A SURVEY ON TOOLS USED FOR MACHINE LEARNING

Dr. S. Veena  
Professor,  
Department of CSE  
S.A. Engineering College,  
Thiruverkadu, Chennai,  
Tamil Nadu, India.

T. Shankari  
UG Student,  
Department of CSE  
S.A. Engineering College,  
Thiruverkadu, Chennai,  
Tamil Nadu, India.

S. Sowmiya  
UG Student,  
Department of CSE  
S.A. Engineering College,  
Thiruverkadu, Chennai,  
Tamil Nadu, India.

M. Varsha  
UG Student,  
Department of CSE  
S.A. Engineering College,  
Thiruverkadu, Chennai,  
Tamil Nadu, India.

ABSTRACT:- In this paper, a brief introduction to Machine Learning and its Tools are studied. In the recent developments, most of the Machine learning tools are more advanced and efficient. The various tools learn the machine by using a training set, which predicts the output correctly and efficiently. Machine Learning is applied in different applications such as Agriculture, Data Quality, Information Retrieval, Financial Market Analysis etc., In this paper, we have discussed few tools like Scikit learn, Pytorch, Tensor flow, Amazon Machine Learning, KNIME, Rapid Miner, Keras, and Shogun with its features and its advantages.

Keywords: Machine Learning, Model, Types of Learning, Applications

I. INTRODUCTION

Machine Learning is the Scientific study of Algorithms and statistical models to perform a specific task without explicit instruction instead uses patterns and inference. It is widely seen as a Subset of Artificial Intelligence(AI). The name Machine learning was proposed by Arthur Samuel in 1959. Tom M.Mitchell provided a formal definition of Algorithms as “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.” Numerous Machine Learning Algorithms have been around for a long time, the ability to carry out complex mathematical calculations in big data and deliver faster and more accurate results.

II. TYPES OF MACHINE LEARNING

Machine Learning can be broadly classified into three main categories,

They are,

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Supervised Learning needs to be guided, Unsupervised Learning Learns from the Observation and finds it’s Data Structure but Reinforcement Learning directly interact with the environment and Follows Hit and Trial Method. Machine Learning tools are the Artificial Intelligence-Algorithmic Applications that provide Systems the ability to understand and improve without considerable human Input.

Machine Learning Tools consists of

- Preparation and Data Collection
- Building Models
- Application Deployment and Training.

Machine Learning Tools can be broadly classified based on Platforms And Libraries. A Machine Learning Platform provides capabilities to complete a machine learning project from beginning to end where as the Machine Learning Libraries provides capabilities to complete a part of machine learning project.

In Supervised Learning, the history of data can be sued to make predictions. These predictions are more accurate. Regression and Classification algorithms falls under the Supervised Learning. The Unsupervised Learning used to find the hidden patterns. Clustering and Association rule mining algorithms comes under this type of learning. When the system need to improve the efficiency then the Reinforcement learning can be used.
### COMPARISON OF VARIOUS TOOLS IN MACHINE LEARNING

