

# A Study on Machine Learning Tools

Shivani Sharma<sup>1</sup>, Sudhir Kumar Sharma<sup>2</sup>

<sup>1,2,3</sup>*Institute of Information Technology and Management, New Delhi*  
shivani.sharma9885@gmail.com, sudhir\_sharma99@yahoo.com

**Abstract** - The purpose of this paper is to present an idea of machine learning tools that are currently in use or are being studied globally. This work explains the detailed explanation of what is machine learning. This paper introduces a machine learning study by various authors. This work is summarizing the various machine learning tools and their comparisons and recent studies of machine learning tools respectively. The material presented in this paper is the result of a literature review of different research papers and books. This work compares the various tools of machine learning.

**Keywords** - Machine learning, Scikit, Weka, Tensorflow Accord.net.

## I. INTRODUCTION

Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome. Machine learning is so pervasive today that you probably use it dozens of times a day without knowing it.

There are several tools we have discuss in this paper that can let you play with data, train models, create your algorithms. You can invest your time by using the good tool instead of giving efforts on making your own. Tools can save time and helps in getting the desire result.

### A. Machine learning

Machine learning is the part of AI it is basically the ability of machine to learn from their previous knowledge and analyze automatically and progress from that information to give the best outcome. It focuses on the development of computer systems which can access and process data by own.

The learning problem is divided into three types:

1) *Unsupervised learning*: It gets a model that is similar to the origin of the installation experience and can predict what new experience to expect. The computer is trained on the raw data. Here no teacher, in fact a computer can teach you new things after learning patterns in data, these algorithms are especially useful in situations where a human expert doesn't know what to look for during data.

While the standard algorithm falls under supervised learning there are Neighborhood, Naive Bay, Decision Making, Linear Extraction, Machine Support (SVM), Neural Networks, guided reading, reinforcement learning [1].

2) *Supervised learning*: Separate which item belongs to you after seeing examples of items from each category or compression. Supervised learning algorithms for measuring the relationships and dependencies between target outputs and thus input features such as this will predict the output values of brand spanking new data that support this relationship you learned from the previous data set [1].

3) *Reinforcement learning*: Reinforcement learning can be a form of machine learning, and thus also an AI branch. This type of learning allows machines and software representatives to automatically determine the correct behavior within a specific context, thus increasing its performance. A simple reward answer is needed for the agent to find a code of conduct; this is often referred to as a sign of strength.

Learning Stress is defined by the selected type of problem, and all of its solutions are calculated as Reinforcement Learning algorithms [2].

### B. Importance of Machine learning

Reviving the interest of machine learning thanks to the same things that have made the data processing and Bayesian analysis much popular than before. Things like increasing volume and the types of data available, computer-related processing is cheaper and more powerful, and cheaper data storage.

All of the things mean that it is possible to fastly and automatically generate models which can analyze big, complex data and can deliver fast, accurate results - or to a much greater degree. And by making models, the company is investing in better opportunities - or avoiding some unknown risks [3].

### C. Why did we need machine learning tools?

Thanks to new computer technology, machine learning today is not the same as ML. It was born of pattern recognition and thus the idea that computers can also learn by own without being programmed by us to perform some specific tasks, Researchers want to know about AI they want to know if computers can learn from data. The iterative feature of ML is very important because as the models are exposed here to new data, that they are not ready to adapt independently. They learn from their previous statistics to give reliable, repeatable results and results. It is a science that is not new - but one that has received new impetus.

While many ML algorithms have been around for a long time, the realm of automating complex mathematical calculations into big data - often, quickly and quickly - is recent advances [5].

This paper is further divided into five sections. Section II, introduces literature survey on machine learning. Section III,

summarizes the various machine learning tools. Section IV, compares the various machine learning tools. Conclusions is presented in Section VI.

## II. LITERATURE SURVEY

WEKA provides implementations of learning algorithms that you can easily apply to your dataset. It also includes a variety of tools for transforming datasets, such as the algorithms for variational and sampling (Witten, Ian H., et al.,1999) [10].

Machine learning tools helps to build program capable of learning from data. Instead of creating new tools we should use production ready frameworks.

Scikit-learn is very easy to use it gives the way to learn machine behavior.

Tensorflow has a complex library which makes easy to train and implement large neural networks. (Géron, Aurélien, 2019) [11].

Mastering machine learning tools will let you play with the data, train your models, discover new methods, and create your own algorithms ( Shivashish, 2020) [14].

Machine Learning methods to statistical time series forecasting and compared the correctness of those methods with the correctness of conventional statistical methods and found that the first one is better and out top using the both measures of accuracy. They provide the reason for the accuracy of learning models is less that of statistical models and suggested some other achievable ways (Makridakis et al., 2018) [13].

The machine learning method with genetic algorithm (GA)-SVR with real value GAs, The experimental findings investigates SVR outshines the ARIMA models and BPNN regarding the base the normalized mean square error and mean absolute percentage error (Wang ,2007) [15].

## III. MACHINE LEARNING TOOLS

ML tool can be time- consuming and it helps to agree on the delivery of best results across projects. ML tools are not for using ML algorithms. They will be, but will also provide skills that can simply use at any of step within working process with a ML problem [4].

