Proceedings of the Third International Conference on Electronics Communication and Aerospace Technology [ICECA 2019]
IEEE Conference Record # 45616; IEEE Xplore ISBN: 978-1-7281-0167-5

Comparison of Various Machine Learning Techniques and Its Uses in Different Fields

Aditya Srivastava, Sonia Saini and Deepa Gupta
Amity Institute of Information Technology,
Amity University, Noida, Uttar Pradesh, India
adityasrv23@gmail.com, ssani2@amity.edu, dgupta@amity.edu

***Abstract*— Machine learning uses artificial intelligence and helps systems to develop the ability to do self learning and improve from past experience without much of programming. It can access data and use it for themselves. In this research paper, we will be discussing the different machine learning techniques and how they are used in different day to day applications.**

Keywords*: *Machine Learning Techniques, Supervised learning, Unsupervised learning, Data Mining, Sports Data Analysis

I. Introduction

Machine learning [18] is used to access different data sets, learn from the data and apply to themselves. Machine learning uses artificial intelligence that enables the systems to do self learning and improve their accuracy with each experience without using highly complex programs. Artificial intelligence is used to enable intelligent machines that can perform tasks and behave like humans. Machine learning [21] is the process of learning by looking at patterns within the data and making better future decisions according to the examples that we provide in any application that is used in everyday life. Machine learning allows computers to learn themselves without human intervention [26].

In this research paper, we will be discussing and comparing different machine learning techniques and how they are used in different day to day applications shown in Fig 1 below.

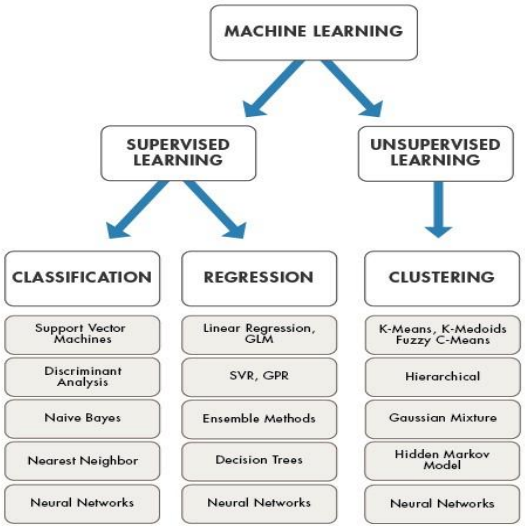

Fig. 1 Diagram of Machine Learning Techniques

Machine learning techniques can be classified into two categories, supervised and unsupervised learning.

Supervised Learning: It teaches the machine to learn and access labeled data and it means that some data is already tagged with correct answer [15].

Unsupervised learning: It teaches the machine to learn and use non-labeled data. The machine groups unsorted data according to similarities, patterns and differences. It finds hidden patterns in the data.

A. Supervised Learning Techniques

Supervised learning can be classified into two techniques:

Classification: The desired output is a category of a given input based on the classes that are defined using labeled data such as “Yes” or “NO” or “Disease” or “No Disease” [3].

Regression: The desired output is continuous. The real value of the output (or target) variable is given for input variables. For example, a

regression algorithm can give the estimate of air travel price between two cities based on the given inputs.

B. Unsupervised Learning Techniques

Clustering: It is used to find groupings and hidden patterns in the data. The data needs to be organized [13]

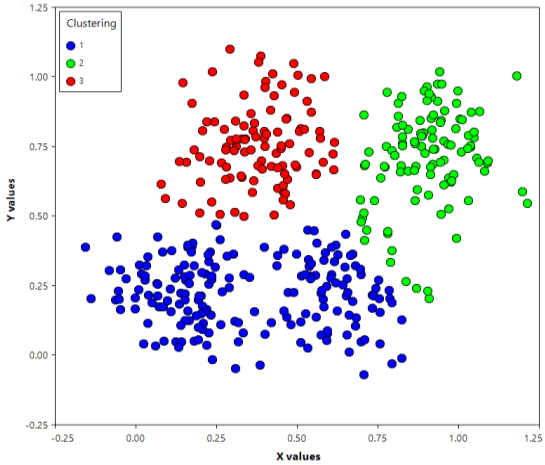

Fig. 2 Grouping in three clusters

Fig 2 shows a graph that groups the data into three different clusters. All the blue data is grouped into cluster 1. All the green data is grouped into cluster 2. All the red data is grouped into cluster 3. This way the data is more organized, the three different groups of data are shown clearly and the hidden patterns can be found [27].

