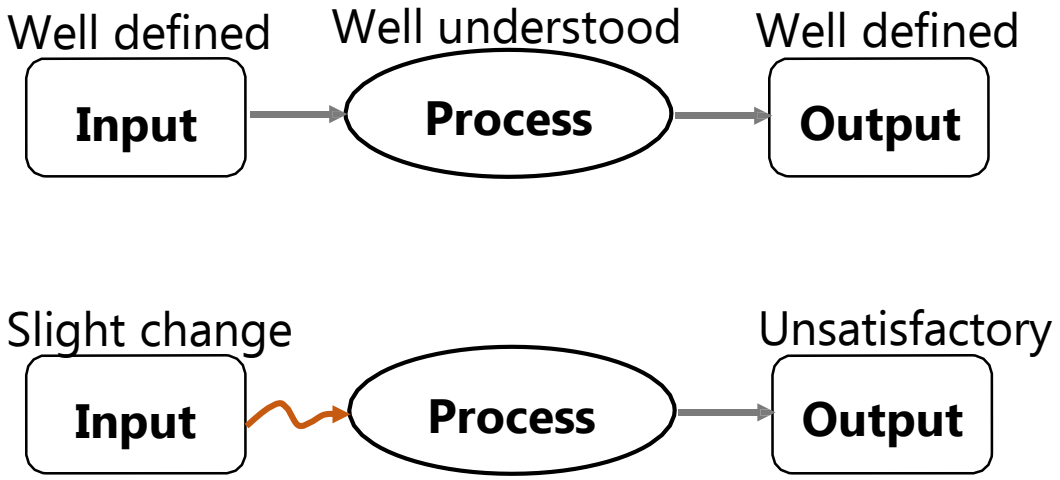
Introduction to Machine Learning

How Traditional Computing Works

Traditional computing or programs are very:

Deterministic programs are rigid

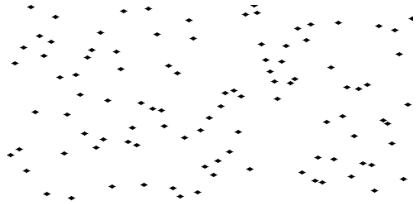


Every time you run the deterministic algorithm for a particular input, you'll get the same output

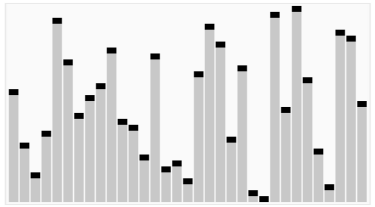
Traditional Computing Example 1

Series of numbers 20 5 3 55

Algorithms:



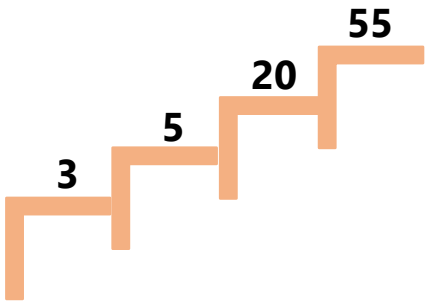
Bubble Sort



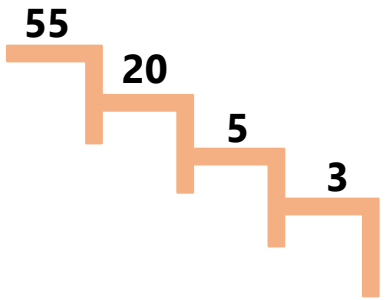
Quick Sort

Result

Output

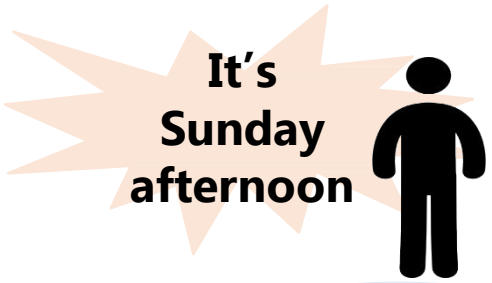


Ascending



Descending

Traditional Computing Example 2



Raghav is getting ready for a weekend party

He does not like going out usually

But his friends are forcing

Raghav

Raghav went to engineering

His high school friend Ramesh

Ramesh found Mathematics interesting

Traditional Computing Example 2

Algorithm to count the number of times ‘Raghav’ appears in the text

- Initialize a variable called count to 0;
- Split the paragraph into words
- Check if each word is same as ‘Raghav’
- If yes, increment count by 1
- Same result, regardless of programming language
- Result = 3

Failure of Traditional Programming

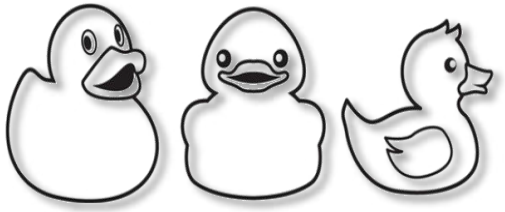
How do you identify a marketing message?

