

A Brief Study on Machine Learning Tools

^[1] J. Arockia Jeyanthi

^[1] Research Scholar, Department of Computer Science, St.Xavier's college (Autonomous and Affiliated to Manonmaniam Sundaranar University, Tirunelveli), Palayamkottai, Tamil Nadu, India.

Dr. S. Chidambaranathan

^[2] Head and Associate Professor, Department of Computer Applications, St.Xavier's college (Autonomous and Affiliated to Manonmaniam Sundaranar University, Tirunelveli), Palayamkottai, Tamil Nadu, India.

Abstract: Machine learning (ML) which is an application of artificial intelligence (AI) provides the system the ability to automatically learn and improve from experience without being programmed explicitly. ML focuses on the development of computer programs that can access data and use it to learn for themselves. As per the saying “The best trained soldiers can't fulfill their mission empty-handed”, data scientists are given their own weapons known as ML tools. These tools help the data scientist to deliver results with ease in a machine learning project. Tools are a big part of machine learning and selecting the right tool can be as important as working with the best algorithms. The objective of this paper is to take you to a closer look at popularly used machine learning tools and discover their importance and enable you to choose the types of tools among many.

Keywords: Machine Learning, Tools, Types, Right tool .

I. INTRODUCTION

Machine Learning (ML) has emerged as the most important technology of the 21st century. For building your very own machine learning model you will take a look at some of the highly popular software among the many prolific algorithms that can be used for designing machine learning solutions. The learning process starts with observations or data, such as examples, direct experience, or instruction, so as to look for patterns in data and make enhanced decisions in the future based on the examples that we give. The main aim is to allow the systems to learn by itself without human intervention or support and adjust the actions accordingly. A big part of machine learning are tools and choosing the right tool can be as important as working with the best algorithms. **In this paper you will take**

a closer look at machine learning tools and discover their importance and their types that you could select from.

II. WHY USE TOOLS

The need for the usage of ML tools are described below.

- Faster:** In the applied machine learning process, good tools can automate every step. This means that the time taken from ideas to results is critically reduced. The other approach is that you have to implement each capability yourself from scratch. This can take much time than selecting a tool to use off the shelf.
- Easier:** Instead of researching and implementing techniques you can spend your time choosing the good tools. The other approach is that you have to be an expert in each step of the process in order to implement it. To ensure it is implemented efficiently it requires research, deeper exercise in order to understand the techniques, and

a higher level of engineering.

- **Fun:** For beginners to get good results there is a lower barrier. To get better results or work on more projects you can use the extra time. On the other hand rather than on getting results you will spend most of your time building your tools.

III. TOOLS WITH A PURPOSE

ML tools must serve a solid purpose. These tools facilities to deliver results in a machine learning project. ML tools are not just implementations of ML algorithms. At any step in the process of working through a machine learning problem the tools provide capabilities that you can use.

IV. WHEN TO USE MACHINE LEARNING TOOLS

Machine learning tools helps you to save time and enable consistency to deliver good results across projects. Some examples of using machine learning tools include:

- **Getting Starting:** At your start machine learning tools guide you through the process of delivering worthy results quickly and make you confident to carry on with your following project.
- **Day-to-Day:** Machine learning tools can allow you to focus on the specifics of your problem quickly rather than on the depths of the techniques you need to use to get an answer.
- **Project Work:** For a large project, these tools help you to prototype a solution, figure out the requirements and provide you a template for the system that you may want to implement.

V. 10+ MOST POPULAR MACHINE LEARNING SOFTWARE TOOLS

Among the several Machine Learning Softwares available in the market the most popular ones among them are enlisted below.

1. *Scikit-learn* :

It is a free machine learning library for Python. It provides support to Python numerical and scientific libraries like NumPy and SciPy.

Features:

- It aids in data mining and data analysis.
- It facilitates with models and algorithms for Classification, Regression, Clustering, Dimensional reduction, Model selection, and Pre- processing.

2. PYTORCH:

Pytorch was developed by Facebook. It provides an advanced deep learning framework.

You can develop rapid prototyping for research and can build software pipelines using Pytorch. Using Pytorch, Uber's very own probabilistic programming language is built which is useful to develop dynamic graphs so as to accelerate your machine learning processes. Also it provides your code the ability of data parallelism

Features:

- Through Autograd Module it aids in building neural networks.
- For building neural networks it provides a variety of optimization algorithms.
- It finds its usage on cloud platforms.
- It offers various tools, libraries and distributed training.

3. TENSORFLOW:

TensorFlow is the standard name for Machine Learning in the Data Science industry. Through its extensive interface of CUDA GPUs it facilitates building of both statistical Machine Learning solutions as well as deep learning. A tensor which is a multi-dimensional array is the most basic data type of TensorFlow.

It is an open-source toolkit which finds its usage to build machine learning pipelines so that you can build scalable systems to process data. For various applications of ML such as Computer Vision, NLP and Reinforcement Learning, it provides support and functions. One of the must-know tools of Machine Learning for beginners is TensorFlow.

Features:

- It facilitates in training and building your models.
- With the help of TensorFlow.js which is a model converter you can run your existing models.
- It finds its use in the neural network.

4. WEKA:

Weka is the acronym for Waikato Environment for Knowledge Analysis. It is a ML software written in Java. It consists of various ML algorithms that can be deployed and are ready for use. These algorithms finds its usage in data mining.

Weka being an open-source GUI interface allows easy implementation of machine learning algorithms with minimal programming lines. Without writing any line of code you can perform the functioning of machine learning on the data. Hence, it is most suitable for freshers in machine learning.

