January-February 2020
ISSN: 0193-4120 Page No. 4146- 4150

Machine Learning in Healthcare Industry: Tools and Techniques

Neha Singh[1]
Department of Information Technology,
Babasaheb Bhimrao Ambedkar University, Lucknow, India
chauhannehasingh2408@gmail.com[1]

P. K. Chaurasia[2]
Department of Information Technology,
Babasaheb Bhimrao Ambedkar University, Lucknow, India
pkc.gkp@gmail.com[2]

Article Info
Volume 82
Page Number: 4146- 4150
Publication Issue:
January-February 2020

Article History
Article Received: *18 May 2019*
Revised: *14 July 2019*
Accepted: *22 December 2019*
Publication: *21 January 2020*

Abstract:
From last two decades, technology evolving very quickly and changed the view nearly all aspects of modern life. There is a lot of development made in the area of machine learning. It is the subset of Artificial Intelligence that gives machine the possibility to determine and make decision without using explicit instructions. Today, ML has a tremendous impact in across all over the world and benefited various areas like medical diagnostics, fraud detection, driverless vehicles and security surveillance etc. Machine learning layout an inspiring set of technologies that consist of pragmatic tools for reviewing data and making predictions with the latest headway in artificial intelligence. The major theme of the paper is focused on the various tools and techniques of machine learning and its application in healthcare industry.

Keywords—*Machine Learning, Artificial Intelligence, Fraud Detection, Traditional Programming, Supervised, Unsupervised*

I. INTRODUCTION

Machine learning is determined as an independent process that brings out the patterns from data. Machine learning requires very less human interaction and enable a machine to learn patterns from enormous data [2]. Machine learning is one of the effective techniques that make prediction which are based on past experiences. Machine learning is a subfield of Artificial intelligence that developed gradually from the study of pattern recognition. We may not identify the whole process so that we recognize certain patterns or regularities. It help us to make predictions, about future or near future. This technique of machine learning uses some models [1]. Predictive model is used to make predictions in the future anddescriptive model is used to gain knowledge from data.

Machine learning technique is relying on human knowledge. It can help a machine to learn more logical through techniques like propitious future selection, relay learning and multitask learning [3]. Machine learning is distinct from traditional programming. Traditional programming entails input (data) and programs. But machine learning entails input (data) and output (data) as shown in fig. 1.

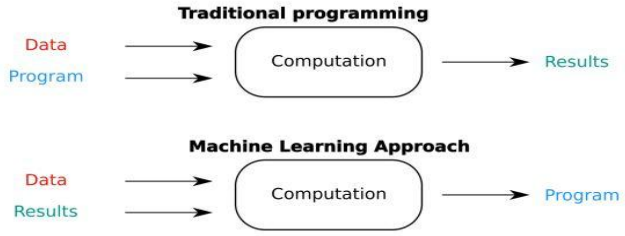

Fig1: Traditional Programming vs Machine Learning

There are numerous machine learning algorithm used to learn the machine [4]. The most methodical machine learning algorithms are those that optimize the decision making process by universalize from well-known examples. Machine learning techniques are classified in 3 wide range of classes as shown in fig. 2

January-February 2020
ISSN: 0193-4120 Page No. 4146- 4150

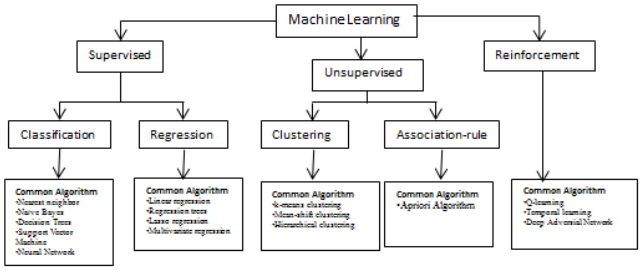

Figure 2: Classification of Machine Learning

SUPERVISED LEARNING

Supervised learning training data is given as an input to train the model. Model is executed to learn the mapping function from input variables (X) to output variables (Y); means Y= f(X). In supervised learning the input which is given as training example is associated with labels as shown in figure 3. The objective of the correct mapping function when any new data is given to the machine [5]. This is why supervised learning is also called as labeled training algorithm. Supervised learning works in an iterative manner that means learns from training data and then apply learning to predict new data. This predicted data is evaluated by using cross validation techniques in which original data is divided into training set and test to train the model and evaluate the model respectively. For example: K-Fold cross validation data is split into k subsets; where one subset is used to test the model and left over, k-1 subsets are used to train the model [17]. This cross validation process is frequently performed k times with label subset and the average accuracy of k folds is selected as endmost accuracy. Supervised learning is further bifurcate into classification and regression algorithms [4].

