

# What is Machine Learning ?

A friendly introduction  
to machine learning for  
data scientists and  
managers

- KUNAL JAIN

# Contents

What is Machine Learning? - 03

How do Machines Learn? - 04

Applications of Machine Learning - 05

Why is Machine Learning getting so much attention? - 11

Machine Learning vs Automation - 12

Tools used in Machine Learning -13

Problems solved using Machine Learning -14

Algorithms used in Machine Learning -17

Steps involved in building a Machine Learning model -18

Challenges in adapting Machine Learning -19

How one can build a career in Machine Learning -20

Skills needed to build a career in Machine Learning -21



# 1. 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 the machines to learn by themselves? Well – it has a lot of benefits. Machines can do high frequency repetitive tasks with high accuracy without getting tired.

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

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 be far better in doing the same job.

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 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 whether the floor needs cleaning or not. And, for how long does it need the cleaning.



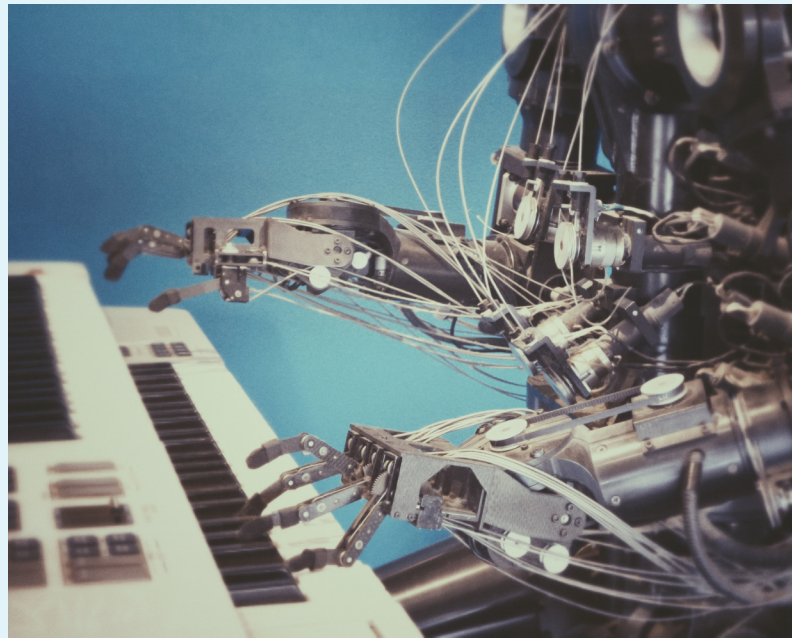
# 2. How Do Machines Learn?

Tasks difficult for humans can be very simple for machines. For example - multiplying very large numbers. On the other hand, tasks which look simple 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.



