



[Home](#) » What is Machine Learning? A Friendly Introduction for Aspiring Data Scientists and Managers

# What is Machine Learning? A Friendly Introduction for Aspiring Data Scientists and Managers

[Article](#)

[Video Book](#)

Machine learning is ubiquitous in the industry these days. Organizations around the world are scrambling to integrate machine learning into their functions and new opportunities for aspiring data scientists are growing multifold.

But we have noticed a huge gap between what the industry needs and what's on offer right now. Quite a large number of people are not clear about what machine learning is.

By end of this page, you will not only understand what is machine learning but also its different types, its ever-growing list of applications, the latest machine learning developments, the top experts in machine learning, among various other things.

This is your one-stop destination for understanding machine learning!

## Topics we will cover

- [What is Machine Learning?](#)
- [How Do Machines Learn?](#)
- [Applications of Machine Learning in our day-to-day-life](#)
- [Why is Machine Learning Getting so Much Attention Lately?](#)
- [How is Machine Learning Different from Automation?](#)
- [What Tools are Used in Machine Learning?](#)
- [How is Machine Learning Different from Deep Learning?](#)
- [How is Machine Learning Different from Statistical Modeling?](#)
- [What Kind of Problems can we Solve using Machine Learning?](#)
- [Different Algorithms used in Machine Learning](#)
- [How Much Data is Required to Train a Machine Learning Model?](#)
- [The Steps Required to Build a Machine Learning Model](#)
- [The Latest Breakthroughs and Developments in Machine Learning](#)
- [Challenges in Adopting Machine Learning](#)
- [Is Machine Learning a Complete Black Box?](#)
- [How can you build a Career in Machine Learning?](#)
- [Preparing for Machine Learning Interviews](#)
- [List of Popular Machine Learning Books](#)
- [Top Machine Learning Researchers and Experts](#)

## What is Machine Learning?

Machine Learning is the science of teaching machines how to learn by themselves. Now, you might be thinking – why on earth would we want machines to learn by themselves? Well – it has a lot of benefits.

**“ Machines can do high-frequency repetitive tasks with high accuracy without getting bored.**

For example – the task of mopping and cleaning the floor. When a human does the task – the quality of outcome would vary. We get exhausted/bored after a few hours of work and the chances of getting sick also impact the outcome.





Depending on the place – it could also be hazardous or risky for a human.



On the other hand, if we can teach machines to detect whether the floor needs cleaning and mopping and how much cleaning is required based on the condition of the floor and the type of the floor – machines would perform the same job far better. They can go on to do that job without getting tired or sick!

This is what Machine Learning aims to do – enable machines to learn on their own. In order to answer the questions like:

- Whether the floor needs cleaning and mopping?
- How long does the floor need to be cleaned?

Machines need a way to think and this is precisely where machine learning models help. The machines capture data from the environment and feed it to the machine learning model. The model then uses this data to predict things like:

- Whether the floor needs cleaning or not, or
- For how long does it need to be cleaned, and so on.



## How do machines learn?

This video snippet from our [popular course AI & ML for Business Leaders](#) will help you understand how a machine learning model works at a high level:



Sadly, things which are usually intuitive to humans can be very difficult for machines. You only need to demonstrate cleaning and mopping to a human a few times – before they can perform it on their own.

But, that is not the case with machines. We need to collect a lot of data along with the desired outcomes in order to teach machines to perform specific tasks. **This is where machine learning comes into play.**

Machine Learning would help the machine understand the kind of cleaning, the intensity of cleaning, and duration of cleaning based on the conditions and nature of the floor.

## Applications of Machine Learning in day-to-day life

Now that you get the hang of it, you might be asking what are some of the examples of machine learning and how does it affect our life? Unless you have been living under a rock – your life is already heavily impacted by machine learning.

