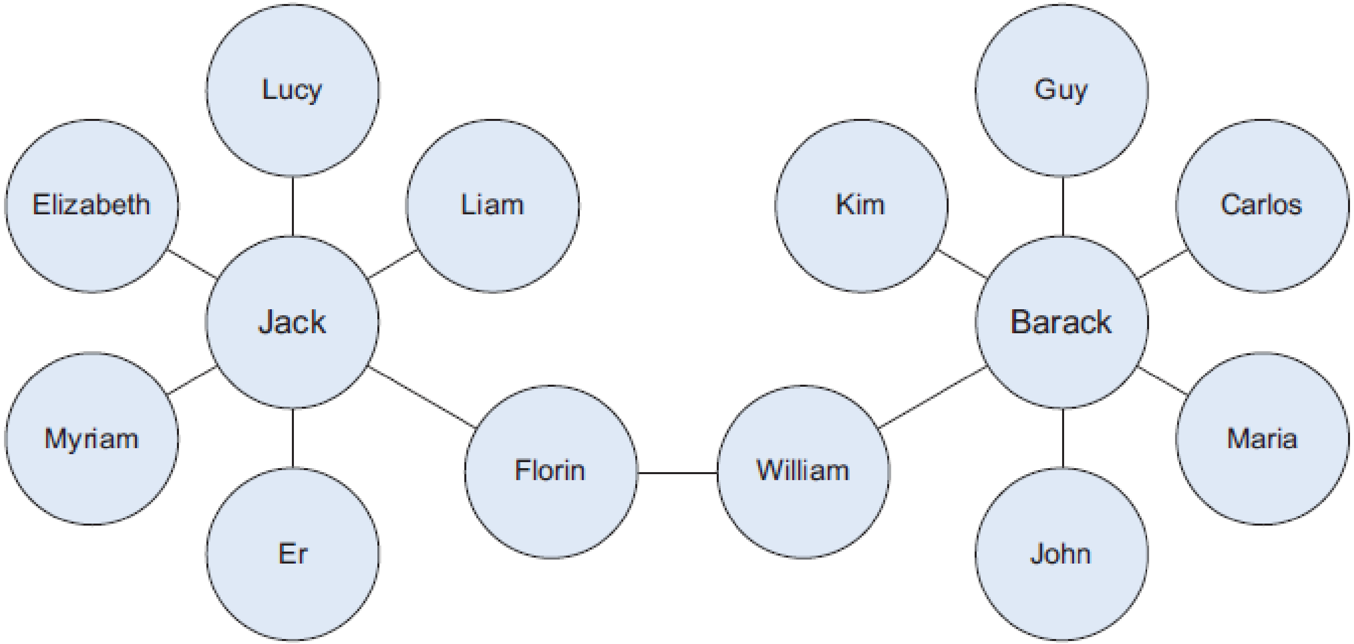


Figure 1.4 **Friends in a social network are an example of graph-based data.**

Audio, Image, and Video

Audio, image, and video are data types that pose specific challenges to a data scientist.

Tasks that are trivial for humans, such as recognizing objects in pictures, turn out to be challenging for computers.

Recently a company called DeepMind succeeded at creating an algorithm that's capable of learning how to play video games. This algorithm takes the video screen as input and learns to interpret everything via a complex process of machine learning.

Streaming data

While streaming data can take almost any of the previous forms, it has an extra property.

The data flows into the system when an event happens instead of being loaded into a data store in a batch.

Although this isn't really a different type of data, we treat it here as such because you need to adapt your process to deal with this type of information.

Examples are the "What's trending" on Twitter, live sporting or music events, and the stock market.

