



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Issue 3, March 2017

## Machine Learning: An Effective and Efficient Way to Solve Unknown Pattern Problem

Dr. Anuprita Deshmukh<sup>1</sup>, Dr. S. D. Gore<sup>2</sup>

Assistant Professor, Sinhgad College of Science, Savitribai Phule Pune University, Pune, Maharashtra, India<sup>1</sup>

Rtd. Professor. Savitribai Phule Pune University, Pune, Maharashtra, India<sup>2</sup>

**ABSTRACT:** Machine learning techniques is widely used to prepare different type of computer algorithm which is supposed to learn from the existing data its result and test it on the similar data and at end it will give predict the unknown result / pattern. The more algorithm works on the data it becomes experienced. And the more it becomes the experienced it will predict more accurate result.

**KEYWORDS:** Data Analysis, Machine Learning, supervised learning, unsupervised learning

### I. MACHINE LEARNING INTRODUCTION

Machine learning is very useful technique of finding different patterns in huge data using computer based statistical software / computer algorithms. First the collected huge and various structured or unstructured data is divided. Some part of the data taken as sample data and other remaining data is used for the testing purpose. If the result / testing are not giving expected output then the same procedure repeated. i. e. resample was collected from the data processed it again tested it on remaining data. There are many techniques / algorithms available for the processing. Once the process gives the result according to the expectation, then this process can be used to predict the future unexpected / unknown result. The process is adjusted in such a way that it will learn automatically, discover the hidden patterns in database. Preparation of such customized process is called machine learning. There are many techniques available in the machine learning which we are discussing in this paper.

In machine learning concept the computer algorithm / program is expected to be learn based upon experience on the data. The definition of machine learning is notable for its defining machine learning in fundamentally operational rather than cognitive terms, thus following Alan Turing's proposal in his paper "Computing Machinery and Intelligence", that the question "Can machines think?" be replaced with the question "Can machines do what we (as thinking entities) can do?"<sup>[5]</sup>

### II. WAYS / TYPES OF MACHINE LEARNING

Machine learning tasks are typically classified into three broad categories, depending on the nature of the learning "signal" or "feedback" available to a learning system.<sup>[5]</sup>

Different types / technique used for machine learning are given below.

- Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.<sup>[5]</sup> In supervised learning the collected data has labels and it also gives the results. From this collected data some data is used for training and some data is used for testing data. Labels mean the data has some heading on it. So there can some prediction / targeted values are given about the output. In supervised learning the every task at every level is supervised. For example, a teacher is teaching to a student, and at every level he checks what is understood by the student. If it is not up the mark the teaching methodology can be changed. The supervisor is committed to the provide target for any new input after sufficient training.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Issue 3, March 2017

- Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).<sup>[5]</sup> In this unsupervised learning the collected data does not have the any labels. i. e. based on existing data the patterns has to prepare from the data. To carry out the example given above, the student is learning on his own without the teacher.
- Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent). The program is provided feedback in terms of rewards and punishments as it navigates its problem space.<sup>[5]</sup> the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions.<sup>[7]</sup>

There are many algorithms available in the market for the machine learning technique. These algorithms are given below.

- Linear Regression

From the collected data the pattern / equation are find out. In this linear only one variable is independent other variable is dependent on it. So the relation is found out and the best fitted line draws.

- Logistic Regression

It is a classification not a regression algorithm. It is used to estimate discrete values ( Binary values like 0/1, yes/no, true/false ) based on given set of independent variable(s). In simple words, it predicts the probability of occurrence of an event by fitting data to a logit function. Hence, it is also known as **logit regression**. Since, it predicts the probability, its output values lies between 0 and 1 (as expected).<sup>[8]</sup>

- Decision Tree

If there are multiple variables (variables can be categorical as well as continuous), which are affecting the prediction. The data can be dividing in homogeneous sets. It is a type of supervised learning algorithm. And it is used for the classification problems.

- SVM (Support Vector Machine)

It is a classification method. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate.<sup>[8]</sup>

- Naive Bayes

It is used to to build mainly for the simple and very large data sets. It assumes that there are many features which are unrelated to the presence many variable.

- KNN (K – nearest neighbor)

This KNN is required very complex computational. In KNN the entire variable must be normalized. It is also used for removing outlier or noise removal.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Issue 3, March 2017

- K-Means

The K-means is used in unsupervised learning mythology. This is mainly used for the clustering problem. It uses simple and easy way to classify a data set in clusters. It selects K number of points from the clusters.

- Random Forest

In this concept the data is divided and prepare multiple trees. After the completion of this trees. All the trees are called forest. Out of this forest the sample trees are collected and based on required task that particular tree structure can be used. Every tree can be expand on its extends possible. There is no pruning.

- Dimensionality Reduction Algorithms

This is very good to build a robust model. The major task in this is to find the major / most significant variable which gives the predictions. So all the variables are studied then the one by one variable is reduced therefore at last only one variable is remaining which can be used.

- Gradient Boost & Adaboost

It is an ensemble learning algorithm which combines the prediction of several base estimators in order to improve robustness over a single estimator. It combines multiple weak or average predictors to a build strong predictor. These boosting algorithms always work well in data science competitions like Kaggle, AV Hackathon, CrowdAnalytix.<sup>[8]</sup>

### III. CONCLUSION

Machine learning is widely used to find a method used to develop complex models as well as algorithms which can be used for the prediction, analysis the data. The predicted output should be produced reliable, repeatable decisions and results. It is used to uncover unknown / hidden pattern. This prediction is used for better take business solutions.

### REFERENCES

- [1] Brandon Heung, Hung Chak Ho, Jin Zhang, Anders Knudby, Chuck E. Bulmer, Margaret Schmidt, An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping
- [2] Flavia FECHETE, Prof.univ.dr. Anisor Nedelcu, Analysis Of The Economic Performance Of A Organization Using Multiple Regression, International Conference Of Scientific Paper, Afases 2014
- [3] <http://www.sciencedirect.com/science/article/pii/S0016706115301300>
- [4] <https://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>
- [5] [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)
- [6] <https://www.quora.com/What-is-the-difference-between-supervised-and-unsupervised-learning-algorithms>
- [7] <https://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>
- [8] <https://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>