

Understanding Machine Learning

We'll cover the following ^

- What Is Machine Learning?
- Main Components of Machine Learning
- Machine Learning Applications

What Is Machine Learning?

Machine Learning is essentially about teaching computers to learn from data:

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.

The idea is that there are **generic algorithms** that can tell you something interesting about a set of data **without having to write any custom code specific to the problem**. Instead of writing explicit code, you **feed data** to the generic algorithm and it builds its own logic based on the data.

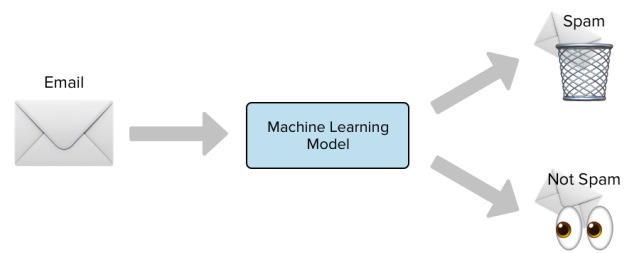
Let's say we want to recognize objects in a picture. In the old days programmers would have had to write code for every object they wanted to recognize, e.g., person, cat, vehicles. This is not a scalable approach. Today, thanks to machine learning algorithms, one system can learn to recognize both by just showing it many examples of each. For instance, the algorithm is able to understand that a cat is a cat by looking at examples of pictures labelled as "*this is a cat*" or "*this is not a cat*", and by being corrected every time it makes a wrong guess about the object in the picture. Then, if shown a series of new pictures, it begins to identify cat photos in the new set **just like a child learns** to call *a cat a cat and a dog a dog*.

This ***magic*** is possible because the system **learns based on the properties** of the object in question, a.k.a. features.

For example, while learning to distinguish between apples and oranges in a very rudimentary way, color could be used as a feature, and all the red colored fruits would then get assigned as “apple” while the ones with an orange color would get labelled as “orange”.



Your spam filter is another example of a machine learning program. There is no explicit algorithm, but given enough examples of spam and non-spam emails, the generic machine learning algorithm can automatically learn to flag spam emails. This is achieved by detecting specific patterns, e.g., occurrence of certain words and phrases in spam emails compared to non-spam examples. The greater the variety in the samples we provide to our algorithms, the easier it is to find relevant patterns and predict correct results.



Main Components of Machine Learning

Based on our examples, can you spot the three main components of machine learning?

Basically, we need three components to train our machine learning systems:

- **Data:** this is why data is being called the new oil! Data can be collected both manually and automatically. For example, users’ personal details like age and gender, all their clicks, and purchase history are valuable data for an online store. Do you recall “ReCaptcha” which forces you to “Select all the street signs”? That’s an example of some free manual labor!

Select all the street signs ! That's an example of some free manual labor!

Data is not always images; it could be tables of data with many variables (features), text, sensor recordings, sound samples etc., depending on the problem at hand.



Image Credits: xkcd.com

- **Features:** features are often also called variables or parameters. These are essentially the factors for a machine to look at — the properties of the “object” in question, e.g., users’ age, stock price, area of the rental properties, number of words in a sentence, petal length, size of the cells. Choosing meaningful features is very important. Continuing with our example of distinguishing apples from oranges, say we take bad features like ripeness and seed count. Since these are not really distinct properties of the fruits, our machine learning system won’t be able to do a good job at distinguishing between apples and oranges based on these features. Remember that it takes practice and thought to figure out what features to use as they are not always as clear as in this trivial example.
- **Algorithms:** Machine learning is based on general purpose algorithms. For example, one kind of algorithm is classification. Classification allows us to put data into different groups. The interesting thing is that the same classification algorithm used to recognize handwritten numbers could

classification algorithm used to recognize handwritten numbers could also be used to classify emails into spam and not-spam without changing

a line of code! How is this possible? Although the algorithm is the same, it's fed different input data, so it comes up with different classification logic. However, this is not meant to imply that one algorithm can be used to solve all kinds of problems! The choice of the algorithm is made based on the type of problem at hand, e.g., *are we working with predicting stock prices or do we want to assign labels like spam or not-spam?* We will learn the details in the coming sections.

The choice of the algorithm is important in determining the quality of the final machine learning model. However, one very important thing to remember is that if the data is crappy, even the best algorithm won't help. **Garbage in, garbage out is what they always say.** This is why acquiring as much data as possible is a very important first step in getting started with machine learning systems.

