IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-4, April 2019 E-ISSN: 2347-2693

A Comparative Study of Different Machine Learning Tools

Himani Maheshwari[1*], Pooja Goswami[2], Isha Rana[2]

[1]Uttaranchal University, Dehradun, India
[2]Master of Computer Application Scholar, Uttaranchal University, Dehradun, India

Corresponding Author: himani_bahmah@yahoo.com

DOI: https://doi.org/10.26438/ijcse/v7i4.184190 | Available online at: www.ijcseonline.org

Accepted: 15/Apr/2019, Published: 30/Apr/2019

***Abstract*—** Machine Learning is the science of making computers learn and act like humans by feeding data and information without being explicitly programmed. It imbibes the philosophy of human learning, i.e. learning from expert guidance and from experience. The objective of this paper is to venture into the arena of machine learning from evolution to types of machine learning. In addition to, this paper also insights the comparison between various programming and non programming tools of machine learning.

***Keywords*—** machine learning, supervised, unsupervised, reinforcement, active, semi supervised

I. INTRODUCTION

It dates to many years ago a computer program defeated the reigning world champion in a game which is considered to need a lot of intelligence to play. The computer program IBM's Deep Blue and IBM's Watson beats human champions which attract most number of people and gave serious attention to a fast evolving field in computer science or more specifically artificial intelligence- i.e. machine learning. In today's world, machine learning is a mature technology area finding its application in almost every sphere of life. The foundation of machine learning stated in the 18th century in year 1763 when Thomas Bayes works on Bayes Theorem, on which a number of machine learning algorithm is based. In 1805, the method of least squares to solve regression problems was developed and in 1913, Andrey Markov came up with the concept of Markov chains. The evolution of machine learning from 1950 is depicted in Table 1 [1]. Tom Mitchell in 1997 quoted that an agent or machine said to learn from experience with respect to some class of tasks, and a performance measure P, if the learner's performance at tasks in the class, as measured by P, improves with experience. In broad terms, Machine Learning is a set of tools that allows us to teach computers how to perform tasks by providing examples of how they should be done. Machine Learning is a combination of computer science (programming approach), software engineering (engineering approach) and statistics (mathematical computational approach) [2, 3].
Figure 1 is a schematic representation of the machine learning process. Machine learning has been adopted by various industry domains such as banking and financial services, insurance, healthcare, life sciences, etc. to solve problems. Before starting to solve any problem using machine learning, it should be decided whether the problem is a right problem to be using it or not.

Table 1: The Evolution of Machine Learning [1]

Year	Evolution
1950	Alan Turing proposes learning machine.
1952	Arthur Samuel developed machine player checkers.
1957	Frank Rosenblatt developed first neural network program.
1967	Nearest Neighbour algorithm created.
1979	Standford University develops self driving cars.
1982	Recurrent neural network developed.
1989	Reinforcement Learning conceptualized.

1995	Random Forest and support Vector Machine algorithm developed.
1997	IBM's Deep Blue beats the world chess champion Gary Kasparov.
2006	Deep Learning conceptualized.
2010	Kaggle launched.
2011	IBM's Watson beats two human champions in Jeopardy.
2016	Google AlphaGo program beats unhandicapped professional human player.