Mails with words like ĥbuyˇ or ĥpurchaseˇ cannot be classified as marketing emails

Even if it were a marketing mail, it's difficult to classify either ĥwantedˇ or ĥunwantedˇ

Failure of Traditional Programming

Scenario



Marketing message promoting toys for kids

If you do not have kids

Spam

If you have kids

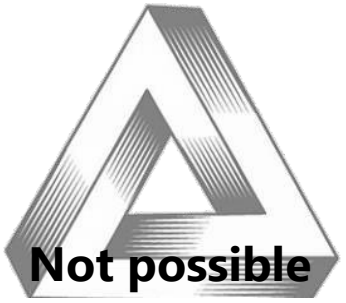
You will welcome such emails if you are looking out for new interesting toys

Such emails become a spam if you're not interested in buying new toys

Failure of Traditional Programming

Hand-writing recognition

Everyone has their own way of writing



To anticipate all the writing styles and write strict rules for identifying different alphabets

Hence an algorithm is required:

- ❑ To accommodate different styles of writing
- ❑ That is precise enough to identify alphabets or numbers like a human

Failure of Traditional Programming

Let's see if we can expect a machine to do a particular task

Write a program to extract a information ˆWhen and where Sachin hit his first century in one day international?ˇ

This involves a technology called ˆNatural Language Processingˇ

Natural Language Processing makes heavy use of Machine Learning to deal with complexity and uncertainty of human language

