International Journal of Advanced Trends in Computer Applications
www.ijatca.com

A Review on Applicability of Machine learning

[1]Nikhil Katoch, [2]Diksha Nagpal
[1,2]Department of Computer Science
Chandigarh University
Gharuan, Punjab
[1]kochi.nikhil@gmail.com, [2]dikshanagpal91@gmail.com

Abstract: *Machine learning is an application of artificial intelligence in which the machines learn themselves and then work accordingly to the instructions. Basically, machine learning works on the data sets. Data is unprocessed raw facts and figures. The machine works on the data, tries to understand and correlate with different fields and then give output. In this paper, we will be discussing the basic knowledge required to build up the machine learning models, the hypes and reality related to machine learning and most importantly how machine learning and interrelated fields are used in various platforms. This is one of the fast-growing fields in the present world, as it is reducing the load of computation and helping companies to strategies accordingly. But as every coin has two sides, machine learning also has its own positive and negative views as day by day it is reducing human efforts. Almost every multinational company is using this technology for solving the problems of society and people. Machine learning is also linked to other branches like artificial intelligence, data science, computational statistics and probability. These all fields are linked with one another, machine learning is all about the mathematics mainly probability and statistics. Analyzing the data depending upon the various factors and then work according to them is a part of machine learning.*

Keywords: Machine learning, Clustering, Regression.

I. Introduction

Machine learning is an application of the artificial intelligence in which the machine learns from the data provided to the machine, and then makes correlations according to the mathematical calculations. It is vastly used nowadays in the field of technology as it is reducing the human efforts and bringing out the best results to the company benefit. International business machine popularly known as IBM is one of the emergent technologies that is using this technology for its benefit. Watson is an open multi-cloud platform, developed by the IBM that helps in building the machine learning models and helps the user to understand the dependency of the data.

Arthur Samuel was the first person who coined the term machine learning in 1959, while he was working in IBM. Machine learning works on the statistical data and then tries to predict the possible outcomes of the conditions [1]. Machine learning is also related to fields like Artificial intelligence and deep learning. These fields work together to provide the most significant result.

Let us understand how machine learning works. The first step is building the data set. The data set consists of the relevant data which is particularly divided into the dependent and independent data set. The dependent dataset consists of the data which depends on the other factors whereas the independent data-set consists of the data which has no connection with each other. Then comes to relate the independent dataset with the dependent data-set, here we use many techniques like clustering, regression and many more. Then after making the machine learn the various relations between the independent and dependent dataset, that machine can be used to predict the output depending on the various inputs.

Now let us understand how machine learning is linked to fields like artificial intelligence [2] and deep learning. Starting with Artificial learning, artificial intelligence is the human intelligence embedded in the machine. So that it can perform continuous actions. Machine learning is an approach to achieve artificial intelligence. Deep learning [3] uses the Artificial Neural Network (ANN) [4] which is generally an algorithm which is used to study the data. So, we can say that deep learning is an important part of the machine

learning as the machine learns the data but it won't be possible without the deep learning and its algorithm.

Suppose we are a growing company and we want to know where we should invest our money. The machine can do this job for us, by comparing the data of the other companies' data and helps us to strategies accordingly. The machine can predict the most probable areas where we invest our money to have the best profitable results for the company.

But letting the machine to take decisions for us, is really not a great idea, as every coin have two sides. It has its own merits and demerits. In this paper we will be going to discuss, the various techniques to build the machine learning models, clustering, regressions and the (5) demerits and merits of the machine learning. Also, how much we should use the machine learning, and should we let the machine to take the decisions for us or not which is one of the main focused questions nowadays.

Tesla, one of the world leading automobile companies uses the machine learning to develop self-driving cars which are generally based on the machine learning models. NASA Curiosity rover also works on machine learning and artificial intelligence models. So basically, two different questions how much we should use and how to use for the better mankind, so that these technologies don't affect humanity. So in this paper we will be having the overview of the machine learning and what techniques we should learn to become a good machine learner and how it is changing the world.

II. STARTING WITH BUILDING THE DATA SET

Before going into detail, we should know little about the data and how it is being made in the present world.