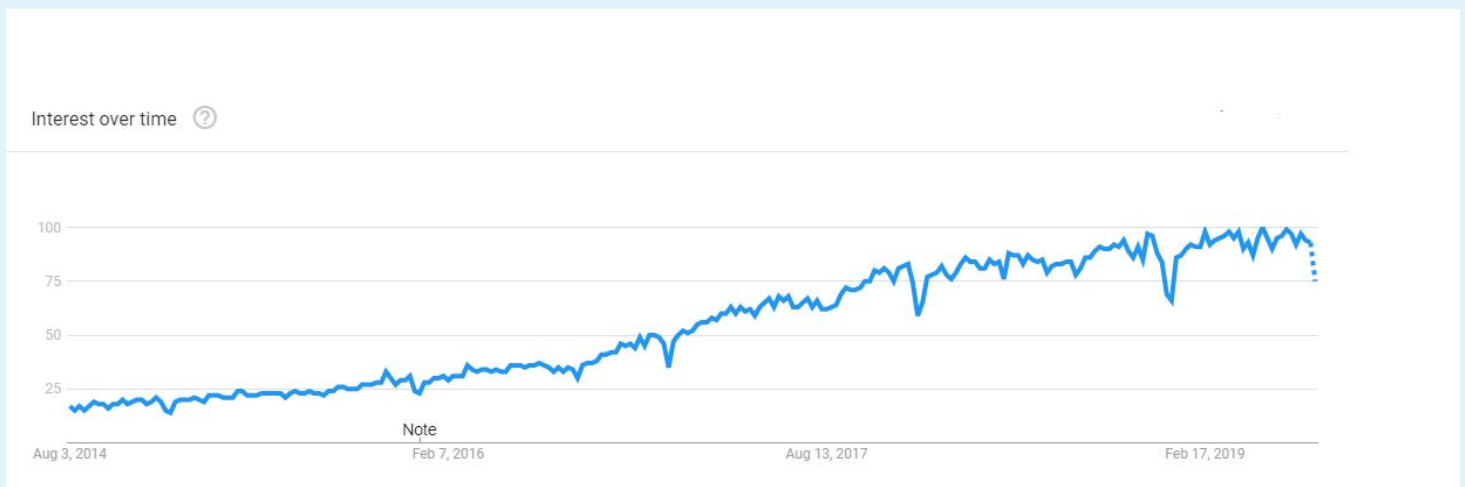
# 3. Applications of Machine Learning

Unless you have been living under a rock – your life is already heavily impacted by applications of machine learning.

Let us look at a few examples where we are already using the outcomes of machine learning:

- Smartphones detecting faces while taking photos or unlocking themselves.
- Facebook, LinkedIn or any other social media site recommending 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.

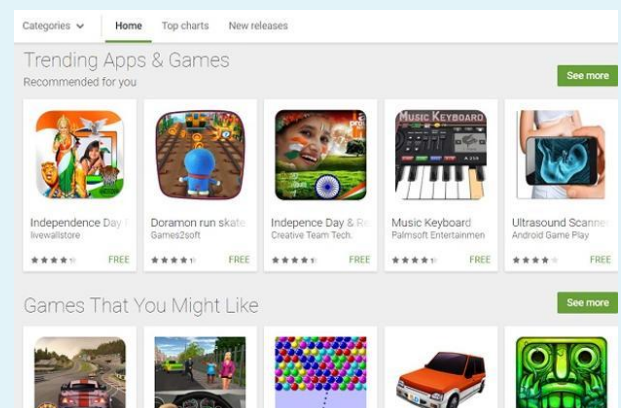
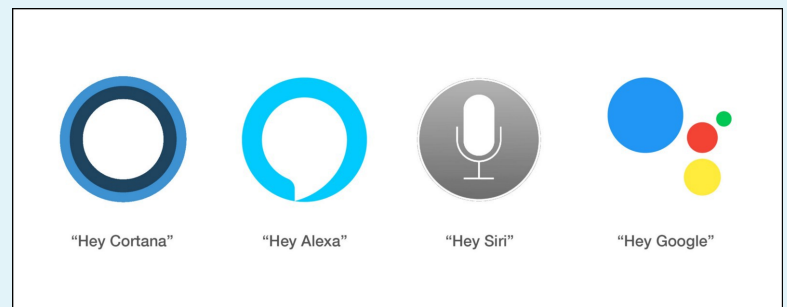
This chart below shows the increase in interest on "Machine Learning" in Google searches:



# 3.1 Machine Learning Use Case in Smart Phones

From the voice assistant that sets your alarm and finds you the best restaurants to the simple use case of unlocking your phone via facial recognition – machine learning is truly embedded in our favorite devices.

