International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 02 , February 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

# MACHINE LEARNING TOOLS-AN OVERVIEW

**V.Vinothina**

*Department of Computer Science, Kristu Jayanti College Autonomous, Bengaluru, (India)*

## ABSTRACT

*Machine learning is technique of making the computer to learn and respond without explicitly being programmed. It is becoming widespread in the field of Adaptive websites, Affective computing, Bioinformatics, Brain-machine interfaces, Cheminformatics, Classifying DNA sequences, Computational anatomy and Computer vision, including object recognition due to its enormous uses and its benefits. Due to the heavy usage of cloud by many users, it is becoming significant to apply machine learning techniques in cloud for effective utilization cloud resources and cloud services. Various algorithms and tools are emerged to support machine learning technique. In this paper, the author analyses most popular tools for machine learning techniques.*

***Keyword: Machine Learning, Cloud, Resource Management, Data Mining, Dataset.***

## I INTRODUCTION

Machine Learning deals with programming the systems to automatically learn and to improve with experience. Using machine learning technique, the machine is able to make intelligent decisions based on learning and experience. Here learning means understanding the complex patterns undergone by the system. It is too complex to describe the set of all possible decisions for the given inputs. To overcome this problem, various machine learning algorithms are developed to find knowledge from specific data and experience based on statistics, probability theory, optimization, reinforcement learning and control theory. Existed machine learning methods are applied in language processing, forecasting, pattern recognition, games, data mining, expert systems and robotics [5].

The milestone of machine learning begins from Turing Test to test ENIAC machine performance in the year 1950, checkers game-playing program in the year 1952, ELIZA a simulation of psychotherapist in 1960's, expert systems, statistical AI in 1990's and in Bigdata to find the complex pattern from huge amount of data. In the upcoming years Machine Learning will play a major role in the discovery of knowledge from the wealth of data that is currently available in a diverse amount of application areas. In cloud, machine learning algorithms are employed to predict the resource demands [6].

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

The development in machine learning benefits industries and business. In designing quadcopter machine learning is applied to explore every possible solution of and new ideas for good flying around and payload [7]. By giving constraints like good payload the computer gives new ideas that designers never examined. Autodesk uses machine learning to reimagine and redesign a new airplane cabin partition on a project with Airbus. In Robotics also machine learning pressed its print. Machine learning is the most preferable choice of 54.6 percent developers to build their robotics applications and 24.6 percent of all developers use it in their projects [8]. The Machine learning related key areas in robotics are computer learning, imitation learning, self supervised learning, Assistive and Medical technologies and Multi-Agent learning. Kinematics, Bayesian models, inverse optimal control and support vector machines are the robotics related machine learning concepts.

The machine learning approach of varying the outcome and behavior based on knowledge or observation is essential for IoT (Internet of Things). Integration of IoT and machine learning best suits in traffic routing which explores different route to destination [9]. There are many companies and government use IoT and ML algorithms for health care, traffic. Etc. With those billions of data gathered from the machines, such kind of application is highly valued. As most of the companies started using Cloud, there also machine learning plays a major role in the prediction of resource provisioning. The rest of the paper is organized as follows: section 2 describes motivation of this paper, section 3 describes machine algorithms ad tools and section 4 describes conclusion of the paper.

## II MOTIVATION

Recently, Machine learning has great scope in industry business, healthcare, transportation etc. Various researchers proposed new innovative ideas using machine learning techniques. As machine learning techniques are applied in areas like robotics, IoT and Cloud, it is becoming necessary to know about various machine learning algorithms and tools used in those areas. Akindele et.al [6] developed and evaluated cloud client prediction models for TPCW benchmark web application using three machine learning techniques: Support Vector Machine (SVM), Neural Networks (NN) and Linear Regression (LR). It provided the client with a more robust scaling decision choice. The research shown that SVV provided the best prediction model.

A system is constructed using supervised learning algorithms to detect the presence of a person in space using information from several home automation devices [10]. The model used extraction of features and labeling step used combination of heuristics to assert the presence of anyone being at home. The results shown possible paths to improve the system accuracy. Even in Healthcare also machine learning approaches are deployed. In paper [11], the authors used machine learning approaches to determine the minimal amount of resources needed to ensure minimal number of bottlenecks in the patient flow not only promotes patient satisfaction but also provides financial benefits to hospitals. The increases of data gathering by healthcare facilities in the last years have brought new opportunities

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

to apply machine learning techniques to tackle this problem. This work given better accuracy in predicting the resource usage. Most commonly users find difficulty in indentifying most relevant web pages or information according to their interests [4]. The need of using machine learning concepts in cloud is agreed by research communities [2] [3].