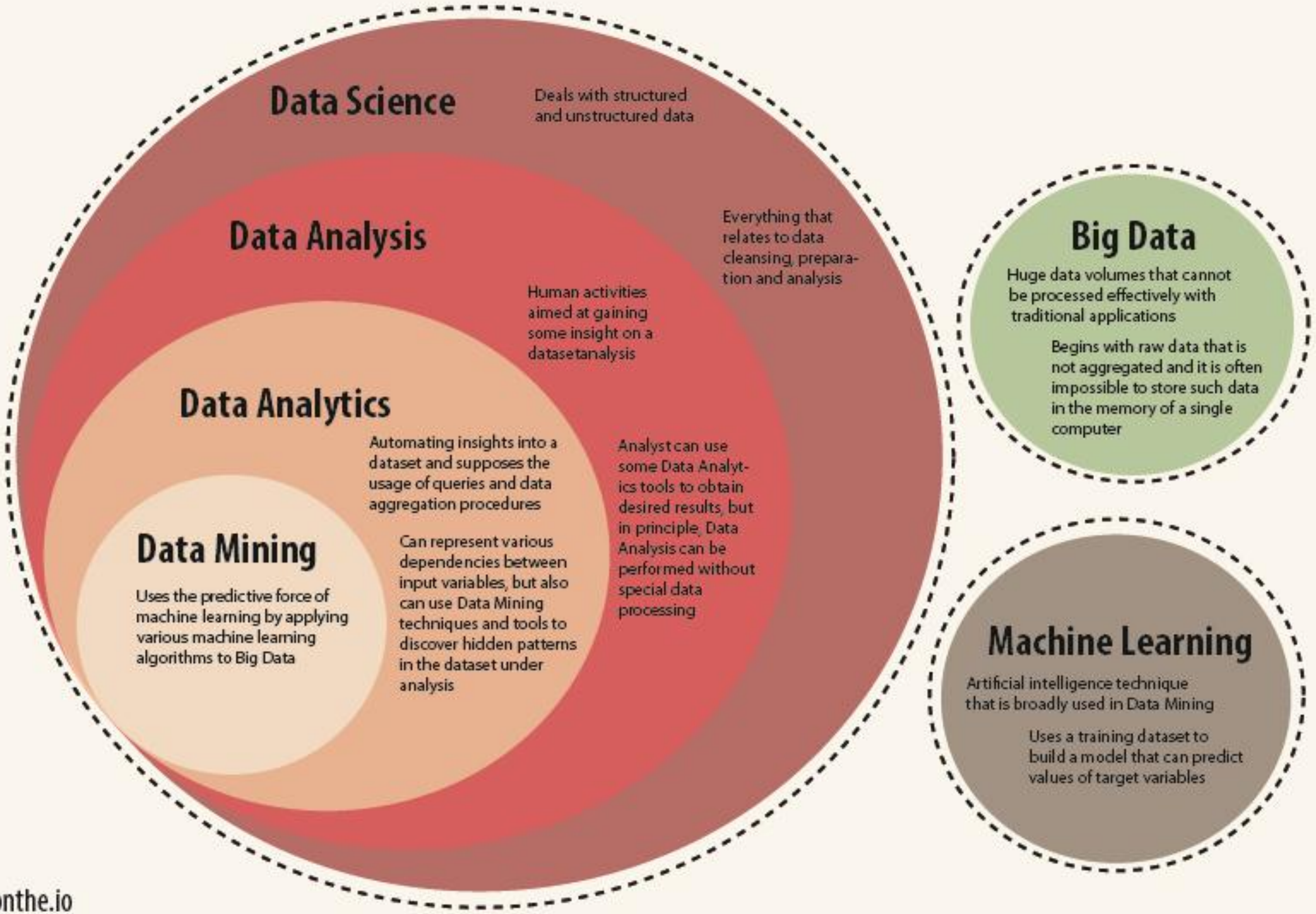
Key Areas to Social Media Analytics

Many social media teams don't know exactly which areas of their social media marketing can really benefit from analysis.

Key areas to social media analytics:

- **Audience analytics**
- **Performance analytics**
- **Competitive analytics**
- **Paid Social Media analytics**
- **Influencer analytics**
- **Sentiment Analysis**
- **Customer Service analytics**

What is the difference between Data Science, Data Analysis, Big Data, Data Analytics, Data Mining and Machine Learning?



Source: onthe.io

Machine learning

Machine Learning is a system that can learn from example through self-improvement and without being explicitly coded by programmer.

The breakthrough comes with the idea that a machine can singularly learn from the data (i.e., example) to produce accurate results.

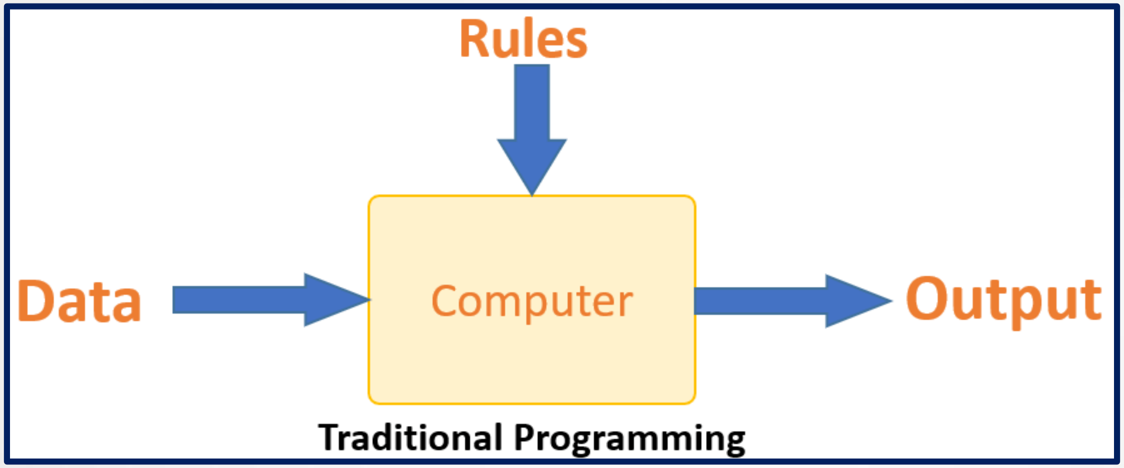
Machine learning

Machine learning combines data with statistical tools to predict an output.

This output is then used by corporate to makes actionable insights.

Machine learning is closely related to data mining and Bayesian predictive modeling. The machine receives data as input, use an algorithm to formulate answers.

Machine Learning vs. Traditional Programming

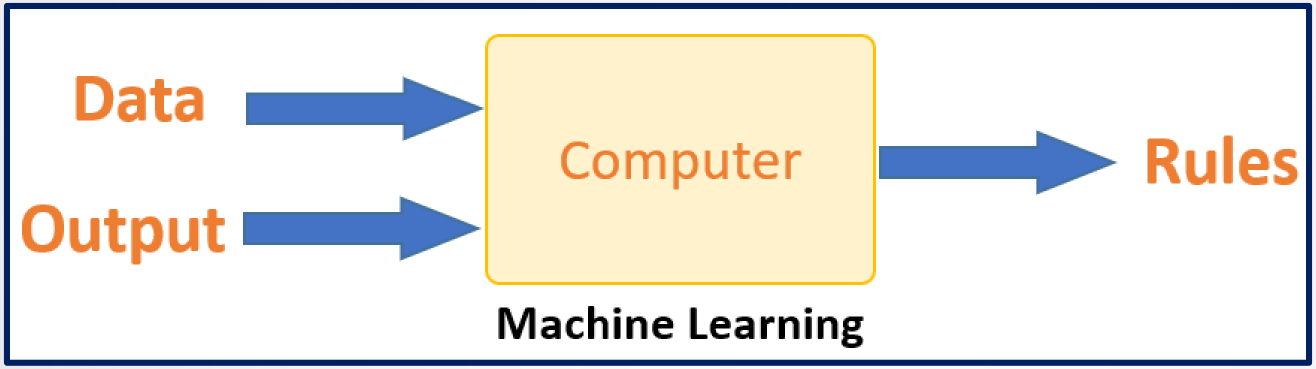


In traditional programming, a programmer code all the rules in consultation with an expert in the industry for which software is being developed.