II. Comparison of different machine learning algorithms

This research paper will focus on the following machine learning algorithms: Support Vector Machine, Decision Tree, K-Nearest Neighbor (KNN), Neural Networks, Naïve Bayes, Linear Regression and Ensemble Methods.

A. Support Vector Machine (SVM)

Support vector machine algorithm is utilized mainly in classification techniques. Support Vector Machine creates a hyperplane which divides the data into different classes. SVM can solve linear or non-linear problem. Support vectors are the data-points which are nearest to the hyperplane [9].

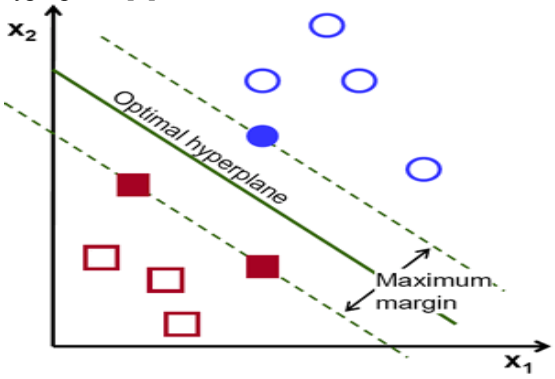

Fig. 3 Support Vector Machine

In Fig 3 it shows the support vector machine with optimal hyperplane in the middle and it's separating the two datasets. The two data sets are squares and circles. The maximum margin is being shown on both classes for each data. It even shows the support vectors in each class. The two squares filled in red and the circle filled in blue are the support vectors because they are the nearest data-points to the hyperplane.

B. Decision Tree

A decision tree is a graphical representation which uses a branching method to show every possible outcome of a decision [5]. Each node represents an input variable and the branches are created depending on the different possible values of the input variables represented with the path from root to leaf [16].

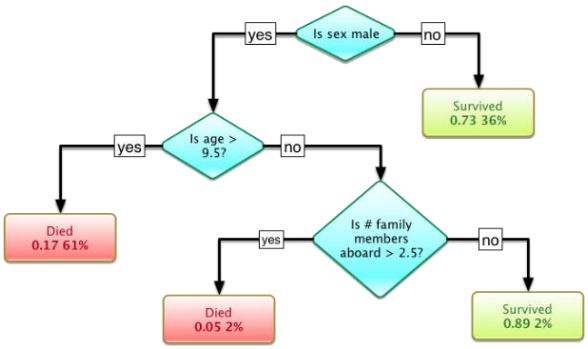

Fig. 4 Decision tree for survival on Titanic

Fig 4 shows a decision tree that shows the survival on Titanic. Each leaf has an outcome for whether the answer is yes or no. It shows the outcomes of who survived on the Titanic and who died.

C. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is a type of clustering algorithm. KNN makes sure that similar things exist in close proximity. It makes sure that similar things are near to each other. It recognizes pattern. It makes sure that the new data is classified and assigned to the right neighboring group. K can be an integer greater than 1.

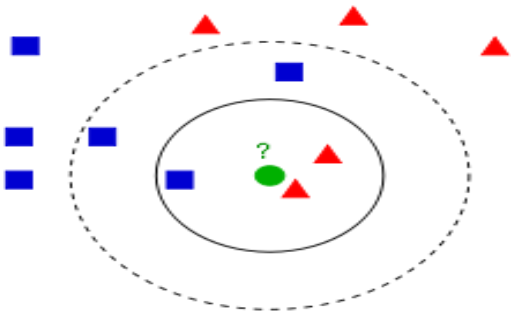

Fig. 5 Example of KNN Classification

In Fig 5 the first class is blue squares and the second class is red triangles and the green circle is the test sample that's classified to one of those two classes. In the solid line circle K=3 and it is assigned to the first class because there are 2 red triangles and 1 blue circle inside it. The dashed line circle is where K=5 is assigned to the second class because there are 3 blue squares and 2 red rectangles overall (inner circle and outer circle combined). The green sample circle would be classified as a red triangle because in the first class there are 2 red triangles and 1 blue square. Because there are more red triangles than blue squares in first class, the green circle is a red triangle.

