

# DATA ANALYTICS WEBINAR



**Dr. Emilie Ramsahai**  
**Data Scientist**



**July 17, 2020**  
**12:00 - 1:00 p.m.**  
**CPE: 1**

[www.engage.isaca.org/trinidadtobagochapter](http://www.engage.isaca.org/trinidadtobagochapter)

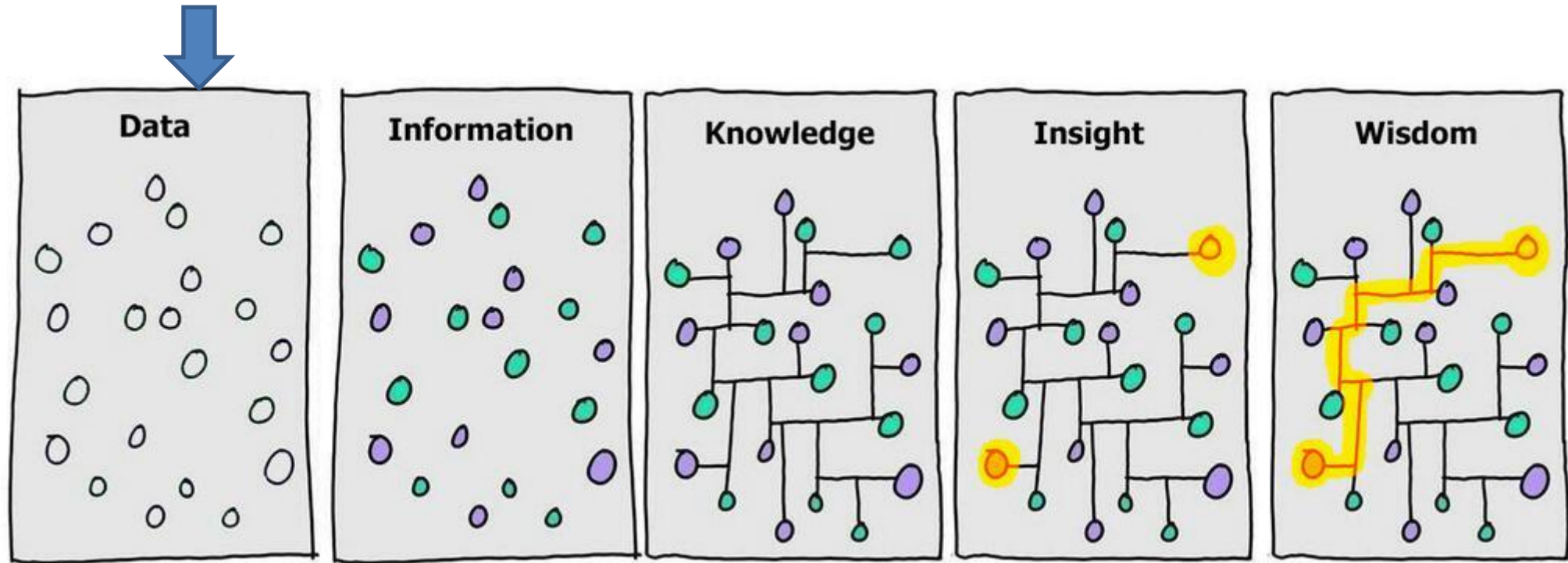


**ISACA®**

Trinidad & Tobago Chapter

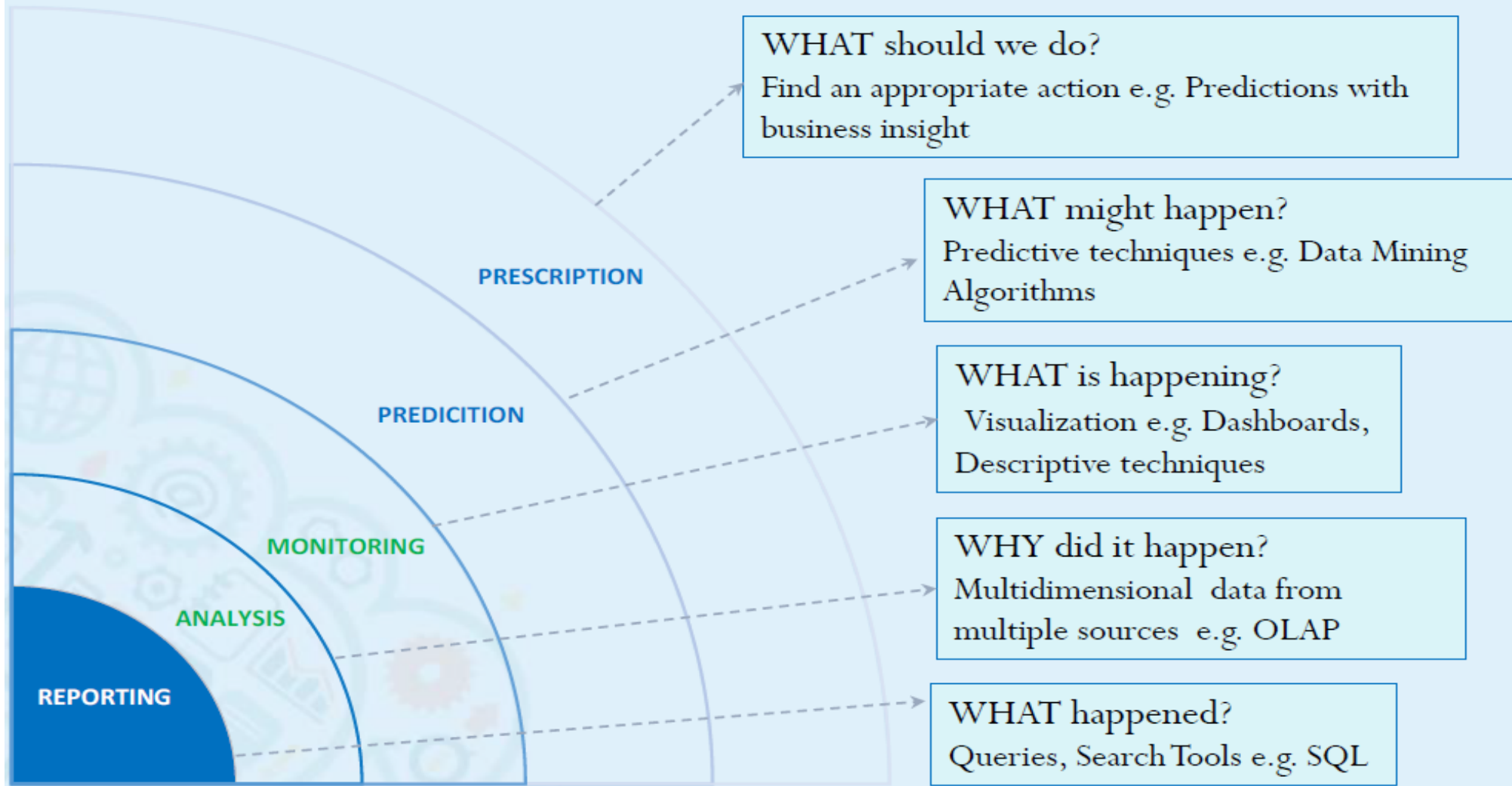
# Introduction with Applications

## The Information Continuum



Cartoon by [David Somerville](#), based on a two pane version by [Hugh McLeod](#)

# Types of Analytics



# Big Data

## Better Predictions

Big data can be characterized by the following:

Volume. Big data is voluminous. And its volume just keeps on expanding over time!

Velocity. Big data is generated at an exponential rate!

Variety. Data will come in from all sorts of formats beginning from structured and numerical data in traditional databases to unstructured text documents, emails, PDFs, videos, and financial transactions.



