International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp.4750– 4762

Study and application of various machine learning algorithms in the contemporary world

Hassan A. Alessa,
Ph.D, Mechanical Engineering, Najran University, Saudi Arabia
drhassanalessa@gmail.com

Abstract

In the field of computer science, the idea of Machine Learning is one of the fascinating innovations that make our jobs simpler in our everyday lives. Machine Learning is being used to solve real-world problems by storing, manipulating, extracting, and retrieving data from significant sources. It's a powerful tool. It refers to the creation of computer programmers that can access data and learn for themselves without human intervention. Various types of machine learning algorithms and their implementations are reviewed in this article.

Keywords:*Random forest,Machine learning, Supervised, Reinforcement, Regression*

1. INTRODUCTION

The modern program relates to a manually constructed system that allows input data that creates and runs virtual machines to produce output. In artificial intelligence, input/ output data is fed into a software creation algorithm. The following schematic Figure-1 demonstrates the difference between machine learning and traditional programming.

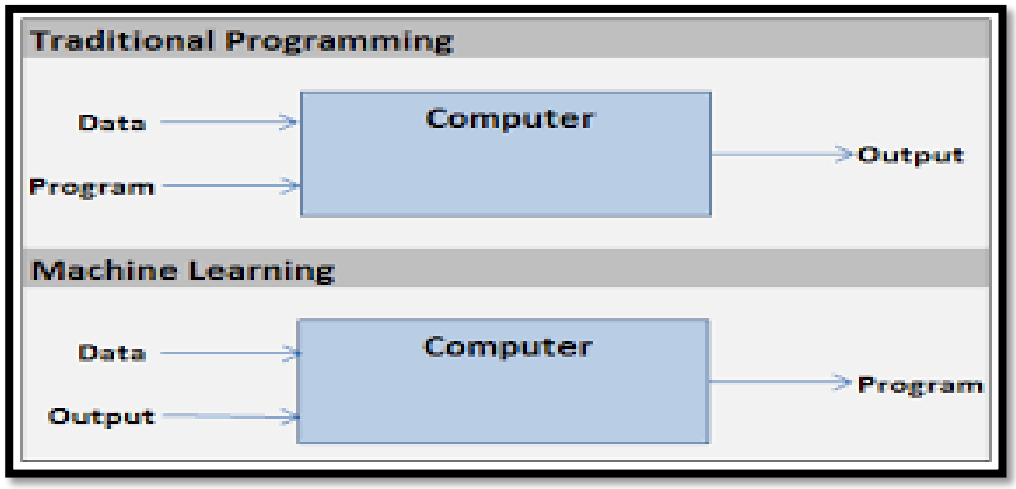

Fig 1.1 Differences between Machine and LearningTraditional Programming

Machine learning aims primarily at creating computer programmes that can handle and use data for themselves. Learning starts with experiences or data, such as examples, direct experience, or guidance, so that in the future, we can look for models in data and make better decisions based on the examples that we have[1][2].

Over the previous years, machine learning has gained momentum: vehicles, effective internet browsing, and language and image detection. Promising findings are increasingly extending to our everyday lives. ML is a class of Ai techniques that allows the machine to work in a self-learning format without being specifically programmed. It has a broad range of applications in different domains such as bio-

ISSN: 2233-7857IJFGCN

engineering, Intrusion identification, Retrieval of information, Game Play, Advertising, malware protection, and Image Deconvolution. It's a very fascinating and dynamic subject that could push the future of technology. [14]

A neural network is a machine learning algorithm focused on the theory of biological neural networks organization and working. It is made up of single units said to be Neuron. Neurons are found in a variety of layers of classes. Neurons are bound to the next layer of neurons in each layer. Information is collected from the input neurons on certain compounds to the output layer. Each node carries out an essential mathematical estimation. It then transfers its data to all the nodes with which it is linked.