The modern world is full of technology, with the passage of the time and technological advancements; there comes the problem of the keeping record of each and every detail. Taking an example of the cricket, modern cricket has advanced as much, that we can predict the winning percentage based on the previous records, which player is going to have a major impact in the leagues.

So where the data came from, actually it is not collected in a day or a week. It is being collected over the passage of the time and when data is collected over a vast period of time it can then be used to predict the situations. A machine does not provide the accurate information, or it is not like the prediction made by the machine is always true, generally, it gives us the trend about the situations, like how things are going on for the time being and what are the steps that should be taken to make the situation better.

Now to build the data set [5] we need to look for the two main things.

1) Independent data set
2) Dependent data set

Independent data set

Independent data set are those data sets which have their own existence and they are not depending on the other values. Those values have their significance and during building, the data set it is the most important part.

Dependent data set

A dependent data-set are those data sets which depend on the other factor. They can generally consider as the output of the inputs.

After dividing the data set, the next step is to analyze the data before giving to the machine, like is our data is consistent or inconsistent. By saying consistent means, data should not contain any missing values and any other type of discontinuity.

Note: it is not advised to delete the column from the data set, because it is treated as the false machine learning practice.

Dealing with the missing values and other problems in the data

The basic techniques to overcome this problem is to strategies like mean, median of the columns so that it can be computed rather than deleting the entire column. It is also advised to rearrange the data in the machine understandable code, as machine understands only binary so the data that can be converted in form of the binary data can be converted to the binary. The process of converting the data in the form of the binary is called as label encoding. For doing the label encoding, python has inbuilt functions like Label Encoder which is present in the sklearn library followed by the advanced techniques like One HotEncoder which is also used to represent the data in the form of the binary but in a more organized way.

Dividing the data in the training set and test set

Before splitting the data into sets, we should know about what the need is for splitting the data. In general, the more data the machine analyses, the better is the prediction. Now the question arises that if that is the case, so we should give our all the data to

the machine as input. However this is not true, we should divide the data so that we should know whether the machine is giving the right output or not and accuracy of the output is compared with the test set, so we should divide the data into two parts, i.e training set and test set.

THE DETAILED PROCESS OF THE SETTING UP THE DATA:

All the operations are performed on the training set. All the processes mentioned above are applied to whole data like label encoder, one hot encoding, imputer, standard scalar and many more are applied to the whole data-set. Then comes the step of the splitting of the data, for that python have the inbuilt library learn, from that, we have to import the train_test_split functions for splitting the data. While splitting the data we specify the size, which generally means the percentage of the data-set we are giving to the training set and remaining to test set.

Note: It is advised to keep the training size as max as we can, generally the test size is kept within 20 per cent to 30 per cent. So that machine has enough data to understand.

Now simply just fit the model (predictor) in the test set and predict the value accordingly. For checking whether the machine is predicting the right values or not, use the predictor function and pass the input as test independent set and compare the predicted values and original output values. For comparing we can use the matplot, pyplot function which allows us to draw graphs and as graphs are much better for comparing the two values so we can check the accuracy of our machine learning model and if the predicted values are close to the real values so the machine learning model is working good, and if not, we should again reconsider the data and try a different approach.

III. Clustering

Clustering [6] is the process of reassembling the scattered data. During the time of machine learning models sometimes we see that data is scattered in the 2D plane so for collecting the data we use the clustering method.

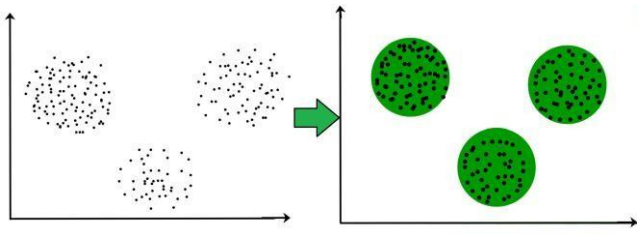

Figure 1: Clustering

This picture represents the clustering of the data the left part of the picture represents the data in the 2d plane without clustering and the right part of the picture represents the data after the clustering.