# Different type of Data on the Internet

## 2019 *This Is What Happens In An Internet Minute*

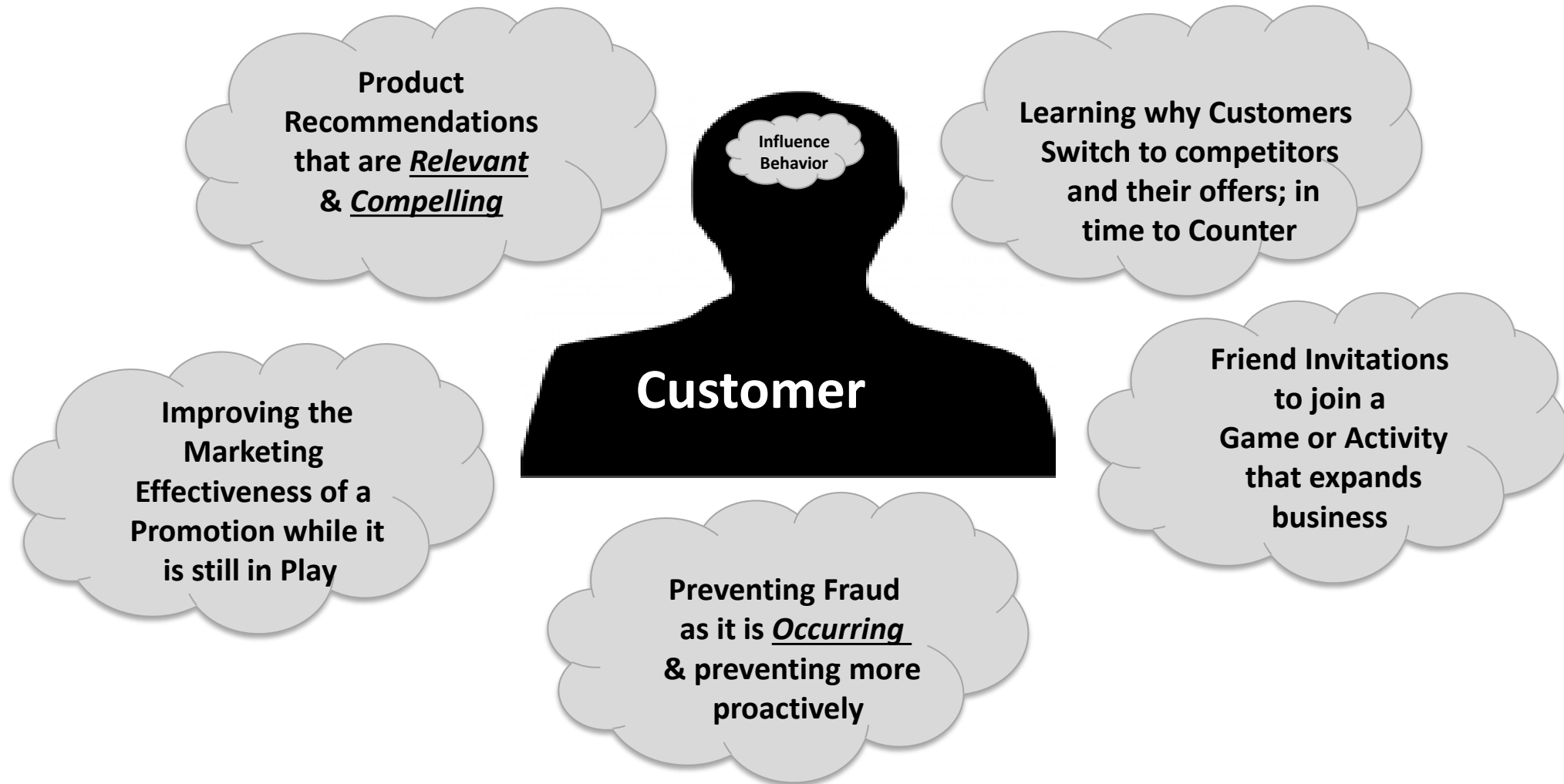


Real Time:  
analysis of  
fast changing  
streaming  
data





# APPLICATION Real-Time Analytics/Decision Requirement



## SENTIMENT ANALYSIS



**Discovering people opinions, emotions and feelings about  
a product or service**



# CORRELATION

1. Collect Big Data or obtain access to a repository.
2. Perform data analysis to explore patterns (pattern recognition, predictive analytics).
3. Identify potential correlations.



# Autonomous Vehicles

## Under the bonnet

How a self-driving car works