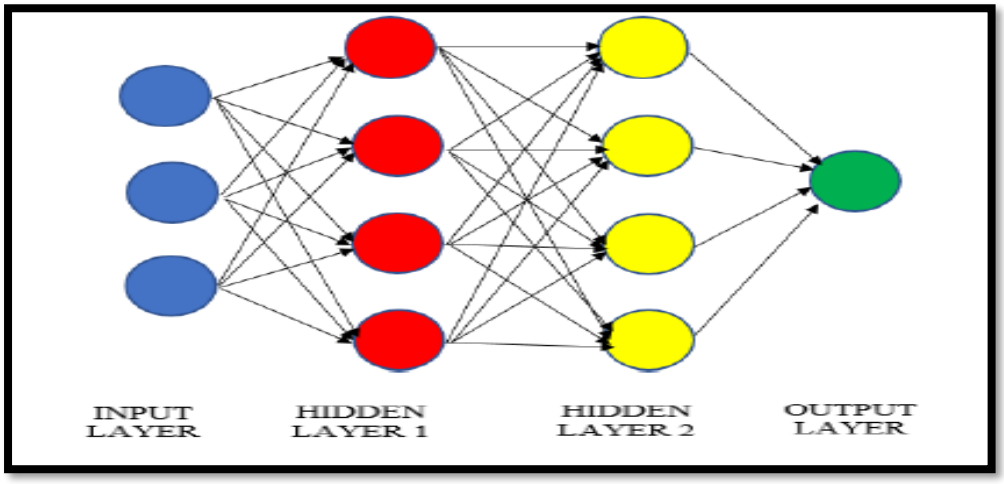

Figure 1.2 Neural Networks

The last wave of neural networks was related to the rise in computing capacity as well as the development of experience. This brought profound learning, in which the technological architectures of neural networks are becoming more dynamic and have the ability to solve a broad range of tasks that were not effectively resolved before [15].

Deep learning is a form of machine learning which utilizes computer systems to iterative manner calculation patterns on their own [16]. It seeks to learn attribute hierarchies with characteristics from higher hierarchies created by the configuration of lower tiers. Automatically learning functions at various abstraction levels enable a framework to learn complicated tasks to map the entry explicitly from the data, without entirely depending on human-made features [17].

One of the significant issues of the deep learning category is poor performance due to over-fitting. The fit occurs when the model matches the training set too well. The model then finds it hard to extrapolate new examples which were not used in the learning. To combat this issue and enhance performance, it is vital to have a large amount of data supplied by using CNN [18] [19].

The deep learning system employed during image classification is the Convolutionary Neural Network (CNN) [19]. CNN learns from the input image explicitly and removes manual extraction features [20].

A deep learning neural network is focused on Kera, and tensor flux, which uses python for image classification, is built in this article. Here we are using the array of various images, i.e., a database that includes two animal types, including cat and dog that used to train the device.

ISSN: 2233-7857IJFGCN

Fig 1.3 Sample Dataset

Multiple classifiers, including Sigmoid, softmax, and the CNN activation function Relu, are contrasted in this study. Tensor flow is an open-source software library that provides several tasks for data-flow programming. This is used for deep learning image recognition and applications in machine learning, including neural networks.
Keras is a python-written open-source neural network library. It is built to assist the fast testing of deep neural networks.

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

1.2. MACHINE LEARNING USES

- **Face recognition**-Facial recognition is one of the enormous characteristics that only machine learning develops
- **Videos Surveillance-**It involves monitoring violence or missed incidents before they occur. It can help track people's unusual acts, like sitting on their banks and staying still for an extended period of time, slipping, respectively. They will immediately alert people who are all posted there and helped avoid issues.
- **Social Networking Platform-**Social media can be used to provide improved social media feeds and publicity in users' interests, mainly through machine learning. Examples include input from friends, comments on Facebook accounts, albums, and videos online.
- **Spyware and malicious software-**Email clients are increasingly using different junk mail philters modified and primarily used in machine learning. Similarly, viruses were found and detected predominantly by system security programmers that are mostly enabled only by machine learning.
- **Search Engine**- Search Engine uses a wide range of machine learning algorithms.
- **Apps / Companies**-Many applications and companies have used computer education to do every day work since this is quite reliable and accurate than human action. These were all Facebook, Google Maps, Gmail, Google Search., etc. [3] [4] [5].

