Machine Learning In 8 Minutes

Farhad Malik
Jul 20, 2018 · 8 min read ★

Machine learning is the present and the future. All technologists, data scientists and financial experts can benefit from Machine Learning.

Please read FinTechExplained disclaimer.

This article focuses on outlining an easy-to-understand explanation of what machine learning is. Then it introduces how machine learning works and finally, it details a number of machine learning applications.

Let's Start — What Is Machine Learning

Machine learning is a branch of Artificial Intelligence. Machine learning is a set of algorithms that learn to discover trends and patterns in data to gain insights. These algorithms then become self-sufficient to make decisions on the data.

Machine learning algorithms are now utilised in nearly all sectors — from healthcare to financial organisations to anti-fraud companies to shopping websites.

How is it different from traditional programming?

Traditionally, programmers code a number of software procedures or rules. These procedures are also known as rules or methods. The set of instructions take certain inputs and produce expected outputs. Instructions can also execute other functions.

Machine learning is about getting machines to learn data and then make decisions on similar data. Machine learning is about using *predictive algorithms* to forecast the behaviours of data so that calculated decisions can be taken.

Machine learning algorithms are built on statistical features.

Why Is Machine Learning So Popular Nowadays?

It is probably worth knowing that machine learning is not a new concept. You might have heard the buzzwords artificial intelligence/deep learning/machine learning/big data/data scientist in the near past and possibly more recently.

Machine learning's growing popularity is primarily due to an increase in data availability and advancements in technology. Faster machines and smarter algorithms are implemented daily. Subsequently, cloud computing is introduced where we can load a large quantity of data. The amount of data stored in the servers is growing at an exponential rate. This data is valuable and can help us make better decisions in the future.

> *We Have a Vast Amount Of Factual Historic Data Stored That Can Help Us Understand Behaviour Better*

Nearly all major organisations from Google to Microsoft to Amazon to IBM have started to adopt machine learning algorithms and it is possibly the hottest topic right now.

Machine learning is getting popular day by day

What Is Involved In Building A Machine Learning Application?

A range of skills is required to build machine learning applications. We need subject matter experts who understand data and can interpret data in an efficient manner. These experts are occasionally known as data scientists or domain experts. Often domain experts are teamed up with technical experts who help build intelligent algorithms that predict future and make calculated decisions.
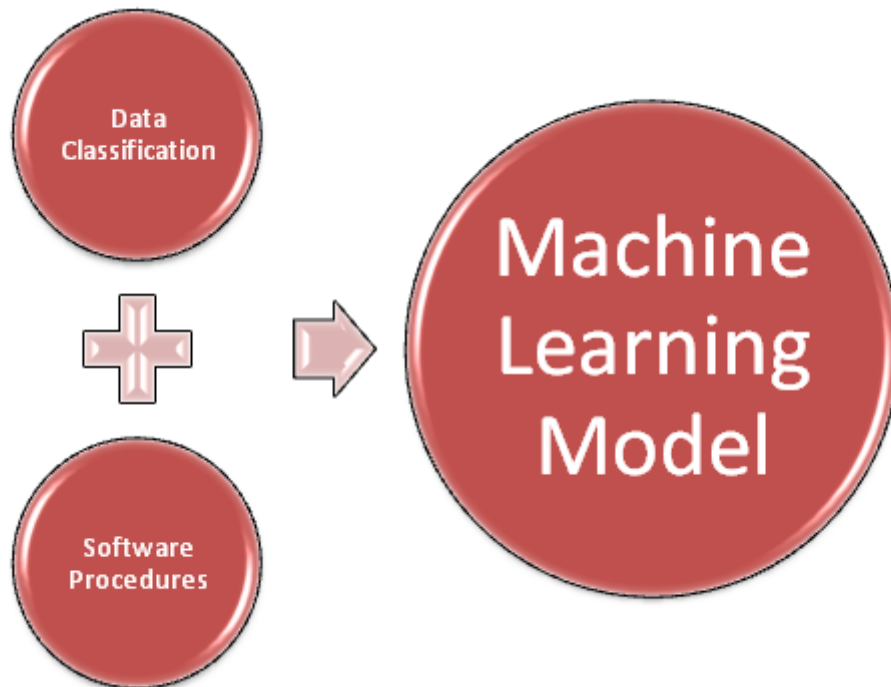

Data Scientists = Domain Experts + Programmers

To intricate, programmers along with domain experts build algorithms that understand data, structure and classify data. Once the algorithms are trained then these algorithms start forecasting and start making intelligent decisions. These decisions help analysts perform predictability analysis and help organisations improve productivity.