## III MACHINE LEARNING TOOLS

Machine learning Tools can be discussed from the perspective of Library, Graphical User Interfaces(GUI), Command Line Interfaces (CLI), Application programming interfaces (API), Local Machine Learning (LML) and Remote Machine Learning (RML) tools which are listed in table 1.

### 3.1 Machine Learning Library (ML-L)

ML library contains capabilities such as configuration data, documentation, help data, message templates, pre-written code and subroutines, classes, values or type specifications to develop a project. It may provide modeling

**Table.1 Classification of Machine Learning Tools**

| ML-L | ML-GUI | ML-CLI | ML-API | ML-LT | ML-RT |
|---|---|---|---|---|---|
| scikit | KNIME | Waffles | Pylearn2 | Shogun Library | Google Prediction API |
| JSAT | Rapid Miner | WEKA | DeepLear-ning4j | GoLearn | AWS Machine Learning |
| Accord | Orange | | LIBSVM, Sys-Weka | | Microsoft Azure Machine Learning |

algorithms to do users task. The machine learning libraries are scikit-learn, JSAT and ACCORD Frame work.

Scikit is open source machine language library for Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, *k*-means and DBSCAN. JSAT (Java Statistical analysis Tool) is java library for machine learning. This library is useful for research and specific needs as it contains many algorithms such as data transforms, predictive, tree based, vector quantization based and meta algorithms [12]. Accord is for .Net platform written in c#. It is a complete framework for building production-grade computer vision, computer audition, signal processing and statistics applications even for commercial use**.**

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Features of machine learning libraries are:

- Provides specific functionality for one or more steps in a project
- Interface Requires programming
- Customized for specific use case, problem type or environment.

### 3.2 Graphical User Interface (ML-GUI)

Machine learning tools provide a graphical user interface including windows, point and click and a focus on visualization. The benefits of a graphical user interface are:

- Allows less-technical users to work through machine learning.
- Focus on process and how to get the most from machine learning techniques.
- Structured process imposed on the user by the interface.
- Stronger focus on graphical presentations of information such as visualization.

KNIME, RapidMiner and Orange are examples of machine learning tools with a graphical interface. KNIME (Konstanz Information Miner) is an open source data analytics, reporting and integration platform. It integrates various components for machine learning and data mining through its modular data pipelining concept. A graphical user interface allows assembly of nodes for data preprocessing (ETL: Extraction, Transformation, Loading), for modeling and data analysis and visualization. KNIME has been used in pharmaceutical research, CRM customer data analysis, business intelligence and financial data analysis [13].

RapidMiner is an extension for OpenML, an open science platform for machine learning. It allows researchers to submit datasets, algorithms, workflows, experiments and their results to a single platform. From that all users can share machine learning datasets, algorithms and experiments. In order to share machine learning experiments as easily as possible, it is being integrated into various popular data mining and machine learning tools, including RapidMiner.

Orange is a data mining and machine learning suite for data analysis through Python and visual programming. It is hierarchically organized tool box and its main components branches are data management and preprocessing, classification, regression, association, ensembles, clustering, evaluation and projection [15]. Hence new functionality can be added easily. It simplifies the assembly of data analysis workflow and the designing of new data mining approaches from existing components.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

### 3.3 Command Line Interface (ML-CLI)

ML-CLI provides a command line interface including command line programs, command line parameterization and a focus on input and output. It features are as follows:

- Non programmers can work through machine learning projects.
- Ability to provide program modes for specific subtasks of machine learning project.
- Ability to store commands and command line arguments.

Waffles and WEKA are some examples of machine learning tools for a command line interface. It is collection of cross-platform command-line tools for researchers in machine learning. It automates the scripting and all functionality also available in C++ [16]. The operations supported by Waffles are classification, clustering, data transformations, evaluating the data in reduced dimensionality, training the model and visualization. It provides "Wizard" tool that guides the user through a series of forms to construct a command that will perform the desired task.

