International Journal of Electrical Electronics & Computer Science Engineering
Special Issue - ICSCAAIT-2018 | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

Tools and Techniques of Machine Learning paradigm: A Study and Analysis

Tapan Kumar[1], Pooja Aneja[2]
Assistant Professor, Dept. of IT, GIBS (Affiliated to GGSIPU, Delhi)
[1]tapanjha91@hotmail.com, [2]puja52.hce@gmail.com

Abstract: ***This is era of Artificial intelligence and Machine Learning. Searching, Sorting and finding relevant data becomes very tedious task, thus not only us but machine must have enough knowledge to understand and interpret the data. Today all tech giants like Google, Apple, Microsoft and Facebook standing on the bleeding edge of Artificial intelligence and Machine Learning. They are actively innovation and investing in the democratization of artificial intelligence. Recently these companies providing many Artificial intelligence and Machine Learning tools as open sourced as well as their commercial offerings. In this paper we are presenting a study and analysis of various tools and techniques of machine learning paradigm.***

Keywords: Artificial intelligence, Machine Learning, Apache Spark MLlib, Caffe, ai-one, Protégé, IBM Watson, DiffBlue, Google's TensorFlow, Amazon Web Services.

I. INTRODUCTION

Learning is a process by which a system improves performance from experiences. Learning also denotes changes in a system that enable a system to do the same task more efficiently the next time. Machine learning has been studied and defined by different author but some widely accepted definition is as follows:

According to Arthur Samuel: "Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed."[1]

According to Tom M. Mitchell: 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."[2]

According to Ron Kohavi*: Foster Provost:* In Knowledge Discovery, *machine learning* is most commonly used to mean the application of induction algorithms, which is one step in the knowledge discovery process. [3]

According to Alan Turing: The definition of machine learning based on class of tasks are given by Tom M. Mitchell can be used to proposed various characteristics possessed by a thinking machine.[4]

Machine learning paradigms are broadly classified into three types, based on the nature of the learning 'signal' or 'feedback' available to a learning system.

Supervised Learning: Let 'x' be the set of input variables and 'Y' be the set of output variables then, if we use an algorithm to map 'x' and 'y' such that y= F(x) and the aim is to achieve the efficient mapping such that we are able to predict the value of y for given set of input variables x. Thus 'x' can be termed as input object (vector) and y is the desired output value (supervisory signal). Classification and Regression are two different types of supervised learning.

Unsupervised Learning: This is a machine learning activity of inferring a function to explore hidden structures from unlabeled data. It is similar to the problem of density estimation in case of statistics.

Reinforcement Learning: This is an area of machine learning and also a branch of Artificial Intelligence. It is based on behaviorist psychology. It allows both machine and software agents to relevant behavior so as to maximize its performance. It is applicable to various fields like game theory, information theory, swarm intelligence, statistics and genetic algorithms. In machine learning, an agent selects best action to be taken based on dynamic programming approach form its current state and this process repeated for finding optimal solution, such process is known as Markov decision process (MDP).

II. TOOLS

There are various Machine Learning and Artificial Intelligence tools available, but most famous 15 tools are listed below:

1. Apache Spark MLlib
2. Apache Mahout
3. Amazon Web Services
4. AI-ONE
5. Caffe
6. DiffBlue
7. Google's TensorFlow
8. IBM Watson
9. KNIME
10. Microsoft Azure
11. Nervana Neon
12. OpenNN
13. Oryx 2
14. Protégé
15. Veles

1. Apache Spark MLlib: Apache Spark MLlib[5] is a library for Machine Learning; its goal is to make Machine Learning scalable and easy. It provides various learning algorithms and utilities, such as classification, regression, clustering, collaborative filtering,

dimensionality reduction as well as lower-level optimization primitives and higher-level pipeline APIs.

2. Apache Mahout: Apache Mahout[6] is a free and open source project of the Apache Software Foundation, has a aim to develop free distributed or scalable machine learning algorithms for different field like collaborative filtering, clustering and classification. it has Java libraries and Java collections for various kinds of mathematical operations. It is implemented on top of Apache Hadoop using the MapReduce paradigm. It provides the data science tools to find meaningful patterns in Big Data sets. It Process 'big information' quickly and easily.

3. Amazon Web Services (AWS): It is a subsidiary of Amazon.com. It offers on-demand cloud computing platforms to individuals, companies and governments, on a paid subscription basis with a free-tier option available for 1 year. According to survey, In 2017, AWS offers more than 90 services including computing, storage, networking, database, analytics, application services, deployment, management, mobile, developer tools, and tools for the Internet of Things (IoT). Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3) are among most popular services offered by AWS.