IV. Regression

It is a mathematical term which is used to measure the relation between the mean and the values. Python has inbuilt regression functions like discussed above: linear regressions [7], multiple linear regression and many more.

Figure 2: Regression

This is a type of the linear regression which was built on the model of the salary vs. age which can predict the values of the salary depending upon the experience of the employee.

V. Merits and demerits of machine learning

5.1 Demerits [8]

1) As we all know the machine is taking place of the humans, which results in the downfall of the job vacancies in the companies.
2) With the advancements in the technology, giving the machine the ability to understand and think, take choice according to the situations, this is not a positive site for the future.

Example:
[7]When Mark Zuckerberg was working in the field of machine learning and artificial intelligence he developed the chat bots with the motive to provide convenience to the society and users, but while he was working on the chat bots, the chat-bots started coordinating with each other and developed their own language, which is not understandable to humans. So if the machines are allowed to take the decisions, saying

that one day we will work under machines won't be wrong.
SOFIA the world first fully AI robot having a citizenship of the country, when asked in the interview about the relationships of the machines and humans, replied that humans will be going to work under the machines.
So, we shouldn't be using machine learning where it can cause damage to the society, excessive use of the machine learning practice will be going to have a major impact on the society, no doubt the skilled labor demand will be increased for particular fields related to the machine learning, but in spite of all the pros and cons we should rethink about the situations and then work accordingly.

5.2 Merits

1) Reduces the human efforts, now human don't have worry about the minute issues, now a machine can be programmed to work in place of them, and they can devote their time in some other matters, so indirectly more amount of work is done in short span of time.
2) Recent figures show that the amount of energy used in the supercomputers, the transferring of data when the machine is allowed to handle the situations it becomes more efficient.
3) The world is changing at a rapid pace and with the passage of the time, more advancements in the technology is sure to come. To handle the amount of data of the users in the future is not a simple task for the humans, so handling the data by the machine where the machine will be going to do the work for us, is handier and will be going to take the burden off from the shoulders of the employees.

VI. Myths and reality of machine learning

1) The machine will be totally going to take over from humans
2) The jobs will completely be going to have vanished in future
3) There won't be any place for the people in the future as machines will be going to learn everything.

6.1 Reality is actually quite the opposite thing

1) It will increase the job opportunities in the particular sector, as people will be needed for instructing the machines at timely intervals.
2) If the machine is taught up to a limit its good as said excess of everything is bad, so using the machine learning for having the control over everything is bad.
3) Humans will always be needed because all the algorithms till date are made by the humans, and it's the machine that learns algorithms so developing the algorithms will always be under the task of the humans.
With this, it's not like machine learning practice is good or bad for the society , its depend upon the way we use the technology, as said every coin have 2 sides, similarly technology have also two sides, one that helps the society for solving the problems other that bring more challenges in front of the society. So, the passage will continue as always, it's upon us how we use the technology.

VII. Applications of the machine learning in various Fields

The modern world is changing at a rapid pace. With the modern development there comes the problem of managing large amount of data, which is difficult to be handled by human beings. So we made machines to work for us to reduce the human workload.
1) In field of agriculture: The use of machine learning can be used in field of agriculture to analyze the yield of the crops. The agricultural drone nowadays is helping farmers to increase in the crop production and monitor crop growth [10].
2) In industries: Companies like Facebook, google, amazon and many more are using the machine learning models for developing the chat-bots. Predicting the choice of user, like predicting the music choices and the interest of the users, is achieved with the help of the machine learning models.
3) Self driving cars: Companies like tesla, uber had been building the self-driving cars also known as robot cars which also work on the machine learning model, that senses the data from the surrounding and then move the car accordingly.
4) In day to day life: Machine learning in coordination with artificial intelligence, helps us to develop virtual personal assistants like Alexa, Siri, Google and many more. These technologies help us in our daily life by assisting us to perform task. We can set up alarms, set up reminders and can do everything just from the voice control.
5) Prediction while computing: We all are using the GPS navigation services. While we do that, our current location and velocities are being saved at a central server for managing traffic. This helps to re-route the traffic so that there won't be any traffic issues.
6) Social media services: Like we all know about the chat bots, chatbots work on the principle of the machine learning and artificial intelligence. Other services like "People you may know" or features like facial recognition are possible with the help of the machine learning and artificial intelligence [9].
7) Online customer support: Number of websites nowadays gives the feature of chat with the customer support. However not every site has a live executive to

answer the queries. So the companies work on the chatbots and till the time executive gets free the chatbots handle the customer in place of the executive.