1.3. Machine Learning and Its Types
Following figure shows the types of Machine learning.

ISSN: 2233-7857IJFGCN

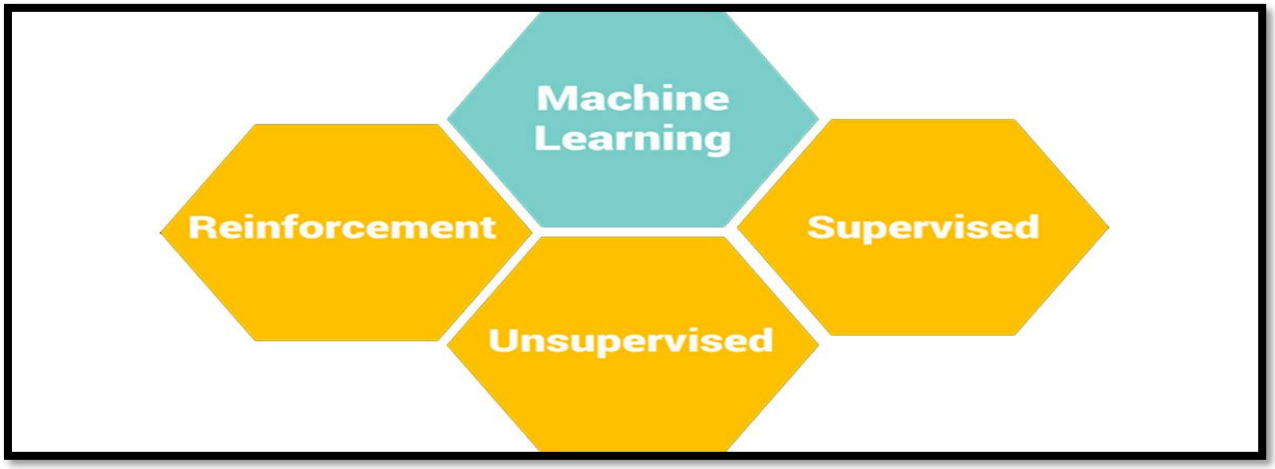

Fig 1.4 Machine learning types

1. Supervised learning: it helps the model to predict future outcomes after training, dependent on previous experience. It means that the computer learns from the data identified. An example of supervised learning is that we have to remember the vehicle, and we also have data on cars such as tyres, air suspension, engines, steeringand lighting.Still, only vehicles must be identified, so additional features like vehicle height, seating capacity, and vehicle model are identified. This information is called if it has these attributes, it is likely to be auto. We must map our performance in supervised learning. This previous information or details are referred to as a database. Dataset is divided into two classes, such as training and validation collection. The training set includes information on all aspects. From the above, we understand that we try to derive features from the trained data in supervised learning. According to the learning algorithm, the training model is used to predict future results if we know the right responses from past data.Using a statistical improvement method that is based on trial and error, the machine progressively improves accuracy by checking outcomes against a collection of values given by the supervisor.

2. Unsupervised learning: Unsupervised learning is a class without information on the machine. The algorithm can work on the story with no prior experiences. We must figure out that secret patterns of data Unsupervised learning in exploratory data analysis are beneficial, as hidden patterns can be identified automatically from data. Training data includes input values without associated output or objective values, according to the specified pattern. The aim of uncontrolled learning is to cluster the data across distinctively different classes. Data may be organized in various groups based on the way they are grouped. Clustering is linked in broad sensory data to find similar groups or similar individuals. We can explain, with an example, the idea of unsupervised learning. Assume that we're looking at news online through Google searches like news.google.com; there are many hypertext links open to us. When we pick a news link, it can contain a different link that splits the news into categories such as political, climate, style, and many links. If we tapped on all links, the websites of this news would be different. This single hyperlink contains a variety of sub-hyperlinks, in which this approach is merely clustering. Clustering is related to data groups. [6].

3. Reinforcement learning:
Reinforcement learning is supervised learning. But as we know about supervised learning, we don't see the performance. We know only in reinforcement strategies to achieve the best activities we have. This method is also said to be a method of reward. This definition relies on the model's behavior. Reinforcement learning components are 1) Environment 2) Agent 3) Condition 4) Reward 5) Behavior. We would use a self-driving car to grasp reinforcing learning principles. The car system's just an

ISSN: 2233-7857IJFGCN

employee: path, traffic, and all these so-called ecosystem signals. The state is like a traffic signal means we have to stop when it is red. And action has to be taken after displaying the green signal car. If our auto drive is run on a green signal, recognized as a positive reward, it will be a detrimental reward if it does not run on a Green signal. The agent's main objective is to enhance the anticipated total award.He delayed returns they generate. Strengthening learning is referred to as sequential decision-making. [7].

2. ALGORITHMS of MACHINE LEARNING

The ML algorithm is a normal algorithm evolution. Anything "smarter" our strategies by making them recognize the data we have provided instantly. The algorithms for ML are ranked as follows.

2.1 Supervised Learning Algorithm

This algorithm is composed of a variable target/output (or dependent variable) to be predicted by several predictors (independent variables). We create a function using these variables to map inputs to the expected output. The training procedure goes on until the model is accurate on the training data.
Examples: linear regression, decision trees, random forests, KNNs, logistic regression, etc.

1. Linear Regression

Linear regression is a supervised machine learning algorithm that plots a linear relationship of a target variable against other variables. The model plots the available points from the dataset on a graph and finds a statistical relationship between the scalar variable with the target variable [4]. A statistical relationship differs from a deterministic relationship in the sense that it cannot be always accurately predicted. The basic idea of this algorithm is to plot a line that can be used to represent the trend of the target variable against other variables with minimum error. An error in this model can be defined as the closest distance from the point to the line. Once the line is plotted, it can be used to predict the value of the target variable for future independent variables.