D. Neural Network

Neural Networks are used to solve artificial intelligence (AI) problems. It is a network or circuit of neurons which consist of artificial neurons and nodes. It is used to process the output of the information through these neurons and nodes in biological systems. It is used for function proximation, regression analysis and prediction of the analysis. It is used to classify the pattern and sequence of a data. [10]. It can also be used for decision making. It includes input layer, hidden layers and output layer as shown in the architecture below.

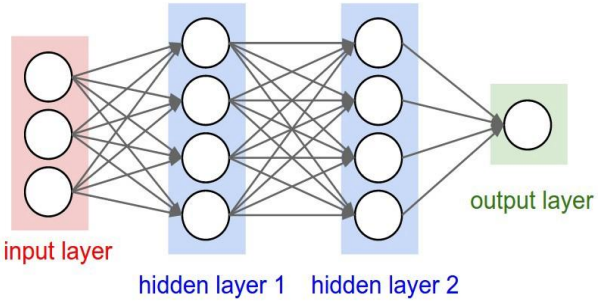

Fig. 6 Neural network Architecture

E. Naïve Bayes

The Naïve Bayes Classifier is established on the conditional probability and is utilized when the complexity of the inputs is quite high. It illustrates the probabilities of the outcomes [15]. It is used by applying the Bayes' Theorem which is as follows:

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

where

- **P(C|A)** is the probability of hypothesis h given the data d. This is known as posterior probability.
- **P(A|C)** is the probability of data d given that the hypothesis h was true.
- **P(C)** is the probability of hypothesis h being true (regardless of the data). This is known as prior probability of h.
- **P(A)** is the probability of the data (irrespective of the hypothesis).

F. Linear Regression

Linear regression is the process used to estimate the relationships among the variables. In this one of the variables is dependent on one or more independent variables. It is part of regression analysis.

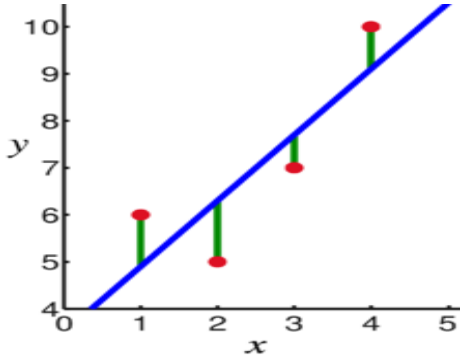

Fig. 7 Linear Regression model

As shown in Fig 7, The observations are marked in red and are the result of random deviations (marked in green) from the underlying relationship (marked in blue) between the independent variable(x) and the dependent variable (y).

G. Ensemble Methods

Ensemble methods is a technique that combines many algorithm models to produce one optimal predictive model [16]. A common type of ensemble method is "Random Forest". It is a method of classification and regression that constructs many decision trees at training time and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random Decision Forests correct the decision trees' habit of overfitting to their training set.

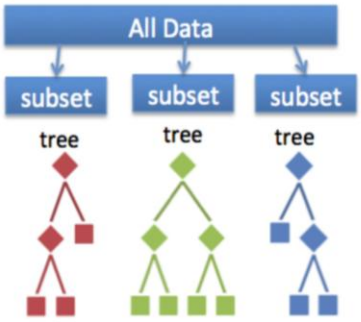

Fig. 8 Random Forest architecture

As shown in Fig 8, a random forest takes a random subset of features from data and creates n random trees from each subset. Trees are connected at the end.

After comparing all these different machine learning algorithms in this research paper, I found that each algorithm has different ways of getting all the data and information in machine learning.