Each rule is based on a logical foundation; the machine will execute an output following the logical statement.

When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain.

Machine Learning vs. Traditional Programming

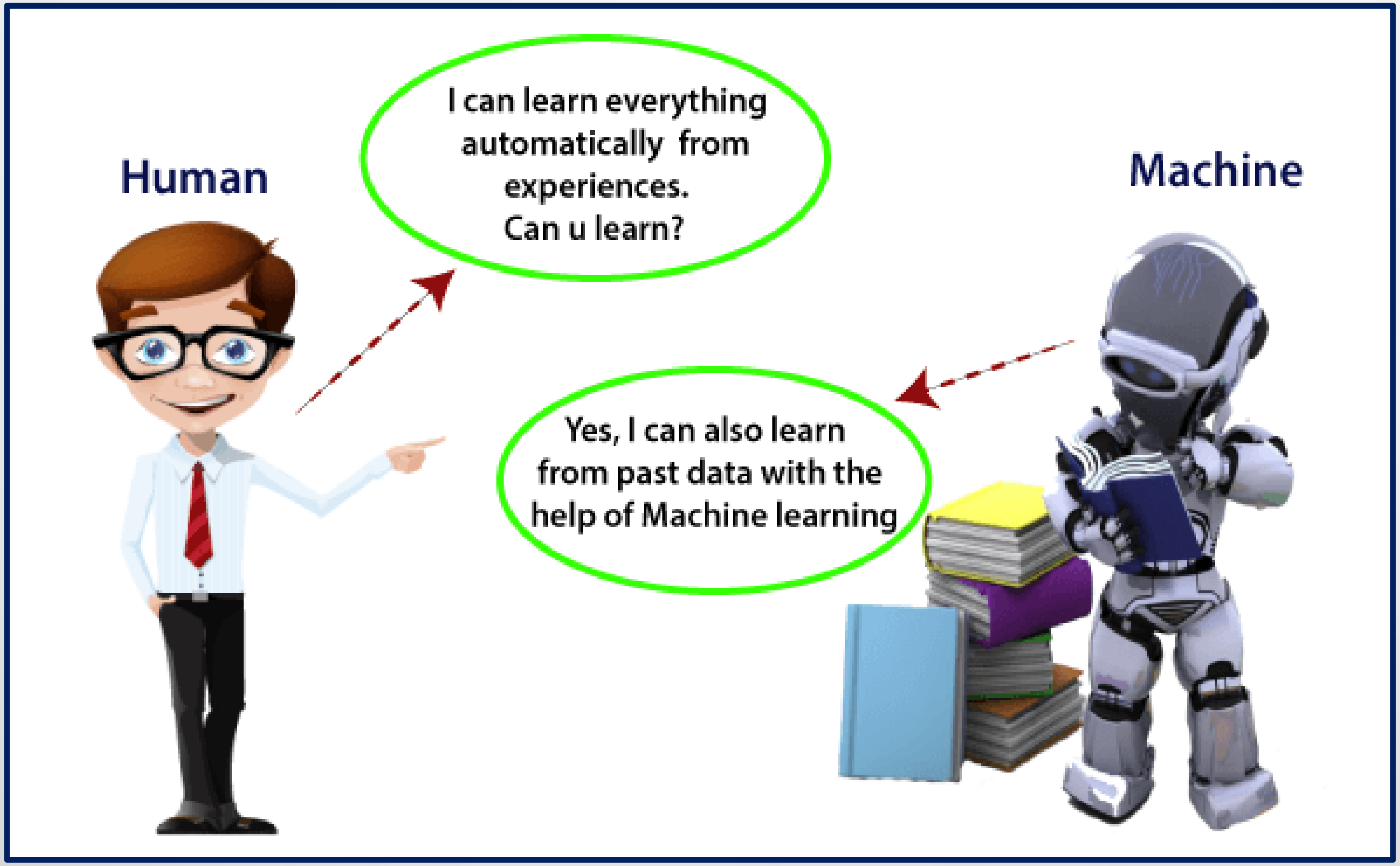


Machine learning is supposed to overcome this issue.

The machine learns how the input and output data are correlated and it writes a rule.

The programmers do not need to write new rules each time there is new data. The algorithms adapt in response to new data and experiences to improve efficacy over time.

Machine Learning



Types of Machine Learning

Types of Machine Learning

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Supervised Learning

Supervised learning as the name indicates the presence of a supervisor as a teacher.

Basically supervised learning is a learning in which we teach or train the machine using data which is well labeled that means some data is already tagged with the correct answer.

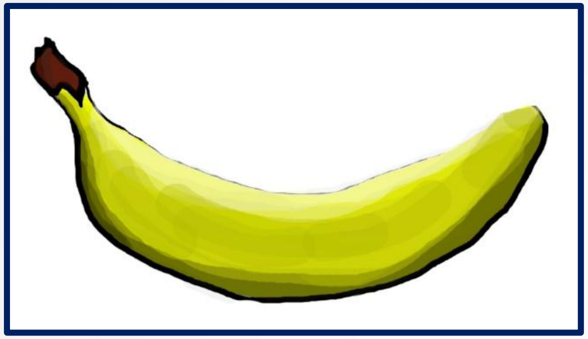
After that, the machine is provided with a new set of examples(data) so that supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labeled data.

Supervised Learning

suppose you are given a basket filled with different kinds of fruits. Now the first step is to train the machine with all different fruits one by one.

Now suppose after training the data, you have given a new separate fruit say Banana from basket and asked to identify it.

Supervised Learning



Since the machine has already learned the things from previous data and this time have to use it wisely.