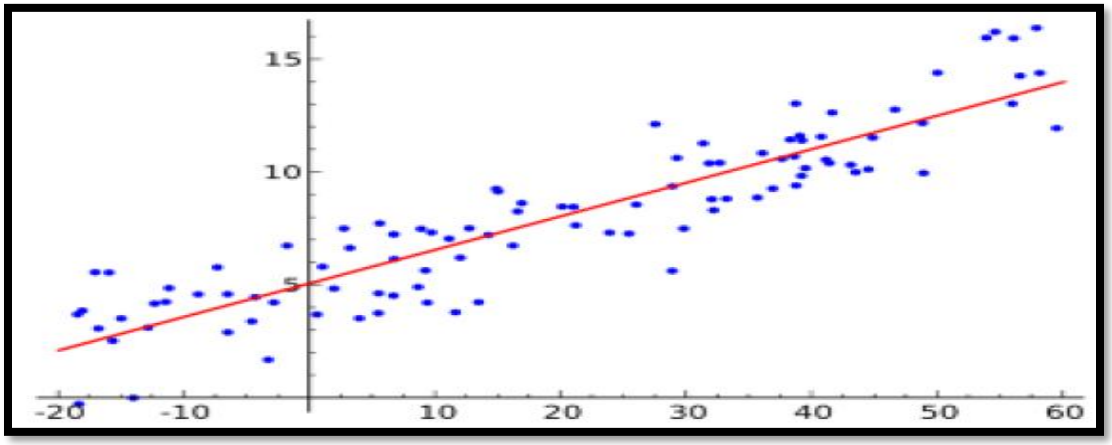

Fig 2.1 Example of Linear Regression

The equation of Linear Regression line can be given as follows:

$$Y= B_0+B_1X$$

Here, 'Y' represents the target variable to be predicted, X is the independent variable or predictor. 'B0' is the intercept and B1 is the coefficient which represents the slope of the line. To check the accuracy of our model, the R-square metric can be formulated as follows:

$$R^2 = 1- \frac{RSSS}{TSS}$$

ISSN: 2233-7857IJFGCN

Here, 'RSS' is the Root Mean Square which is the squared difference between the observed value and the predicted value. 'TSS' is the Total Sum of Squares which is defined as the squared difference between the observed value and the mean. Typically, a high R2 value is associated with a good model, but it often depends on what the application considers as a fit model. Linear Regression is popular with applications that have a linear relationship but not exactly predictable. For example, changes in house prices with respect to the area or height of a person concerning his or her age.

2. Decision Tree

This is a type of supervised learning algorithm employed primarily for problems with classification. For example, if you want to buy shampoo on the market. Next, you can examine whether you still need shampoo. If you're out, you'll need to purchase it from the store. You will also look outside and analyze the temperature. That's to say, if it rains, you won't go, and if it doesn't, you won't. We can intuitively imagine this scenario with the accompanying Fig-3.

Fig 2.2 Decision Tree Example

Decision Trees can be further classified into two types – Categorical Variable Decision Tree and Continuous Variable Decision Tree. This is based on the nature of the target variable. For example, making financial decisions can only have outcomes like 'Yes' and 'No'; there cannot be a 'Maybe' outcome as it does not fit this application. But some qualities, like the willingness of a customer to buy a product, can be visualized on a continuous scale. Tree-based algorithms empower prognosticative models with high accuracy and easy interpretation. They are used to map non-linear relationships as well. Techniques like decision trees, random forest and gradient boosting arebeing popularly used in all kinds of data science problems. They are useful to solve classification as well as regression problems.

We may construct a hierarchical tree with the same concept in order to achieve our production through several decisions. The decision-making tree is built in two ways – induction and pruning. We create a Decision Tree in Induction and in pruning simplify the Tree by eliminating various complexities [8][9][10].

3. Random forest

A supervised learning algorithm for regression and classification is Random Forest. Nevertheless, it is primarily used for classification issues. As we know, there are trees in a forest and more trees, which mean more robust forest. Random forest algorithms also create decision trees for data samples, then predict each sample and then choose the optimum solution via a vote. It is a better method than a single tree, as averaging the outcome prevents over fitting.

ISSN: 2233-7857IJFGCN

4. kNN (k- Nearest Neighbors)

kNN is one of the supervised machine learning algorithms for data and machine learning that we use. This classifier then learns the patterns in the classifier rely upon similar results. It is a non-parametric and lazy algorithm for learning. In lazy loading, the training of data points for model generation is not needed. The training data are used in the test phase, causing the test phase to be slower and more expensive than the training phase.

KNN makes predictions based on the outcome of the 'K' neighbors closest to that point. Usually odd value of 'K' is selected to avoid ties during voting.