Domain experts help technical experts understand the data so that efficient and accurate algorithms can be built and models can be trained. Technical experts help

domain experts by programming efficient routines that break and structure large amounts of data in a fast manner.

Though the process is iterative in nature but once done right, it can yield great benefits.

The machine learning model is a set of rules or instructions or software procedures implemented by a combination of a domain and technical experts. These experts are also known as Data Scientists.
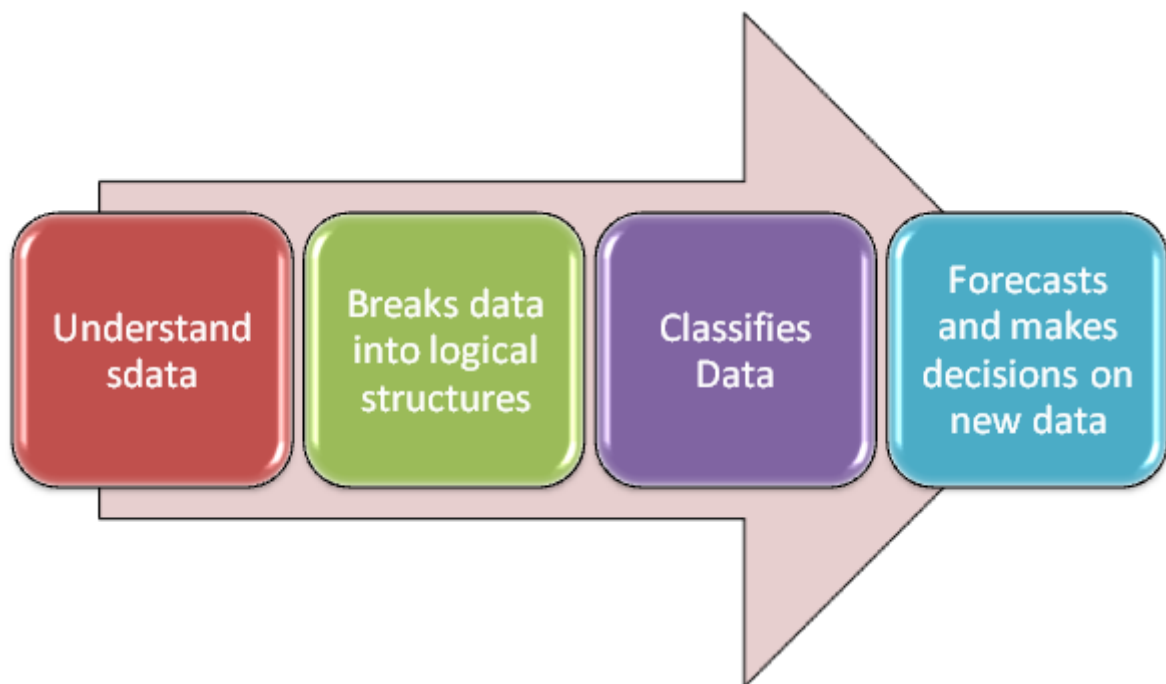
Machine Learning Process

Writing an efficient and accurate model is the key to increasing chances of a successful machine learning process.
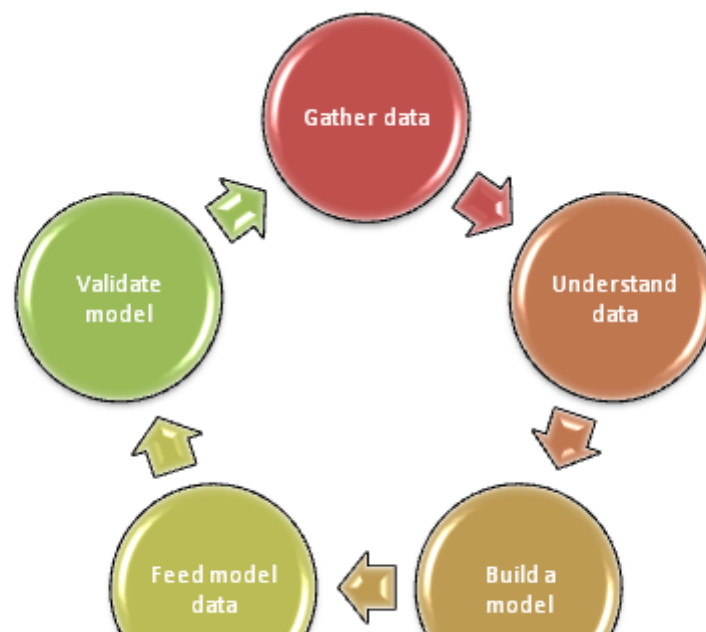
At a high level, the process is:

1. Gather and clean data (sample) to represent large data (population) — this step can at times take the longest time.

2. Learn and understand data to figure out trends and patterns

3. Build a model that understands the data and makes decisions on data

4. Feed the model 70%-80% of sample data. This set of data is known as **Training Data**.

5. Validate the model with the rest of the data. This set of data is known as **Test Data**.

6. Based on results, repeat the steps if required.

The process is iterative in nature. It requires deep analysis of variables behaviour.