Let us look at a few examples where we use the outcome of machine learning already:

- Smartphones detecting faces while taking photos or unlocking themselves
- Facebook, LinkedIn or any other social media site recommending your friends and ads you might be interested in
- Amazon recommending you the products based on your browsing history
- Banks using Machine Learning to detect Fraud transactions in real-time

The applications of machine learning are immense. You can read this article for a comprehensive list of applications driven by machine learning, which we use in our day-to-day life:

- [Popular Machine Learning Applications and Use Cases in our Daily Life](#)

## Why is Machine Learning getting so much attention recently?

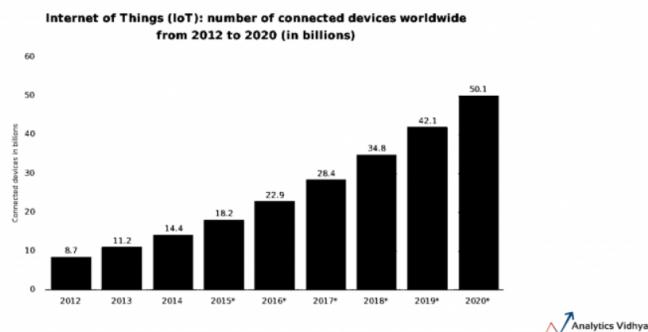
Sounds exciting! But this idea of teaching machines has been around for a while. Remember Asimov's Three Laws of robotics? Machine Learning ideas and research have been around for decades. However, there has been a lot of action and buzz recently.

The obvious question is why is this happening now when machine learning has been around for several decades?

This development is driven by a few underlying forces:

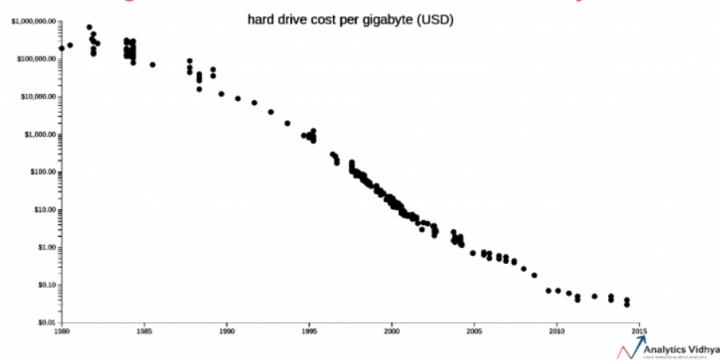
- The amount of data generation is increasing significantly with a reduction in the cost of sensors (**Force 1**)
  - Every time you take an action on any website, including Facebook and YouTube – you create data for these companies
  - All connected devices including fitness bands, smartwatches, and connected equipment are generating data

## There will be 50 Bn connected devices by 2020



- The cost of storing this data has reduced significantly (**Force 2**).

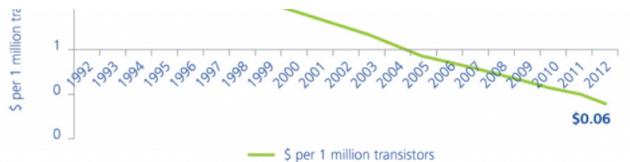
## The storage costs are a fraction of what they were



- The cost of computing has come down significantly (**Force 3**).

## The computational costs continue to fall





- Cloud has democratized Compute for the masses (**Force 4**).

These 4 forces combine to create a world where we are not only creating more data, but we can store it cheaply and run huge computations on it. This was not possible before, even though machine learning techniques and algorithms were well known.

## How is machine learning different from automation?

If you are thinking that machine learning is nothing but a new name of automation – you would be wrong.

Most of the automation which has happened in the last few decades has been rule-driven automation. For example – automating flows in our mailbox needs us to define the rules. These rules act in the same manner every time. On the other hand, machine learning helps machines learn by past data and change their decisions/performance accordingly.

Spam detection in our mailboxes is driven by machine learning. Hence, it continues to evolve with time.