WEKA is a workbench for machine learning that is particularly used in agriculture and Horticulture domain. It provides domain specific environment and interactive tools for data manipulation, result visualization, database linkage, and cross-validation and comparison of rule sets, to complement the basic machine learning tools. Using WEKA, machine learning tools written in c, c++ and LISP can be accessed.

### 3.4 Application Programming Interface (ML-API)

ML-API provides us the flexibility to decide what elements to use and exactly how to use them within our own programs. The features of application programming interface are:

- embedding machine learning into other software projects.
- creating own machine learning tools by creating our own processes, automations and including new methods with existing methods and libraries.

Few examples for application programming interfaces are Pylearn2 for Python, Deeplearning4j for Java and LIBSVM for C. Researchers can add features as they need them A machine learning toolbox for easy scientific experimentation. All models/algorithms published by the LISA lab should have reference implementations in Pylearn2. Pylearn2 may wrap other libraries such as scikit-learn when this is practical. Pylearn2 aims to provide great flexibility and make it possible for a researcher to do almost anything. It provides Dataset interface for vector,

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

images, video, etc. It allows to reuse sub-component of Pylearn2. It also supports cross-platform serialization of learned models [18].

Deeplearning4j is the open-source, distributed deep-learning library written for Java and Scala. Integrated with Hadoop and Spark, DL4J is designed to be used in business environments on distributed GPUs and CPUs [19]. Deep learning can be applied to any data type such as sound, text, images, time series and video. It can be applied to solve almost any problem of machine perception, including classifying data, clustering it, or making predictions about it. Common use cases for deep learning include sentiment analysis, classifying images, predictive analytics, recommendation systems, anomaly detection and more.

SysWeka which extends Weka capabilities and provide a software interface for usage by higher application for managing resources on cloud systems. "Flextic" is a new job execution environment that exploits scalable static scheduling techniques. Thomas A. Henzinger et. Al [24] evaluates Flextic on top of the Amazon EC2 cloud. Also they choose jobs from different domains: gene sequencing, population genetics, machine learning, and image processing [1].

### 3.5 Local and Remote Tools (ML-LT and ML-RT)

One of the metrics to compare machine learning tools is whether the tool is local or remote. A local tool is downloadable, installable and usable locally where as remote tools are usable only on third party server. This might be a useful distinction to help us understand and choose a machine learning tool. Feature of Local tools are as follows:

- Fit for in-memory data and algorithms.
- Control over run configuration and parameterization.
- Integrate into our own systems to meet your needs

Few examples for local tools include Shogun Library for C++ and GoLearn for Go. SHOGUN is for unified large-scale learning for a broad range of feature types and learning settings. It offers a considerable number of machine learning models such as support vector machines, hidden Markov models, multiple kernel learning, linear discriminant analysis, and more [20]. SHOGUN is implemented in C++ and interfaces to MATLAB, TM R, Octave, Python, and has a stand-alone command line interface.

GoLearn is local machine learning library for GoLang/Go. It implements k-nearest-neighbours classification and regression, with some useful CSV parsing. The objective of the tools is simplicity and customisability [21].

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

A remote tool is hosted on a server and called from your local environment. These tools are often referred to as Machine Learning as a Service (MLaaS). The features are as follow:

- runs on larger datasets, across multiple systems, multiple cores and shared memory.
- supports Fewer algorithms because of the modifications required for running at scale.
- provides less control over run configuration and algorithm parameterization.
- supports remote procedure calls to integrate into our environment.

Google Prediction API, AWS Machine Learning and Microsoft Azure Machine Learning are few examples for remote tools. Google Cloud Prediction API provides an API to build Machine Learning models. Prediction's cloud-based machine learning tools can help analyze data to add various features to applications, such as customer sentiment analysis, spam detection, recommendation systems, and more [22]. It provides pattern-matching and machine learning capabilities. After it learns from training data, Prediction API can predict a numeric value or choose a category that describes a new piece of data. With these capabilities, we can create applications to perform tasks such as predicting what movies or products a user might like, categorizing emails as spam or non-spam, assessing whether posted comments have positive or negative sentiment, or guessing how much a user might spend on a given day**.**

AWS machine learning is a service which makes it easy for all developers for machine learning technology. As it automatically trains and tests a lot of complex models, tuned with different parameters, so that the best one will be chosen for the final evaluation. It will also handle all input normalization, dataset splitting and model evaluation work. As long as we provide a valid data source, AWS Machine Learning can solve most of your low-level problems [23]. Microsoft Azure machine learning is powerful cloud-based analytics tool to enable predictive maintenance. There are other tools such as Apache Mahout for Hadoop, MLlib for Spark and PredictionIO that we can use to set-up our own remote solution and integrate into our environment as a service.