4. AI-ONE: This tool is very helpful in big data translation into opportunity using machine learning approach. It offers biological inspired intelligence to find relevant patterns and relationships. It offers APIs that enable us to develop machine learning applications.

5. Caffe: Caffe is a framework for deep learning .it is based on the core idea of expression, speed, and modularity. It is developed by Berkeley AI Research (BAIR) and by community contributors[10]. Yangqing Jia created the project during his PhD at UC Berkeley. Caffe is released under the BSD 2-Clause license.

6. DiffBlue: It is a world leader and has capability to understand code[12]. It has ultimate aim to automate all traditional coding activities like bug fixing, test case writing, finding and fixing bugs, refactoring code, translating from one language to another, and creating original code to map the requirement and specifications.

7. Google's TensorFlow: It is an open source library which uses data flow graph (DFG) for numerical computations [7]. The nodes of DFG denote mathematical operations and edges denote multidimensional data arrays (tensors) communicated between them. It offers flexible architecture which allows us to deploy various types of computations. It was originally developed by a group of Engineers and researchers on Google brain team for the purpose of conducting machine learning and deep neural network research.

8. IBM Watson: It is released recently after 12 months of Beta Testing [8]. It is mainly design to offer services to Data Scientists and Developers, who can deploy and monitor machine learning model for various field.

9. KNIME: It is GUI based tool for Machine Learning. It offers various functionality from I/O to data manipulation, transformation and data mining. It consolidates whole process into single workflow. It is best tool for beginners to build their own machine learning model[16].

10. Microsoft Azure: It is developed on top of the machine learning capabilities of various products and services offered by Microsoft[9]. It offers the real-time predictive analytics of the new personal assistant in Windows Phone known as Cortana. It has been found useful in providing some services to Xbox and Bing. It has more than 25 machine learning APIs. Microsoft Azure has been successfully implemented for various applications including Bing Synonyms API, Bing Speech Recognition Control, Bing Search API and Microsoft Translator.

11. Nervana Neon: It is an open source library for developing frameworks that can efficiently run deep learning computations on a variety of compute platforms. Nervana Systems offers Neon as an open-source, Python-based, deep learning framework[10].

12. OpenNN: OpenNN (Open Neural Network) is an open source class library, which is written in C++ programming language[11]. It offers high performance library for advance analytics. It also implements neural networks efficiently. To provide GUI support to OpenNN another software tool named as Neural Designer has been introduced, which simplifies the data entry and result interpretation.

13. Oryx2: Oryx 2 is based on Lambda architecture. It has capability to process real-time large scale machine learning applications[13]. It offers functionality such as building applications and includes packaged, end-to-end applications for collaborative filtering, classification, regression and clustering. General Lambda architecture tier: it offers batch, speed and serving layers Specialization on top which offers machine learning abstraction to parameter selection. End-to-end implementation of the same standard machine learning algorithms as an application (ALS, random decision forests, k-means) on top.

14. Protégé: It is a free, open-source platform that offers ontologies development environment [14]. It offers variety of suit of tools, which helps in construction of domain models and knowledge-based applications. It provides support for editing OWL 2 ontologies. Multiple format exits for upload and download of ontologies (supported formats like RDF/XML, Turtle, OWL/XML, OBO, and others).

15. Veles: It is Python based distributed platform, which offers machine learning and data processing services[15]. Veles Machine Learning algorithms are

International Journal of Electrical Electronics & Computer Science Engineering
Special Issue - ICSCAAIT-2018 | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

Neural Networks as Znicz plug-in and Genetic Algorithm as Genetic in Veles core. Znicz is a sub module of Veles. All users of VELES are divided into three groups: (a) Casual User-Entry (b) Models Constructor – Medium (c) Unit Developer – High.