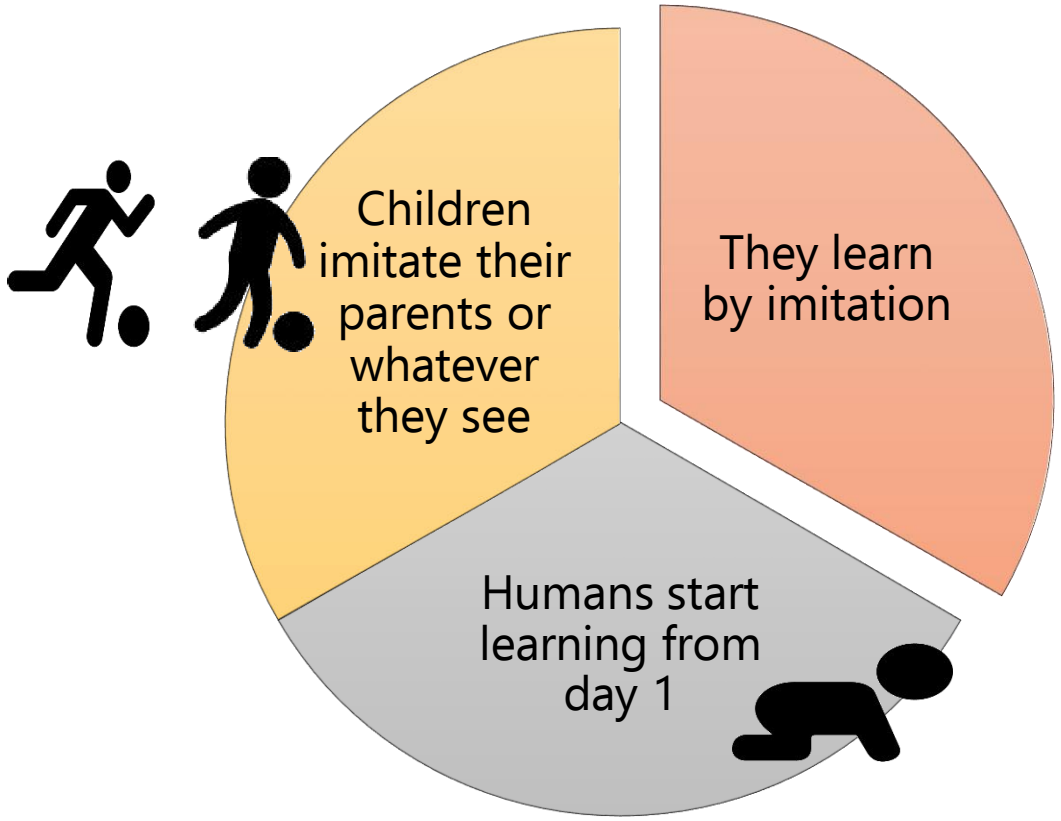
Requirements for New Approach

- Acknowledge inability to write precise rules to solve a particular problem or to do a particular task
- Accommodate possible variations or noise in the input data
- Able to handle fairly complex problems
- Okay to get approximately good results on an average
- Able to learn on its own

Tasks that can't be done by precise rules

We humans can do some amazingly complex tasks by ˹learning˺ things without being ˹programmed˺

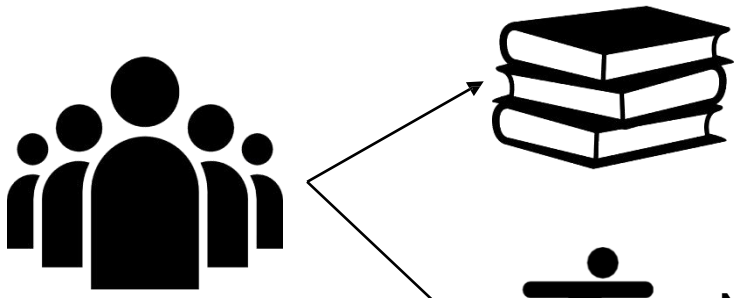
Learning by Humans



- This learning is different from computer programming
- We also undergo a formal ˆtrainingˇ or education like in school or college

Example: Learning Cycling

We will see how learning cycling is related to Machine Learning

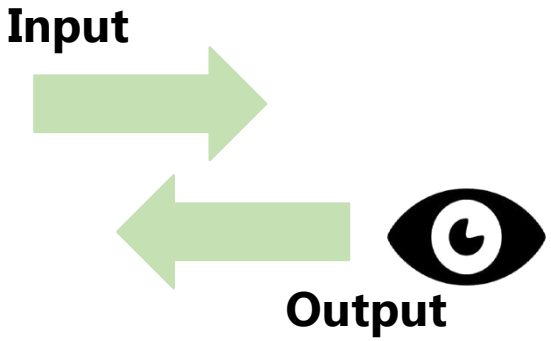


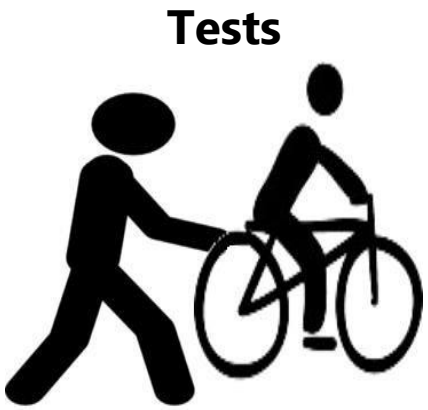
Did not read books

Nor we learned Newton's laws of motion

- We observed how others were doing, and we started with the help of a family member or a friend
- We gradually learned how to control handle, how much pressure to apply, how fast to peddle, etc by trial and error

Example: Learning Cycling





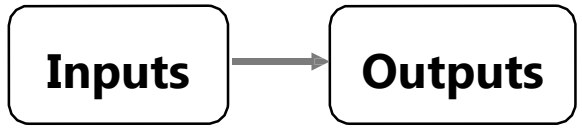
Friends pushing

Set free

If we had learned our lessons properly, we would ride safely, otherwise we would stop or fall down, which meant we needed some more practice