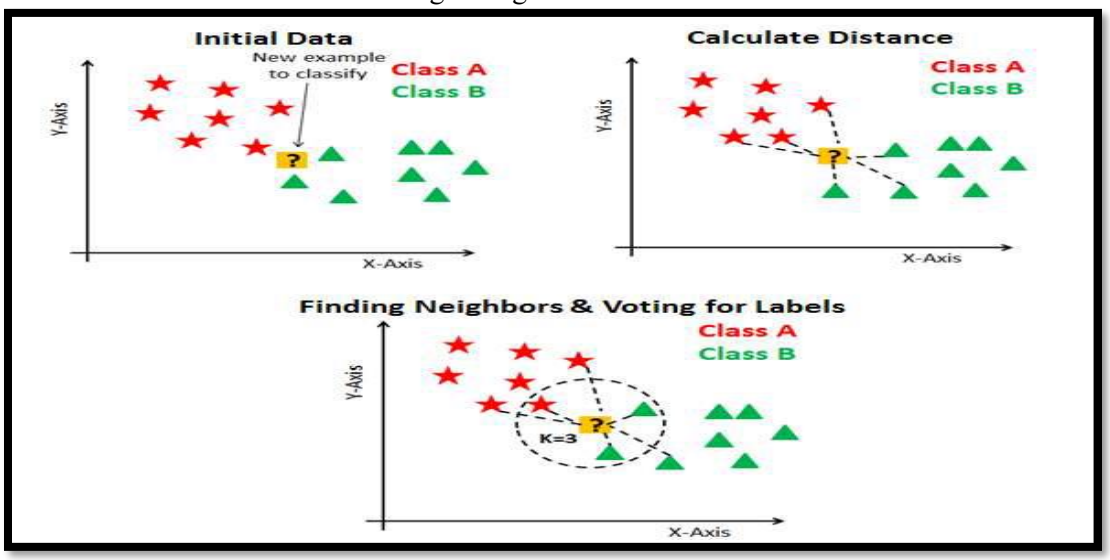

Fig 2.3 Steps for KNN Algorithm

KNN can out-perform many other classifiers in selected applications that cannot entirely rely on features of their datasets. Sometimes it becomes impossible to fill in the missing data into the dataset. Hence, plotting the available features on a graph and classifying new points based on their proximity to original points becomes a better alternative. Applications like text mining and classification, and fingerprint and pattern recognition use KNN to achieve their goals optimally.

5. Logistic Regression

It is a classification, not an algorithm for regression. It is used for estimating discrete values based on a set of independent variables (binary values like 0/1, yes/no, tr / false). It predicts the likelihood of an event occurring by adding data to a logistic function in simple words. Therefore, logistical regression is also known. Since the probability indicates, its performance values range from 0 to 1.

2.2 Unsupervised Learning Algorithm

I this algorithm, we have no predict / evaluate goal or result variable. It is used to cluster people into various categories, which are commonly used for consumer segmentation in different groups for specific intervention. Unattended learning examples: Apriori algorithm, K-means.

1. Apriori Algorithm

Apriori is used primarily for the sorting of large quantities of data. Association also rules sort of data. Regulations help to demonstrate which aspects that set of data have in common. These association rules can then be used to construct categories. Data in classifications allow users and algorithms to recognize new patterns and structure data sets. Over time, they will be better able to pick out patterns. An artificial neural network utilizing Apriori can also weigh various categories to increase or decrease

ISSN: 2233-7857IJFGCN

their significance. This helps an artificial neural network to process data, recognize trends, and establish patterns that otherwise would have been missed. Apriori is hugely beneficial in many areas for data analysts. Its significance only continues to grow as more and more regions make sense of large data sets using artificial intelligence. Apriori will remain a crucial tool in advancing machine learning & artificial intelligence in the coming years.

2. K-Means

The classical version of the KMeans algorithm used Euclidean distance as a proximity measure among data points plotted on a two-dimensional plane. Once all the points are allotted to some cluster, newcentroids are calculated as the barycenter of each cluster, and the process is repeated until a fixed criterion is met, significant difference in the clusters [6]. Although the algorithm can be proved to terminate in every case, it does not guarantee optimal partitioning of data into clusters. Also, there are no guides for selecting the number of clusters and the initial representatives which lead to varied results with different values.

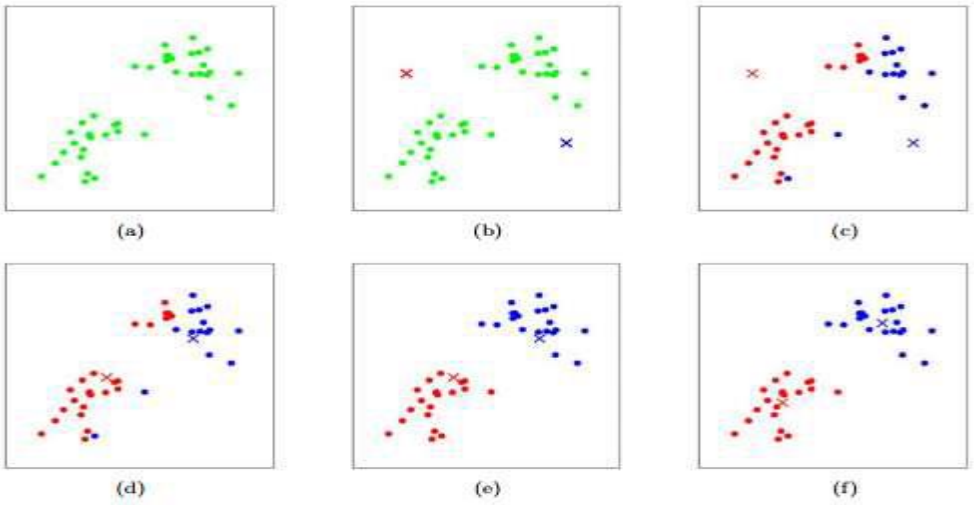

Fig 2.4 Steps for K-Means Algorithm

2.3 Reinforcement Learning Algorithm

The machine is trained to make precise decisions with this algorithm. This works: the machine is subjected to an environment in which trial and error are continuously used. This machine learns from previous experience and aims to obtain the best information possible to make accurate business decisions [11] [13].