Signals from **GPS (global positioning system)** satellites are combined with readings from tachometers, altimeters and gyroscopes to provide more accurate positioning than is possible with GPS alone

**Lidar (light detection and ranging)** sensors bounce pulses of light off the surroundings. These are analysed to identify lane markings and the edges of roads

**Video cameras** detect traffic lights, read road signs, keep track of the position of other vehicles and look out for pedestrians and obstacles on the road

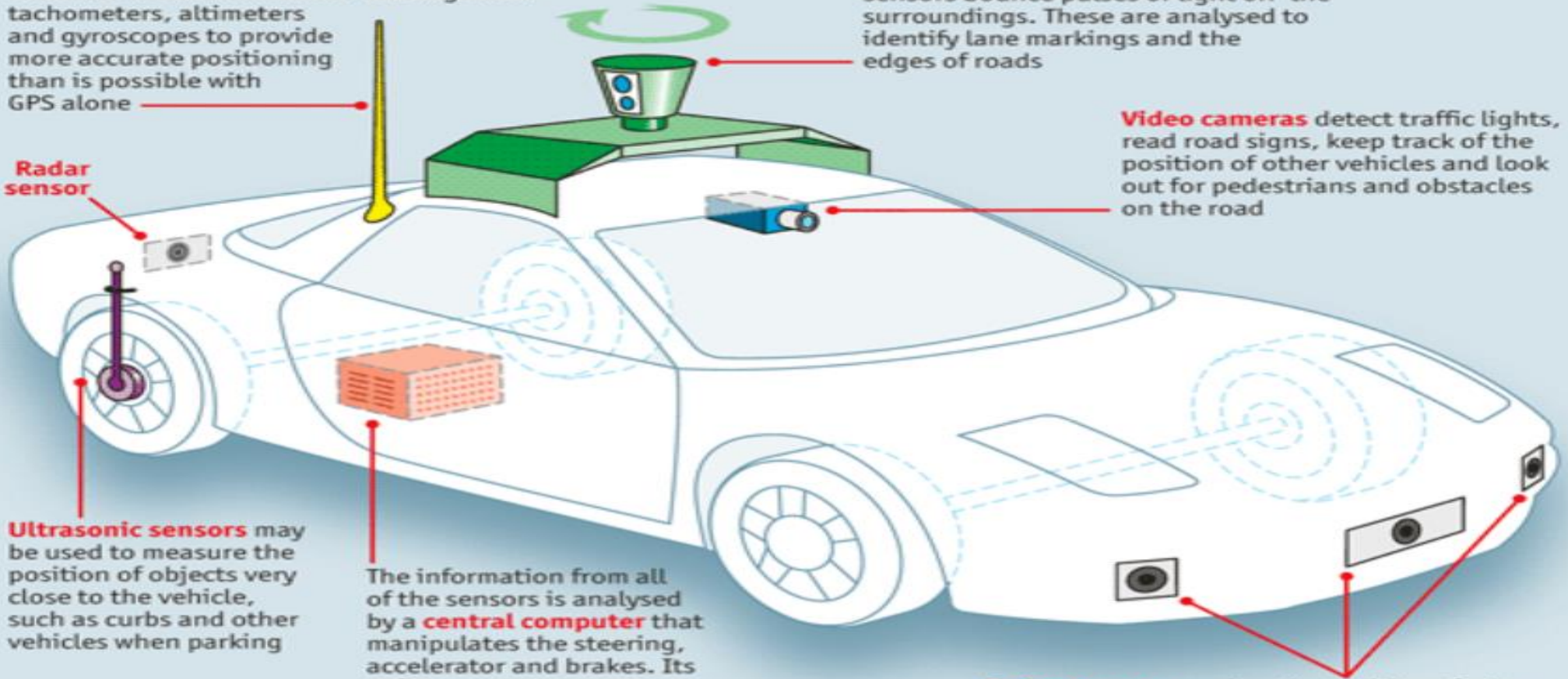
**Radar sensor**

**Ultrasonic sensors** may be used to measure the position of objects very close to the vehicle, such as curbs and other vehicles when parking