“

*The only relation between the two things is that machine learning enables better automation.*

## What tools are used in Machine Learning?

There are several tools and languages being used in machine learning. The exact choice of the tool depends on your need and scale of operations. But, here are the most commonly used tools in machine learning:

- Languages:
  - R
  - Python
  - SAS
  - Julia
  - Java
  - Javascript
  - Scala
- Databases:
  - SQL
  - Oracle
  - Hadoop
- Visualization tools:
  - D3.js
  - Tableau
  - QlikView
- Other tools commonly used:
  - Excel
  - Powerpoint

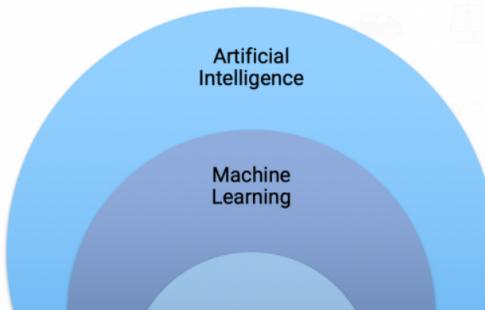
Check out the below articles expounding on a few of these popular tools (these are great for making your ultimate choice!):

- [6 Useful Programming Languages for Data Science You Should Learn \(that are not R and Python\)](#)
- [Python vs. R vs. SAS – which tool should I learn for Data Science?](#)

## How is Machine Learning Different from Deep Learning?

Ah! So you have heard about Deep Learning!

Deep learning is actually a sub-field of Machine Learning. So, if you were to represent Machine Learning and Deep Learning by a simple Venn-diagram – it will look like this:





You can read this article for a detailed deep-dive into the differences between deep learning and machine learning:

- [Deep Learning vs. Machine Learning – the essential differences you need to know!](#)

## How is Machine Learning Different from Statistical Modeling?

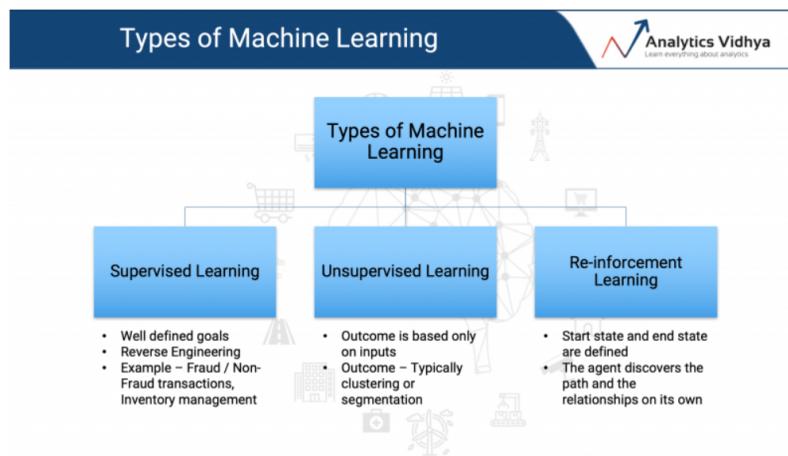
If you are thinking that machine learning and statistical thinking are the same – again you are wrong! Read this article to understand the differences between Machine Learning and Statistical Learning:

- [Difference between Machine Learning & Statistical Modeling](#)

## What are the kind of problems which can be solved using machine learning?

Machine Learning problems can be divided into 3 broad classes:

- **Supervised Machine Learning:** When you have past data with outcomes (labels in machine learning terminology) and you want to predict the outcomes for the future – you would use Supervised Machine Learning algorithms. Supervised Machine Learning problems can again be divided into 2 kinds of problems:
  - **Classification Problems:** When you want to classify outcomes into different classes. For example – whether the floor needs cleaning/mopping is a classification problem. The outcome can fall into one of the classes – Yes or No. Similarly, whether a customer would default on their loan or not is a classification problem which is of high interest to any Bank
  - **Regression Problem:** When you are interested in answering how much – these problems would fall under the Regression umbrella. For example – how much cleaning needs to be done is a Regression problem. Or what is the expected amount of default from a customer is a Regression problem
- **Unsupervised Machine Learning:** There are times when you don't want to exactly predict an Outcome. You just want to perform a segmentation or clustering. For example – a bank would want to have a segmentation of its customers to understand their behavior. This is an Unsupervised Machine Learning problem as we are not predicting any outcomes here
- **Reinforcement Learning:** Reinforcement Learning is said to be the hope of true artificial intelligence. And it is rightly said so because the potential that Reinforcement Learning possesses is immense. It is a slightly complex topic as compared to traditional machine learning but an equally crucial one for the future. [This article](#) is as good an introduction to reinforcement learning as any you will find



