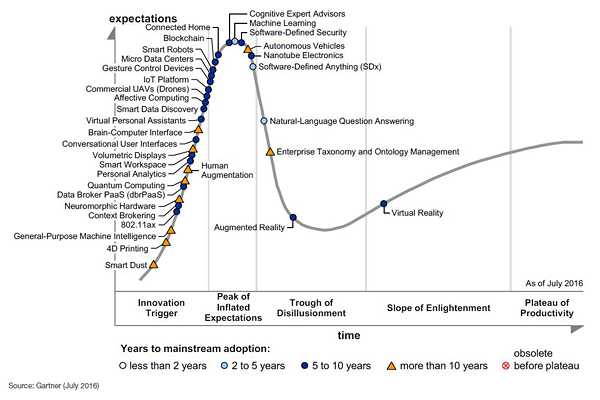
**Blog Post 1:**

**Machine Learning**

According to Garner’s [2016 Hype Cycle](http://www.gartner.com/newsroom/id/3412017) published in August, Machine Learning is at very peak.



Over the course of next few blogs, we’ll get to know about Machine Learning and few basics related to the field & how it is applicable in various industries.

What is Machine Learning?

Let’s start with this famous definition of Machine Learning

*Tom M. Mitchell: A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."*

We’ll look at an example to understand the definition. Suppose we have program that performs the function of identifying spams in email. Then,

E -> Learning about the email characteristic when it’s marked as spam/not spam. T -> Performing the task of marking email as spam/not spam.

P -> Measurement of number of correct classifications and improvements

Is it being done in real life?

You have already been a beneficiary of Machine Learning if you have used Amazon, Google, Netflix, and Facebook.

Product recommendations in Amazon.com

Google Photos - Image recognition and thereby tagging people automatically. Machine Learning programs write Facebook’s Trend headlines.

Imagine the Product recommendation that Amazon does. It will be practically impossible to code all products/categories & then check if you had bought x, recommend y. The program can’t be written with if’s & else’s. There are lots of parameters in contention towards a product recommendation. Machine Learning makes this job easier and more importantly - as accurate as possible.

Example - Industry use cases

There are common use cases that can be used across industries –

*Example: Fraud Detection – Claims submission. This applies to Insurance, Healthcare.*

*Product Recommendation - Applies to Finance and Insurance.*

*Authorization - Loan/Credit card based on credit history.*

**Blog Post 2**

**Types in Machine Learning**

Machine Learning categories can be broadly classified into following types but not limited to only these:

Supervised Learning

Unsupervised Learning

Supervised Learning

This type of Machine Learning addresses problems, where you have information (answer) about a prediction that you would like to make. You then feed that information to a Machine Learning program. The program learns from it, then predicts for new data that we supply.

These types of Machine Learning are classified into two different areas: Classification & Regression.

Classification

Classification addresses problems, which can answer - Is it A or B? Is it A, B, C or D? It predicts a certain category your prediction classifies into, hence classification problem.

*Example: Kaggle\* runs a competition which challenges developers to predict survivors of Titanic. As an initial step we are provided with a dataset where details of passengers along with Survived/Not Survived information. Then we were provided a dataset where we have new set of passengers whose status we don’t know. This problem has only two outcomes for each passenger –*

*1. Survived 2. Not Survived. These types of problems are called* ***Binary******classification*** *problems.*

\*Kaggle is a website where Machine Learning competitions are hosted. We strongly recommend you to check it out.

Regression

Regression addresses problems, which can answer - How much? How many? It predicts a certain value.

*Example: Kaggle runs a competition, which challenges developers to predict housing price in a given area. Similar to the Classification example, we are provided with a dataset with details about houses and its price. We have another dataset where we don’t have the prices, which we have to predict*

Unsupervised Learning

These types of Machine Learning addresses problems where we either don’t have a target variable to predict or our need is to find patterns/groups within data. Clustering is one type of Unsupervised learning

*Example: Market segmentation of customer based of their spending pattern. Suppose we have details of customers and their spending dollars. We can group customers based on various parameters available to us.*

**Blog Post 3:**

What is a typical workflow of a Machine Learning solution?

Key to solve a problem successfully is having a clear problem statement. Machine Learning problems also starts with such a problem statement.

Exploratory analysis of dataset. If dataset is not available, identifying right dataset is one step. We call this Training dataset.

*What happens in this phase: Each record/column is analyzed for Data Quality, Statistical fitness for Models etc., Example: Data Type, Normal distribution.*

Data Munging/Manipulation: Modifying data as per our need to solve the problem

*What happens in this phase: We typically fix the data to suit our need. Eg., Removing Outliers, Fixing Missing Values*

Feature Selection & Engineering: Selecting columns that are considered to be a good fit our model

Identifying or Preparing Test dataset: Typically we split the training dataset into Train and Test data

Model Selection & Training: Select a Machine Learning model that will helps solve our problem.

*For Eg., To solve a Regression problem, we can use Linear Regression to predict an target variable*

Evaluation of Model - Train data is fed into the Model to learn then Test data is fed to compare Predicted vs Actual. Different models have different evaluation parameters to validate.

Based on Model performance in Test data, we go back to the beginning of the workflow and refine each step for better performance.

What are the typical toolset that is available to solve problems using Machine Learning?

R & Python are popular programming languages that are available.

Azure Machine Learning and IBM Watson Analytics are cloud based tools which provides GUI based environment to solve Machine Learning Problems