III. Literature Review

There are many applications for machine learning with the most relevant being data mining [15]. Data mining is the process of extraction of relevant information and patterns from the data [2]. Data mining [1] uses the same techniques as machine learning such as decision tree, k-nearest neighbor, classification, regression, artificial intelligence, neural networks etc [2]. We have studied 25 research papers and content available in some websites as listed in references section including the objectives, data collection, results and conclusions.

IV. Uses in Different Fields

Machine learning techniques can be used in many different applications. In this research paper, I will be discussing how the machine learning techniques are used in some of the important applications.

A. Sports : Machine learning techniques can be used in sports data analysis to predict the outcome of a game, predict performances of individual players or teams, building strategies for upcoming games, deciding the price of a player if a team was to sign, trade or cut the player and connecting players to brand and sponsors [28].

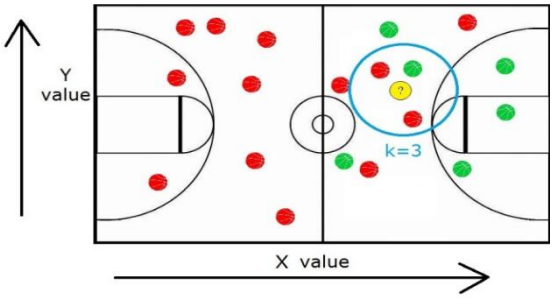

Fig. 9 Basketball Shot Chart

In Fig 9 it shows the basketball shots by a player on the court. red balls are misses and green balls are makes. In this the KNN algorithm is used to predict whether the shot with the question mark was a make or a miss by looking at the 3 nearest neighbors in the circle which are 2 red balls and 1 green ball therefore K=3. Because there are 2 red balls and 1 green ball in the circle it would be predicted that the shot with the question mark is a red ball and was a miss [28].

B. Robotics : Machine learning techniques can be used in robotics to program the robots to do work that human usually do. The robots become programmed with general instructions and behaviors. These robots are programmed to make the robots recognize the objects and the obstacles. Artificial intelligence is used in robots to give the robots the intelligence to work diligently and not mess up in their work. Artificial intelligence gives robots a mind that is way more intelligent than that of humans.

Fig. 10 Flexible Manufacturing Robot

A flexible manufacturing robot is used in work to manufacture multiple items for a business. It is programmed to manipulate new types of parts without damaging them. It can repair those items [29].

C. Climatology : Weather can be analyzed over a period of time and future weather patterns can be predicted. Natural calamities can be determined

and weather forecast can be done using machine learning techniques [11].

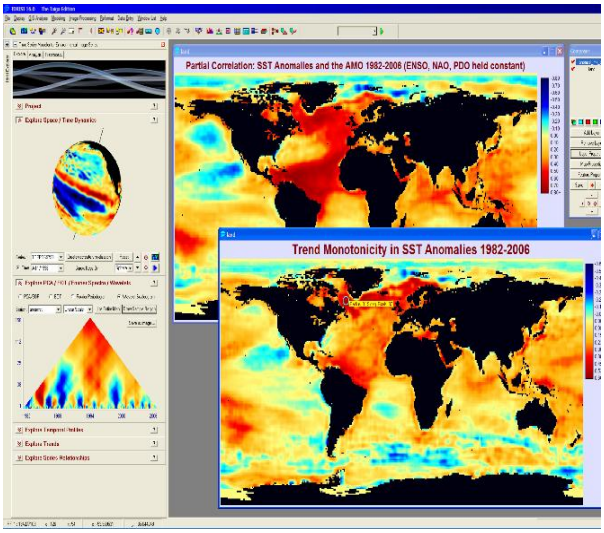

Fig. 11 Weather analysis

D. Aviation : Machine learning methods can be used in aviation to forecast the weather and to predict if the conditions are safe for pilots to fly the airplane [23]. Machine learning methods can be used to facilitate flight tracking [24]. It can also be used to facilitate airport related insights, airline related insights, market related insights, passenger related insights, air safety related insights and aircraft related data [25].

Fig. 12 Flight Tracking

E. E-commerce : Machine learning techniques are used in e-commerce to analyze search patterns to promote up sale, cross sale and the price of the goods or services.