## What are the Different algorithms used in Machine Learning?

- Supervised Learning
  - Linear Regression
  - Logistic Regression
  - k-nearest neighbors
  - Decision Trees
  - Random Forest
  - Gradient Boosting Machines
  - XGBoost
  - Support Vector Machines (SVM)
  - Neural Networks
- Unsupervised Learning
  - k means clustering
  - Hierarchical clustering
  - Neural Network
- Reinforcement Learning

For a high-level understanding of these algorithms, you can watch this video:





For knowing more about these popular algorithms along with their codes – you can look at this article:

- [Commonly used Machine Learning Algorithms \(with Python and R Codes\)](#)

## How much data is required to train a machine learning model?

There is no simple answer to this question. It depends on the problem you are trying to solve, the cost of collecting incremental data and the benefits coming from incremental data. But here are some guidelines:

- In general – you would want to collect as much data as possible. If the cost of collecting the data is not very high – this ends up working fine
- If the cost of capturing the data is high, then you would need to do a cost-benefit analysis based on the expected benefits coming from machine learning models
- The data being captured should be representative of the behavior/environment you expect the model to work on

## What kind of data is required to train a machine learning model?

“

*Everything which you see, hear and do is data. All you need is to capture that in the right manner.*

Data is omnipresent these days. From logs on websites and smartphones to health devices – we are in a constant process of creating data. In fact, 90% of the data in this Universe has been created in the last 18 months.

Data can broadly be classified into two types:

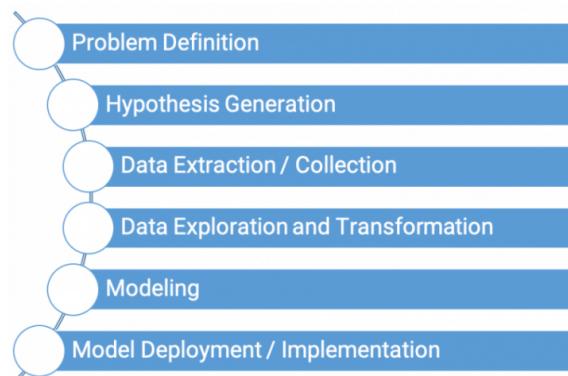
1. **Structured Data:** Structured data typically refers to data stored in a tabular format in databases in organizations. This includes data about customers, interactions with them and several other attributes, which flow through the IT infrastructure of Enterprises
2. **Unstructured Data:** Unstructured Data includes all the data which gets captured, but is not stored in the form of tables in enterprises. For example – letters of communication from customers or tweets and pictures from customers. It also includes images and voice records.

Machine Learning models can work on both Structured as well as Unstructured Data. However, you need to convert unstructured data to structured data first.

## What are the steps involved in building machine learning models?

Any machine learning model development can broadly be divided into six steps:

- **Problem definition** involves converting a Business Problem to a machine learning problem
- **Hypothesis generation** is the process of creating a possible business hypothesis and potential features for the model
- **Data Collection** requires you to collect the data for testing your hypothesis and building the model
- **Data Exploration** and cleaning helps you remove outliers, missing values and then **transform** the data into the required format
- Modeling is where you actually build the **machine learning models**
- Once built, you will **deploy the models**