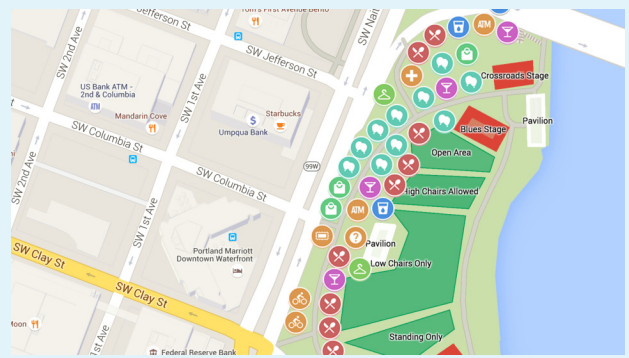
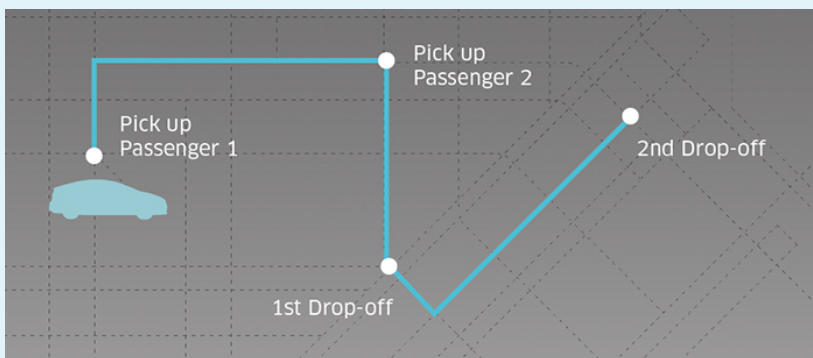
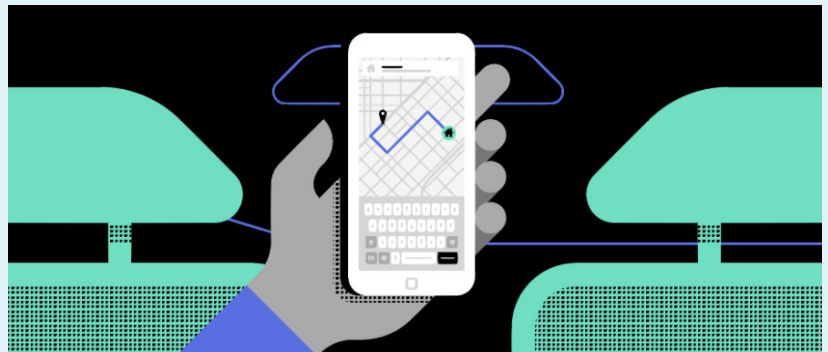
- Voice Assistants
- Smartphone Cameras
- App Store and Play Store Recommendations
- Face Unlock



# 3.2 Applications in Transportation

The application of machine learning in the transport industry has gone to an entirely different level in the last decade. This coincides with the rise of ride-hailing apps like Uber, Lyft, Ola, etc. These companies use machine learning throughout their many products, from planning optimal routes to deciding prices for the rides we take. So, let's look at a few popular use cases in transportation which use machine learning heavily.

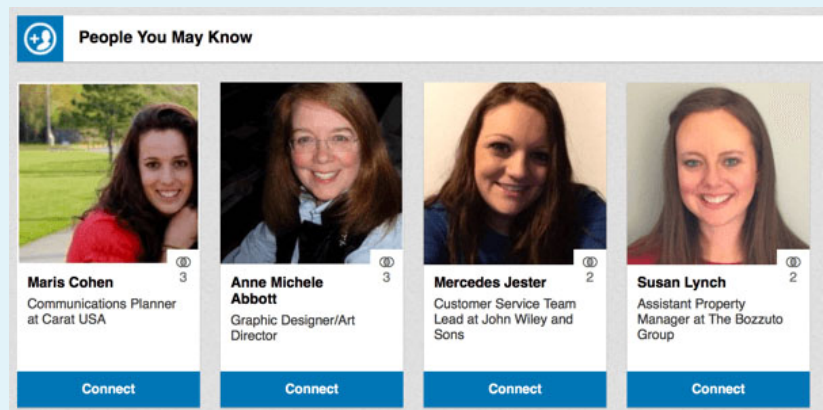
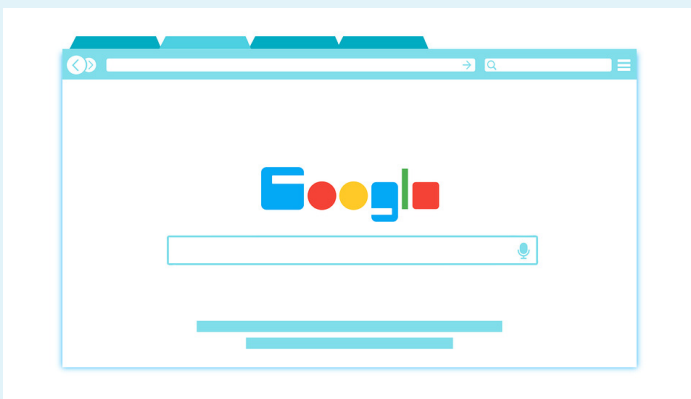
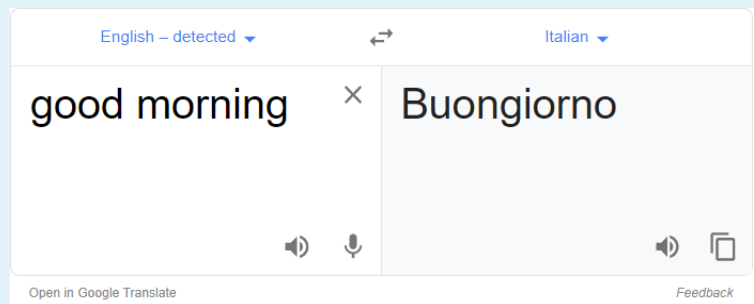
- Dynamic Pricing in Travel
- Transporting and Commuting - Uber
- Google Maps