8) Search engine result refining: Google and other companies use machine learning to refine the output of the search results. Machine keeps the records the most viewed websites and based on the search input it tries to give the best result based on what machine has learned.

9) Online customer support: Number of websites nowadays gives the feature of chat with the customer support. However not every site has a live executive to answer the queries. So the companies work on the chatbots and till the time executive gets free the chatbots [11] handle the customer in place of the executive.

10) Search engine result refining [12]: Google and other companies use machine learning to refine the output of the search results. Machine keeps the records the most viewed websites and based on the search input it tries to give the best result based on what machine has learned.

VIII.Conclusion

Machine learning is a technique that can be used to reduce the human effort, although it has its own advantages and disadvantages. If we have the small amount of data and well labeled, we can go for supervised learning. But on the other hand, if we have the large amount of data, definitely the unsupervised learning will give the better performance. In the next upcoming years, machine learning will be going to play a major role in our day to day life. From smart-phones to smart-cities, machine learning and artificial intelligence will revolutionize the regular way of living. Thus, to maintain the proper working of the all the systems, we need a huge work force, thus creating the job opportunities in the future. Though there will be more demand of the skilled labour in the future, more towards to the technical department but according to me, the future is not just toward the machine and artificial intelligence, Its more about collaboration of the technology with the mechanical industry to develop AI-Robots that can take place of the humans in the difficult jobs, like mining, secure country borders, researching deep inside the oceans, interplanetary missions and many more. So, machine learning is really useful if being used for the welfare of our society and people, as every coin has 2 sides, technology also have 2 sides positive and negative, it totally depends on the way we want to use it.

References

[1]. Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2

[2]. Poole, David; Mackworth, Alan; Goebel, Randy (1998). Computational Intelligence: A Logical Approach. New York: Oxford University Press. ISBN 978- 0-19-510270-3.

[3]. Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview". Neural Networks. 61: 85– 117. arXiv:1404.7828

[4]. Bethge, Matthias; Ecker, Alexander S.; Gatys, Leon A. (26 August 2015). "A Neural Algorithm of Artistic Style". arXiv:1508.06576

[5]. Kohavi, Ron (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection" (PDF). International Joint Conference on Artificial Intelligence.

[6]. Coates, Adam; Lee, Honglak; Ng, Andrew Y. (2011). An analysis of single-layer networks in unsupervised feature learning (PDF). Int'l Conf. on AI and Statistics (AISTATS).

[7]. Alpaydin, Ethem (2010). Introduction to Machine Learning. MIT Press. p. 9. ISBN 978-0-262-01243-0

[8]. Garcia, Megan (2016). "Racist in the Machine". World PolicyJournal. 33 (4):111117. doi:10.1215/074027753813015. ISSN 0740-2775

[9]. Francesco Ricci and Lior Rokach and Bracha Shapira, Introduction to Recommender Systems Handbook, Recommender Systems Handbook, Springer, 2011, pp. 1- 35

[10]. Reina, Giulio (2018). "A multi-sensor robotic platform for ground mapping and estimation beyond the visible spectrum". Precision Agriculture. 20 (2): 423–444. doi:10.1007/s11119-018-9605-2

[11]. Luka Bradeško, Dunja Mladenić. "A Survey of Chabot Systems through a Loebner Prize Competition" (PDF). Retrieved 28 June 2019.

[12]. Vaughan, Liwen; Mike Thelwall (2004). "Search engine coverage bias: evidence and possible causes". Information Processing & Management. 40 (4): 693–707. CiteSeerX 10.1.1.65.5130. doi:10.1016/S0306-4573(03)00063-3.