It will first classify the fruit with its shape and color and would confirm the fruit name as BANANA and put it in Banana category.

Thus the machine learns the things from training data(basket containing fruits) and then apply the knowledge to test data(new fruit).

Supervised Learning

Supervised learning classified into two categories of algorithms:

Classification: **A classification problem is when the output variable is a category, such as “Red” or “blue” or “disease” and “no disease”.**

Regression: **A regression problem is when the output variable is a real value, such as “dollars” or “weight”.**

Unsupervised Learning

Unsupervised learning is the training of machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance.

Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore machine is restricted to find the hidden structure in unlabeled data by our-self.

Unsupervised Learning

Thus the machine has no idea about the features of dogs and cat so we can't categorize it in dogs and cats.

But it can categorize them according to their similarities, patterns, and differences.

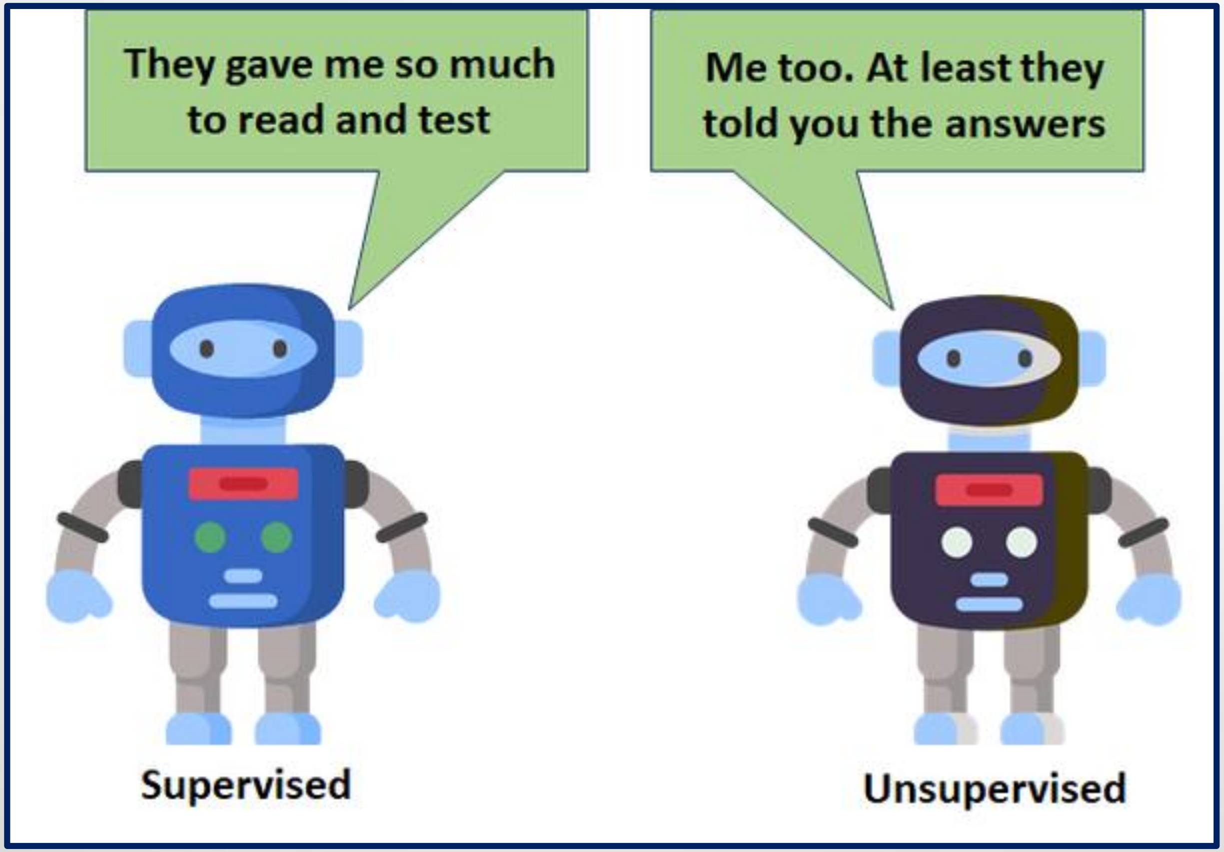
Unsupervised Learning

Unsupervised learning classified into two categories of algorithms:

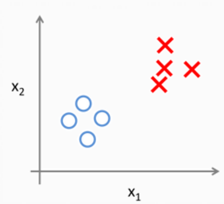
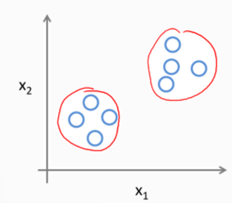
Clustering: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

Association: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Machine Learning



Machine Learning

Supervised	Unsupervised
Input Data is labelled	Input Data is Unlabelled
Uses training Dataset	Uses just input dataset
Data is classified based on training dataset	Uses properties of given data to classify it.
Used for prediction	Used for Analysis
Divided into two types Regression & Classification	Divided into two types Clustering & Association
Known number of classes	Unknown number of classes
	
Use off-line analysis of data	Use Real-Time analysis of data

Reinforcement Learning

It is a Machine Learning algorithm that allows software agents and machines to automatically determine the ideal behaviour within a specific context to maximize its performance.

It does not have labelled dataset or results associated with data so the only way to perform a given task is to learn from experience.

For every correct action or decision of algorithm, it is rewarded with positive reinforcement whereas, for every incorrect action, it is rewarded with negative reinforcement.

In this way, it learns which actions are needed to perform and which are not.

Reinforcement Learning

Whenever you encounter a problem, wherein your computer has to make a decision based on the training that you have given it, it involves Reinforcement Algorithms.

Your temperature control system, when it has to decide whether it should lower the temperature of the room, or increase it.

Rather than teaching the computer what to do, you let it decide what to do, and at the end of that action, you give either a positive or a negative feedback.