## IV CONCLUSION

Machine learning is nothing but giving training to system to react/respond to various situations based on data pattern. It involves various processes of data preprocessing, classification, regression, clustering, visualization and prediction etc. In this paper, tools are used for implementing machine learning techniques are discussed. This is useful for academicians, students and researchers who keen on machine learning techniques. Based on researcher's use case, specific tool has to be chosen and implemented to achieve the better result.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

**REFERENCES**

1. Elham Hormozi, Hadi Hormozy , Mohammad Kazem Akbari and Morteza Sargolzai Javan, "Using of Machine Learning into Cloud Environment (A Survey), Seventh International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, DOI 10.1109/3PGCIC.2012.69,2012, 363-368.
2. David Burford, "Cloud computing: a brief introduction,",LAD NTERPRIZES, February 20, 2010.
3. Gurmeet Singh, "Scope of machine learning in cloud computing,", Under the guidance of Dr. Diganta Goswami, 8-11-2010.
4. A. Serengul guven smith1 and ann blandford," MLTutor: An Application of Machine Learning Algorithms for an Adaptive Web-based Information System", International Journal of Artificial Intelligence in Education, 2002.
5. http://www.mlplatform.nl/what-is-machine-learning, 2007-2011 SNN Adaptive Intelligence.
6. Akindele A. Bankole and Samuel A. Ajila," Predicting cloud resource provisioning using machine learning techniques", 2013 26th IEEE Canadian Conference on Electrical and Computer Engineering, 2013.
7. https://redshift.autodesk.com/machine-learning/
8. http://techemergence.com/machine-learning-in-robotics/
9. http://www.cs.wustl.edu/~jain/cse570-15/ftp/iot_ml/
10. Rui Madeira and LuisNunes, "A machine learning approach for indirect human presence detection using IOT devices", 2016 Eleventh International Conference on Digital Information Management (ICDIM), 2016.
11. Daniel Vieira and Jaakko Hollmen,"Resource Frequency Prediction in Healthcare: Machine Learning Approach", IEEE 29th International symposium on Computer-Based Medical Systems(CBMS), 2016.
12. https://github.com/EdwardRaff/JSAT
13. Tiwari, Abhishek; Sekhar, Arvind K.T. (October 2007). "Workflow based framework for life science informatics". Computational Biology and Chemistry. **31** (5-6): 305–319.
14. Jan N. van Rijn, Venkatesh Umaashankar, Simon Fischer, Bernd Bischl, Luis Torgo, Bo Gao, Patrick Winter, Bernd Wiswedel, Michael R. Berthold and Joaquin Vanschoren, "A RapidMiner extension for Open Machine Learning".
15. Janez Demsar et.al, "Orange: Data Mining Toolbox in Python", Journal of Machine Learning Research 14 (2013), 2349-2353.
16. Mike Gashler, "Waffles: A Machine Learning Toolkit", Journal of Machine Learning Research, 12 (2011) , 2383-2387.
17. G.Holmes, A.Donkin, I.H Witten, "WEKA: a machine learning workbench", Proceedings of Second Australian and New Zealand conferences on Intelligent Information System, 1994.
18. Pylearn2 Documentation Release dev, LISA lab, University of Montreal, 2015.
19. https://deeplearning4j.org/deeplearningforbeginners.html

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

20. S¨oren Sonnenburg et.al, “The SHOGUN Machine Learning Toolbox”, Journal of Machine Learning Research 11 (2010) ,1799-1802.
21. http://126kr.com/article/yucgkiovd
22. https://cloud.google.com/prediction/docs/
23. http://cloudacademy.com/blog/aws-machine-learning/
24. Thomas A. Henzinger, Anmol V. Singh, Vasu Singh, Thomas Wies,Damien Zufferey, “Static Scheduling in clouds”