# 3.3 Applications in Web Services

We interact with certain applications every day multiple times. What we perhaps did not realize until recently – most of these applications work thanks to the power and flexibility of machine learning.

- Email Filtering
- Google Search
- Google Translate
- Facebook and LinkedIn Recommendations





# 3.4 Machine Learning Use Cases in Sales and Marketing

Top companies in the world are using machine learning to transform their strategies from top to bottom. The two most impacted functions? Marketing and Sales!

These days if you're working in the marketing and sales field, you need to know at least one Business Intelligence tool (like Tableau or Power BI). Additionally, marketers are expected to know how to leverage machine learning in their day-to-day role to increase brand awareness, improve the bottom line, etc.

- Recommendation Engine
- Personalized Marketing
- Customer Support (Chatbots)



# 3.5 Machine Learning Use Cases in Financial Domain

Most of the jobs in machine learning are geared towards the financial domain. And that makes sense – this is the ultimate numbers field. A lot of banking institutions till recently used to lean on logistic regression (a simple machine learning algorithm) to crunch these numbers.

- Fraud Detection
- Personalized Banking



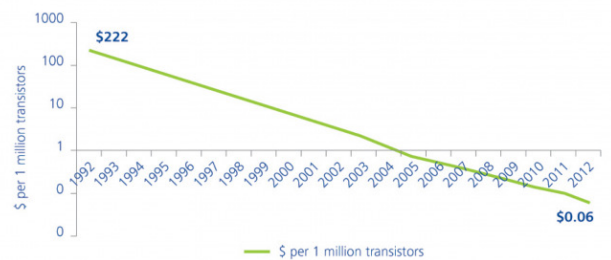
# 4. Why is Machine Learning Getting Attention Recently?

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.

This development is driven by a few underlying forces:

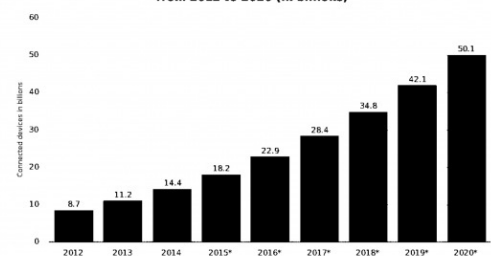
- The amount of data generation is increasing significantly with reduction in the cost of sensors
- The cost of storing this data has reduced significantly
- The cost of computing has come down significantly
- Cloud has democratized Compute for the masses

The computational costs continue to fall



There will be 50 Bn connected devices by 2020

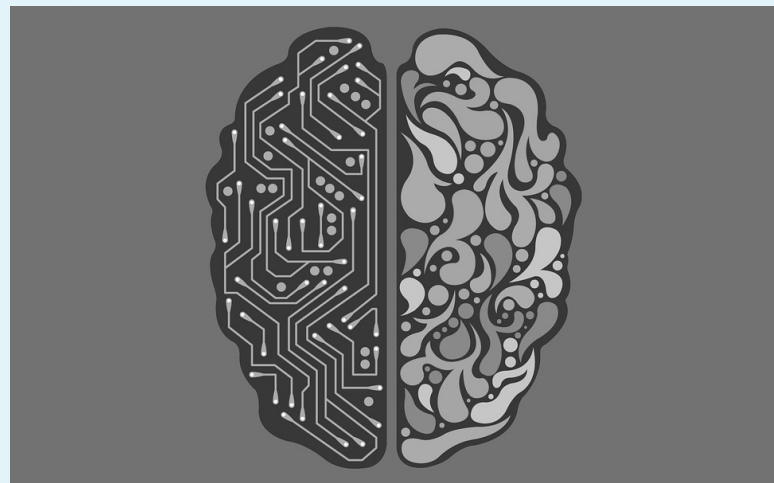
Internet of Things (IoT): number of connected devices worldwide from 2012 to 2020 (in billions)



# 5. Machine Learning v/s Automation

Machine Learning and Automation are two different things. The only way they are related is that machine learning helps us unlock more levels of automation.