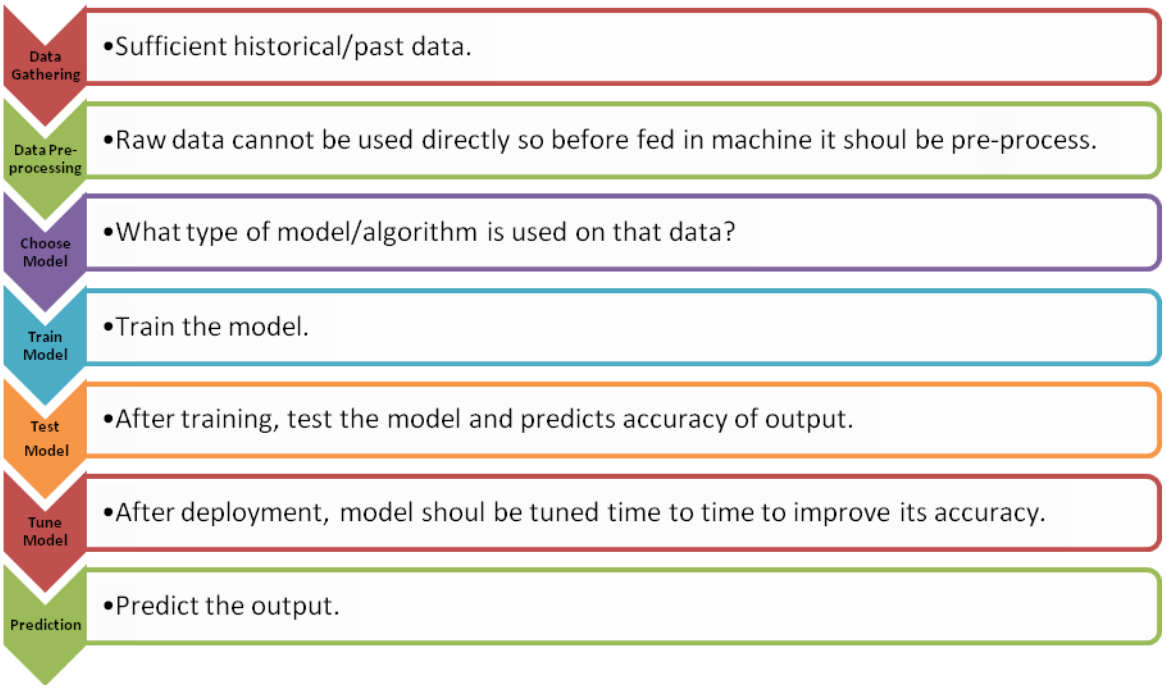

Figure 1: Machine learning process

II. TYPES OF MACHINE LEARNING

Machine Learning can be classified into three broad categories: supervised, unsupervised and reinforcement learning as depicted in Figure 2. Supervised Learning as the name suggests, works under supervision in this machine is trained with data which is well labeled and then predicts with the help of trained labeled data. Data for which target answer is known is called a labeled training data. Figure 3 is a simple depiction of the supervised learning process. Labeled training data containing past information as an input and based on this machine builds a predictive model that can be used on test data to assign a label for each record in the test data. Supervised learning used in predicting whether a tumor is malignant or benign, predicting the results of a game, classify emails as spam or non-spam, etc. In Unsupervised Learning there is no labeled training data to learn from and no prediction to be made. In this, the algorithm is trained using data that is unlabeled as input and try to find natural groupings or patterns within the data and gives the response. Therefore, unsupervised learning is also termed as descriptive model and supervised learning termed as approximation model. Figure 4 is a simple depiction of the unsupervised learning process. Reinforcement Learning is used when there is no idea about the class or label of a particular data [1, 4]. The model has to do the classification, it will get rewarded if the classification is correct, else gets punished. The model learns and updates itself through reward or punishment. It is also known as reward based learning. Figure 5 shows a simple depiction of the reinforcement learning process.