Machine learning applications need technical experts who can implement computationally intensive and intelligent algorithms efficiently and domain experts who can understand data, classify it and figure out trends and patterns.

Let's go over a Machine Learning Use case

- Assume you are a bank manager. You have a team of 2 staff members, Alex and Bob. Both Alex and Bob have a different set of skills.

- You have decided to build a machine learning algorithm that can help improve customer's satisfaction level with the bank's services.

- Therefore, every time a customer comes into your bank, you want a machine to automatically assign the customer to an appropriate staff member.

- To help you design the algorithm, you hire a domain and a technical expert.

The domain expert advises you to start collecting information about your customers such as Age, Nationality, Gender, whether they are new customers and the reason for visiting the bank. These customers' attributes are known as *data features*. After each customer is dealt with a staff member, the customer's satisfaction level is also recorded.

> *As it is near to impossible to collect this data for all customers, therefore, a handful of customers are chosen at random.*

This chosen set of customers is the *sample* that will represent the entire population of customers.

Sample Dictates Goodness Of Model

A poor sample will yield poor results. Imagine over time your collected data looks somewhat like this:

Age	Nationality	Gender	Existing	Reason	Client	Staff	←These

			Client	For Visiting	satisfaction	member	are known as features
45	British	Male	Yes	Withdraw money	8	Alex	
34	American	Male	Yes	Deposit money	2	Bob	
46	British	Female	No	Opening new account and deposit money	8	Bob	
45	Chinese	Female	No	Query	8	Bob	

The efficiency of machine learning algorithms is dependent on the quality and quantity of data. If your sample is weak, your algorithm is going to be weak.

From the data above, we can notice a pattern that Bob is good with new clients and he does not have appropriate skills to deal with existing customers. Notice, this might be coincident and if we were to increase sample size then we might experience a different behaviour.

We might also conclude that Alex is great at customer service however again; this might not always be the case if we increased the sample size. Therefore to increase machine learning algorithm's efficiency, it is important to train it properly with valid data.

> *We could also take staff hair colour as a feature but it will prevent our algorithm to generalize data. This problem is known as overfitting. Hence why a domain expert is required to establish a good machine learning model.*

Machine Learning Algorithm Variants

Machine learning algorithms variations are based on how the algorithms learn. Working out the rules and programming them is also a hard task. This is where your technical expertise will help. Machine learning algorithms can be classified into three categories:

  1. Supervised, 2. Unsupervised and 3. Reinforcement Learning.

1. Supervised:

The algorithm is given a set of inputs (known as labelled or tagged data) and then instructed what the expected output is. In supervised algorithms, concepts are taught, data is familiarised to an extent that decisions can be made on new data.

> *Supervised learning is like performing a task, which you were taught before and you have a fairly good idea about the expected result for the given set of inputs.*

The supervised algorithm works out rules that can give expected output for the given inputs. This decision-making ability then enables the algorithm to forecast new inputs and make decisions. When new data is encountered, existing data and rules are utilised to understand and to make decisions on the data.

Supervised algorithms are expected to forecast on data.

Think Of Supervised As Guided Learning By Human.

Examples: random forests, decision trees etc.

2. Unsupervised:

Unsupervised learning algorithms are purposed to model structures, data distributions and workout the results themselves.

Inputs are given without expected outputs.

> *Unsupervised learning is like performing a task that you have not experienced before and you start the experience by gathering as much information as possible. Imagine learning a language without knowing the basics of the language.*

When new data is encountered, data is classified first and then categorized into clusters or groups. Finally, decisions are made on the new data.

Think Of Unsupervised Algorithms As Self-Taught Algorithms.

Classification information is given and then large data is fed into the algorithm so that it can classify the data into appropriate groups and then make decisions.