Example: Learning Cycling

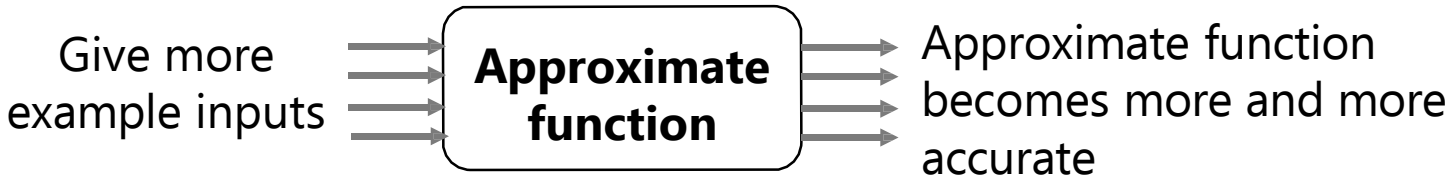
Machine Learning is also quite similar



We don't have to know the rules that produce these ôutputsˇ

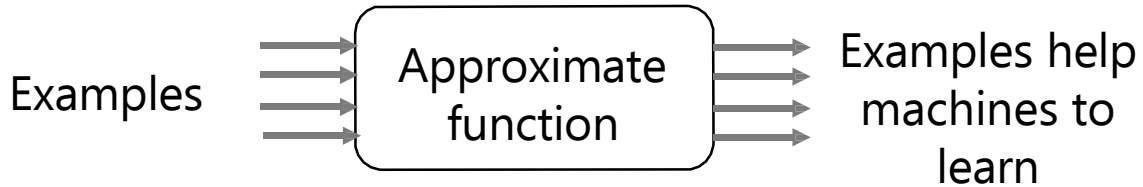
Create a program that can approximate a function and can produce outputs given some inputs

Core idea of Machine Learning



Learning by Machines

Machines or computers can be programmed to ˹learn˺



- TRAINING

- Machine Learning is:
  - ✓ Not a hypothetical technology
  - ✓ It is a reality today

- Examples of Machine Learning are:
  - ✓ Separating spam emails from non-spam
  - ✓ Online-banking security systems
  - ✓ Recommendation systems that suggest products to buy in
  - online ecommerce sites etc

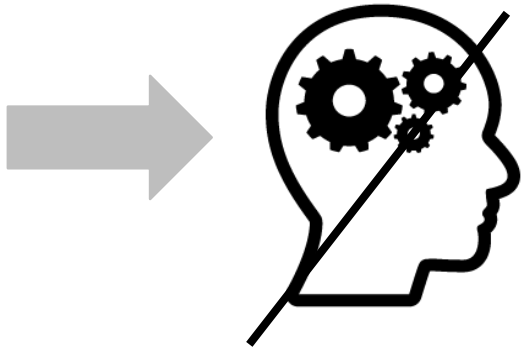
Learning by Machines

- Machine Learning algorithms are usually not 100% accurate because they are statistical in nature.
- The output of such Statistical Learning or Machine Learning is a ˆmodelˇ that approximates a phenomenon or a function
- This ˆmodelˇ is used for prediction or to perform some task and there will always be a parameter that represents accuracy of the model
- Accuracy level of the ˆmodelˇ and where it works well should be known
- Never blindly trust a model without proper testing

Data in Machine Learning

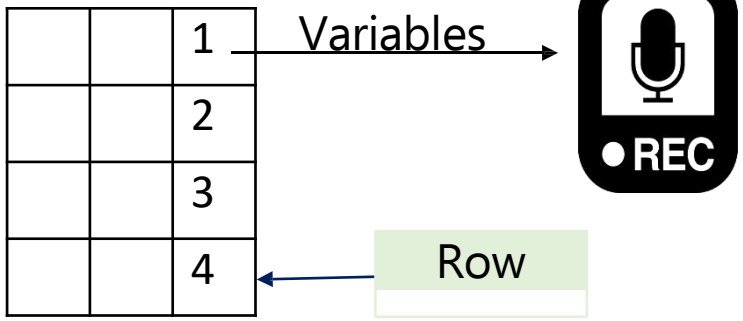
Data is central to Machine Learning algorithms

No Data



No Machine Learning

Machine Learning is a process of learning from examples and examples are fed to the algorithm through data



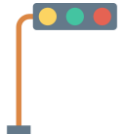
Data is usually in table format

Machine Learning Tasks

Regression: Predicting the value of continuous variable based on how a variable is related to some other variables

Example –

a. An e-retailer like Amazon or Flipkart can predict how much money a customer will spend in next 1 month based on his/her purchase history and user interaction on their mobile apps



Classification: Predicting the class of a given data point, given certain attributes

Example –

a. Based on the credit history, a bank can predict if a given person will pay his full due
b. An expert system controlling a self-driving car while determining the speed limit by following traffic signs