1. Markov Decision Process

The Markov Decision Method, more commonly known as the MDP, is an approach for enhancement learning in the grid world. A grid world environment is made up of grid states. The MDP attempts to grid a world by splitting it into states, behavior, models, and rewards.

The agent shall observe a state and carry out an operation that decreases intermediate costs. The expense and the form of succession depend only on the current situation and the acts chosen. For instance, sometimes, an action cannot achieve the desired target state but stays with a small probability in the current state [12].

2. Evolutionary algorithm

The evolutionary algorithm begins entirely at random with code generation. Each of these codes is checked to see if the desired target is achieved. This is how the code evolves. Over time, it gets better and can be better than any human coder designs after several generations if conditions are right. It involves

ISSN: 2233-7857IJFGCN

preserving a distribution across network weight values, which consists of using parameters sampled from a wide range of agents. Each agent works in its own setting, and once several episodes or steps of an episode have been completed, a cumulative award shall be returned to the algorithm for health. With this score, the distribution of parameters can be shifted to that of the most active and not successful agents. By following this strategy millions of times, the distribution of weight with hundreds of agents transfers to space, which provides agents with an excellent policy to solve this task [13].

3. APPLICATIONS

1. Health Care

Machine learning techniques have found a way into medical diagnosis. Many trained models are able to detect cancer tumours [14] and detect the probability of pneumonia from chest X-rays [10] has shown below in figure 7. 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.

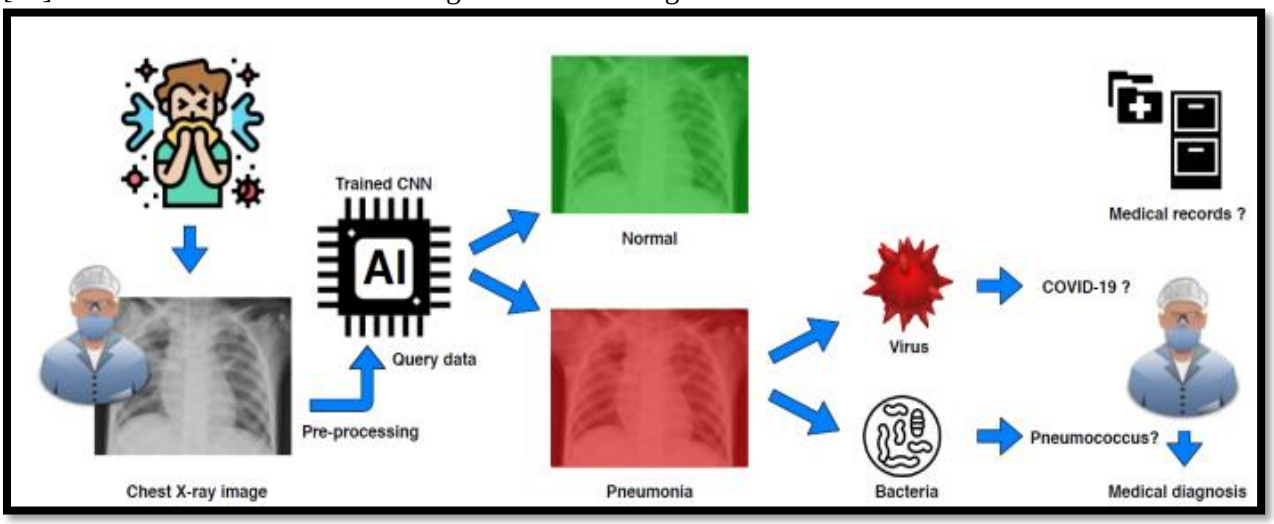

Fig 3.1 Machine learning model to processing chest X-rays

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 [24].

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.