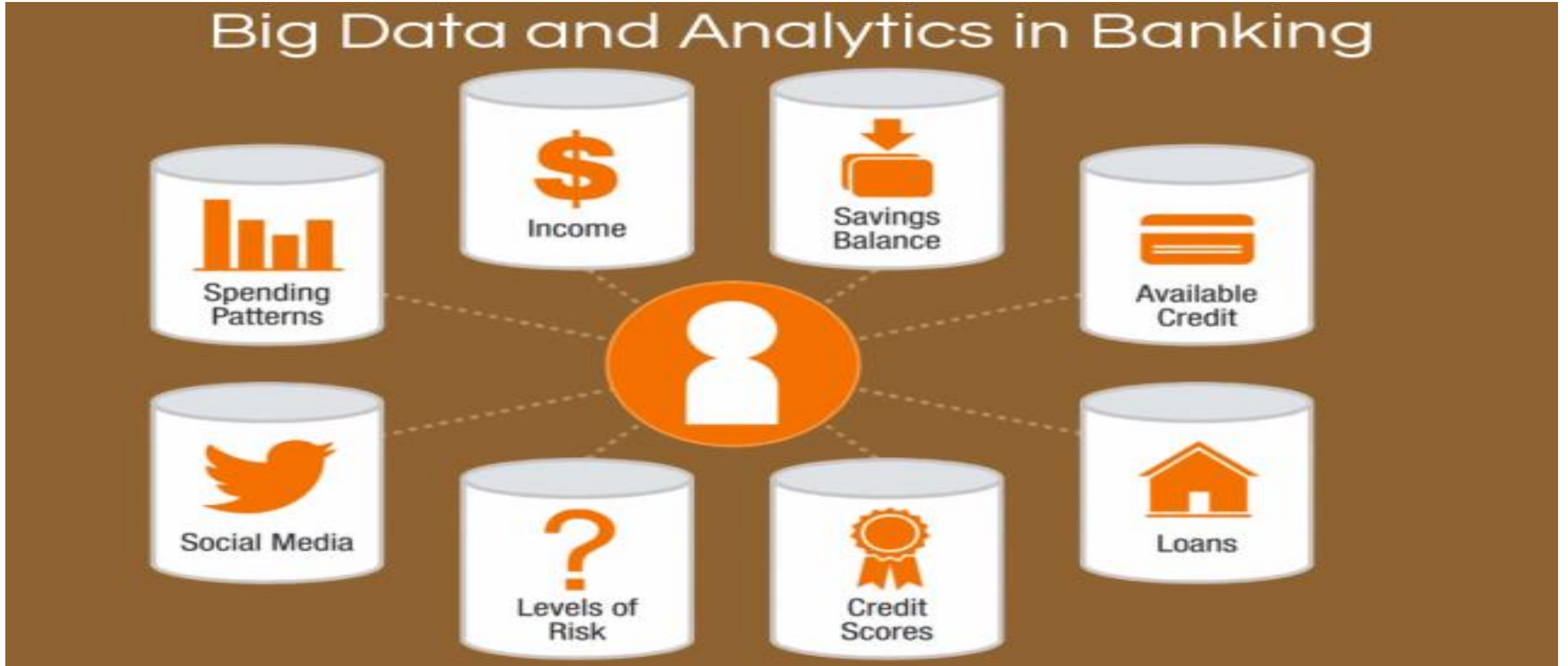
The information from all of the sensors is analysed by a **central computer** that manipulates the steering, accelerator and brakes. Its software must understand the rules of the road, both formal and informal

**Radar sensors** monitor the position of other vehicles nearby. Such sensors are already used in adaptive cruise-control systems

Source: *The Economist*



# LOAN APPROVAL





# Facebook Scandal Timeline

**March 21, 2018**

Fears of increased regulation over social media firms triggered Facebook's shares to tumble more than 9 per cent in the past week, losing \$60 billion.



**March 28, 2018**

Facebook announces changes to privacy settings to make them easier to find and use.



**April 10, 2018**

Mark Zuckerberg testifies to Congress. Facebook begins blocking apps from accessing user data 90 days after non-use. It also rolls out the earlier trailed updates to its bug bounty program.



**March 25, 2018**

Facebook apologizes for the data scandal with a full page ad in newspapers in the U.S. and U.K.



**April 9, 2018**

Facebook says it will begin informing users if their data was passed to Cambridge Analytica from today by dropping a notification into the News Feed.



Data-driven behavior

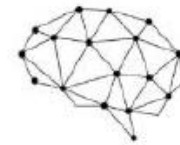


# Netflix's 'The Great Hack' highlights Cambridge Analytica's role in Trinidad & Tobago elections



THEY TOOK YOUR DATA.  
THEN THEY TOOK CONTROL.