Unsupervised Learning: Finding similar data points in a dataset

Example –

a. Grouping segments of customers who are similar on certain sets of attributes like demography, buying behaviour etc.
b. Building recommendation engines, which are also based on the notion of similarity

Regression Task

Historical data

Gender	Income	Age	Amount Spent
Male	40000	30	1000
Female	35000	26	500
Female	50000	32	2500
Male	50000	40	5000
Female	65000	35	5000

Regression task

Build an algorithm

Predict the amount spent

Algorithm will learn to predict Amount Spent of a new customer with the available information about their gender, age and income

Regression Example 2

Temperature	Average Rainfall	Crop Yield
34°C	135 inches	80 Units
33°C	140 inches	65 Units
38°C	137 inches	60 Units
31°C	152 inches	82 Units
42°C	120 inches	55 Units

Data used for regression

Regression is basically about predicting a number from given inputs

The Machine Learning problem for this data set would be:

Given: Temperature and Rainfall
Predict: Crop Yield

Classification Task

Historical data

Gender	Income	Age	Good
Male	40000	30	Yes
Female	35000	26	No
Female	50000	32	Yes
Male	50000	40	No
Female	65000	35	No

Classification

Bank

Train an algorithm

Classify its current customers into good or bad customers

Help the bank in future to predict if a customer will be a good or not given his demographic data such as gender, income and age

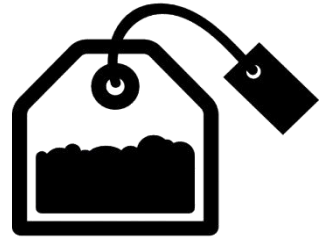
Labelled Data: Crop Yield

When predicting the crop yield, ‘High’ or ‘Low’ are considered as labels that we attach to each observation or each row

Temperature	Average Rainfall	Crop Yield
34°C	135 inches	High
33°C	140 inches	Low
38°C	137 inches	Low
31°C	152 inches	High
42°C	120 inches	Low

Naturally

This data is called labelled data



Given: Temperature and Rainfall
Predict: Crop yield either ‘High’ or ‘Low’

Classification Task

Image labels and corresponding signs

Image	Sign
Image 1	Stop
Image 2	U Turn
…	…
…	…
Image N	Parking

Classification

Expert system recognizing images of different traffic signs

Image labels

Train an algorithm

Recognize the sign corresponding to the image label

Labelled Data: Crop Yield

Labeled data:

✓ Input variables

✓ Output variables

We can predict output variables given input variables

Labeled data is used in supervised learning

Supervised Machine Learning

Classification Task

Gender	Income	Age	Good
Male	40000	30	Yes
Female	35000	26	No
Female	50000	32	Yes
Male	50000	40	No
Female	65000	35	No

Image	Sign
Image 1	Stop
Image 2	U Turn
...	...
...	...
Image N	Parking

Regression Task

Gender	Income	Age	Amount Spent
Male	40000	30	1000
Female	35000	26	500
Female	50000	32	2500
Male	50000	40	5000
Female	65000	35	5000

Contained a column that was needed to be predicted

Supervised Machine Learning

Whenever an algorithm is trained in a manner where the variable required by the algorithm to predict is present in the training data

Unlabeled Data

It is the data that records some specific observations and here it is not necessary to identify ˇinputˇ and ˇoutputˇ variables

Age	Gender	Income	Location
45	Male	5.78 lakh	Bangalore
32	Female	8.81 lakh	Chennai
33	Male	11 lakh	Mumbai
71	Male	20 lakh	Mumbai
28	Female	8 lakh	Delhi

Unlabelled data does not have ˇinputˇ and ˇoutputˇ variables

- You cannot possibly use one or more variables to predict another variable
- Unsupervised learning algorithms uses this type of data, which can be further used for clustering, etc

Unsupervised Machine Learning

Recommendation engine

Customer	Item 1	Item 2	Item 3	Item 4
C1	Yes	Yes	No	No
C2	No	No	Yes	Yes
C3	Yes	?	No	No
C4	No	No	?	Yes

User behaviour with respect to product offerings

C3 and **C4** - not bought some items and their preferences are not known

C1 and **C3** – similar in terms of product preferences

Recommend **Item 2** to **C3**

C3 will like **Item 2**

C1 - liked **Item 2**

Unsupervised Machine Learning

Recommendation engine

Customer	Item 1	Item 2	Item 3	Item 4
C1	Yes	Yes	No	No
C2	No	No	Yes	Yes
C3	Yes	?	No	No
C4	No	No	?	Yes