Rather than defining what is right and what is wrong in your system, you let your system "decide" what to do, and in the end give a feedback.

Applications of Reinforcement Learning

Robotics: Reinforcement learning is used in the advancement of robotics. These models are used to train robots so that they can learn from their experience which is a belief of reinforcement learning.

Traffic Light Management System: Reinforcement Learning model that was applied to the traffic management system provided better results in comparison to the traditional method for the congestion problem.

Machine Learning

TYPES OF MACHINE LEARNING

Supervised Learning

Train an algorithm to perform classifcation and regression with a labelled data set.

Unsupervised Learning

Train an algorithm to find clusters and associations in an unlabelled data set.

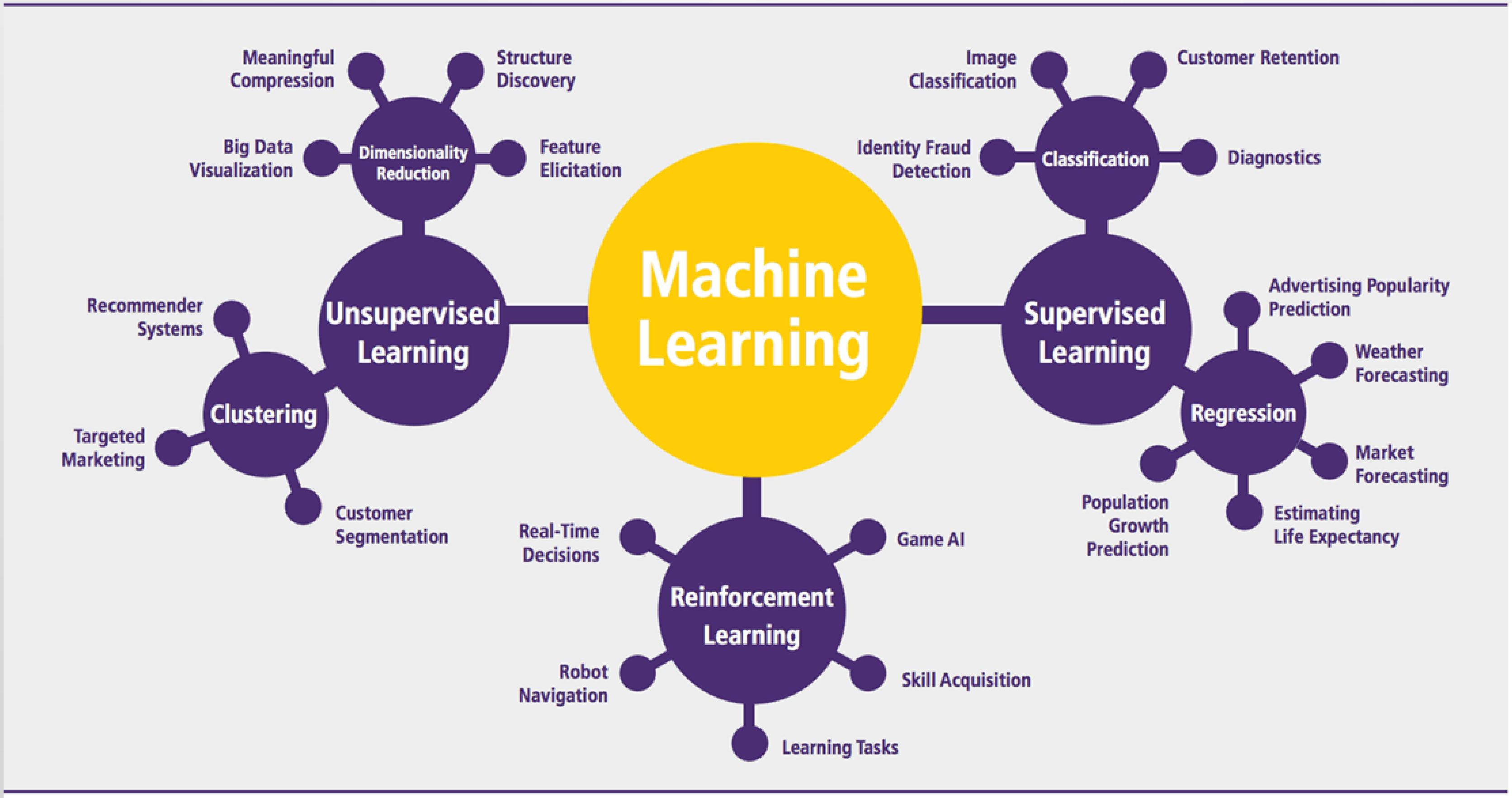
Reinforcement Learning

Train an agent to take certain actions in an environment without a data set.

Machine Learning

Supervised Learning	Labeled data; Direct feedback; Predict outcome/future
Unsupervised Learning	No labels; No feedback; Find hidden structure in data
Reinforcement Learning	Decision process; Reward system; Learn series of actions

Machine Learning



Machine Learning Applications

Virtual Personal Assistants:

Names like Siri and Alexa bring to mind the capabilities of virtual assistants.

We can ask Siri to **make a call** for you or play music.
You can request Alexa for today's **weather forecast**.
You can even **set an alarm** or **send an SMS**.

Only need to speak to it and it will listen to your command.
This comes in handy for **differently abled people**.

Such assistants take note of how you interact with them and use that to make your next experience with them better.

Machine Learning Applications

Social Media Services:

you would have noticed several features of Facebook- ‘People You May Know’ and ‘Face Recognition’.

It uses machine learning to monitor your activity- what profiles you visit, which people to send requests to, which ones you accept requests of, the people you tag, among much more.