FACEBOOK DATA BREACH



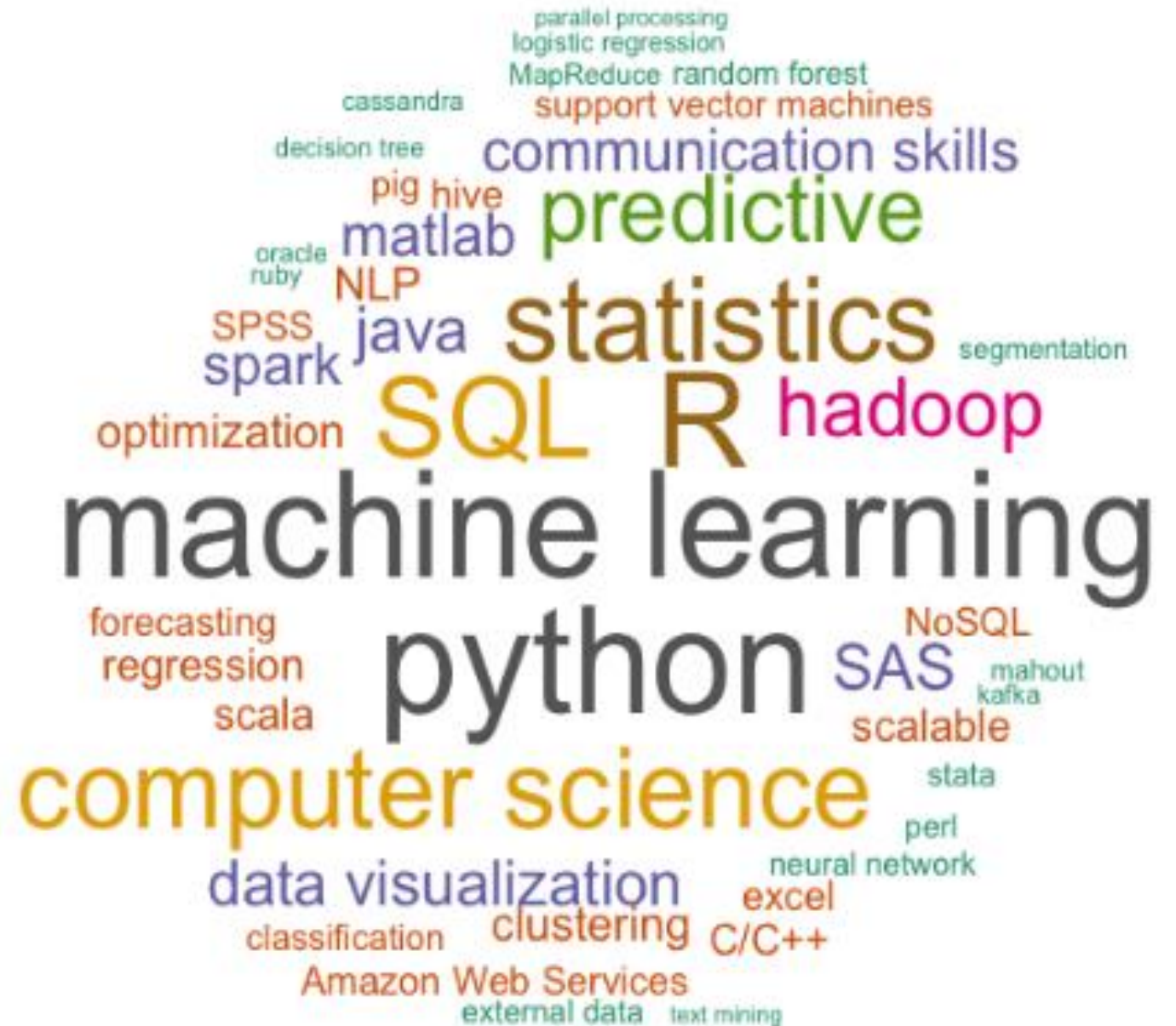
Cambridge  
Analytica



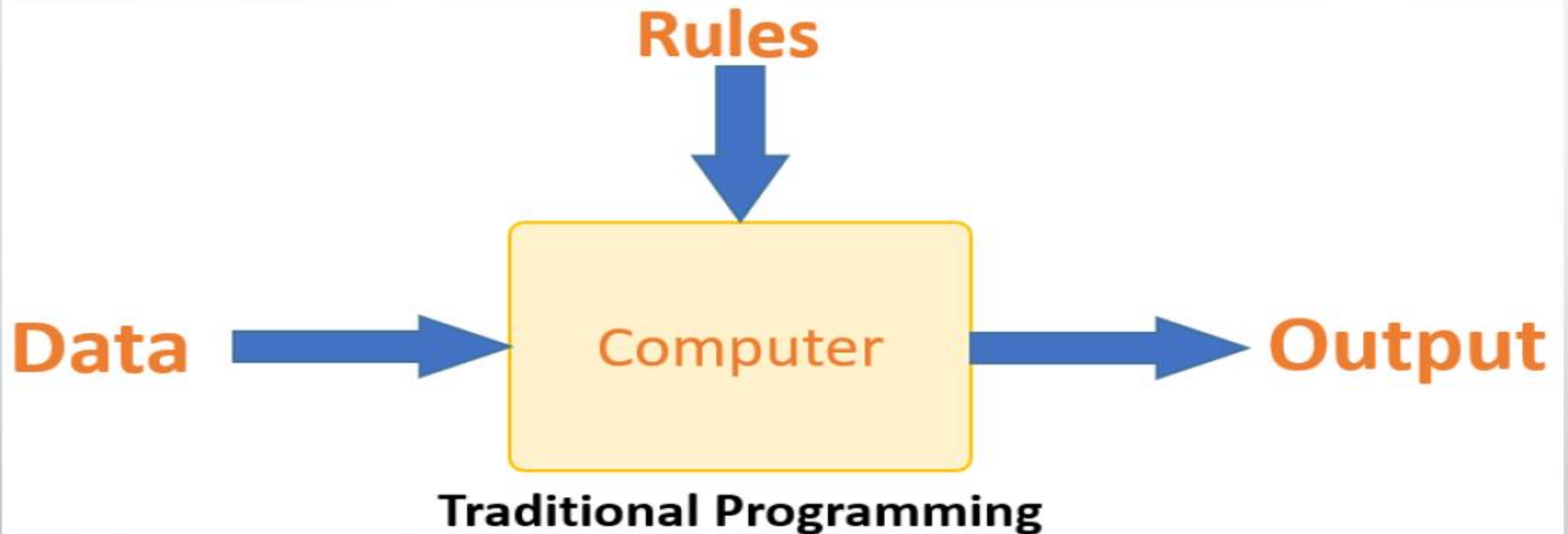
INFO ON 50M FACEBOOK USERS

JULY 24 | **NETFLIX**

# WORD CLOUD: SUMMARY OF MANY DOCUMENTS



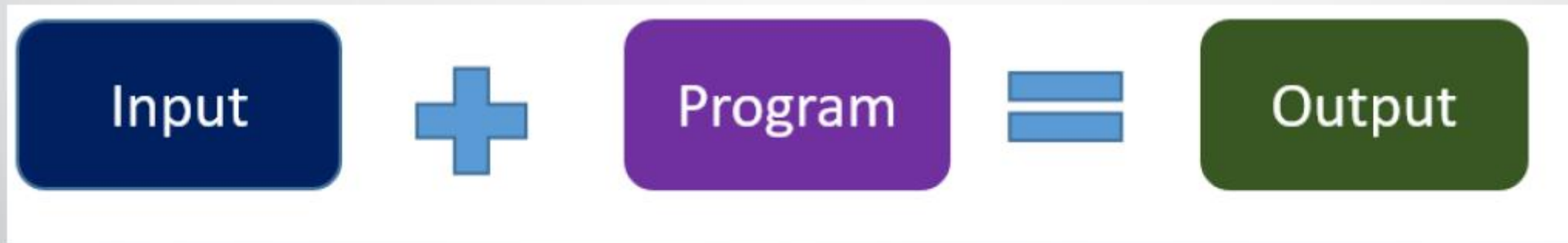
# Traditional Computer Programming





# Traditional Computer Programming

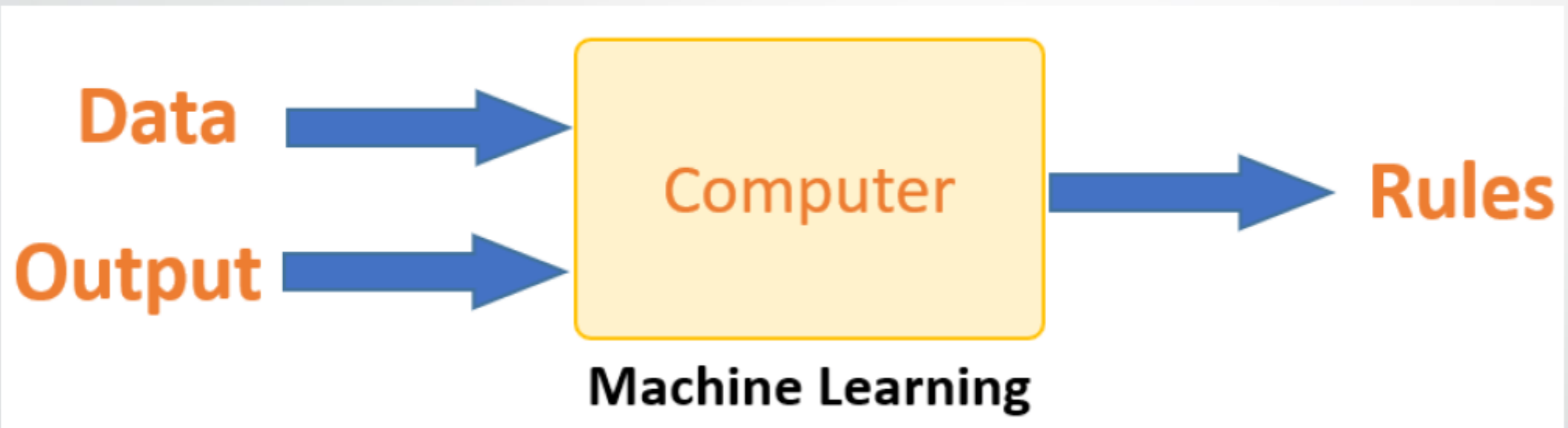
- In traditional programming, the rules are hardcoded in a program using a programming language such as Python, C++, or Java
- When the program is run with the input data, appropriate output is produced:



- Consider the problem of finding out how many days there are in a certain month. How many inputs are required?