### A. Scikit

Python has become a go-to language for learning math, science, and math because of its easy accessibility and thus the wide range of libraries available in almost any application. Scikit - learn how to extend this range by building over multiple Python packages - NumPy, SciPy, and Matplotlib - for maths and science work. Used libraries are often used for workbench applications or integrated into other software and then reused. The kit is under BSD license, so then it is open and reusable [4].



Fig. 1 Scikit

### B. Accord.Net

Accord, studying machine and framework for .Net features, is an extension of the previous project to the same vein. Accord includes a group of libraries for processing audio signals and streaming images (such as videos). Its visual processing algorithms are often used for facial recognition tasks, image placement, or tracking of moving objects.

Accord also been includes libraries that has provide a standard gamut machine learning activities, from the neural networks to the decision-making organizations [4].



Fig. 2 Accord.Net

### C. H2O

H2O algorithms are designed for business processes-trick or track prediction, for example, rather than, say, image analysis. H2O can interact during stand-alone fashion through HDFS stores, over YARN, on MapReduce, or directly in the case of Amazon EC2. Hadoop mavens can use Java to communicate with H2O, but the framework also provides Python, R, and Scala bundles, allowing you to interact with all the libraries available on those platforms as well [4].

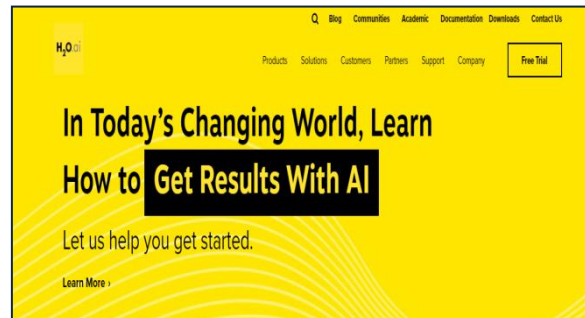


Fig. 3 H2O

#### D. Weka

It may be a set of Java machine learning algorithms designed specifically for data processing. This GNU GPLv3 licensed collection features a package program to enhance its functionality, by including both the formal and informal packages. Setup even comes with a manual to explain both of the software so the techniques are used.

While the layout is not specifically focused on Hadoop users, the most recent versions are often used with Hadoop due to a bunch of wrappers. Note that the layout no longer supports Spark, only MapReduce [4].



Fig. 4 Weka

#### E. TensorFlow

It was developed by Google itself, it is a popular and basic tool who start exploring machine learning tool or we can say that machine learning lovers must know about it.

It gives features such as:

- Building of neural networks.
- Lets you build and train your data models.
- It gives model converter named as Tensorflow.js by which its easy to convert your model that already exists.

It can be use in two methods once is by script tags or another is through install NPM [4]

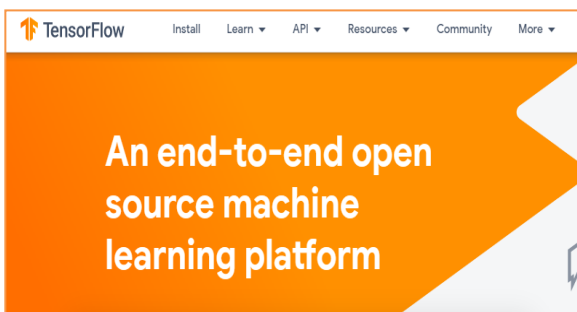


Fig. 5 TensorFlow

#### F. PyTorch

It is developed by Facebook and based on torch which is Lua based framework.

It provides neural network and optimizing algorithm features. It helps in building dynamic graphs and gives more interactive framework which makes you feel more communicative with tool [6].

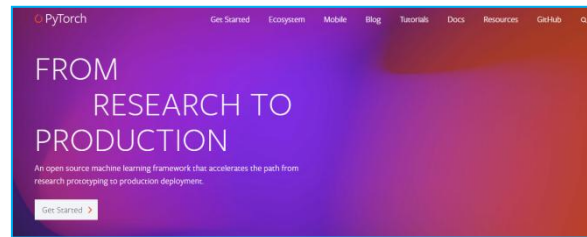


Fig. 6 PyTorch

#### G. Knime

It is a free open source software and first started development on Jan 2004 by software engineer team.

It has interactive drag and drop feature of analytics simple and robust in nature, it uses the pipeline feature to combine the data.