<table>
<thead>
<tr>
<th>TOOL NAME</th>
<th>FEATURES</th>
<th>ALGORITHMS</th>
<th>ADVANTAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCIKIT LEARN[7]</td>
<td>• It provides machine learning development libraries for python language</td>
<td>• Classification</td>
<td>• Easily understandable documentation</td>
</tr>
<tr>
<td></td>
<td>• It is also used in data mining and data analysis.</td>
<td>• Clustering</td>
<td>• Parameters for any specific algorithms can be altered</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Regression</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Dimensionality reduction</td>
<td></td>
</tr>
<tr>
<td>PYTORCH[10]</td>
<td>• Pytorch is a torch based machine learning library in python which is</td>
<td>• Neural Network</td>
<td>• It helps in creating computational graphs.</td>
</tr>
<tr>
<td></td>
<td>used on cloud platforms.</td>
<td></td>
<td>• It has hybrid front end which makes usage comfortable.</td>
</tr>
<tr>
<td></td>
<td>• It is used for building neural networks through autograde module.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• It provides optimization algorithms for neural networks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOOGLE CLOUD ML ENGINE[11]</td>
<td>• It is a machine learning tool to create and run the optimum quality</td>
<td>• Naive Bayes Classifier</td>
<td>• Most suitable for enterprises.</td>
</tr>
<tr>
<td></td>
<td>models.</td>
<td></td>
<td>• It can be used to train a complex model</td>
</tr>
<tr>
<td></td>
<td>• It provides two services namely,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Prediction</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Training</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TENSOR FLOW[12]</td>
<td>• Tensor flow is a Java script library for machine learning.</td>
<td>• Libraries for Data flow</td>
<td>• It can be used either as script tags or by installing through NPM.</td>
</tr>
<tr>
<td></td>
<td>• It works as a model converter as it is used to run existing models.</td>
<td>• Deep learning</td>
<td>• It helps for human pose estimation.</td>
</tr>
<tr>
<td>AMAZON MACHINE LEARNING[13]</td>
<td>• Amazon machine learning is a cloud based and robust machine learning</td>
<td>• Linear Regression,</td>
<td>• Permits user to create data source object form MySQL.</td>
</tr>
<tr>
<td></td>
<td>tool</td>
<td>Logistic Regression,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fundamental modes are ML models, Real time prediction, evaluations and</td>
<td>Multinominal Logistic</td>
<td></td>
</tr>
<tr>
<td></td>
<td>batch predictions</td>
<td>Regression</td>
<td></td>
</tr>
<tr>
<td>KNIME[7]</td>
<td>• KNIME is a tool for data analytics, reporting and integration platform.</td>
<td>• Text Mining</td>
<td>• It is an alternative for SAS.</td>
</tr>
<tr>
<td></td>
<td>• It uses data pipelining concept.</td>
<td>• Image Mining</td>
<td>• Easy to learn</td>
</tr>
<tr>
<td></td>
<td>• It is used to integrate the programming languages like C, C++, R, Python,</td>
<td></td>
<td>• Easy to deploy and install.</td>
</tr>
<tr>
<td></td>
<td>JAVA.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• It is a Graphical user interface</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOOL NAME</td>
<td>FEATURES</td>
<td>ALGORITHM</td>
<td>ADVANTAGES</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>-----------</td>
<td>------------</td>
</tr>
</tbody>
</table>
| **RAPID MINER**[14] | - Rapid miner is a platform for machine learning, deep learning.  
- It helps in designing and analytical workflows through graphical user interface (GUI).  
- Model validation and optimization is done. | - Logistic Regression  
- Linear Regression  
- LARS | - Extensible through plugins.  
- Ease of use  
- No programming skills required. |
| **KERAS**[7] | - Keras is an API for neural networks which is a subset of machine learning.  
- It supports convolution and recurrent networks.  
- It can be run on CPU and GPU. | - Convolutional networks  
- Recurrent networks | - It is user friendly  
- It is extensible and modular. |
| **SHOGUN**[7] | - It provides various algorithms and data structures for machine learning.  
- It supports in implementing Hidden Markov models.  
- It support vector machines. | - Classification  
- Clustering  
- Regression  
- Support Vector Machine  
- Dimensionality Reduction | - It is a unified and efficient machine learning library.  
- It can process large set of data.  
- User friendly and provides good customer support. |
| **WEKA**[7,16] | - WEKA is Waikato Environment for Knowledge Analysis.  
- It is a software program which has a collection of machine learning algorithms to perform data mining tasks.  
- It is an open source Java based platform  
- It consists of a visualization tool to perform data analysis and predictive modeling. | - Classification  
- Clustering  
- Regression  
- Association Rule Mining  
- Visualization | - The package in WEKA has an integration of the three entities such as data preparation, feature selection and data mining algorithms.  
- It is a open source under GPU.  
- It is portable as it is written in Java. |
IV. APPLICATIONS

Web Search Engine, Photo Tagging Applications and Spam Detector are the major applications of machine learning. It also used in the field of Medical Diagnosis, Speech and Image recognition, Statistical Arbitrage, Learning Association, Prediction, Classification, Extraction and Regression.

V. CONCLUSION

Machine learning is an incredible breakthrough in the field of artificial intelligence. Although machine learning has been transformative in some fields, but its programs fail to deliver expected results in some way. This may cause due to lack of respective data, lack of access to the data, data bias, privacy problems, badly chosen tasks and algorithms, wrong tools and people, lack of resources and evaluation problems.

VI. ACKNOWLEDGEMENT

We would like to thank our Management, Advisor and Principal of S.A.Engineering College for constantly motivating us to do research in our work environment and encourages us to publish more research papers.

VII. REFERENCES

15. https://towardsdatascience.com/data-mining-tools-f701645e0f4c