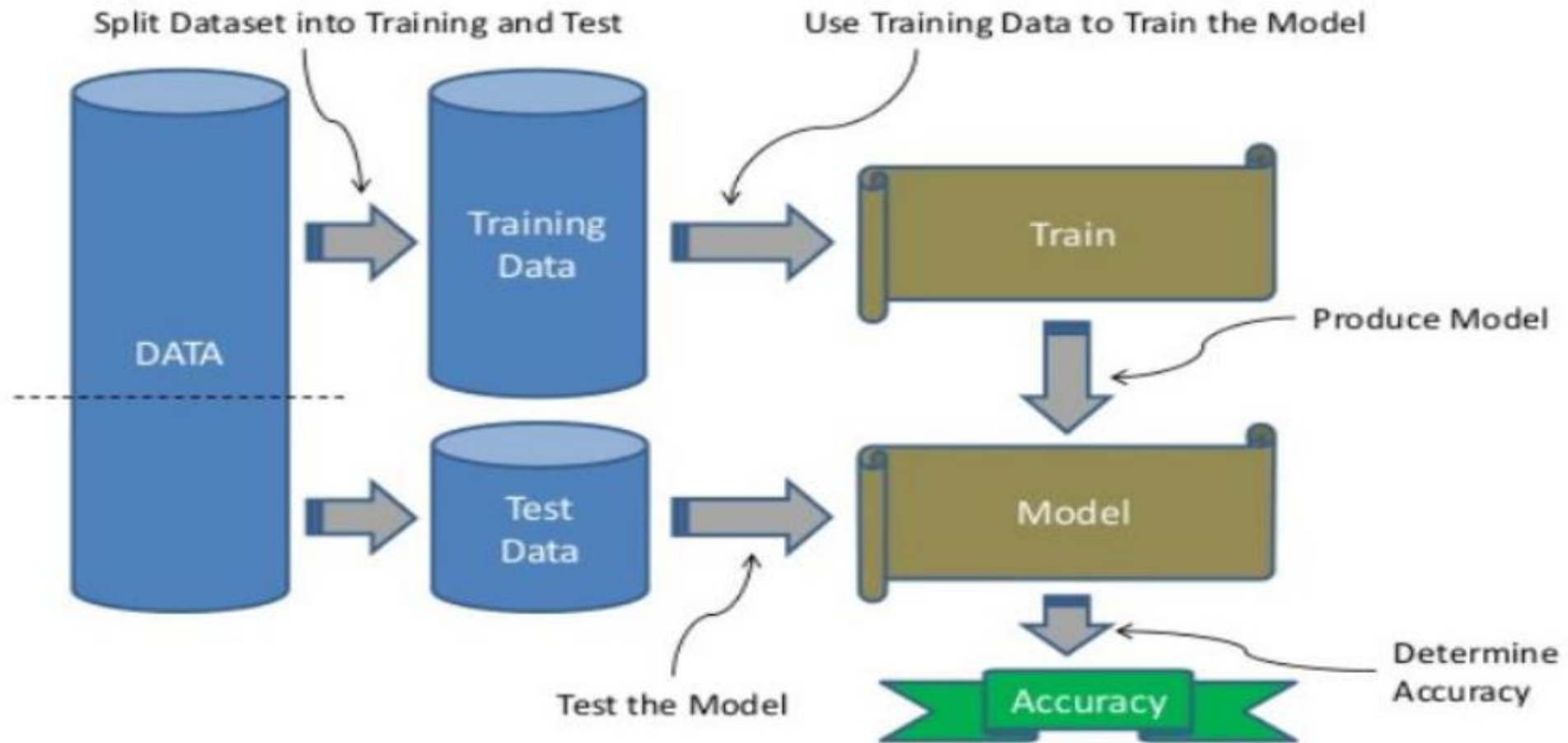
# Machine Learning

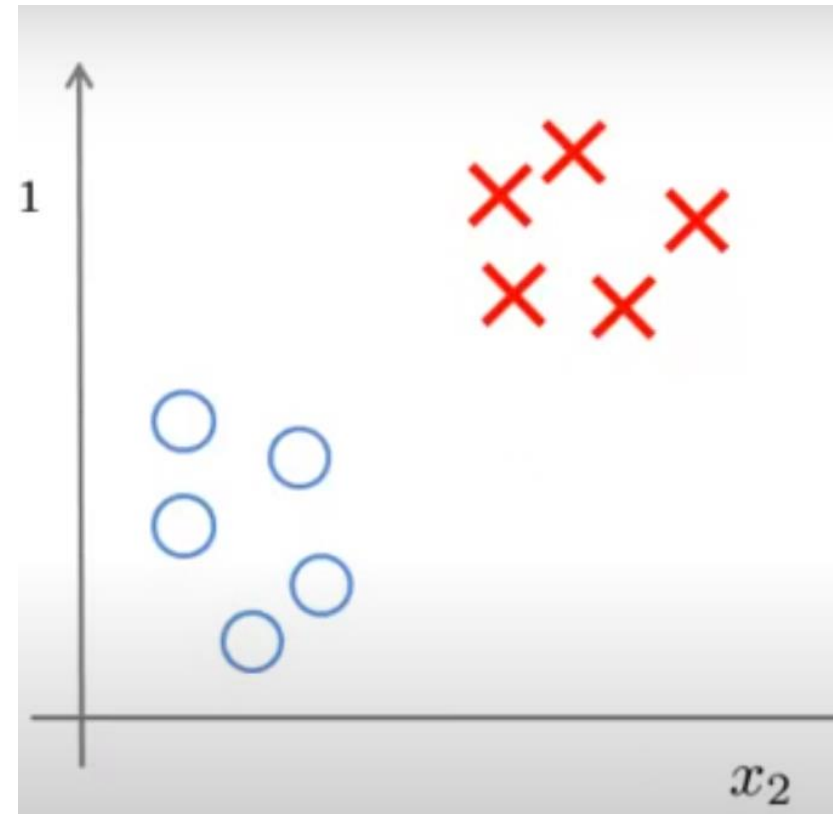
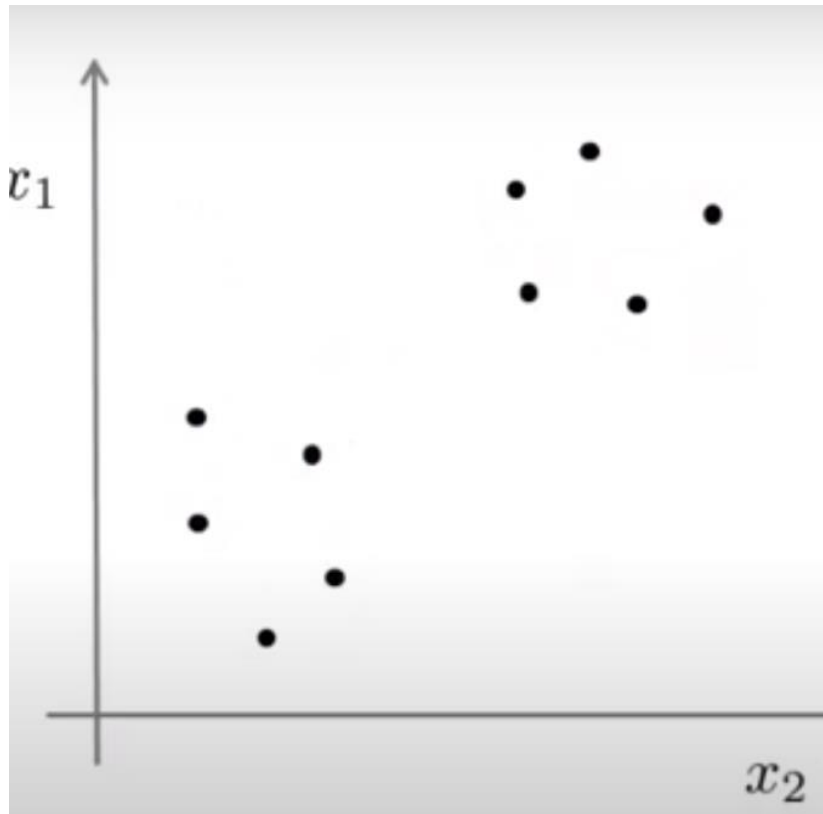


