



# Different Tools for Implementing ML Algorithms in Data Mining Tasks

W Sarada<sup>1</sup> | Dr.P.V.Kumar<sup>2</sup>

<sup>1</sup>Dept. of Computer science, RBVRR, Research Scholar, Rayalaseema University, Kurnool, India

<sup>2</sup>Dept. of Computer science and Engineering, Osmania University, Hyderabad, India

## To Cite this Article

W Sarada and Dr.P.V.Kumar, "Different Tools for Implementing ML Algorithms in Data Mining Tasks", *International Journal for Modern Trends in Science and Technology*, 6(8S): 23-27, 2020.

## Article Info

Received on 16-July-2020, Revised on 15-August-2020, Accepted on 25-August-2020, Published on 28-August-2020.

## ABSTRACT

In this paper I would like to discuss about the different tools which can be used for implementing machine learning algorithms for the tasks in data mining using java, etc. There are various tools which are available as open source or for free to use for executing or implementing various algorithms depending up on our project requirement. We are discussing about various machine learning tools here, about the features of each tool, about various algorithms related to each tool, the languages they support, the flexibility, the pros and cons of each tool, the cost if any etc. The main objective of this paper is to how efficiently and effectively we have to choose the tools for implementing or executing or deploying applications depending up on our requirement based on their features of the tools.

**KEYWORDS:** Machine learning, Data Mining, Tools, Features, Implementation

## I. INTRODUCTION

Today many of them are showing enthusiasm towards the field of machine learning and its various tools. Everybody can learn machine learning now a days whether he knows coding or not, as there are various apps like DataScience101 app for learning different ML algorithms like KNN, Linear regression, SVM and develop apps which also provides the codes for using it, Elevate app focused on the thinking skills, speed in processing, develops analytical and communicative skills based on the users previous conduct or behavior.

Lumosity app which sharpens the brains mathematical and logical thinking, NeuroNation app which improves the user's intelligence and logical thinking, Math Workout app which increases the performance in working with math's or numbers.

Qpython for improving the python language skills, Basic Statistics app for improving statistical knowledge and obtain an understanding of large datasets, probability distribution app similar to the Basic Statistics app, programming Hub app which provides manuals for various programming languages and also it is interactive and very useful for researchers, learn python app similar to Qpython app which is very useful for python learners [1].

So depending up on our skills and, where we have to explore them to upgrade ourselves we can choose the best app available for improving our skills and performance in the respective area which will be helpful in developing applications.

## **II. DIFFERENT MACHINE LEARNING TOOLS**

Here we are discussing different types of machine learning tools where one can utilize for executing the algorithms or for deploying applications. [2]For example apps like Amazon Lex for deploying chatbots; Auto-Weka which is a DM tool used for investigating by selecting and performing hyper-parameter optimization of various algorithms and specifies the most preferred one to the user.

Big ML is an ML platform which covers classification regression etc. and provides unlimited predictive applications in various fields, Data Robot is an automated ML platform which uses distributed algorithms for deploying models in a fraction of time.

Driverless AI is another similar AI automatic ML platform for difficult ML models which exports them as modules of Python and Java, Data wrapper is an open source platform where no coding or skills in designing are required which generates charts, maps and graphs for visualization.

Fussio is a database tool where you can build, manage and share information without writing a single line of code, Google Auto ML is a following of ML products which trains and deploys models based on your data and stores the data in cloud storage, IBM Watson studio provides tools and it is easy to work with, train and build models faster.

Microsoft Azure studio is a machine platform for building models irrespective of the level of the skills

and coding knowledge by importing a dataset etc., ML Jar is humanfirst platform of ML, in this first you have to upload, then build and it selects best match based on your data. Paxata is a concern where working process is simple, uses spark for organizational data, performs data exploration and cleaning using normalization, NLP etc.,

Rapid miner is an automated ML tool for analyzing common problems and identifying missing values and generates predictive models, Tableau is a BI tool where various data sources are connected where creation of graphs, charts etc. can be done and can also represent various views by applying filters, creating sets and do forecasting without any knowledge of coding skills.

Trifacta is a free software which provides inherent GUI for cleaning data and provides a summary of statistics through column by column of automatic transformations by taking data as input.

## **III. EXAMINING AND DIFFERENTIATING VARIOUS TOOLS**

In this section we are trying to compare the various ML tools based on their features, their support for algorithms, the languages and the platform they support, the cost if any.

The tools which have been compared vary in the features or the algorithms they support and the platform they are run or the language it supports.

COMPARING THE VARIOUS TOOLS BASED ON THE FEATURES, PLATFORM, LANGUAGES & COST

Sl.n o.	About the tool	Features/Algorithms	Pros/cons/cost	Platform/ language
1	Scikit Learn- is a ML development in python, provides python libraries	Assists in data mining and data analytics Gives models & algorithms Parameters can be changed while calling objects  Classification Regression Clustering Pre-processing Model Selection Dimensionality reduction.	Pros: Very easy to understand the documentation free	Linux, Mac OS, Windows  Python, Cython, C, C++
2	PyTorch-is a torch based python. ML library, a Lua based	Builds neural networks Optimizes algorithms Used on cloud platforms	Pros: Creates computational graphs	Linux, Mac OS, Windows