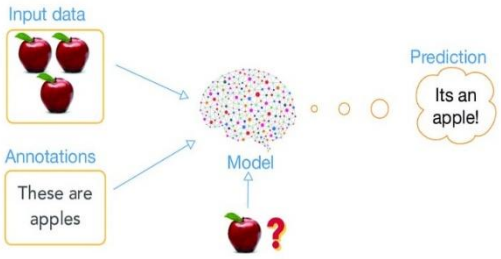

Fig 3: Supervised Learning

A. CLASSIFICATION

Classification is a supervised machine learning technique which applies on where outcomes are a discrete label [22]. It is used to categories dataset into desired and distinct number of classes. We can assign label to each class as shown in fig 4. It is a two step procedure. First stage is known as learning stage where model is trained by pre-arranged data and second stage is known as testing step where model is used to project the new data and test the validity of classifiers. Binary classifier which deals with two distinct classes or with two possible results. For example classification of gender: male or female, classification of positive or negative sentiments, classification of spam email or non-spam email. While multiclass classifier deals with classification of more than two distinct classes. For example types of soil, classification of music and crop etc.

B. REGRESSION

Regression learning algorithm applies where desired output have a numerical or continuous value as represented in figure 4 [9]. On the contrast of classification learning, regression learning predicts output by a quantity rather than set of possible labels [10]. For example, if we want to predict price of house, speed of wind, rain in centimeters, predicting age of a person, nationality of a person etc. There are many types of regression algorithm in machine learning like linear regression, support vector machines, random forest etc.

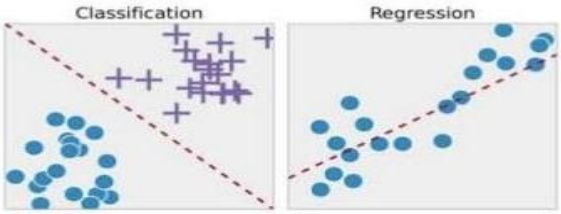

Fig4: Classification Vs Regression

UNSUPERVISED LEARNING

In unsupervised learning only input variable has to be given and no corresponding output needed to given [7]. It is different from supervised learning in a way that in an unsupervised learning there is no teacher, which means no training will be given to the machine. Therefore, machine is restricted to collect the hidden structure by our-self [19]. Unsupervised learning are the types of machine learning algorithm which is mainly used in pattern detection and descriptive modeling. It detects patterns in input data because there are no output categories or labels. In this, the challenge is to collect disordered information according to similarities in patterns without any recommendations of data [18]. For example, an image having apples, bananas and oranges as shown in figure 5. Machine has to categorize these three fruits separately but machine has no idea about the feature of apple, banana and orange. So, unsupervised learning it separates them into the groups by its patterns and similarities, hence we can easily categories the picture into three parts. First part contains all the pictures of apples in it, second part contains all the pictures of oranges in it and third part contains all the bananas in it. As seen we didn't learn anything before, means no training data or example needed in unsupervised learning. Unsupervised learning

January-February 2020
ISSN: 0193-4120 Page No. 4146- 4150

problems are grouped in two major classes called clustering and association problems.

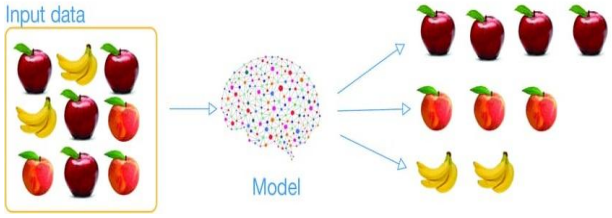

Figure 5: Unsupervised Learning

A. *CLUSTERING*

Clustering is the techniques that group the data on the basis of similarities [12]. Clustering is an unsupervised learning algorithm that finds the similarities in the data and group according to that similarity as shown in figure 6. For example in any market where we dealing with so many customers, this algorithm is very helpful to group different profile of people into separate groups so that it will be easy to target that group according to their needs and it will maximizes the revenue of that market. The advantage of clustering is that it reduces the dimensionality of that data where number of variables to deal. Clustering technique uses K-means clustering and hierarchical clustering algorithms.