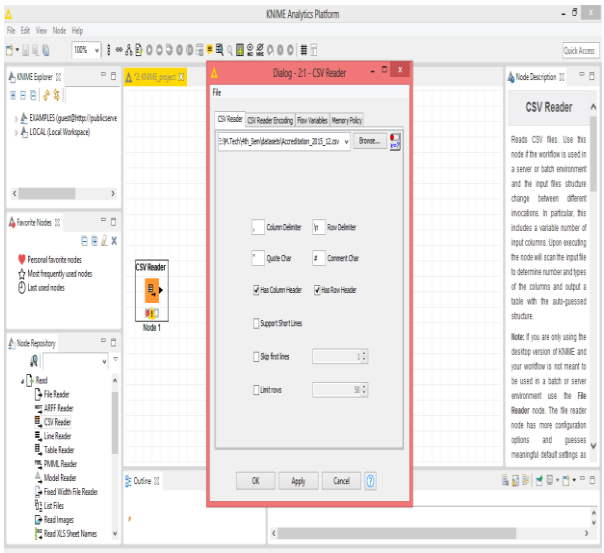
III. IMPLEMENTATIONS

Implementation of Machine Learning Using KNIME: It works on Node- Connector paradigm.

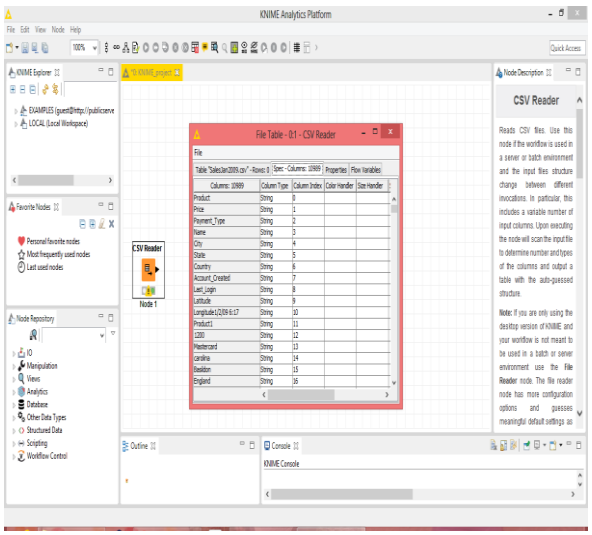
We are uploading NAC_Test.csv file as a dummy data sets, to show implementation of web usage mining using KNIME tool [16]. NAC_Test.csv file have sufficient rows and columns for such purpose.

Working with KNIME Analytics Tools involves following steps: -

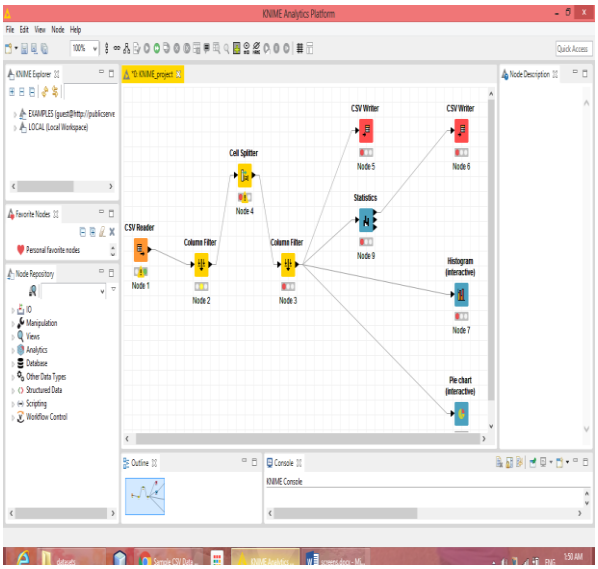
Step I: Drag CSV Reader Node and click to Configure option. A dialog box will open Adding NAC_Test.csv file to workspace into CSV Reader Node.



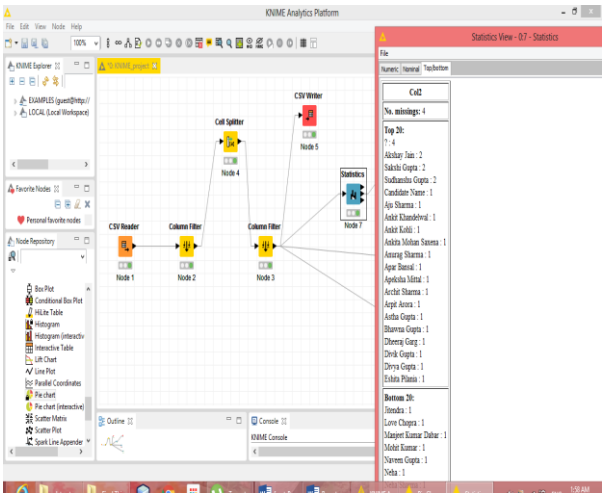
Step II: Right click on CSV reader to see the uploaded file i.e. File Table.



