Machine Learning – An Empirical Science

Mahima Mohini[1], CharveeChoubisa[2],AviralKakkar[2],AnidhyaAthaiya[3]

1 Students, Department of Computer Science, Global Institute of Technology, Jaipur

2 Students, Department of Computer Science, Global Institute of Technology, Jaipur

3 Assistant Professor, Department of Computer Science, Global Institute of Technology, Jaipur

Abstract-All sciences are empirical in nature and so is Machine learning, giving machinesthe ability to learn based on observations and experiences. Machine learning has become an integral part of today's world in many ways, such as commercial applications and research projects which ranges from medical science to social networking to entertainment. It is closely related to the computational statistics. Machine learning is all about getting information and insights from the data, as a result future prediction are made using computers.

I. INTRODUCTION

Machine learning is the branch of Advanced Analytics known as Predictive Analysis, used to make predictions about possible future events. It is the study of mathematical optimization which delivers methodologies, its theory and its various applications. Data Mining is a subfield of Machine Learning and statistics which extracts information from the raw data collected in mining or discovering of data. Machine learning is a subset of Artificial Intelligence. The algorithms of machine learning builds mathematical model empirically on sample data and enables a machine to make prediction or decision without manual help.

II. METHODOLOGIES

Machine learning is a technology using which information can be turned into knowledge. In the past few decades, there has been an abundance of data which is useless unless analysed, so as to obtain insights about data, to generate accurate and effective predictions. There are various forms of machine learning – supervised, unsupervised and reinforcement learning. Each of the learning forms follow different set of methods and approaches to achieve the desired goal. Under supervised learning comes regression and classification problems which work upon continuous values and categorical features respectively. The unsupervised learning deals with the transformations of dataset and clustering problems. An example of transformation of data is dimensionality reduction. While the reinforcement learning deals with the deep learning part of the machine learning involving neural networks.

Most commonly used methods to implement machine learning are –

- Regression – This algorithm is a supervised learning algorithm which helps in explaining and predicting the continuous data values. This algorithm is further classified in various regression algorithm of which linear regression and polynomial regression is most popular.
- Classification – Classification algorithm is another supervised learning algorithm which helps in handling and predicting the categorical data in a dataset. The simplest classification algorithm is logistic regression algorithm.
- Clustering – This is an unsupervised learning algorithm whose aim is to group data of similar characteristics. It does not use the output information, but rather it lets the algorithm find its output. Commonly used methods of clustering are K- means clustering, hierarchical clustering, agglomerative clustering, etc.
- Dimensionality Reduction – The name itself suggests that the dimension of the data is reduced by eliminating the redundant values from the data. Principal Component Analysis PCA is the most popular method of dimensionality reduction.

III. WORKING OF MACHINE LEARNING

The working of machine learning involves several sub-processes, each process performing its unique functionality. The working starts when raw data obtained from the data providers are pre-processed using various pre-processing tools. The data after pre-processing is then structured, such that various machine learning models can be applied on it and desired results can be obtained. The best model is selected for deployment and used in various applications – spam filtering, healthcare bioinformatics, fraud detection, etcetera. In other words, when good quality data is fed to the machine, iterating with best learning algorithm then a golden model is obtained which is implemented in numerous important organisations for analysis of data, future predictions, etc.

The working of machine learning follows the steps in sequence as given below –

- Identification of a well-defined problem
- Acquiring of sufficient data
- Processing, cleaning and visualisation of data
- Generation of a model
- Testing and refining of the model
- Integrating the model
- Deployment of model

IV.   ADVANTAGES

Machine learning is a very powerful tool which can revolutionise the world. The major advantages are :

- It enables us to find the patterns and trends in the data very quickly.
- Human intervention is not needed.
- With experience, the machine learning algorithms are improving day by day rapidly.
- It can handle multi-dimensional and multi-variety data very easily.
- Machine learning has vide range of applications – virtual personal assistants, video surveillance, prediction while computing, etc.

V.   LIMITATIONS

With all its advantages and powerfulness machine learning is not actually perfect. Data acquisition is still a major problem because generation of new data is a difficult task and maintaining its authenticity and quality. A machine takes much time to let the algorithm learn and develop enough to fulfil its purpose for relevant and accurate result. The accuracy obtaining after the interpretation of result is a challenge as well because in order to get an efficient result the selection of the model should be done carefully otherwise it can result in error, causing high susceptibility of errors.

VI.   FUTURE SCOPE

Future of machine is as vast as the limits of human mind. One can always keep learning and teach the computers how to learn. With advancement of time and technology there will arrive a time when no human intervention will be required and the era of automation will commence.

VII.   CONCLUSION

Machine learning has a vast potential but what happens is people fail to realize its potential. The reason is that machine learning works mostly on the mathematical computation and due to lack of knowledge of mathematics the working of machine learning becomes difficult to understand and comprehend. Hence, awareness about this technology should me made because machine learning will be a major part of our upcoming future

VIII.   REFERENCES

[1]   Introduction to Machine Learning with Python – Sarah Guido, Andreas Muller.

[2]   www.towardsdatascience.com

[3]   www.sciencedirect.com

[4]   www.wikipedia.com

[5]   www.geeksforgeeks.com