



# Machine Learning

Third Eye Consulting Services & Solutions LLC.

# Introduction to Machine Learning

- ❑ Machine learning is a method of data analysis that automates analytical model building. Machine learning is a type of artificial intelligence (AI) that, according to Arthur Samuel in 1959, gives "computers the ability to learn without being explicitly programmed." Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look.

As I am a Data Scientist One may ask me a questions like

- Why you should care about Data ?
- Can't you just take a representative sample and do statistical computation on it ?

Answer Is:~

- Machine Learning
- ML focuses on learning by example - and the more example you have, the better the learner.
- It has been said that “more data usually beats better algorithms”.

# Why machine learning is important?

- ❑ Machine learning has several very practical applications that drive the kind of real business results – such as time and money savings – that have the potential to dramatically impact the future of your organization.
- ❑ Things like growing volumes and varieties of available data, computational processing that is cheaper and more powerful, and affordable data storage are main causes for the importance of machine learning.
- ❑ **Data availability:** Today, the amount of digital data generated through smart devices and Internet of Things is huge . This data can be used for analysis to make intelligent decisions and Machine Learning helps in doing so.
- ❑ **Computation power:** Moore's law has ensured that the current hardware has the capability to reliably store and analyze the massive data and perform massive amount of computations in a reasonable amount of time. This allows to build complex Machine Learning models with billions of parameters.
- ❑ Moreover it can be said that is provides High-value predictions that can guide better decisions and smart actions in real time without human intervention.

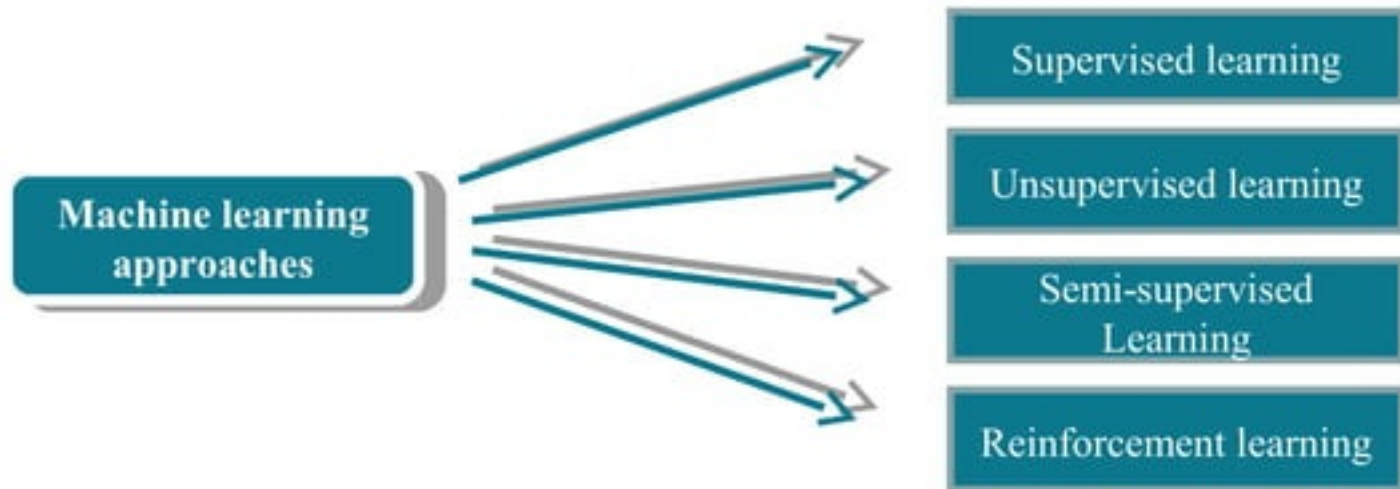
# Fields of Application

Some of the fields where machine learning is used are as follows:

- ☐ Financial services
- ☐ Government agencies
- ☐ Health care
- ☐ Marketing and sales
- ☐ Oil and gas
- ☐ Transportation
- ☐ Telecom
- ☐ Retail etc.



# Machine learning Approaches



# Supervised Learning

- ❑ This kind of a learning is possible at instances when the inputs and the outputs are clearly identified, and algorithms are trained using labeled examples.
- ❑ The learning algorithm receives a set of inputs along with the corresponding correct outputs, and the algorithm learns by comparing its actual output with correct outputs to find errors. Based on this, it would further modify the model accordingly. This is a form of pattern recognition, as supervised learning happens through methods like classification, regression, prediction and gradient boosting, supervised learning uses patterns to predict the values of the label on additional unlabeled data.
- ❑ Supervised learning is commonly used in applications where historical data predicts likely future events.

# Real life example of supervised learning

- ❑ **Loan Status Prediction in banking sector** :A Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. These details are Gender, Marital Status, Education, Income, Loan Amount, Credit History and others. To automate this process, they have given a problem to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers.

- ❑ **Skills:**

- Concordance, Information
- Value, Weight of Evidence,
- C-Statistic, H-L Stat, Gini,
- K-S, Somer's D, RMSE, CP.