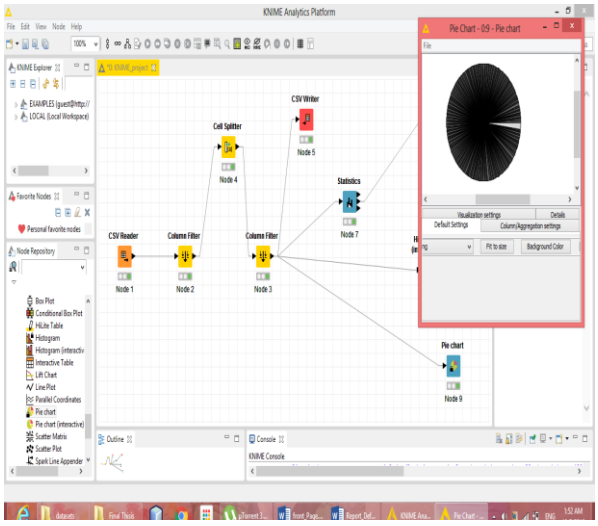
Step III: Place the respective nodes at appropriate position and connect them as it work on Node-Connector paradigm [17].



Step III: click to respective nodes for Result:



[Figure: Statistics View]



[Figure: Pie Chart View]

IV. CONCLUSION

In this research paper a study and analysis of tools of machine learning paradigm has been studied using KNIME by some illustrations.

Machine learning and Artificial intellegence tolls has got and continues to get popularity in various applications. As tech giants are bleeding edge of Artificial intelligence and Machine Learning thus the current system is expected to add more functionality and dependency according to requirement changes and technology, proper working tools has been kept in mind to make it easier for future enhancements.

V. REFRENCES

[1] Supposedly paraphrased from: Samuel, Arthur (1959). "Some Studies in Machine Learning Using the Game of Checkers". IBM Journal of Research and Development. 3

[2] Mitchell, T. (1997). Machine Learning. McGraw Hill. p. 2. ISBN 0-07-042807-7.

[3] R. Kohavi and F. Provost, \Glossary of terms," Machine Learning, vol. 30, no. 2-3, pp. 271-274, 1998.

[4] Stevan Harnad (2008), "The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence", in Epstein, Robert; Peters, Grace, The Turing Test Sourcebook: Philosophical and Methodological Issues in the Quest for the Thinking Computer, Kluwer

[5] Shoro, Abdul Ghaffar, and Tariq Rahim Soomro. "Big data analysis: Apache spark perspective." Global Journal of Computer Science and Technology 15.1 (2015).

[6] Walunj, Sachin Gulabrao, and Kishor Sadafale. "An online recommendation system for e-commerce based on apache mahout framework." Proceedings of the 2013 annual conference on Computers and people research. ACM, 2013.

[7] Abadi, Martín, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems." arXiv preprint arXiv:1603.04467 (2016).

[8] High, Rob. "The era of cognitive systems: An inside look at ibm watson and how it works." IBM Corporation, Redbooks(2012).

[9] Mund, Sumit. Microsoft azure machine learning. Packt Publishing Ltd, 2015.

[10] Bahrampour, Soheil, et al. "Comparative study of caffe, neon, theano, and torch for deep learning." (2016).

[11] Wang, Wenduo, Yi Murphey, and Paul Watta. "A computational framework for implementation of neural networks on multi-core machine." Procedia Computer Science53 (2015): 82-91.

[12] Liang, Xiao, et al. "Energy efficiency formation optimization of a fleet of AUVs based on multi-island genetic algorithm." Control Conference (CCC), 2017 36th Chinese. IEEE, 2017.

[13] Zheng, Jiang, and Aldo Dagnino. "An initial study of predictive machine learning analytics on large volumes of historical data for power system applications." Big Data (Big Data), 2014 IEEE International Conference on. IEEE, 2014.

[14] Lehmann, Jens. "DL-Learner: learning concepts in description logics." Journal of Machine Learning Research 10.Nov (2009): 2639-2642.

[15] velesnet.ml/docs/: An open Source & developed by samsung

[16] Tapan Kumar, Tapesh Kumar "Web Usage Mining Using Semantic Web Approach: A Study, Survey and Analysis", 2016 ICRISEM.

[17] Beisken, Stephan, et al. "KNIME-CDK: Workflow-driven cheminformatics."BMC bioinformatics 14.1 (2013): 1.