The machine learns how the input and output data are correlated and 'creates' a rule. Programmers do not need to write new rules each time there is new data.

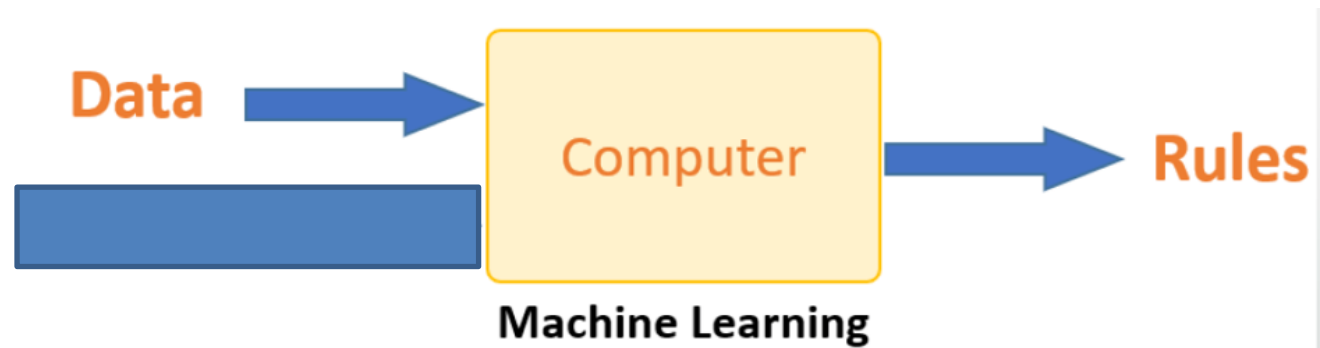
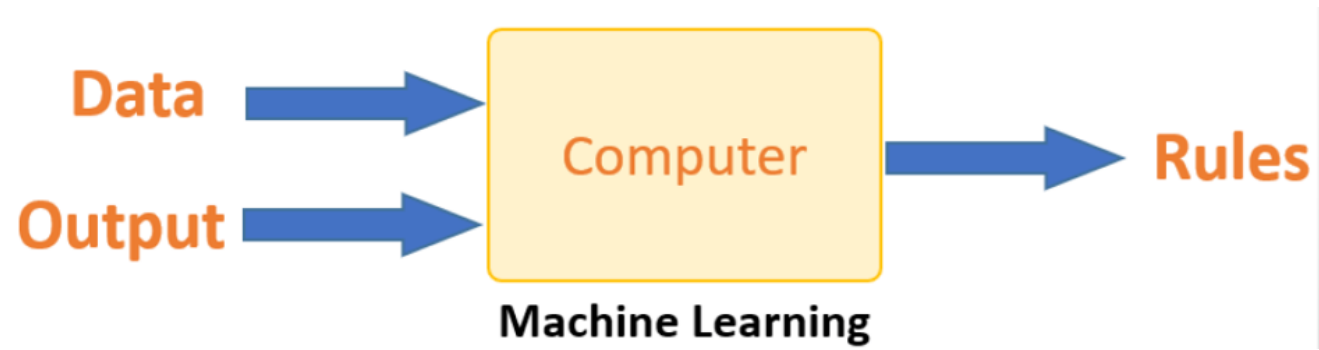
# Training Data

Machine Learning is about using data to train a model





Unsupervised vs Supervised  
Classification Example



Supervised vs Unsupervised



**supervised learning** algorithms, training dataset are labelled with classes/outcomes.

When predicting the outcome of new data, it makes use of the outcome of the training data

The algorithm doesn't learn on its own as some other ML methods, which are a group of Unsupervised algorithms e.g. Clustering

# Kaggle Projects

<https://www.kaggle.com/blastchar/telco-customer-churn>

<https://www.kaggle.com/pavanraj159/telecom-customer-churn-prediction>

<https://www.kaggle.com/mlg-ulb/creditcardfraud>

<https://www.kaggle.com/charumakhijani/credit-fraud-deal-with-imbalanced-data>