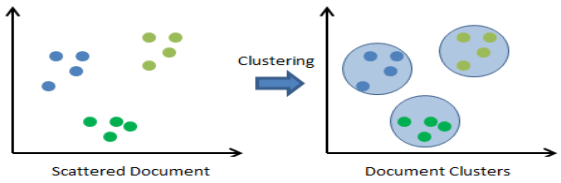

Figure 6: Clustering Technique

B. *ASSOCIATION RULE*

Association rule are the if-then relationship between data items. At basic level it uses the machine learning models to analyze data for patterns in a database. An association rule has two parts one is antecedent and a second is consequent. Antecedent found within dataset and consequents are related with antecedent. This algorithm is very useful in the field of super market so that retailers able to discover links between different items purchased and use this information to predict the likelihood of different products being purchased together [20]. Example relationship between bread and butter, that means the customer who go to a store to buy bread are also likely to buy butter. By using support and confidence values, it detects instance in data. If there is a pair of items to create an instance, if it is calculated, that support is greater than minimum support threshold. The value of support and confidence is calculated by following method as shown in figure 7.

$$Rule: X \Rightarrow Y$$

$$Support = \frac{frq(X,Y)}{N}$$

$$Confidence = \frac{frq(X,Y)}{frq(X)}$$

$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

Figure 7: Association Rule

REINFORCEMENT LEARNING

It is different from supervised learning in a way that in supervised learning labeled output is present so modeled data can train itself according to output but in reinforcement learning, there is no answer is defined but an agent is present who decides what to perform with given task. It learns from experiences. In reinforcement learning one is input (data) which provided at an initial stage from where the model will start. Second is output, as there are many possible solutions of a single problem. And third entity is training which is based upon the input. The model will return reward which is based on output. The best solution decided by the maximum reward.

II. HEALTHCARE INDUSTRY

There are huge amount of data generated regularly by health department which is very complex and large to process and analyze by traditional methods [17]. Healthcare industries have facing many challenges in recent years. Machine learning applications can greatly benefited all the areas of healthcare. For example: It helps to discover effective treatment and best practice that make it easier for patients to find better and more affordable healthcare services. It improves patient-hospital relationship as it is essential at all centers where patients directly interacts with management like physicians' offices, OPD, billing department, patient services, ambulance care facility etc; machine learning also helps healthcare insurers detect fraud and abuse. Some of the highlights of the machine learning application in healthcare industry that emerged recently-

- **Diagnosis**: With the abundance of textbooks, research papers and case studies, a doctor cannot recall all the details of similar cases [6] Yet it is possible to develop machine learning based system that have extensive database which used to predict accurate disease [11] like to diagnose heart disease, cancer disease etc.

January-February 2020
ISSN: 0193-4120 Page No. 4146- 4150

- **Customized Treatments**: On the basis of individuals heath data machine provides personalized medicine for more effective treatment [22]. Supervised learning is used in this application that will receive the symptoms and check them against its database. such as; family health history, medical records, daily habits, heart rate, cholesterol level, allergies and more.
- **Behavioural Changes**: In this model, recognizes hand to mouth gestures in order to help people better understand their behaviour and make changes in them.
- **Drug Discovery**: It is a powerful technology to discover drugs. It can help to discover cancer precision treatment also and reduces the drug discovery time and expenses.
- **Electronic Health Records:** Health related documents are classified with the help of support vector machine and optical character recognition i.e. Change handwriting into digitized character. It uses MATLAB with its ML handwriting recognition technologies.
- **Clinical Trial Research:** Machine learning has various applications which can help directly in clinical trial research. Better sampling techniques would take less time and less cost for trials. Machine learning applications used to find best sample size and medical records to reduce data errors.
- **Epidemic Outbreak Prediction**: Data sources are updated by the environment, satellites, real time updates etc. For example Dengue and malaria like diseases are predict on the basis of temperature, average monthly rainfall, positive no. of case etc.

III. Conclusion

Today, machine learning is the fastest growing field for the researcher which is the base of artificial intelligence and branch of computer science. ML is developing software models in various fields like digital image processing, speech recognition, computer vision and other applications under the umbrella of AI. The main objective of ML is to predict model beyond the training set. Prediction of model required big datasets and should have very high degree of accuracy. Basically, ML works on the datasets and each instance of dataset represented the same features. It may be continuous or discrete. If instances are represented with labels, then the learning is called supervised, while in unsupervised learning, instances are unlabeled. It is reinforcement learning where the training information provided by the environment. ML application in healthcare can have enormous power and utility. Today, healthcare data are used not only for prescription or physicians test, but it has high potential to expand the range and nature of healthcare. This paper can make a contribution to understand the basics of machine learning techniques and enlighten a path to more research in the field of healthcare machine learning.