F. Cloud Computing : Different kinds of machine learning techniques can be used in cloud computing to create different search patterns and applications for finding any information which remains hidden in the unstructured data. Cloud computing is fast, reliable, efficient and secure and it uses machine learning techniques to reduce cost of infrastructure of an individual [14].

G. Detecting Car Crashes : With in excess of 8 million individuals, the RAC is one of the UK's biggest motoring associations, giving roadside help, protection, and different administrations to private and business drivers. To empower fast reaction to roadside occurrences, diminish crashes, and alleviate protection costs, the RAC built up a locally available accident detecting framework that utilizes propelled AI calculations to recognize low-speed impacts and recognize these occasions from increasingly regular driving occasions, for example, rolling over hindrances or potholes. Autonomous tests demonstrated the RAC framework to be 92% precise in distinguishing test crashes.

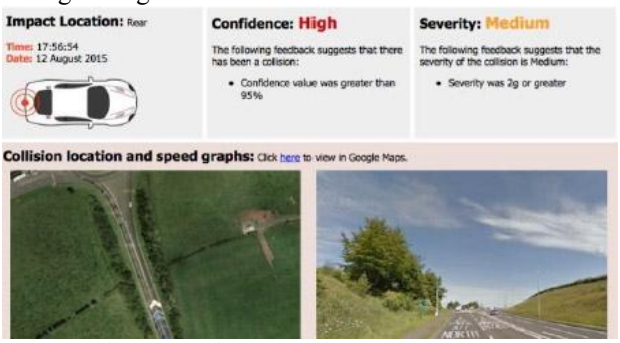

Fig.13 Detecting Car Crash

V. Conclusion

This paper has explored the different Machine Learning Techniques like “Supervised Learning” and “Unsupervised Learning”. It has explained various machine learning algorithms - Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbor (KNN), Neural Network, Naïve Bayes, Linear Regression and Ensemble Methods. We have studied the uses of these Machine Learning techniques in Sports, Robotics, Climatology, Aviation, E-Commerce, Cloud Computing and Detecting Car Crashes.

VI. Future Work

This research paper could be extended in future to explore the use cases for the different machine learning techniques in other fields like Healthcare [12]. One of the key areas where lot of research has been done is Diagnosis of Heart Diseases [7] including other cardio-vascular diseases [19]. Another important area is Breast Cancer [17]. Specific Sports like Basketball, Soccer [20]. A lot of research has been done in USA on Football [8]. There are worldwide research papers on Video Games [22]. Some of these information have been collected in terms of the available research papers as listed in the references.

VII. References

[1] Hilal Almarabeh and Ehab F. Amer, “A Study of Data Mining Techniques Accuracy for Healthcare”, 2017

https://www.researchgate.net/publication/317612611

[2] Mrs. Bharati M. Ramageri, "Data Mining Techniques and Applications", 2010
http://www.ijcse.com/docs/IJCSE10-01-04-51.pdf

[3] Akarshita Tripathi and Amit Kumar, "Analysis of Educational Data Mining Techniques", 2019
http://www.academia.edu/38080804/Analysis_of_Educational_Data_Mining_Techniques

[4] Nikita Jain and Vishal Srivastava, "Data Mining Techniques: A Survey Paper", 2013
https://pdfs.semanticscholar.org/1c59/71351969d6729e0a27715934bf771e1d833c.pdf

[5] Abhijit A. Sawant and P. M. Chawan, "Comparison of Data Mining Techniques used for Financial Data Analysis", 2013
https://www.researchgate.net/publication/316511061

[6] Luís C. Borges, Viriato M. Marque and Jorge Bernardino, "Comparison of Data Mining techniques and tools for data classification", 2013
https://www.researchgate.net/publication/258994903

[7] Serdar Aydin, Meysam Ahanpanjeh and Sogol Mohabbatiyan, "Comparison and Evaluation of Data Mining Techniques in the Diagnosis of Heart Disease", 2016
https://wireilla.com/papers/ijcsa/V6N1/6116ijcsa01.pdf

[8] Carson Leung and Kyle W. Joseph, "Sports Data Mining: Predicting Results for the College Football Games", 2014
https://www.researchgate.net/publication/275538669