ISSN: 2233-7857IJFGCN

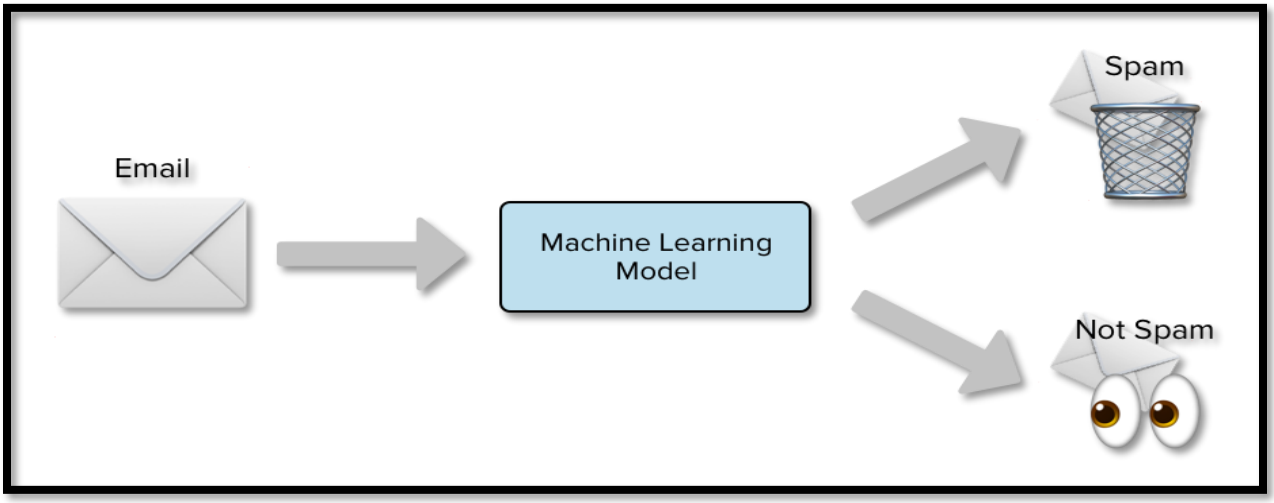

Fig 3.2 Spam Filtering

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 [23]. Financial enterprises utilize 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.

Fig 3.3 Online Fraud Detection

4 Sports Analytics

Machine learning can be used in the sports analytics domain to measure the performance of various sportsmen [21]. This can be utilized 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. 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.

ISSN: 2233-7857IJFGCN

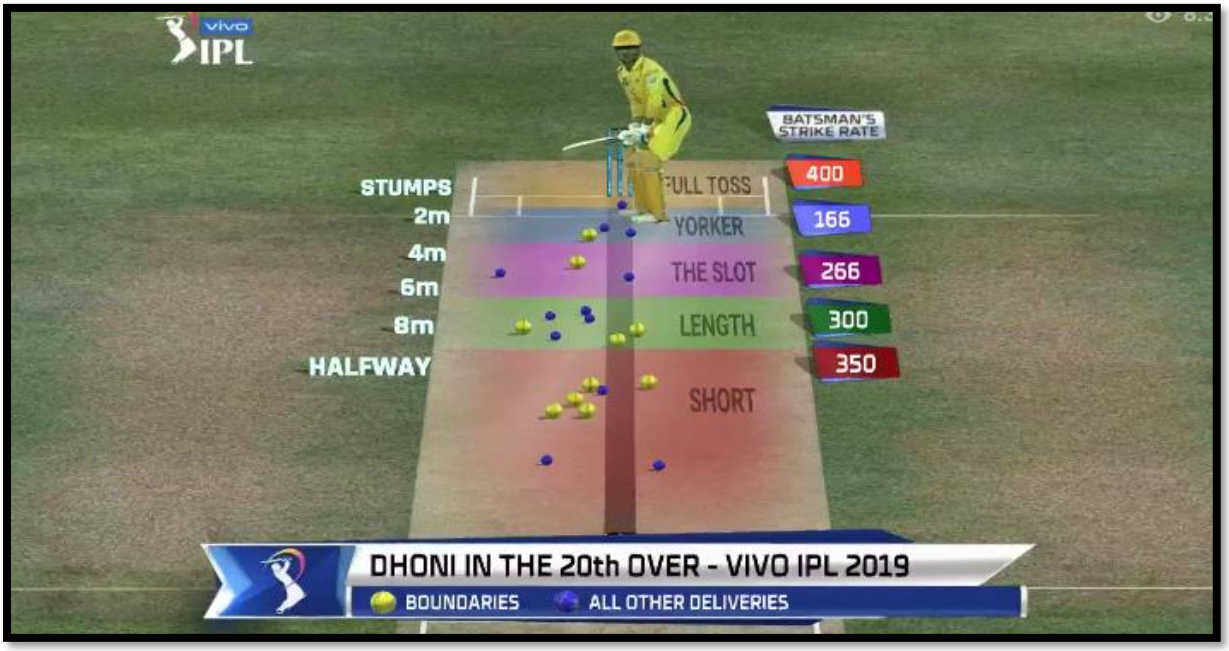

Fig 3.4 Batsman strike rate at different lengths of pitch

5 Climatology

Weather predictions for future can be done efficiently by machine learning algorithms [22]. For efficient predictions, a huge amount of historical data needs to be processed by the model which is practically impossible for a human being. 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.