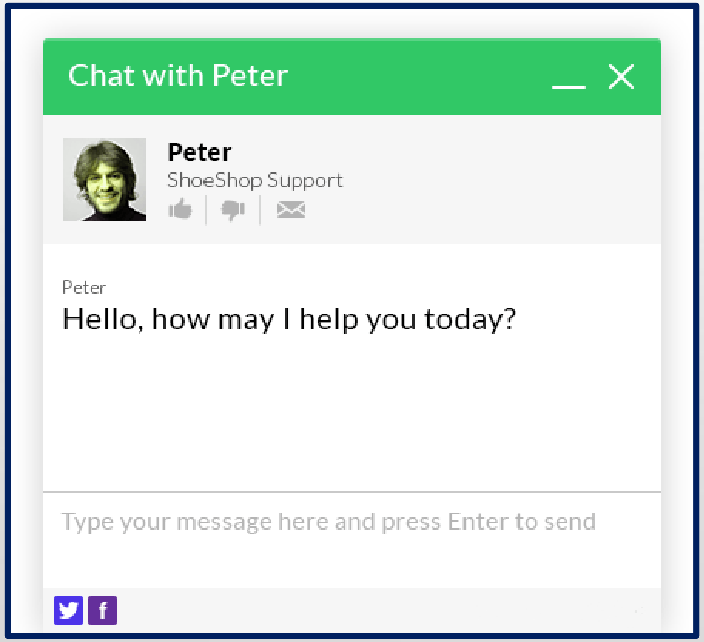
With this, Facebook hopes to provide you with a richer experience on its platform so you will use it regularly.

Machine Learning Applications

Online Customer Support:

Websites like educators and shopping platforms will often pop a live chat up to help you with your questions.

Some websites use a chatbot instead to pull information to the website and try to address the customer's queries.



Machine Learning Applications

Online Fraud Detection:

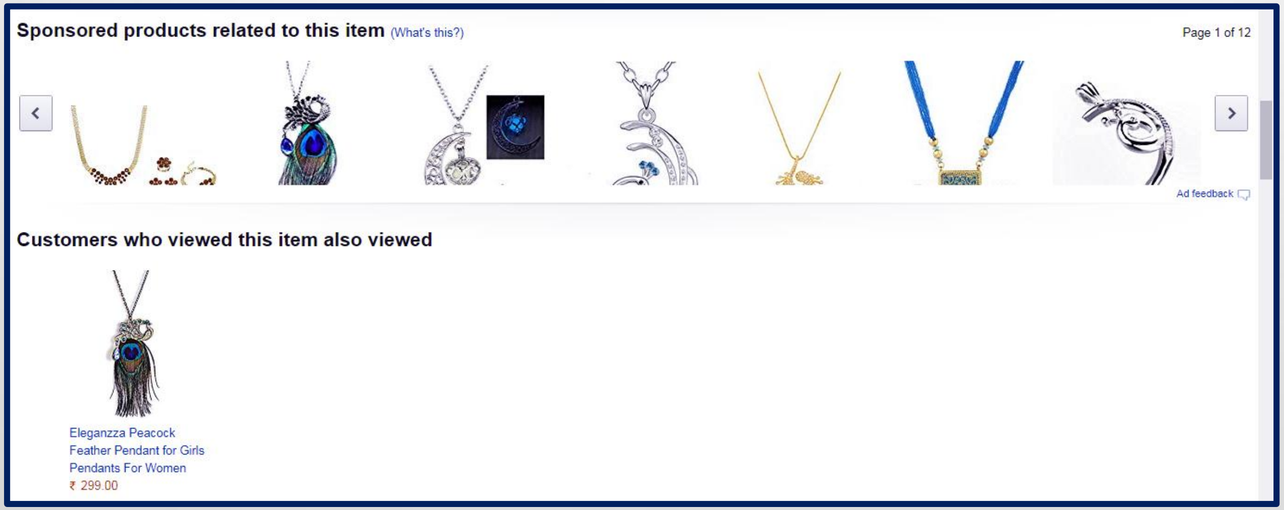
If you're familiar with PayPal, you realize your trust with it. It uses machine learning to stand in defense against illegal acts like money laundering.

By comparing millions of transactions, it can find out which ones are illegitimate.

Machine Learning Applications

Product Recommendations:

Shopping platforms like Amazon and Flipkart notice what products you look at and suggest similar products to you. If this gets a favourite product across to you and results in a purchase you make with them, it's a win for them.
For this, it also uses your wishlist and cart contents.

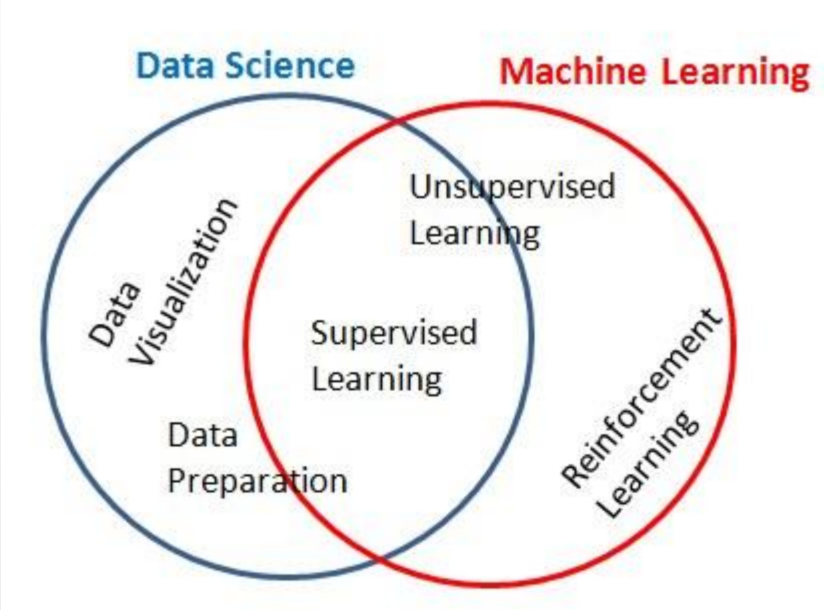


Data Science vs. Machine Learning

Data Science and Machine Learning can be thought of as close cousins.