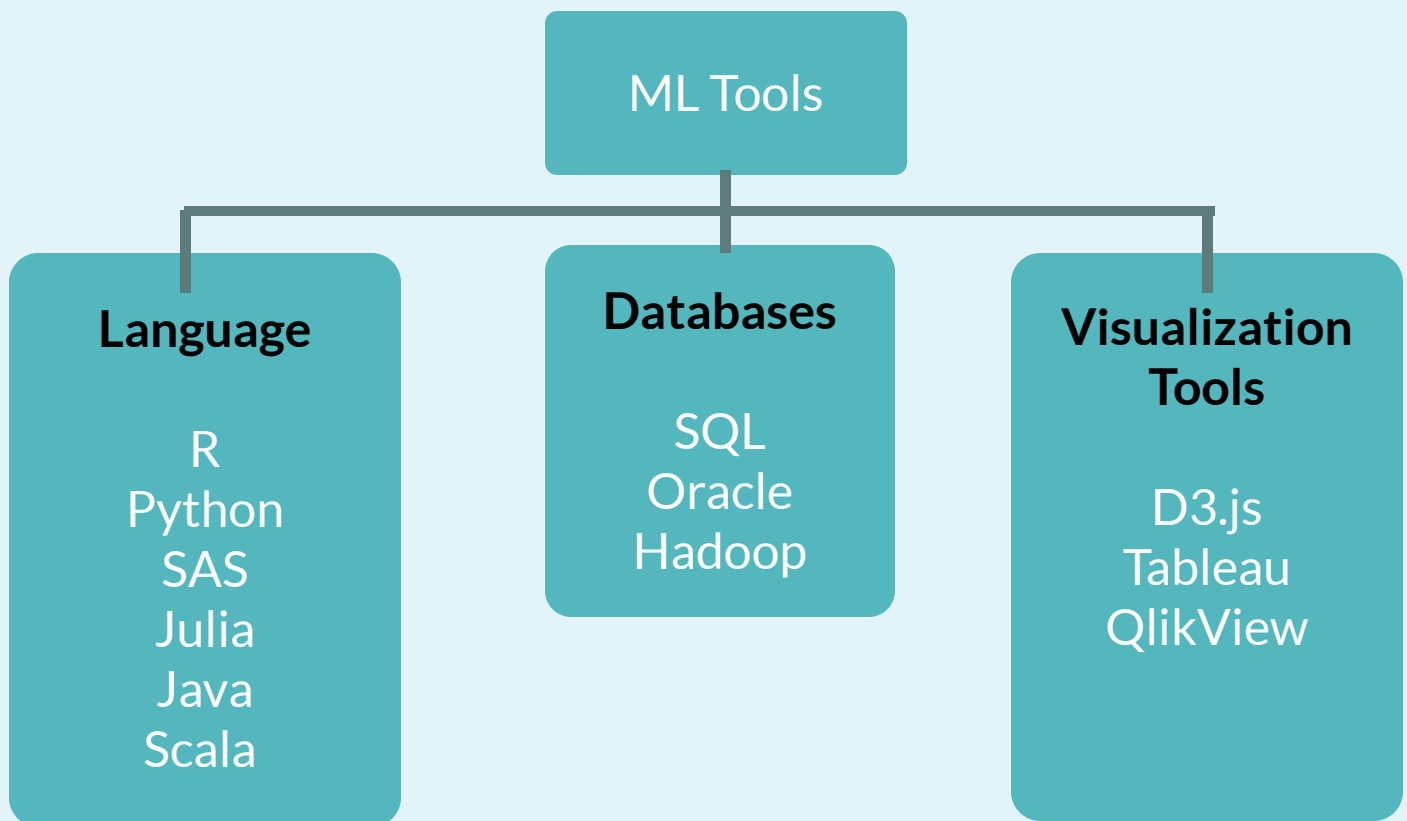
If you are thinking that machine learning is nothing but a new name for automation – you would be wrong. Most of the automation which has happened in the last few decades has been rule-driven automation. For example – automating flows in our mailbox needs us to define the rules. These rules act in the 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.