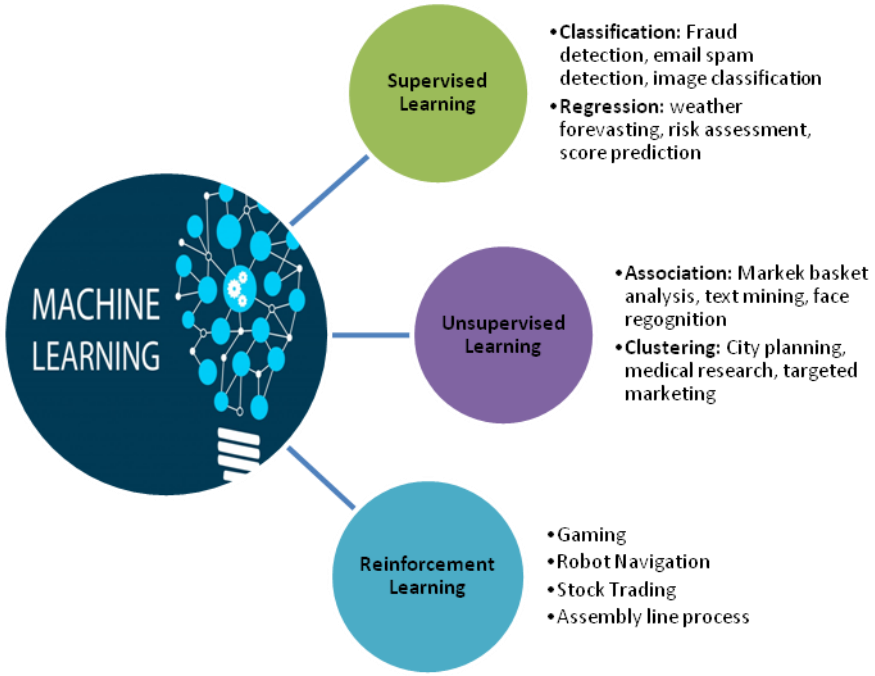

Figure 2: Types of machine learning

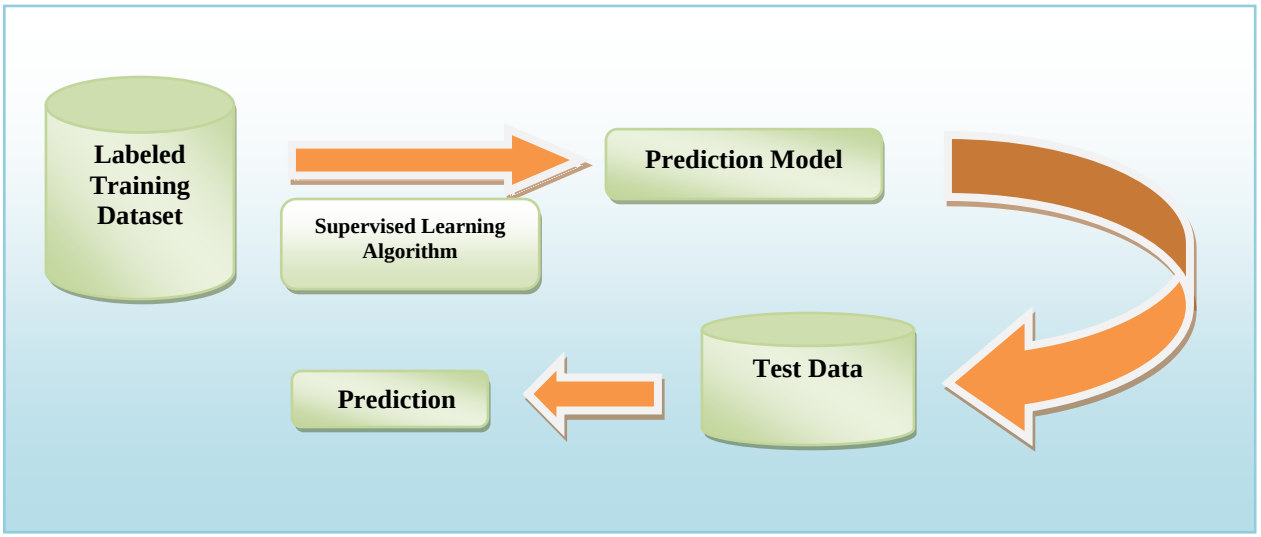

Figure 3: Supervised Learning

There are many other types of machine learning algorithm like semi-supervised learning in which only a subset of the training data is labeled, active learning in which a learning algorithm can interactively query (question) the user to obtain the desired outputs at new data points, anomaly detection learning which is used for fault detection in factories and in surveillance, instance based learning (also known as memory based learning) in which the algorithm compares new problem instances with instances already seen in training, which have been stored in memory for reference rather than performing explicit generalization and other one is representation learning which is also known as feature learning, in this extraction from the overall unlabeled data happens using a neural network [5, 6].