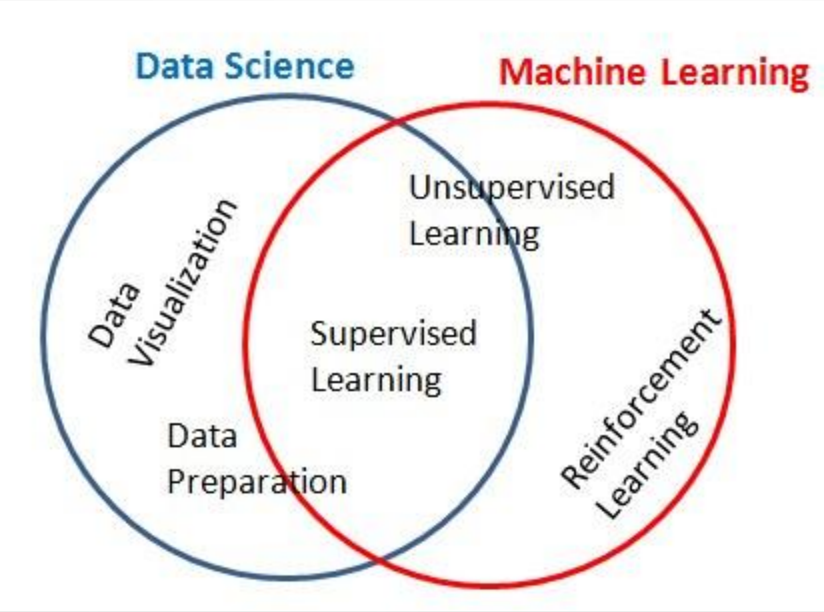
What they have in common is supervised learning methods - learning from historical data.

However, Data Science is also concerned with Data Visualization and presenting results in the form understandable to people. Data Science has much bigger focus on Data Preparation and Data Engineering.



Data Science vs. Machine Learning

Machine Learning main focus is on the learning algorithms - it is not concerned, for example, with data visualization.
Machine Learning studies not only learning from historical data, but also learning in real-time.
A major part of ML are the algorithms for agents acting in the environment and learning from their actions. This is called Reinforcement Learning (RL).



Data-Driven Decision Making

Data-driven decision making (or DDDM) is the process of making organizational decisions based on actual data rather than intuition or observation alone.

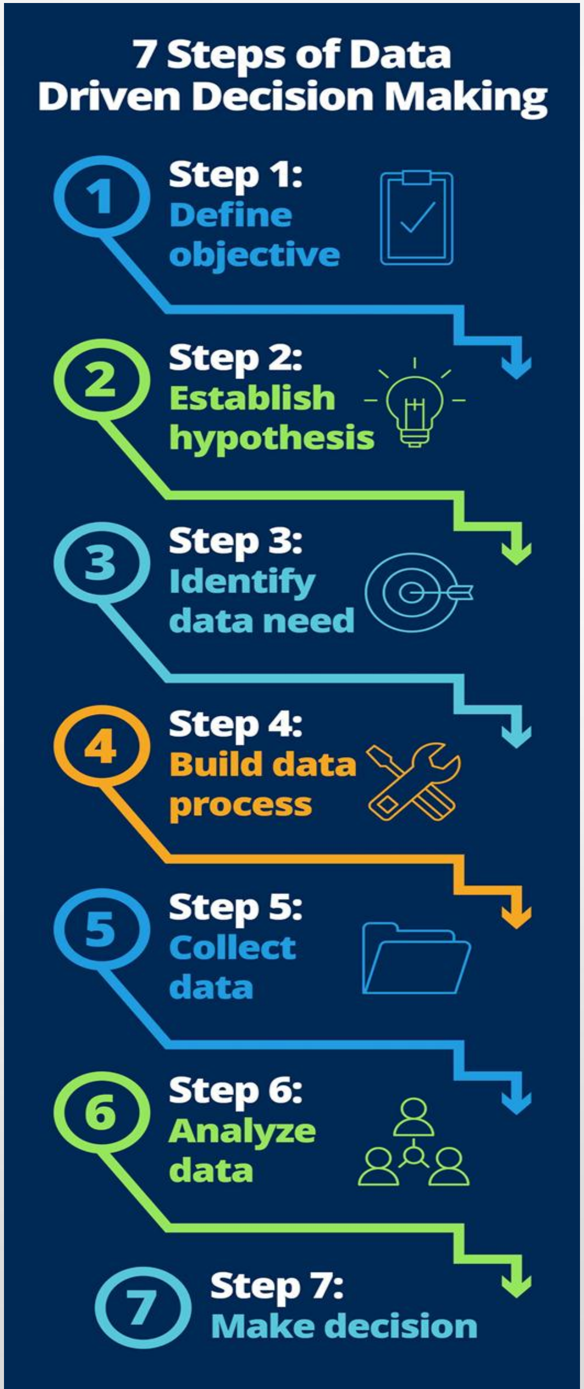
It involves collecting data based on measurable goals or KPIs, analyzing patterns and facts from these insights, and utilizing them to develop strategies and activities that benefit the business in a number of areas.

Why Data-Driven Decision Making

The importance of data in decision lies in consistency and continual growth.

It enables companies to create new business opportunities, generate more revenue, predict future trends, optimize current operational efforts, and produce actionable insights.

Data-Driven Decision Making Process



Data-Driven Decision Making Process

Step 1: Start by Defining the Objective:

Identify your business goals, and then build a broad strategy around them. At this stage, it's fine if the strategy isn't completely defined. But, you need to know how you think big data will help achieve your targeted outcomes.