Unsupervised learning can be used to solve problems that are very complex in nature as the algorithms can learn to solve the problem themselves.
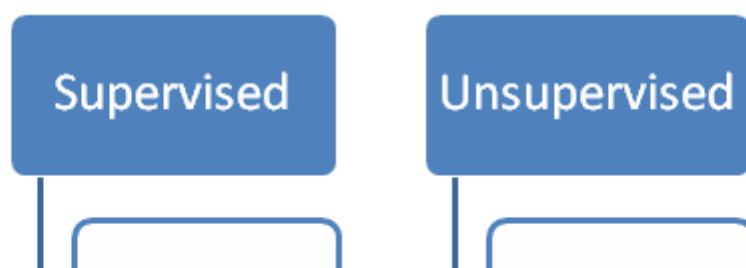
Examples: K means, clustering etc.
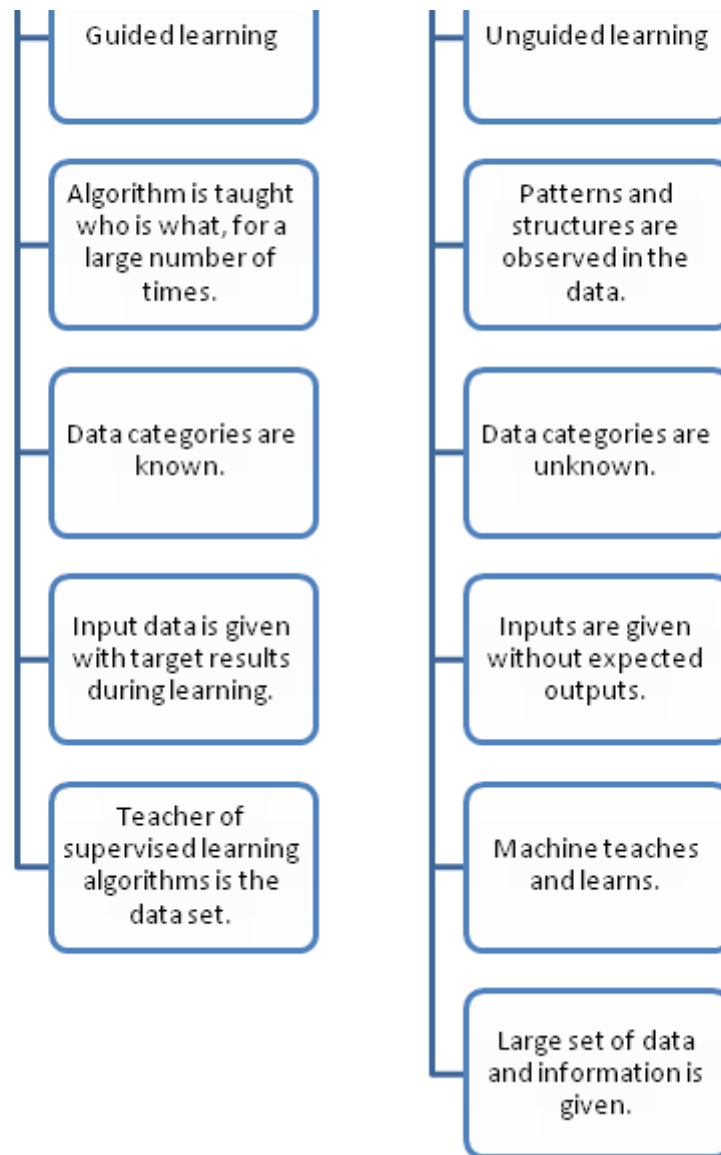
3. Reinforcement Learning:

Inspired by behaviour psychology, the algorithms are mainly used in Game theory and simulation optimisation methods. Reinforcement learning concept revolves around agents taking actions based on the reward of their previous actions.

Example: SARSA

Supervised Vs Unsupervised — A Chart

This chart shows the key characteristics of the two classifications:

Guided learning	Unguided learning
Algorithm is taught who is what, for a large number of times.	Patterns and structures are observed in the data.
Data categories are known.	Data categories are unknown.
Input data is given with target results during learning.	Inputs are given without expected outputs.
Teacher of supervised learning algorithms is the data set.	Machine teaches and learns.
	Large set of data and information is given.

Machine Learning Applications

Financial organisations have started to invest heavily in machine learning.

A number of applications now exist, for example:

- Risk Management — Applications predict credit risk and counterparties default, market data anomaly detection

- Finance — Transactions anti-fraud, financial data trend analysis, the building of exchange rates, implementation of short term interest rates, an automated trader that maximises return and minimises risk.

- Customer Service — Training employees

- Technology — Email filtering

- Health care — Detection of health problems

- Automobile — Pattern and image recognition, self-driving cars

- Telecoms — Facial recognition, security checking

Finally

This article focused on what machine learning process is and introduced how machine learning works. It also outlined the three main variations of machine learning algorithms. Finally, roles of data scientist, domain expert and technical expert were detailed.

Please let me know if you have any feedback.

Machine Learning   Artificial Intelligence   Data Science   Deep Learning   Fintech

About   Help   Legal

Get the Medium app