# 6. Tools Used in Machine Learning

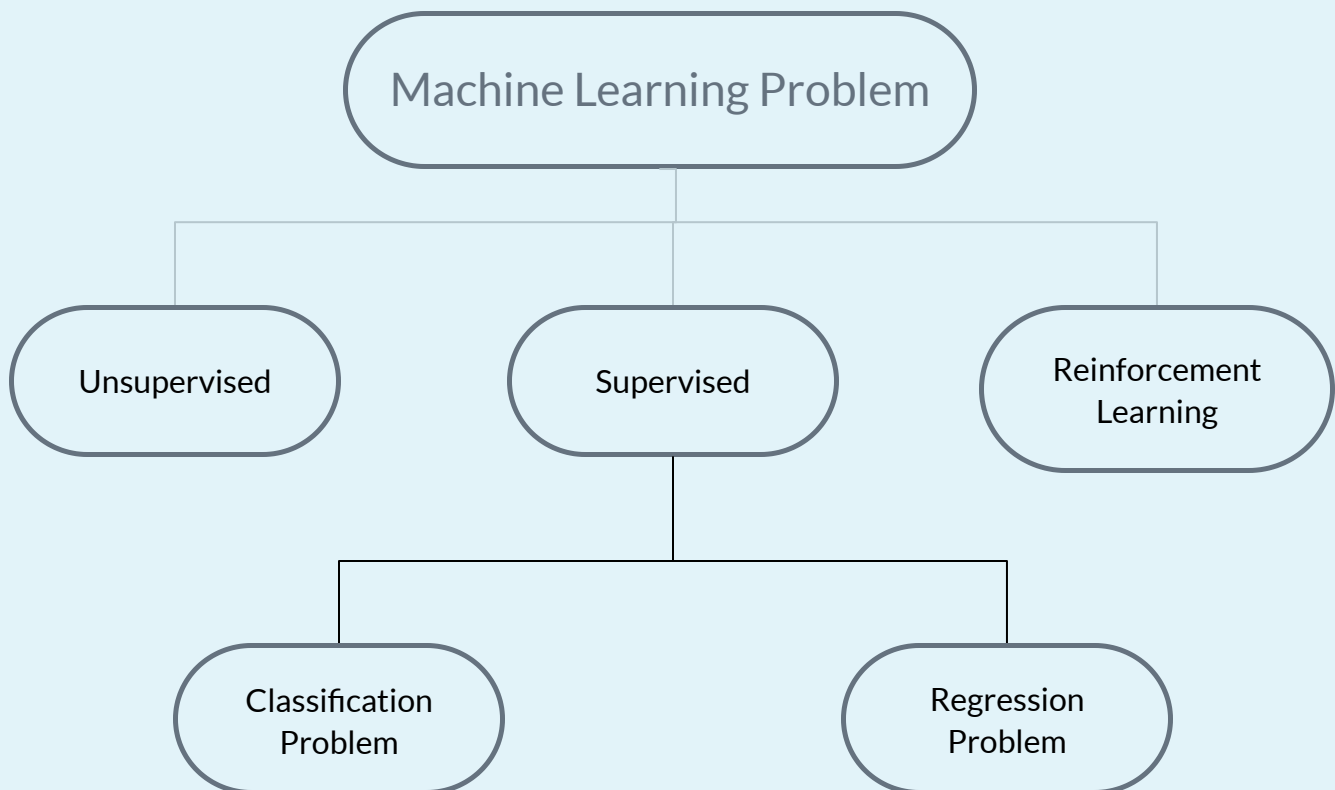
With the amount of development happening in AI & ML, there are plenty of tools coming out regularly. However, there are a few tools that are more popular than others.



# 7. Problems Solved Using Machine Learning

How many different types of problems does a human brain solve? While machines are nowhere close to that - we are moving in that direction.

Machine Learning problems can be divided into three broad classes:



# 7.1 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 be classified into 2 kinds - Classification Problems and Regression 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 be called Regression problems. 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.



# 7.2 Unsupervised Machine Learning

It is astonishing to see the capacity of the human brain. A 2-year-old child can perform Unsupervised learning in a far better manner than the most powerful computer out there!

**Unsupervised Machine Learning** - There are times when you don't want to exactly predict an outcome. You just want to perform 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.