## What are some of the latest achievements and developments in machine learning?

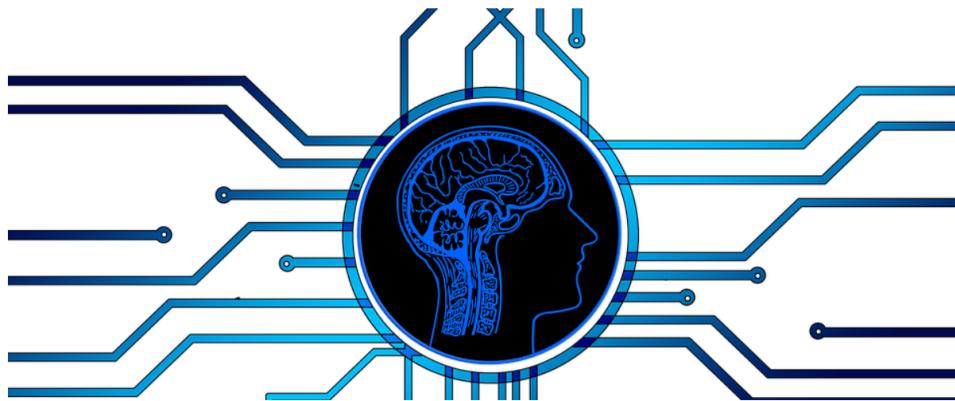
Some of the latest achievements of machine learning include:

- Winning DOTA2 against the professional players (OpenAI's development)
- Beating Lee Sidol at the traditional game of GO (Google DeepMind's algorithm)
- Google saving up to 40% of electricity in its data centers by using Machine Learning
- Writing entire essays and poetry, and creating movies from scratch using Natural Language Processing (NLP) techniques (Multiple breakthroughs, the latest being OpenAI's GPT-2)
- Creating and generating images and videos from scratch (this is both incredibly creative and worryingly accurate)
- Building automated machine learning models. This is revolutionizing the field by expanding the circle of people who can work with machine learning to include non-technical folks as well
- Building machine learning models in the browser itself! (A Google creation – TensorFlow.js)

We actually wrote a comprehensive article on the major AI and machine learning breakthroughs in the past year which everyone should go through:

- [A Technical Overview of AI & ML \(NLP, Computer Vision, Reinforcement Learning\) in 2018 & Trends for 2019](#)

## How good are the machines currently?



At the current level of technological advancements, machines are only good at doing specific tasks. A machine that has been "taught" cleaning can only do cleaning (for now). In fact, if there is a surface of new material or form which the machine has not been trained on – the machine will not be able to work on it in the same manner.

This is usually not the case with humans. So, if a person is responsible for cleaning and mopping, he/she can also be a security guard. He/she can also help in planning logistics.

This phase of artificial intelligence is typically referred to as "[Artificial Narrow Intelligence](#)".

## What are some of the Challenges in the adoption of Machine Learning?

While machine learning has made tremendous progress in the last few years, there are some big challenges that still need to be solved. It is an area of active research and I expect a lot of effort to solve these problems in the coming time.

- **Huge data required:** It takes a huge amount of data to train a model today. For example – if you want to classify Cats vs. Dogs based on images (and you don't use an existing model) – you would need the model to be trained on thousands of images. Compare that to a human – we typically explain the difference between Cat and Dog to a child by using 2 or 3 photos
- **High compute required:** As of now, machine learning and deep learning models require huge computations to achieve simple tasks (simple according to humans). This is why the use of special hardware including GPUs and TPUs is required. The cost of computations needs to come down for machine learning to make a next-level impact
- **Interpretation of models is difficult at times:** Some modeling techniques can give us high accuracy but are difficult to explain. This can leave the business owners frustrated. Imagine being a bank, but you cannot tell why you declined a loan for a customer!
- **New and better algorithms required:** Researchers are consistently looking out for new and better algorithms to address some of the problems mentioned above
- **More Data Scientists needed:** Further, since the domain has grown so quickly – there aren't many people with the skill sets required to solve the vast variety of problems. This is expected to remain so for the next few years. So, if you are thinking about building a career in machine learning – you are in good stead!