	computing framework, scripting language.	Distributed training, tools & libraries  Autograd Module Optim Module nn Module	Ease of use	Python, C++, CUDA
3	TensorFlow-is a JavaScript library helps in ML. Using APIs you can build and train the models and on Node.js	Implements dataflow graphs and helps in the neural network.  Can run existing models with the help of TensorFlow.js which is a model converter.	Pros: Can be used in two ways(1) script tags or (2) installing NPM  Con: Difficult to learn  Free	Linux, Mac OS, Windows  Python, C++, CUDA
4	Weka-is an ML group in university of Waikato	ML algorithms help in data mining  Provides access to SQL databases  Data preparation Classification Regression Clustering Visualization and Association rules mining.	Pros: Easy to understand Training Online courses  Free	Linux, Mac OS, Windows  Java
5 *	KNIME- a tool for data analytics, reporting and integration platform. Uses data pipelining concept, combines different components for ML and DM.	Used for BI, financial data analysis, and CRM.  Integrates the code of various progg. Languages such as C, C++, Java, Python, R, JavaScript etc.	Pros: Easy to learn, deploy and install.  Cons: Difficulty in building complicated models, visualization is limited & exporting capabilities  Free	Linux, Mac OS, Windows  Java
6	Colab-It is a free jupyter notebook, no setup required, runs entirely in cloud	Google Colab a cloud service, supports python, Can build the ML apps using the libraries of PyTorch, Keras, TensorFlow, and OpenCV  Assists in ML education & Research	Pros: Can use from our google drive.  Free	cloud service
7	Apache Mahout- To create scalable performance Apps.	Pre-processors, Regression, Clustering, Recommenders, and Distributed Linear Algebra.  Follows Distributed Linear Algebra. Java libraries for common	Pros: works for large data sets. Simple Extensible  Cons: Needs more documentation help Some algorithms	Cross-platform  Java Scala

		math operations	missing Free	
8	Shogun-provides algorithms and data structures for ML, ML libraries used for research & education	Regression Classification Clustering Support vector machines. Dimensionality reduction Online learning etc.	Pros: Easy to use, can process large data-sets. Provides good customer support. Offers good features and functionalities.  Free	Windows Linux UNIX Mac OS  C++
9	Keras.io- assists in quick research, written in python, API for neural networks	Easy and fast prototyping. It supports convolution networks. It assists recurrent networks. It supports a combination of two networks. It can be run on the CPU and GPU.	Pros: User-friendly Modular Extensible  Cons: Need TensorFlow, Theano, or CNTK. To use keras  Free	Cross-platform  Python
10	Rapid Miner-. It is a platform for ML, DL, data preparation, text mining, and predictive analytics Used for research, education and application development. It comes in three modules-studio, server, Radoop	Using GUI, assists in designing and implementing analytical workflows. Result Visualization. Model validation and optimization. Data loading & Transformation Data pre-processing & visualization.	Pros: Extensible through plugins. Easy to use. No programming skills are required.  Cons: costly  Free plan Small: \$2500 per year. Medium: \$5000 per year. Large: \$10000 per year.	Cross-platform  Java
11	Accord.Net- a framework which provides machine learning libraries for image and audio processing	Classification Regression Distribution Clustering Hypothesis Tests & Kernel Methods Image, Audio & Signal. & Vision	Pros: The source code is available from libraries and through executable installer & NuGet package manager.  Cons: supports only .Net supported languages.  Free plan Small: \$2500 per year. Medium: \$5000 per year.	Cross-platform  C#

			Large: \$10000 per year	
12	Orange- Most suitable for ML and DM	Components based software also called widgets Data visualization Reading, showing DT(Data tables),training predictors	Pros: Interactive, fast formatting, easily moved, fast decisions and comparison  Open source	Python
13	Rattle-GUI based DM tool	Generated datasets can be viewed and edited Can review the code	Pros: can extend code without restrictions Used for numerous purposes  Open source	R
14	Mlpy-Machine learning python	Provides ML methods for problems and helps in finding reasonable solution	Pros:Multi -platform  Open source	Python
15	R Tool-It is a free software environment, To perform statistical computing, DM & Graphics	Supports linear, non-linear modelling, clustering, classification, Time Based Data Analysis	Pros: used in academia, research, industrial applications  Free & Open source	C Fortran
16	Tanagra-It is an easy to use DM software and useful for research purposes	It contains some supervised learning like clustering, feature selection,	Free & open source	It is an academic project used in French-speaking universities

#### IV. CONCLUSION

In this paper, we have inquired into and discussed in detail about the machine learning tools. But choosing a tool depends up on our requirement for the algorithm, your experience level, and also the cost, whether it is available free or open

source or it is a paid tool for use. Except rapid miner most of the tools are free .More popular ones are sci-kit and pytorch

as both of them support python programming language. Tensor flow and Keras are used for neural networks.

[3] Dr.R. Shankar, Dr.S. Duraisamy "Analysis of Data Mining Tasks, Techniques, Tools, Applications and Trends", IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 20, Issue 5, Ver. II (Sep - Oct 2018), PP 12-19.

[4] <https://www.softwaretestinghelp.com/data-mining-tools/>

[5] S.D.Ghaware, A.S.Kejkar, S.M.Tondare" Data Mining: Task, Tools, Techniques and Applications",International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 10, October 2014

[6] <https://towardsdatascience.com/data-mining-tools-f701645e0f4c>

[7] <https://www.learntek.org/blog/data-mining-tools/>

[8] <https://www.tdktech.com/tech-talks/business-intelligence-data-mining-machine-learning>

#### REFERENCES

- [1] <https://analyticsindiamag.com/10-interesting-apps-for-data-scientists-to-enhance-their-skills/>
- [2] <https://analyticsindiamag.com/15-machine-learning-tools-for-ml-enthusiasts-to-hone-their-skills/>