It provides the features such as:-

- Data mining
- Integration
- Analytics
- Machine/ deep learning
- Pipelining[6]
- 

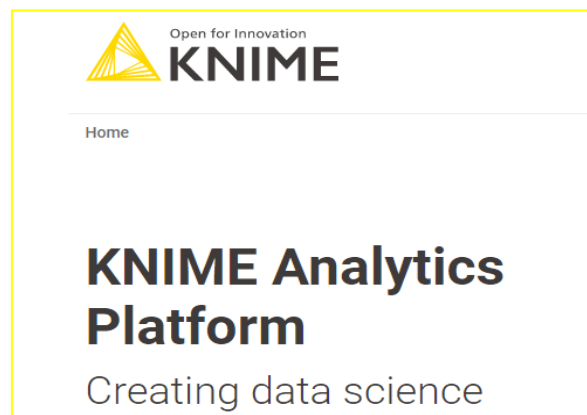


Fig. 7 Knime

#### H. Mllib

It is an Apache Spark machine learning library. Its motive is to build effective ML techniques that are not hot and precise. It can run on hadoop, mesos etc. It contains standard learning algorithms and resources, including segmentation, reorganization, integration, collaborative filtering, dimension reduction, and low-priority features and advanced pipeline APIs.[6]

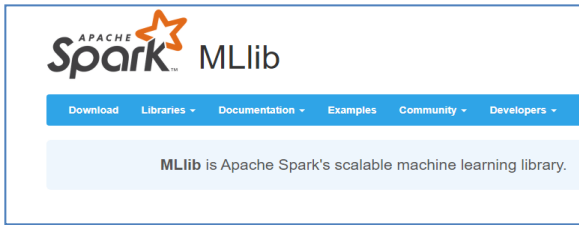


Fig. 8 MLlib

### I. Shogun

It was released in 1999, developers are Soeren Sonnenburg and Gunnar Raetsch. It is a free open source tool. It provides wide range of algorithms and data structure for learning of machine.

This tool is developed in C++ language and it provides support for different languages such as LUA, Python, Ruby, Octave, R, Java etc.

It focuses on the following:

- Regression problem
- vector machine for classification
- large scale data set

It is easy and simple to use. It gives great features for user and its libraries can be used in educational and research purpose [6].



Fig. 9 Shogun

### IV. Comparison of Machine learning Tool

This section compares nine machine learning tools. The details are presented in Table 1.

Table 1. Comparison of Machine Learning Tools

S.no.	Tools	Language	Platform	Open source
1	Scikit	Python, Cython, C, C++	Linux, Windows, Mac OS	Yes
2	Accord.net	C sharp(C#)	Cross-platform	Yes
3	H2O	Java	Linux, Windows, Mac OS	Yes

4	Weka	Java	Linux, Windows, Mac OS	Yes
5	TensorFlow	Python, CUDA, C++	Linux, Windows, Mac OS	Yes
6	PyTorch	Python, CUDA, C++	Linux, Windows, Mac OS	Yes
7	Knime	Java	Linux, Windows, Mac OS	Yes
8	MLlib	Python, Java, R	Linux, Windows, Mac OS	Yes
9	Shogun	C++	Linux, Windows, Mac OS, Unix	Yes

### V. CONCLUSION

This work presented the basic concepts of machine learning. This paper compared the nine machine learning tools widely used now a day. All tools explored in this study are open source. TensorFlow is extensively used by researchers for Deep learning algorithms. A python library known as Scikit learn is very popular among the students for making student projects. Almost all machine learning algorithms are available in the Scikit learn package. These machine learning tools can save time and help in synergy by delivering good results across projects.

### REFERENCES

- [1] IH Witten, E Frank, LE Trigg, MA Hall, G Holmes "Practical machine learning tools and techniques "
- [2] IH Witten, E Frank, MA Hall, CJ Pall "Data Mining: Practical machine learning tools and techniques"- 2016
- [3] Artur Kiulian- "How To Do Business with Artificial Intelligence."-2017
- [4] Yegulalp, Serdar. 2020. "11 Open Source Tools To Make The Most Of Machine Learning".
- [5] "Christian Hissibini – .NET Mobile Development". 2020.
- [6] Brownlee, Jason. 2020. "Machine Learning Tools".
- [7] L.-J. Li and L. Fei-Fei. What, where and who? classifying events by scene and object recognition. In Proc. ICCV, 2007.
- [8] J. Sanchez, F. Perronnin, T. Mensink, and J. Verbeek. Image classification with the fisher vector: Theory and practice. Int'l Journal of Computer Vision, 2013.
- [9] B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. Guibas, and L. Fei-Fei. Human action recognition by learning bases of action attributes and parts. In Proc. ICCV, 2011.
- [10] Witten, Ian H., et al. "Weka: Practical machine learning tools and techniques with Java implementations." (1999).
- [11] Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2019.
- [12] P. Agrawal, R. Girshick, and J. Malik. Analyzing the performance of multilayer neural networks for object recognition. In Proc. ECCV. 2014.
- [13] Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward", PLoS ONE 13, 2018
- [14] Shivashish, "Top 15 Most Used Machine Learning Tools By Experts!!", 2020
- [15] Kuan-Yu Chen, Cheng-Hua Wang, "Support vector regression with genetic algorithms in forecasting tourism demand", Tourism Management 28, pp 215–226, 2007

- [16]S.B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica. pp 249-268, 2007.
- [17]Yogesh Singh, Pradeep Kumar Bhatia & Omprakash Sangwan "A review of studies in machine learning technique". International Journal of Computer Science and Security, vol.1, pp 70 – 84, 2007.
- [18]Petersp, "The Need for Machine Learning is Everywhere" March 10, 2015.
- [19]Pedro Domingos, "A Few useful Things to Know about Machine Learning".2012.