Features:

- Data preparation

- Classification
- Regression
- Clustering
- Visualization and
- Association rules mining.

5. KNIME:

KNIME is the acronym for Konstanz Information Miner. It is an open-source data analytics, reporting as well as an integration platform. One can carry out the various components of machine learning and data mining with the help of KNIME. It is intuitive and is continuously integrating new development features to it. It helps the users to understand the data and design the data science workflows using reusable components which are accessible to all.

Knime makes use of a modular data pipelining concept. It can blend several data sources to carry out data modeling, analysis, and visualization without the need for extensive programming with the help of GUI and JDBC.

Features:

- The code of programming languages like C, C++, R, Python, Java, JavaScript etc can be integrated using Knime.
- It is helpful for financial data analysis, business intelligence and CRM.

6. COLAB :

Google Colab which is a cloud service supports Python. Using the libraries of PyTorch, Keras, TensorFlow, and OpenCV it will help you in building the machine learning applications.

Features:

- It aids in machine learning education.
- Helps in machine learning research.

7. APACHE MAHOUT:

Apache Mahout is an open-source Machine Learning focused on collaborative filtering as well as classification. These implementations are an extension of the **Apache Hadoop** Platform. While it is still in progress, the number of algorithms that are supported by it have been growing significantly. Since it is implemented on top of Hadoop, it makes use of the Map/Reduce paradigms.

Features:

- It facilitates with algorithms for Pre-processors, Clustering, Recommenders, Regression, and Distributed Linear Algebra.
- For common math operations, Java libraries are included.
- Distributed linear algebra framework is followed.

8) ACCORD.NET :

For image and audio processing, Accord.Net provides machine learning libraries.

Features:

It provides algorithms for:

- Numerical linear algebra.
- Numerical optimization
- Statistics
- Artificial Neural networks.
- Image, audio, & signal processing.
- It also extends support for graph plotting & visualization libraries.

9. SHOGUN :

Shogun is a popular, open-source machine learning software.

Features:

- It provides support vector machines for regression and classification.
- It aids in implementing Hidden Markov models.
- It gives support for many languages like – Octave, R, Ruby, Python, Java, Scala, and Lua.

10. KERAS :

Keras that provides support for Python is an open-source neural network library. It is well known for its speed, modularity and ease of use. It provides user-friendliness that enables the users to readily implement neural networks without dwelling over the technical jargon when compared to more widely popular libraries like TensorFlow and Pytorch, Keras.

Features:

- It can be used for easy and fast prototyping.
- It provides support to convolution networks.
- It aids recurrent networks.
- It provides support to a combination of two networks.
- It can be run on the CPU and GPU.

11. RAPID MINER :

Rapid Miner provides a comprehensive and integrated environment for carrying out several tasks like data preparation, machine learning, text mining, deep learning as well as predictive analytics. It is popular for its lightning-fast speed to reduce costs ,drive revenue and avoid risks.

One of it's most significant features is its GUI based drag and drop feature that allows the users to intuitively build data processing workflows which can be selected from over 2000 available nodes. One can also optimize the model performance through bagging, boosting and building the model ensembles apart from building machine learning models.

Features:

- It aids in designing and implementing analytical workflows through GUI,.
- It assists with data preparation.
- Result Visualization.
- Model validation and optimization.

VII. COMPARISON CHART

	Platform	Cost	Written language	in Algorithms or Features	
Scikit Learn	Linux, OS, Windows	Mac	Free.	Python, Cython, C++	Classification, C, Regression, Clustering, Preprocessing, Model Selection, Dimensionality reduction.
PyTorch	Linux, OS, Windows	Mac	Free	Python, CUDA	C++, Autograd, Optim, nn Module, Module
TensorFlow	Linux, OS, Windows	Mac	Free	Python, CUDA	C++, Provides a library for dataflow programming.
Weka	Linux, OS, Windows	Mac	Free	Java	Data preparation, Classification, Regression, Clustering, Visualization, Association rules mining
KNIME	Linux, OS,	Mac	Free	Java	Can work with large data volume, Supports text mining & image

Colab	Windows Cloud Service	Free	-	mining through plugins Supports libraries of PyTorch, Keras, TensorFlow, and OpenCV
Apache Mahout	Cross-platform	Free	Java Scala	Preprocessors Regression Clustering Recommenders Distributed Linear Algebra.
Accors.Net	Cross-platform	Free	C#	Classification Regression Distribution Clustering Hypothesis Kernel Image, Audio & Signal. & Vision
Shogun	Windows Linux UNIX Mac OS	Free	C++	Regression Classification Clustering Support vector Dimensionality reduction Online learning etc.
Keras.io Rapid Miner	Cross-platform Cross-platform	Free Small: per Medium: \$5000 year. Large: \$10000 per year.	Python plan Java \$2500 year. per	API for neural networks Data loading & Transformation Data preprocessing & visualization.

VIII. CONCLUSION

In this paper, you have explored machine learning and the top machine learning software in a detailed manner.

Depending on your requirement for the algorithm, your expertise level, and the price, you can select the appropriate tool.

You can use the Machine learning library with ease. Except Rapid Miner most of these libraries are free. TensorFlow which is more popular in machine learning except Rapid Miner except Rapid Miner but it has a learning curve. The two popular tools for machine learning - Scikit-learn and PyTorch both support Python programming language. For neural networks - Keras and TensorFlow are good.