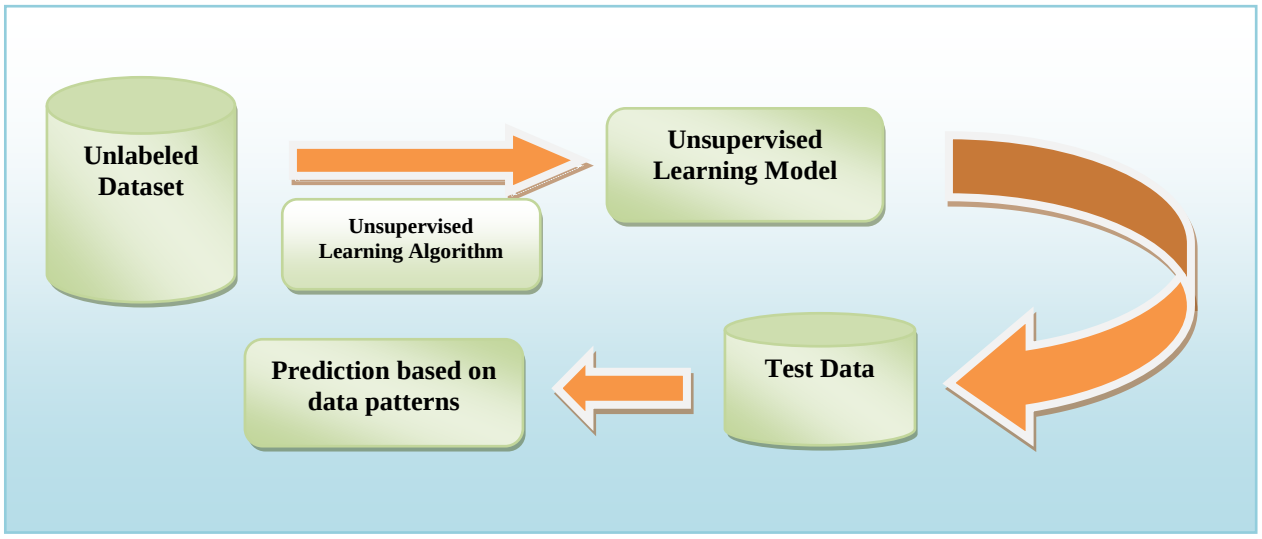

Figure 4: Unsupervised Learning

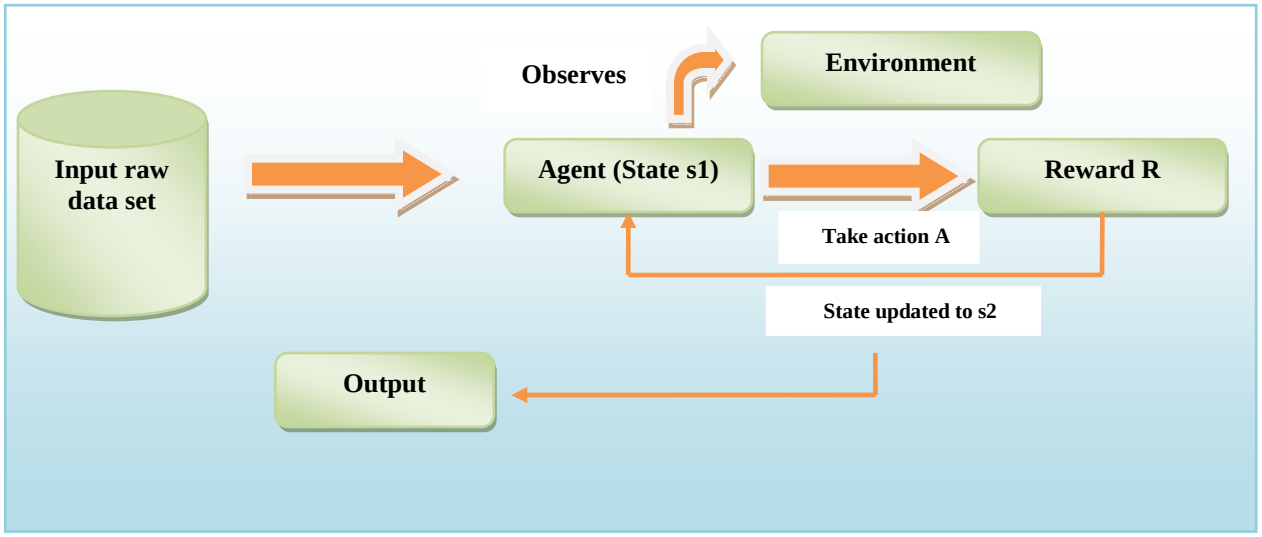

Figure 5: Reinforcement Learning

III. MACHINE LEARNING TOOLS

The algorithms related to machine learning tasks are known to all and can be implemented using any language or platform. It can be implemented using a Java platform or C/C++/Python language or in .NET. However, there are certain languages and tools which have been developed with a focus for implementing machine learning which requires intelligence as a part. Many tools for machine learning are freely available in the market for affordable access. These tools are written in certain programming languages for building of application on computers. These tools support for working with a GPU interface. This GPU feature is a very essential feature for machine learning as it provides the facility for accelerating huge matrix operations. Few of them which are most widely used are compared below in Table 2.

Table 2: Comparison of different machine learning tools [7, 9]

Tool Name	Platform	Language Used	Type	Details
Theano	Cross-Platform	Python	Library	It is a library type machine learning tool. This tool gives a benefit of accelerating the matrix operations done on the huge scale.lt has the license of 'The 3-clause BSD license. It is developed by MILA (Montreal Institute for Learning Algorithms).
Torch 7	Linux, MacOS, Windows	Lua	Framework	It is actually a scientific algorithm compilation for flexible and speedy process. Torch is a tool for machine learning which is already used in Facebook, Google, Twitter, etc.