Machine Learning Applications

Can you think of some examples of Machine Learning that you use everyday?

Here are some popular applications:

- **Virtual Personal Assistants:** Siri, Cortana, Alexa, Google Now
- **Finance:** Fraud detection, prediction and execution of trades at speeds and volumes that humans can't compete with.
- **Social Media:** Face Recognition, People You May Know, Pages You Might Like.
- **Retail:** Product Recommendations; maximization of revenue by learning customers' habits.
- **Online customer support:** Customer support representatives are being increasingly replaced by chatbots.
- **Medicine:** Medical diagnosis, drug discovery, understanding of risk factors for diseases in large populations.
- **Search Results:** When you search on Google, the backend keeps an eye on whether you clicked on the first result or went on to the second page — the data is used to learn from mistakes so that relevant information can be found quicker next time.

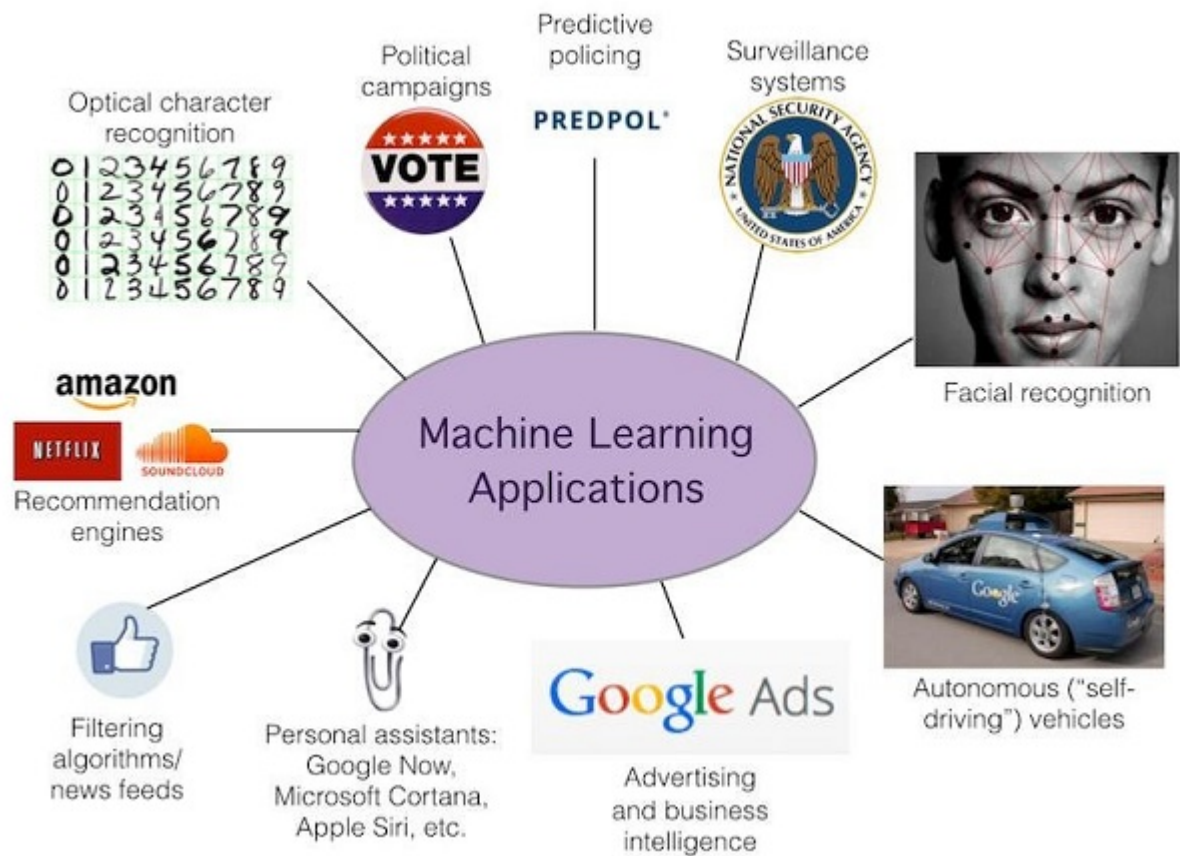


Image Credits: Introduction to Machine Learning - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Machine-Learning-Application_fig1_323108787