Acknowledgment

Author is very thankful to Prof. Sanjay Singh, Vice-Chancellor of BBAU, Central University, Lucknow, India for good environment of computation facilities with excellent lab in the University campus. Thanks are also due to my mentor Prof. R. A Khan, BBAU, Lucknow heading of the department for motivating good research which is benefited for the society.

References

[1] R. Bekkerman, M. Bilenko, J. Langford. “Scaling Up Machine Learning: Parallel and Distributed Approaches”. Cambridge: Cambridge University Press, 2011.

[2] A. Gandomi, M. Haider. “Beyond the Hype: Big Data Concepts, Methods, and Analytics” International Journal of Information Management, Vol. 35, no. 2, pp. 137–144, April 2015.

[3] Y. Bengio, “Foundations and Trends in Machine Learning 2”, Now Publishers, Boston, 2009), pp. 1–127.

[4] S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques”, Informatica 31, pp. 249-268, 2007.

[5] E. Alpaydın, “Introduction to Machine Learning”, 2nd ed., The MIT Press Cambridge, Massachusetts London, England, 2010.

[6] A. H. Seh, P. K. Chaurasia, “A Review On Heart Disease Prediction Using Machine Learning Techniques”, International Journal of Management, IT & Engineering, Vol.9, April 2019.

[7] P. Sharma, P. K Chaurasia (2018), “Proposed Algorithm for Secured Transaction using 3-Tier Architecture”, International Journal of Computer Sciences and Engineering, (IJCSE), ISSN: 2347-2693, Vol. 6, Issue 6, pp.316-321.

[8] J. Han, M. Kamber and J. Pei, “Data Mining Concepts and Techniques”, 3rd ed., USA: Morgan Kaufmann Publishers, 2012.

[9] J. Wiens and E. S. Shenoy, “Machine Learning for Healthcare: On the Verge of a MajorShift in Healthcare Epidemiology” Healthcare Epidemiology, Aug 2017, pp. 149-153.

[10] S. J. Russell and P. Norvig, “Artificial Intelligence A Modern Approach”, 2nd ed., New Jersey: Pearson Education Inc., 2003.

[11] B. Venkata Lakshmi, and M. V. Shivsankar, “Heart Disease Diagnosis Using Predictive Data Mining”, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Special Issue 3, March-2014.

[12] K. S. Divya, P. Bhargavi, S. Jyothi, “Machine Learning Algorithms in Big data Analytics”, International Journal of Computer Sciences and Engineering, E-ISSN: 2347-2693, Vol. 6, Issue 1, pp.-64-70, 2018.

[13] R. Bekkerman, M. Bilenko, and J. Langford. “Scaling Up Machine Learning: Parallel and Distributed Approaches”. Cambridge: Cambridge University Press, 2011.

[14] A. Gandomi and M. Haider. “Beyond the Hype: Big Data Concepts, Methods, and Analytics.” International Journal of Information Management, Vol. 35, no. 2, pp. 137–144, April 2015.

January-February 2020
ISSN: 0193-4120 Page No. 4146- 4150

[15] K. Murphy, "Machine Learning: A Probabilistic Perspective", (MIT Press, Cambridge, MA, 2012).

[16] Y. Bengio, "Foundations and Trends in Machine Learning 2, Now Publishers, Boston, 2009), pp. 1–127.

[17] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica 31, pp. 249-268, 2007.

[18] E. Alpaydın, "Introduction to Machine Learning", 2nd ed., The MIT Press Cambridge, Massachusetts London, England, 2010..

[19] Min Chen, , Yixue Hao1, Kai Hwang,Lu Wang, and Lin Wang "Disease Prediction by Machine Learning OverBig Data From Healthcare Communities", Vol. 5, June 2017, pp. 8869-8879, .

[20] J. Han, M. Kamber and J. Pei, "Data Mining Concepts and Techniques", 3rd ed., USA: Morgan Kaufmann Publishers, 2012.

[21] Regression vs. Classification algorithms webpage on Data Science [online]. Available: https://www.datascience.com/blog/regression-and-classification-machine-learning-algorithms.

[22] Stuart J. Russell and Peter Norvig, Artificial Intelligence A Modern Approach, 2nd ed., New Jersey: Pearson Education Inc., 2003.