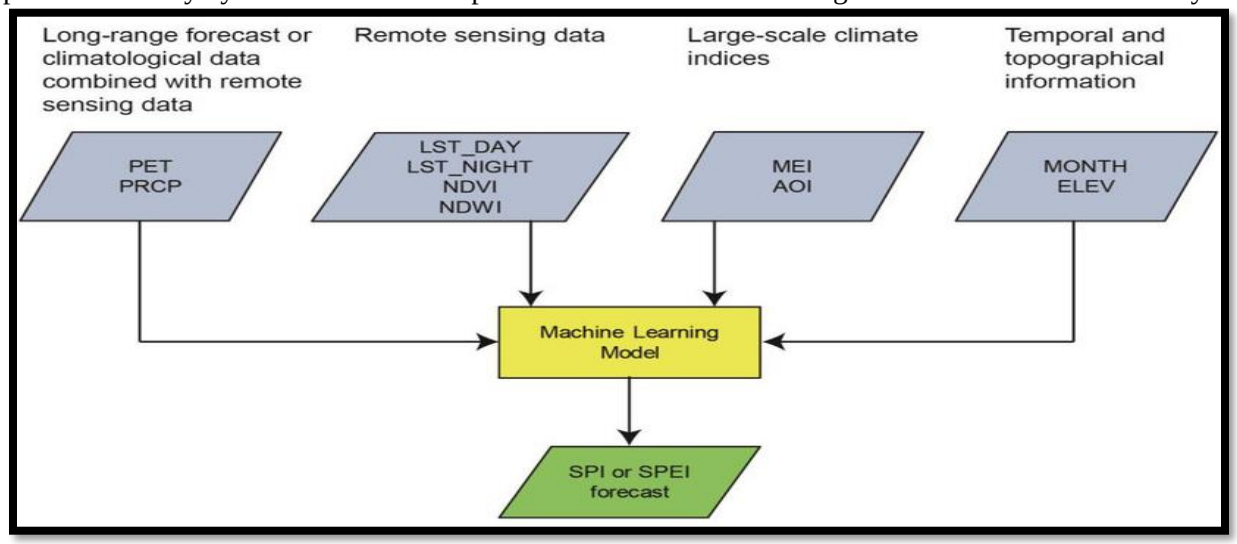

Fig 3.5 Weather Predictor system

4. CONCLUSION

Artificial intelligence is the new electricity. Just as electricity invention changed the face of the industrial age, the futuristic world is moved by artificial intelligence guided by machine-learning techniques. This article addressed the fundamental categorization of machine learning methods – supervised learning, unsupervised learning, and reinforcement learning. I also discussed the six important machine-learning

ISSN: 2233-7857IJFGCN

algorithms, such as Linear Regression, SVM, Decision Trees, K-Nearest Neighbour (KNN), Naive Bayes, and Neural Networks. The paper concludes by explaining the different uses of machine learning in the real world.

REFERENCES

1. https://expertsystem.com/machine-learning-definition/
2. AyonDey, Machine Learning Algorithms: A Review, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-1179
3. https://www.educba.com/applications-of-machine- learning/
4. https://www.educba.com/uses-of-machine-learning/
5. https://www.outsource2india.com/software/articles/machine-learning-applications-how-it-works-who-uses-it.asp
6. https://dzone.com/articles/machine-learning-in-plain-english
7. https://www.geeksforgeeks.org/what-is- reinforcement-learning/
8. https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/
9. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/
10. https://data-flair.training/blogs/machine-learning-algorithms/
11. https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_ with_python_algorithms.htm
12. https://www.sciencedirect.com/topics/ computer- science/markov-decision-process
13. Anushree Raj , Rio D. Souza , A Review on Machine Learning Algorithms , International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 7 Issue VI, June 2019, ISSN: 2321-9653
14. TurkiTurki. “An empirical study of machine learning algorithms for cancer identification”. IEEE 15th International Conference on Networking, Sensing and Control, March 2018.
15. KavithaS;VarunaS;Ramya R. “A comparative analysis on linear regression and support vector regression”. Online International Conference on Green Engineering and Technologies, November 2016.
16. 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.
17. 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.
18. PrajwalShimpi; Sanskruti Shah; MaitriShroff ;AnandGodbole, “A machine learning approach for the classification of cardiac arrhythmia”. International Conference on Computing Methodologies and Communication, July 2017.
19. SiddalingeshwarPatil; Umakant Kulkarni. “Accuracy Prediction for Distributed Decision Tree using Machine Learning approach”. 3rd International Conference on Trends in Electronics and Informatics, April 2019.
20. 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.
21. 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.
22. Nitin Singh; SaurabhChaturvedi ;Shamim Akhter. “Weather Forecasting Using Machine Learning

ISSN: 2233-7857IJFGCN

Algorithm". IEEE International Conference on Signal Processing and Communication, December 2019.

ISSN: 2233-7857IJFGCN