## Is Machine Learning a complete black box?

You heard me there!

No – it is not. There are methods or algorithms within machine learning which can be interpreted well. These methods can help us understand what are the significant relationships and why has the machine taken a particular decision.

On the other hand, there are certain algorithms that are difficult to interpret. With these methods, even if we achieve a very high accuracy, we may struggle with explanations.

The good thing is that depending on the application or the problem we are trying to solve – we can choose the right method. This is also a very active field of research and development.

## How can I build a career in Machine Learning?

Now you are asking the perfect questions! Given the shortage of talent in this domain, it definitely makes sense to look at building a career in data science and machine learning. But before you decide, you should keep the following things in mind:

- You would need to be comfortable with coding in order to build a career as a data scientist. Sure, there are GUI tools available – but data scientists need to code their own algorithms to be up to speed with the latest developments in the domain
- You do not need a background or a Ph.D. in mathematics. You can always pick up the things you need. If you are from this background – it helps, but it is not mandatory

- For those of you transitioning from any other domain or field – plan for at least 18 months of transition. If you get a break before – consider this as a bonus

If you are ready to build a career in data science after reading the tips above – we have a plan for you. You can check out the [FREE learning path to become a data scientist](#) by Analytics Vidhya. If you need guidance and mentorship – check out our [AI & ML Blackbelt program](#).

In addition, you can do the following:

- Take MOOCs or University Courses
  - Coursera
  - Analytics Vidhya Courses
- Attend Meetups, Webinars, Conferences
- Solve problems yourself and learn along the way
- Become part of data science communities and learn from experts

## What are the skills needed to build a career in Machine Learning?

- **Structured thinking, communication, and problem-solving**
  - This is probably the most important skill required in a data scientist. You need to take business problems and then convert them to machine learning problems. This requires putting a framework around the problem and then solving it. Check out [this course](#) to build and hone your structured thinking skills
- **Mathematics and Statistics**
  - You need mathematics and statistics to understand how the algorithms work and what are their limitations
- **Business Understanding**
  - At the end of the day, you will be solving business problems using machine learning. So, you would need to have a good understanding of the current processes, limitations, and options
- **Software Skills**
  - Data Scientists not only need to build algorithms, but they also need to code them and integrate them into the products seamlessly. That is where software skills come into play

## How can I prepare better for Data Science and Machine Learning Interviews?

Check out our awesome course "[Ace Data Science Interviews](#)" for a detailed and Structured preparation module. Here is a comprehensive guide you might want to look at as well:

- [Don't Miss this Comprehensive 7 Step Process to Ace Data Science Interviews!](#)

## Which books should I read about Machine Learning?

As a beginner with no background in machine learning, you can read the following books (the links are affiliate links):

1. [Superintelligence](#)
2. [The Master Algorithm](#)
3. [Life 3.0](#)
4. [AI Superpowers](#)
5. [Moneyball](#)
6. [Scoring Points](#)
7. [The Singularity is Near](#)

On the other hand, if you already have the required background and want to learn machine learning – these are the books you should read:

- [Must-Read Books for Beginners on Machine Learning and Artificial Intelligence](#)

## Who are the top researchers and influencers in machine learning?

As mentioned multiple times – Machine Learning is a very active field of research. From Andrew Ng to Peter Norvig, the contributions of top experts and researchers cannot be spoken about enough.

You can read this article to get a list of top researchers in the field of machine learning:

- [Heroes of Machine Learning – Top Experts and Researchers you should follow](#)

You've chosen the right career at exactly the right time. Happy learning!

This article is quite old and you might not get a prompt response from the author. We request you to post this comment on Analytics Vidhya's [Discussion portal](#) to get your queries resolved



Download App



**Analytics Vidhya**

About Us

Our Team

Careers

Contact us

**Data Science**

Blog

Hackathon

Discussions

Apply Jobs

**Companies**

Post Jobs

Trainings

Hiring Hackathons

Advertising

**Visit us**