Caffe	Linux, MacOS, Windows	C++	Framework	It was developed at University of California. It has a python interface. it is 'BSD' licensed.lt is efficient in application used for classification and image segmentation. According to sources, there are articles which tells that in 2017, it was announced by Facebook that Caffe2 being driven in market which featured Recurrent Neural Network.
Keras	Linux, MacOS, Windows	Python	Framework	It is a Python framework for deep learning. The advantage of Keras is that it uses the same Python code to run on CPU or GPU. It has been developed by François Chollet, a researcher at Google. Prominent companies like CERN, Yelp, Square or Google, Netflix, and Uber are using Keras.
TensorFlow	Linux, MacOS, Windows, Android	C++, Python, CUDA	Library	It is Google open source project and the most famous deep learning library. It was released in late 2015. TensorFlow is used in prominent organizations like Airbus, Google, and IBM and so on to produce deep learning algorithms.
Scikit-Learn	Linux, MacOS, Windows	Python	Library	It provides a number of supervised and unsupervised algorithm which can be directly implemented. It is focused on modeling (clustering, cross validation, datasets, supervised models) data and not on loading, manipulating or summarizing data.
LIBSVM	Cross-Platform	Java and C++	Library	It is BSD licensed. It is used for classification and sort of SVM. The last version is 3.23 as per July 2017 and is available in zip file or tar.gz file. It is used for supervised learning.
R	Linux, MacOS, Windows	R	Environment/ Language	It is easy to use and efficient for data ware housing and mining. It is supported by R foundation for statistical computing.

As the demand for machine learning and artificial intelligence rise, leading tech giants realized the need to give developers access to tools to build and deploy models. There are five open source machine learning tools that can use without any coding which are depicted in Table 3.