- ❑ **Statistical model:**

- Logistic Model
- Decision Tree
- Random Forest etc.



# Real life example of supervised learning

## ➤ Churn Prediction:

Telecommunication market is expanding day by day and thereby due to growing competition companies are facing loss of customers and thereby a severe loss in revenue. The customers who are leaving the company and moving to the other telecom companies are called Churn.

### ❑ **Skills:**

- Concordance, Information
- Value, Weight of Evidence,
- C-Statistic, H-L Stat, Gini,
- K-S, Somer's D, RMSE, CP.

### ❑ **Statistical model:**

- Logistic Model
- Decision Tree
- Survival Analysis

# Unsupervised Learning

- ❑ Unlike supervised learning, unsupervised learning is used against data that has no historical data. The goal is to explore the data and find some structure within.
- ❑ Unsupervised learning works best on transactional data.
- ❑ Popular techniques include self-organizing maps, nearest-neighbor mapping, k-means clustering and singular value decomposition. These algorithms are also used to segment text topics, recommend items and identify data outliers.

# Real life example of unsupervised learning

## ➤ Market Basket Analysis:

Nowadays all of we are familiar with online retailers like flipkart, Amazon etc. Now what they do, they suggest some relevant products on purchase of some particular product. Identifying products and content that go well together. Using it, retailers get a window into customers' purchasing behavior.

## Statistical Skills:

Association Rule . (Support, Confidence, Lift)

## Statistical Algorithm:

Apriori Algorithm

Collaborative filtering

# Semi-supervised Learning

- ❑ As the name suggests, semi-supervised learning is a bit of both – supervised and unsupervised learning and uses both labeled and unlabeled data for training. In a typical scenario it would use small amount of labeled data with large amount of unlabeled data, the reason being that, unlabeled data is less expensive and takes less effort to acquire.
- ❑ This type of learning can be used with methods such as classification, regression and prediction.
- ❑ Semi-supervised learning is useful when the cost associated with labeling is too high to allow for a fully labeled training process.

# Real life example of semi-supervised learning

- One real world application for semi-supervised learning, is webpage classification. Say you want to classify any given webpage into one of several categories (like "Educational", "Shopping", "Forum", etc.). This is a case where it's expensive to go through tens of thousands of webpages and have humans annotate them (imagine how boring and strenuous it would be). However, in terms of availability, webpages are abundant. Simply write a Python/Java/etc. crawler, and you can collect millions of pages in a few hours.
- **In-depth analysis of product reviews in retail:** Suppose a manufacturing company want to analyze the rating and review of a certain product to get a view about the popularity of the product for improving the quality of the product or launch a better product.
- ❑ **Statistical model:**
  - Latent semantic analysis
  - Support vector machine



# Reinforcement Learning

- ❑ This is a bit similar to the traditional type of data analysis as the algorithm discovers through trial and error and decides which action results in greater rewards.
- ❑ This type of learning has three primary components: the agent (the learner or decision maker), the environment (everything the agent interacts with) and actions (what the agent can do).
- ❑ The objective is for the agent to choose actions that maximize the expected reward over a given amount of time. This is best achieved when the agent has a good policy in hand. Learning the best policy, hence remains to be the goal in reinforcement learning.

# Real life example of reinforcement learning

- **Optimization of anemia management in patients undergoing hemodialysis.**

This is a relevant problem in Nephrology, in which we focus on obtaining the optimal Erythropoietin (EPO) dosages that should be administered for an adequate longterm anemia management.

- **Optimization of a marketing campaign.**

In this case, we used data from a marketing campaign to suggest modifications based on RL to the company policy in order to maximize long-term profits.

## Industry Figure

- Global IT companies are looking into Analytics as well as ML as a next generation growth engine.
- Banking Sector and financial services globally are using ML to support their own business.
- Many Academic sector also uses Data Science, ML for statistical computing.
- ML is not a decision making system , is a decision supporting system.

## Role of a ML Expert

- Documenting the types and structure of the business data (logical modeling).
- Analyzing and mining business data to identify patterns and correlations among the various data points.
- Mapping and tracing data from system to system in order to solve a given business or system problem.
- Design and create data reports and reporting tools to help business executives in their decision making.
- Perform statistical analysis of business data.

# Limitations of Machine Learning

- ☐ Each narrow application needs to be specially trained
- ☐ Require large amounts of structured training data
- ☐ Learning must generally be supervised: Training data must be tagged
- ☐ Do not learn incrementally or interactively, in real time
- ☐ Poor transfer learning ability, re-usability of modules, and integration



# Pros and cons of machine learning

## Pros

- Good for document level
- High recall
- Robust
- Easy to scale
- Fast development
- Feature learning
- Parameter Optimization

## Cons

- Requires large annotation
- Course-grained
- Difficult to debug
- Fail in short messages
- Only shallow NLP
- Works with continuous loss function
- Limited
- Large data requirement

Thank You