For example, **if you own a big chain of grocery stores, you can examine customers' fruit purchases with the goals of improving the supply chain, the placement of products in the store, the pricing, and even the packaging of store-brand offerings.**
Articulating these goals will shape your data collection and analytics strategies from the outset.

Data-Driven Decision Making Process

Step 2: Focus on a Specific Business Area and Define Questions: Determine which area of your business needs the most attention. Typically, the answer is customers, finances, or operations (or some mixture of the three). Prioritize the area that will have the biggest impact.

Your next step is to identify the specific business questions you want to answer. Formulate this thought in the form of a hypothesis, a proposed if-then statement that can be proven or disproven and that you can use as a starting point for investigation.

For example, "If we add a self-service portal to the website, we will decrease the number of after-hours emergency trouble calls."

Data-Driven Decision Making Process

Step 3: Identify the Data You Need: The data you examine needs to be relevant to your business question. To build on our grocery store example, customer payment receipt data would be useful.

Two broad categories of data -

Qualitative Data: This kind of data is non-numeric and subjective. It is observed, rather than measured. For example, having a store worker ask people a few questions about their shopping experience will generate qualitative data.

Quantitative Data: This kind of data is numerical and objective and is measured, not observed. Quantitative data is what people think about when they think of big data: rows upon rows of data concerning who purchased what, how much they spent on it, when they bought it, how often they buy it, and so on.

Data-Driven Decision Making Process

To get a sense of your data's quality, here are some key questions to answer:

Who collected the data, and is it reliable? Reliability requires that it measure what it claims to, that all data was collected using the same parameters, and that it does not contain false responses (such as a fake email address).

How was the data sampled? Is the data representative of reality? Or, were only certain customers asked to respond?

Does the data include outliers (exceptionally high or low measurements)? How are these affecting the overall data distribution?

Data-Driven Decision Making Process

Step 4: Determine How You Will Get the Data You Do Not Have:

If you do not already have the data, you have to figure out how to get it.

Think of the four Vs of big data: How can you optimize volume, velocity, variety, and veracity? Is it by collecting the data yourself or by getting it from an external source?

Remember that holes in the data are more likely to crop up when different departments use different systems to collect, manage, and report that data. Some systems, which cross-analyze data that comes from different sources, can help solve this problem.

Data-Driven Decision Making Process

Step 5: Collect the Data:

When we talk about big data, we're usually talking about the automated data collection of thousands, millions, or even more cases. But, data analytics can also work with smaller quantities of data that are collected by people.

1) Assign People to Data Collection and Data Management Roles: **These people are the first — and sometimes the only — check on the veracity of the data.**

Data-Driven Decision Making Process

2) Define the Processes and Protocols of Data Collection:

Identify data sources and precise methods of data collection — that is, whether the data is being recorded by a machine or by a human. You should also create a data dictionary to catalog and define each of the recorded variables.

3) Clean and Do a Preliminary Processing of the Data:

Cleaning data is arduous but vital work to ensure integrity and usability. This process looks for duplicated, incorrect, or corrupted data. The analyst may also run preliminary descriptive analyses, such as data distributions, which will inform more advanced analytics.

Data-Driven Decision Making Process

Step 6: Analyze the Data:

To grow your business even to grow in your life, sometimes all you need to do is Analysis!

If your business is not growing, then you have to look back and acknowledge your mistakes and make a plan again without repeating those mistakes. And even if your business is growing, then you have to look forward to making the business to grow more. All you need to do is analyze your business data and business processes.

Data-Driven Decision Making Process

Data analysis tools make it easier for users to process and manipulate data, analyze the relationships and correlations between data sets, and it also helps to identify patterns and trends for interpretation.

Data-Driven Decision Making Process

Step 7: Present the Findings and Make Decisions:

Unless the results are presented to the right people at the right time in a meaningful way then the size of the data sets or the sophistication of the analytics tools won't really matter.

Finally, you need to apply the insights from the data to your decision making, making the decisions that will transform your business for the better.

Big Data Analytics

Big Data analytics is the process of collecting, organizing and analyzing a large amount of data to uncover hidden pattern, correlation and other meaningful insights.

It helps an organization to understand the information contained in their data and use it to provide new opportunities to improve their business which in turn leads to more efficient operations, higher profits and happier customers.

Big Data Analytics

To analyze such a large volume of data, Big Data analytics applications enables big data analyst, data scientists, predictive modelers, statisticians, and other analytical performers to analyze the growing volume of structured and unstructured data.

It is performed using specialized software tools and applications. Using these tools various data operations can be performed like data mining, text mining, predictive analysis, forecasting etc., all these processes are performed separately and are a part of high-performance analytics.

Using Big Data analytic tools and software enables an organization to process a large amount of data and provide meaningful insights that provide better business decisions in the future.

Big Data Analytics

Big Data Analytics has been popular among various organizations.

Organizations like the e-commerce industry, social media, healthcare, Banking, Entertainment industries, etc are widely using analytics to understand various patterns, collecting and utilizing the customer insights, fraud detection, monitor financial market activities etc.

Big Data Analytics

e-commerce industry like Amazon, Flipkart, Myntra and many other online shopping sites make use of big data.

They collect customer data in several ways like -

- Collect information about the items searched by the customer.
- Information regarding their preferences.
- Information about the popularity of the products and many other data.

Big Data Analytics

Using these kinds of data, organizations derive some patterns and provide the best customer service like

- displaying the popular products that are being sold.
- show the products that are related to the products that a customer bought.
- Provide secure money transitions and identify if there are any fraudulent transactions being made.
- Forecast the demand for the products and many more.

Big Data Analytics Vs. Data Science

Comparing Big Data Analytics with Data Science

Criteria	Big Data Analytics	Data Science
Type of Data Processed	Structured	All types
Types of Tools	Statistics and data modeling	Hadoop, coding, and Machine Learning
Domain Expanse	Relatively smaller	Huge
New Ideas	Not needed	Needed

Thank You