Table 3: Machine learning tools that can use without any coding [8]

Tool Name	Details
Lobe	It is an easy-to-use visual mechanism which helps users to build custom deep learning models. It automatically builds its users a custom deep learning model and starts training. User can export the trained model and ship it directly in their app. It is also used with audio files, to create music visualizes.
Google's AutoML	It a point-and-click method for generating machine learning models without any coding background. Google offers pre-trained neural networks available through APIs that can accomplish certain tasks. Google does all the complicated operations behind the scenes, so the client doesn't require comprehending anything about the complexities of neural network design. AutoML uses a simple graphical interface, enabling the user to drag in a collection of images.
Data Robot	It is capable of producing world-class imminent modeling capabilities with automated machine learning, transforming machine learning and AI projects in minutes or days instead of months without having to hire and instruct a data science team. It develops and deploys predictive patterns employing traditional methods without any prior programming knowledge. Used for predictive modeling problems.
Orange	It is developed for mining and analysis on a graphical user interface based workflow. It does not require any knowledge of programming for using it and mine data, test numbers and obtain insights. Users can accomplish tasks ranging from primary visuals to data administrations, conversions, and data mining. It supports high level visuals like silhouettes, heat-maps, geo-maps, and etc.
Konstanz Information Miner (KNIME)	It is an open-source data analytics, reporting and integration program which combines various segments for machine learning and data mining through its modular data pipelining notion. A graphical user interface and application of JDBC enables assembly of nodes blending different data sources, including preprocessing, for modeling, data analysis and visualization without, or with only minimal, programming.

IV. CONCLUSION

Machine learning is a field which is a relatively new and still evolving. Before starting to solve any problem using machine learning, it should be decided whether the problem is a right problem to be using it or not and which algorithm i.e. supervised, unsupervised or reinforcement is best suited to problem. Also, the level of research and kind of use of machine learning tools and technologies varies drastically from country to country based on researcher needs and skills.

REFERENCES

[1] Dutt S., Chandramouli S. and Das A. K. (2019). Machine Learning, Published by Pearson Indian Education Services.

[2] Jawad F., Choudhury T. U. R., Najeeb A., Fariha M., Nusrat Chamon, R., Rashedur M. (2015). Data mining techniques to analyze the reason for home birth in Bangladesh, IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, pp. 225-230.

[3] Kumar A., Alzubi J. and Nayyar A. (2018). Machine learning from theory to algorithms: An overview, Journal of Physics: Conference Series.

[4] Nithya, B. and Ilango V. (2017).Predictive analytics in health care using machine learning tools and techniques, International Conference on Intelligent Computing and Control Systems (ICICCS) (IEEE Explore).

[5] R Sujatha, S. Sree Dharinya, E P Ephzibah, R Kiruba Thangam (2019).International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-8 Issue-3, pp. 405-409.

[6] Raghavendra S, Santosh K J, Raghavendra B. K. (2019). Performance Evaluation of Machine Learning Techniques in Diabetes Prediction, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-3, pp.75-79.

[7] Web Url: https://knowm.org/machine-learning-tools-an-overview/ (Last accessed on: Mar, 2019).

[8] Web Url: https://www.analyticsindiamag.com/5-open-source-ml-tools-you-can-use-without-coding/ (Last accessed on: Mar, 2019).

[9] Web Url: https://www.guru99.com/deep-learning-libraries.html (Last accessed on: Feb, 2019).

AUTHOR PROFILE

Dr. Himani Maheshwari is an Assistant Professor in the Department of Computer Applications, Uttaranchal University, Dehradun. She received her Master of Computer Applications degree with honors from Uttar Pradesh Technical University and PhD degree from Indian Institute of Technology (IIT) Roorkee respectively. Her area of interests is Big Data, Machine Learning, AI and GIS. She has published more than 13 Research Papers in various International Journals and Conferences.