[9] Maral Haghighat, Hamid Rastegari and Nasim Nourafza, "A Review of Data Mining Techniques for Result Prediction in Sports", 2013
https://www.researchgate.net/publication/262560138

[10] Chenjie Cao, "Sports Data Mining Technology Used in Basketball", 2012
https://arrow.dit.ie/cgi/viewcontent.cgi?article=1040&context=scschcomdis

[11] Samiddha Mukherjee, Ravi Shaw and Nilanjan Haldar, "A Survey of Data Mining Applications and Techniques", 2015
https://pdfs.semanticscholar.org/afbd/4ff4157bc9c7f8a94203a98c514765874925.pdf

[12] P. Haripriya and R. Porkodi, "A Survey Paper on Data mining Techniques and Challenges in Distributed DICOM", 2016
https://www.researchgate.net/publication/307888518

[13] Ashish Kumar Dogra and Tanuj Wala, "A Review Paper on Data Mining Techniques and Algorithms", 2015
http://ijarcet.org/wp-content/uploads/IJARCET-VOL-4-ISSUE-5-1976-1979.pdf

[14] Smita and Priti Sharma, "Use of Data Mining in Various Field: A Survey Paper", 2014
https://pdfs.semanticscholar.org/1fbf/5206a6049a3cafb168c4bed9edad420cc7b7.pdf

[15] Osisanwo F.Y, Akinsola J.E.T, Awodele O, Hinmikaiy J.O, Olakanmi O and Akinjobi J, "Supervised Machine Learning Algorithms: Classification and Comparison", 2017
https://www.researchgate.net/publication/318338750

[16] Apoorv Maheshwari, Navindran Davendralingam and Daniel A. DeLaurentis, "A Comparative Study of Machine Learning Techniques for Aviation Applications", 2018
https://www.researchgate.net/publication/325964700

[17] Kanchanamani M and Varalakshmi Perumal, "Performance evaluation and comparative analysis of various machine learning techniques for diagnosis of breast cancer", 2016
https://pdfs.semanticscholar.org/0963/cfff9f4295a15af64db13ebd6238858b54d0.pdf

[18] Aik Choon Tan and David Gilbert, "An empirical comparison of supervised machine learning techniques in bioinformatics", 2003
https://core.ac.uk/download/pdf/335643.pdf

[19] M.Chandralekha and N.Shenbagavadivu, "Performance Analysis Of Various Machine Learning", 2018
http://www.naturalspublishing.com/files/published/24463jgqv03g0m.pdf

[20] Rehana Mahfuz, Zeinab Mourad and Aly El Gamal, "Analyzing Sports Training Data with Machine Learning Techniques", 2016

[21] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia and Juan Gutierrez, "A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease", 2017
http://cobweb.cs.uga.edu/~pouriyeh/ISCC2017.pdf

[22] Ben Geisler, "Integrated Machine Learning for Behavior Modeling in Video Games"
https://www.aaai.org/Papers/Workshops/2004/WS-04-04/WS04-04-012.pdf

[23] Sameer Jasra, Jason Gauci, Alan Muscat and Gianluca Valentino, "Literature Review of Machine Learning Techniques to Analyze Flight Data", 2018
https://www.researchgate.net/publication/328725531

[24] Marco Hrastovec and Frank Solina, "Machine Learning Models for Aircraft Performances", 2014
https://www.researchgate.net/publication/262764214

[25] Nathalie Kuhn and Navaneeth Jamadagni, "Application of Machine Learning Algorithms to Predict Flight Arrival Delays", 2017
http://cs229.stanford.edu/proj2017/final-reports/5243248.pdf

[26] Unknown Author, "Machine Learning Definition"
https://www.expertsystem.com

[27] Unknown Author, "Cluster Analysis", 2016
ichrome.com/blogs/archives/221

[28] Sander Dalm, "Data Science Blog", 2016
https://datascienceprojects.wordpress.com/2016/05/16/how-to-make-the-perfect-basketball-shot/

[29] Lisa Lyons, "Machine Learning and Robotics", 2018
https://www.slideshare.net/butest/machine-learning-and-robotics