C3 and **C4** - not bought some items and their preferences are not known

C2 and **C4** – similar in terms of product preferences

Recommend **Item 3** to **C4**

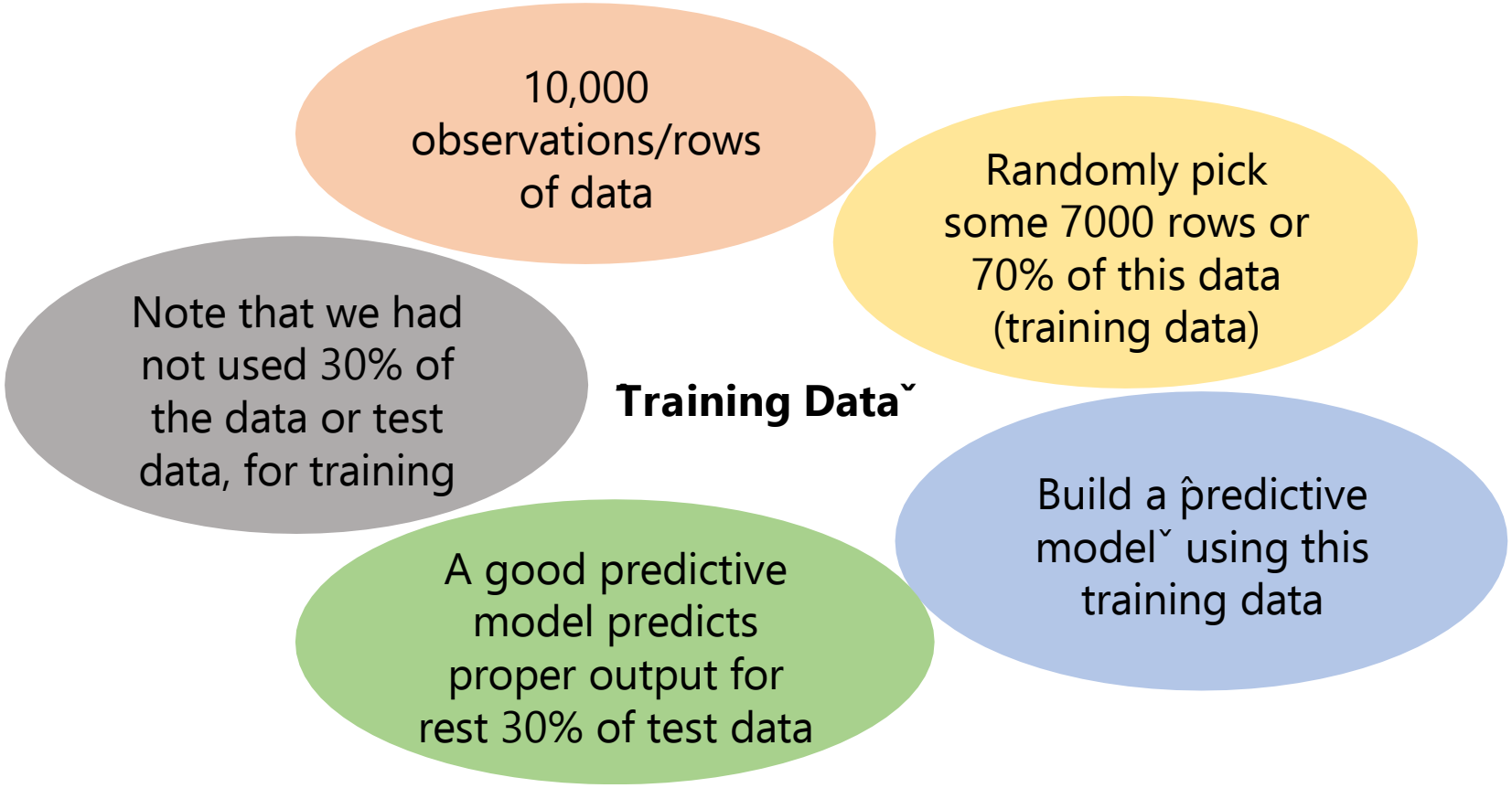
C4 will like **Item 3**

C2 - liked **Item 3**

User behaviour with respect to our product offerings

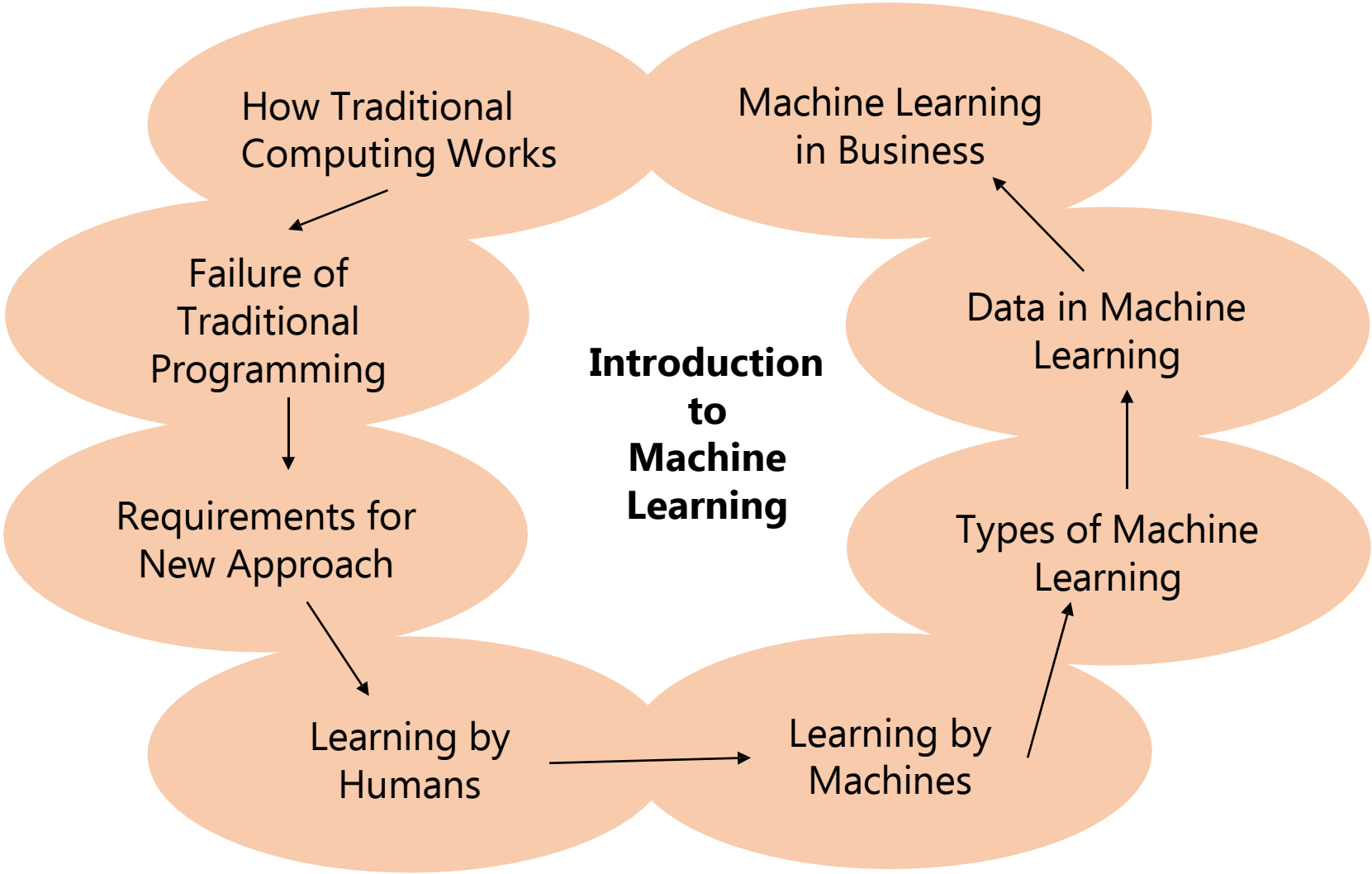
Unsupervised Machine Learning
When no target variable is present in the dataset of the algorithm

Supervised Learning



Very popular and useful Supervised Learning algorithms are Linear Regression, Bayesian Classifier, K-nearest Neighbour, etc

Recap


How Traditional Computing Works
Failure of Traditional Programming
Requirements for New Approach
Learning by Humans
Learning by Machines
Types of Machine Learning
Data in Machine Learning
Machine Learning in Business
Introduction to Machine Learning