# 8. Algorithms Used In Machine Learning

Algorithms are at the heart of machine learning. Not only do they make the whole system work - but when supplied with good data - they can pump out predictions very quickly!

## Supervised Machine Learning

- Linear Regression & Logistic Regression
- k-nearest neighbors
- Decision Trees
- Random Forest
- Gradient Boosting Machines
- XGBoost
- Support Vector Machines (SVM)
- Neural Networks

## Unsupervised Machine Learning

- k-means Clustering
- Hierarchical Clustering
- Neural Networks

## Re-inforcement Learning

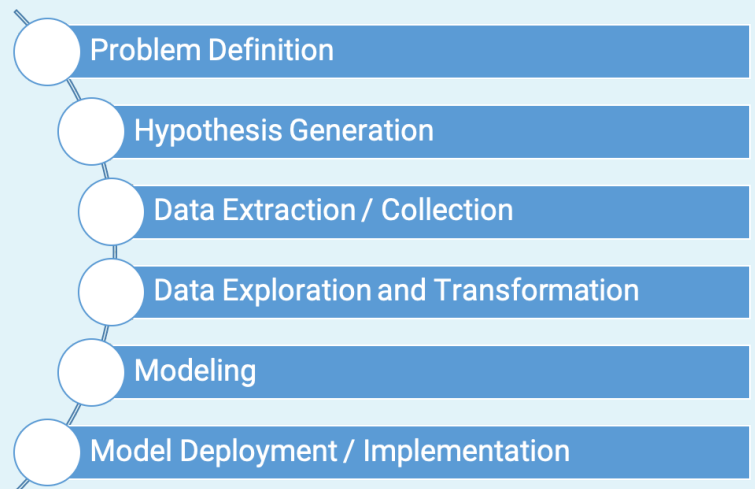
- Q-Learning
- SARSA
- DQN

# 9. Steps Involved In Building a Machine Learning Model

If you think machine learning is all about building predictive models - expect a lot of surprises! More than 80% of the time in machine learning projects goes in activities other than model building.

Any machine learning model development can be divided into 6 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 data** into the required format.
- **Modeling** is where you actually build machine learning models.
- Once built, you will **deploy the models**.



# 10. Challenges In Adapting 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 efforts 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 a Cat and a 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 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.

# 11. How Can One Build a Career In Machine Learning

Given the shortage of talent in this domain, it definitely makes sense to look at building a career in data science and machine learning.

- 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 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 these tips – we have a plan for you.

You can check out the FREE learning path to become a data scientist by Analytics Vidhya [here](#). If you need guidance and mentorship – check out our [AI & ML Blackbelt program](#).

In addition, you can do the following:

- Taking MOOCs or University Courses
- Coursera
- Analytics Vidhya Courses
- Attending Meetups, Webinars, Conferences
- Solving problems yourself and learning along the way
- Becoming part of data science communities and learning from experts



# 12. Skills Needed To Build A Career In Machine Learning

What do you need to be successful as a data scientist?

## 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.

## 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 other 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

# About The Author

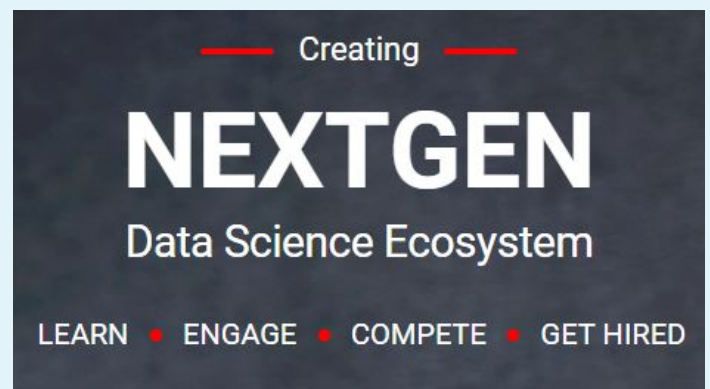


Kunal is the Founder of Analytics Vidhya.

Kunal, alumni of IIT-Bombay, is a data science evangelist and has a passion for teaching practical machine learning and data science. Before creating Analytics Vidhya, Kunal worked in Analytics and Data Science across various geographies and companies like Capital One and Aviva Life Insurance.

## Analytics Vidhya

World's Leading Data Science Community



Visit us: [courses.analyticsvidhya.com](https://courses.analyticsvidhya.com)

Follow us:

