International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

A Study on various Machine Learning Algorithms and their Applications in Contemporary World

Sidharth S Prakash[1]

[1]*M.Tech in Software Systems, Department of Information Technology, School of Engineering, CUSAT*

Abstract - *Machine learning is a subset of the broad domain of artificial intelligence. It is the process of training a machine to think and use that intelligence to make efficient and effective decisions. These decisions are taken by the machine on its own by capturing the various patterns it has deduced when analyzing a huge amount of data provided to it during the training phase. So the rule based programming by a human being can be completely avoided when using machine learning. Hence machine learning has got immense popularity in the current research domains. This paper illustrates the various machine learning techniques and their applications in contemporary world.*

Key Words: **Machine learning, Supervised learning, Unsupervised learning, Reinforcement learning, Artificial intelligence**

1. INTRODUCTION

Artificial intelligence has revolutionised the modern world. Machine learning is a subset of artificial intelligence domain which allows a machine to gain intelligence by analyzing huge amount of data and deriving patterns from it. This gained intelligence is used to make efficient and effective decisions in real world problems without explicit rule based programming by a human being. Hence the human effort in the machine learning process is highly reduced and this is the reason for the huge popularity of this technique [1].

Machine learning is utilized in various real world applications in modern world. This paper aims at discussing the various machine learning techniques and their applications in real world problems.

The rest of the paper is organized as follows. The section 2 describes the classification of machine learning techniques. This is followed by section 3 which discusses some important machine learning algorithms. The section 4 details the application of the machine learning techniques in real world and finally the paper is concluded in section 5.

2. CLASSIFICATION OF MACHINE LEARNING

Machine learning techniques are broadly classified into three categories – supervised learning, unsupervised learning and reinforcement learning [2] as shown in figure 1.

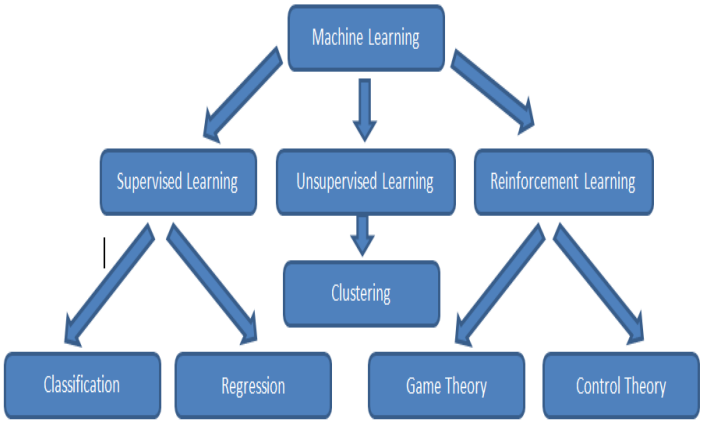

Fig -1 : Classification of Machine Learning Techniques

2.1 Supervised Learning

In supervised learning, the machine is fed with labelled data and training is done using this labelled data set. The model is then validated against a test data set and the accuracy is calculated before the deployment.

The two main use cases of supervised learning is classification and regression. In classification process, the training label will be a category – like "Dog" or "Cat", or "No Disease" or "Disease". The categories are not limited to two, and can go to any number based on the use case. In regression process, the training label will be a continuous value for a set of features and then the model will predict an unknown value for a set of new features.

2.2 Unsupervised Learning

In unsupervised learning, the machine is provided with an unlabelled data and the machine detects patterns from the provided data and draws conclusions.

The main use case of unsupervised learning is clustering applications. Based on the fed data, the machine finds patterns and group the inputs based on the common features into different groups. The main difference of unsupervised learning is the absence of human supervision in the training process.

2.3 Reinforcement Learning

In reinforcement learning a machine is expected to achieve a goal through a set of actions. All actions taken in the situation are categorized as a reward or penalty. The machine aims to attain the goal by maximising the reward and minimising the penalty.

The main use case of reinforcement learning is in game theory or control theory. In game theory, the machine is taught to play a game against a human opponent. In control theory agents are designed to take effective decisions in a control system like power plant.

3. VARIOUS MACHINE LEARNING ALGORITHMS

In this section, I will discuss on the following machine learning algorithms – Linear Regression, Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Decision Trees, Naive Bayes and Neural Networks.

3.1 Linear Regression

Linear regression aims to model the relationship for a dependent variable on one or more independent variables [3]. This process can be used to predict the unknown value of the dependent variable when the set of know values of independent variables are provided.

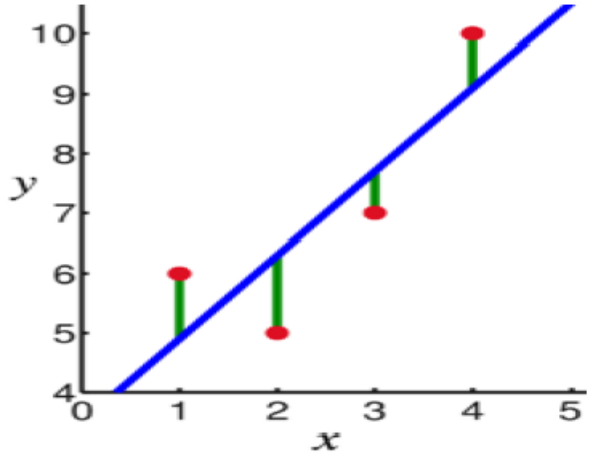

Fig -2 : Linear Regression Model

In the above figure 2, the blue line represents the underlying relationship between the independent variable x and dependent variable y. As a result of random derivations that are marked in green from the relationship, the values of y are predicted and marked in red.

3.2 Support Vector Machine

Support Vector Machine (SVM) is used predominantly for classification tasks. SVM divides the data into various classes by creating a hyperplane [4]. The data points nearest to the hyperplane are called support vectors which predict the category of classification.

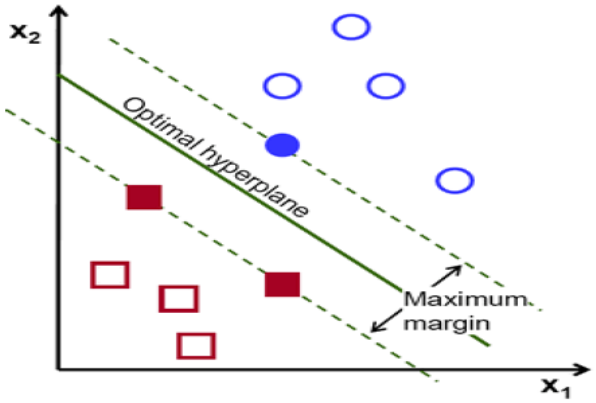

Fig -3 : Representation of Support Vector Machine

In the above figure 3, the two datasets are squares and circles. They are separated by an optimal hyperplane. The circle in blue and two squares in red are the support vectors as they are the closest data points to the hyperplane. The maximum margin that separates the two data is also represented.

3.3 K-Nearest Neighbour

K-Nearest Neighbour is a clustering algorithm which ensures that the similar items in the unlabelled data set are grouped together. This process of keeping the similar data items in close proximity is carried out by the process of pattern matching [5]. When a new data is provided to the machine, it assigns the provided data to the cluster which has more number of neighbours belonging to the specific cluster.

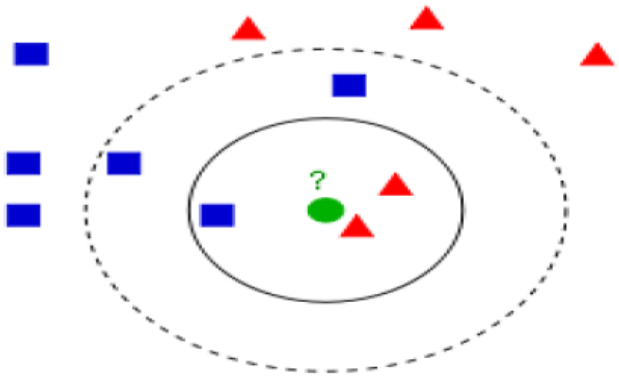

Fig -4 : KNN classification illustration

In the above figure 4, there are two clusters of data – squares and triangles. The green circle is the test data provided to the machine. If the algorithm considers K = 3, the inner circle becomes the point of consideration. In this circle there are 2 triangles and only 1 blue square. Hence the test data will be assigned to the class of triangle. It is also worth noting that if the algorithm considers K = 5, the outer dashed circle become the point of consideration and as the number of squares are more in this, the test data will be classified as a square.

3.4 Decision Trees

A decision tree is constructed graphically using the branching method which specifies the outcome for each of

the conditions. Each node is treated as an input variable and branches are created based on the values of the input variables. The branching paths are represented from root node to the leaf node [6].

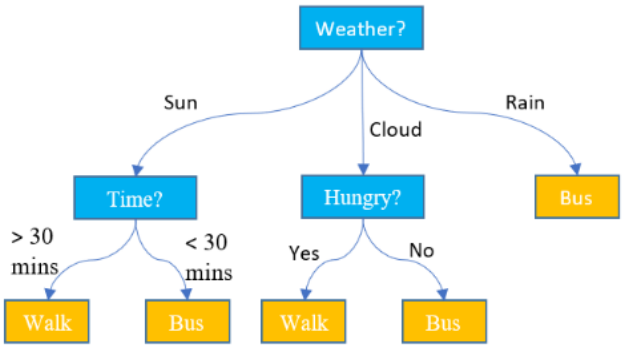

Fig -5 : Decision tree example

In the above figure 5, a decision tree is represented denoting the possible methods of travel for the each situation depicting the weather condition, hunger and time available in hand. Hence if it is a sunny weather and there is enough time, a person may most likely to opt to walk rather than spending money by travelling on a bus.

3.5 Naive Bayes

Naive Bayes is a probabilistic classifier utilising the Bayes theorem for conditional probability [7]. It depicts the probabilities of outcomes and is mostly used when the input complexity is very large. The Bayes theorem is as follows:

$P(C|A) = (P(A|C) * P(C)) / P(A)$
Where

$P(C|A)$ = probability of hypothesis h given data d. This is also called posterior probability.

$P(A|C)$ = probability of data d given hypothesis was true.

$P(C)$ = probability of hypothesis h being true regardless of data. This is also called prior probability of h.

$P(A)$ = probability of data irrespective of the hypothesis.

3.6 Neural Networks

A neural network tries to mimic the working process of the human brain [8]. The human nervous system is made up of interconnected neurons. Similarly an artificial neural network is made of interconnected nodes that process input information as well as provide the output to other nodes if it gets activated.

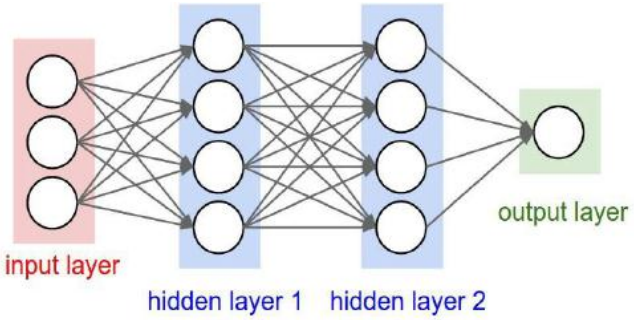

Fig -6 : Neural network model

In the above figure 6, it is observed that three nodes exist in the input layer. These three nodes are connected to the first hidden layer consisting of 4 nodes and these four nodes are connected to the second hidden layer which also consists of four nodes. Finally the second hidden layer is connected to an output node which is capable of computing a binary classification. As the number of classification classes increases, the output layer consists of that many nodes.

4. APPLICATIONS OF MACHINE LEARNING

This section discusses the various applications of machine learning techniques in the real world scenarios.

4.1 Health Care

Machine learning techniques have found a way into medical diagnosis. Many trained models are able to detect cancer tumours [9] and detect the probability of pneumonia from chest X-rays [10] has shown below in figure 7.

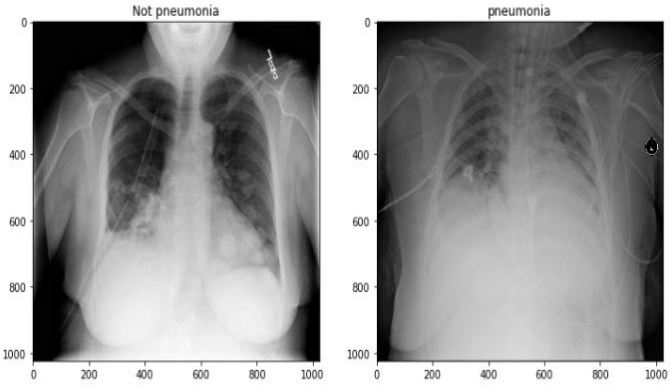

Fig -7 : Machine learning model to processing chest X-rays

These machine learning models have immensely reduced the manual effort from the side of the medical staffs and more sophisticated algorithms are expected to aid the medical diagnosis in the years to come. Moreover, machine learning techniques are employed in the drug discovery process [11] and the time to market a new drug has reduced to a great extent.

4.2 Email Spam Filtering

Decision trees and neural networks are employed for filtering out spam emails. Rule based spam filtering

techniques fail to detect the latest spam email tricks which are constantly evolving. But machine learning algorithms can adapt to these dynamic spam tricks efficiently [12].

Multilayer perceptron model, which is a neural network model, is effectively used by the modern email clients to successfully classify a potential spam email and send it directly to spam folder, without bothering the email client user.

4.3 Online Fraud Detection

Machine learning techniques have the capability to detect anomalies. This feature is employed in finding the fraudulent transactions in the cyberspace [13].

Financial enterprises utilise the power of machine learning and train the models on large amount of legitimate and fraudulent data from history. This knowledge base can be used by the model to flag a fraudulent transaction in the future through the process of pattern matching.

4.4 Sports Analytics

Machine learning can be used in the sports analytics domain to measure the performance of various sportsmen [14]. This can be utilised in the team formations of future games when the sports event is team based or improve the performance of a particular sportsman in case of individual event.

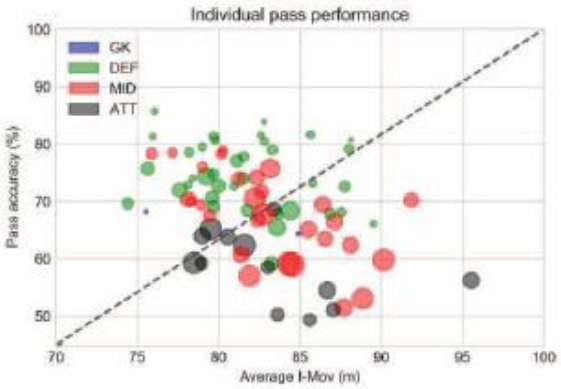

Fig –8 : Determination of pass accuracy in football

Machine learning algorithms can analyse huge amount of data with greater precision and accuracy than a human based analysis. Hence these techniques are deployed by various coaching staffs which are into serious sports.

4.5 Climatology

Weather predictions for future can be done efficiently by machine learning algorithms [15]. For efficient predictions, a huge amount of historical data needs to be processed by the model which is practically impossible for a human being.

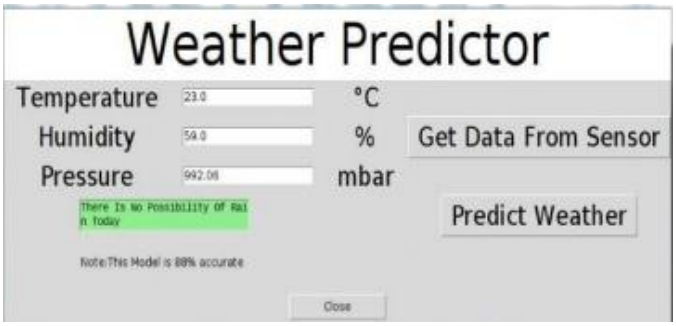

Fig -9 : Screenshot of a Weather Predictor App

Rule based programming is not efficient in climatology as the weather patterns are very dynamic and all the dependant factors cannot be brought into consideration manually.

3. CONCLUSION

Artificial intelligence is the new electricity. Just like how the discovery of electricity revolutionised the modern era, artificial intelligence powered by machine learning techniques will move the future world. In this paper I have discussed the basic classification of machine learning techniques – supervised learning, unsupervised learning and reinforcement learning. Also the six major machine learning algorithms like Linear Regression, Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Decision Trees, Naive Bayes and Neural Networks were also discussed. The paper concludes by discussing the various applications where machine learning is deployed in the modern world.

REFERENCES

[1] R. Kamalian; E. Yeh ; Ying Zhang ; A.M. Agogino ; H. Takagi, "Reducing Human Fatigue in Interactive Evolutionary Computation Through Fuzzy Systems and Machine Learning Systems". IEEE International Conference on Fuzzy Systems, July 2006.

[2] Prajwal Shimpi; Sanskruti Shah; Maitri Shroff ; Anand Godbole, "A machine learning approach for the classification of cardiac arrhythmia". International Conference on Computing Methodologies and Communication, July 2017.

[3] Kavitha S ; Varuna S ; Ramya R. "A comparative analysis on linear regression and support vector regression". Online International Conference on Green Engineering and Technologies, November 2016.

[4] Jianxin Wu; Hao Yang. "Linear Regression-Based Efficient SVM Learning for Large-Scale Classification". IEEE Transactions on Neural Networks and Learning Systems, 26(10), pp. 2357 – 2369, January 2015.

[5] Jie Huang; Yongqing Wei; Jing Yi; Mengdi Liu. "An Improved kNN Based on Class Contribution and Feature Weighting". 10th International Conference on Measuring Technology and Mechatronics Automation, February 2018.

[6] Siddalingeshwar Patil; Umakant Kulkarni. "Accuracy Prediction for Distributed Decision Tree using Machine

Learning approach". 3rd International Conference on Trends in Electronics and Informatics, April 2019.

[7] Jiangning Chen ; Zhibo Dai ; Juntao Duan ; Heinrich Matzinger ; Ionel Popescu. "Naive Bayes with Correlation Factor for Text Classification Problem". 18th IEEE International Conference On Machine Learning And Applications, December 2019.

[8] Yuxuan Huang; Luiz Fernando Capretz ; Danny Ho. "Neural Network Models for Stock Selection Based on Fundamental Analysis". IEEE Canadian Conference of Electrical and Computer Engineering, May 2019.

[9] Turki Turki. "An empirical study of machine learning algorithms for cancer identification". IEEE 15th International Conference on Networking, Sensing and Control, March 2018.

[10] Enes Ayan ; Halil Murat Ünver. "Diagnosis of Pneumonia from Chest X-Ray Images Using Deep Learning". Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science, April 2019.

[11] Harry F. Rickerby ; Katya Putintseva ; Christopher Cozens. "Machine learning-driven protein engineering : a case study in computational drug discovery". IEEE Engineering Biology, 4(1), pp. 7 – 9, April 2020.

[12] Shatabdi M. Bhalerao ; Madhuri Dalal. "Improved social network aided personalized spam filtering approach using RBF neural network". International Conference on Intelligent Computing and Control, June 2017.

[13] Sonal Kataria ; Md Tabrez Nafis. "Internet Banking Fraud Detection Using Deep Learning Based on Decision Tree and Multilayer Perceptron". IEEE 6th International Conference on Computing for Sustainable Global Development, March 2019.

[14] Matthias Kempe ; Floris R. Goes ; Koen A. P. M. Lemmink. "Smart Data Scouting in Professional Soccer: Evaluating Passing Performance Based on Position Tracking Data". IEEE 14th International Conference on e-Science, November 2018.

[15] Nitin Singh; Saurabh Chaturvedi ; Shamim Akhter. "Weather Forecasting Using Machine Learning Algorithm". IEEE International Conference on Signal Processing and Communication, December 2019.