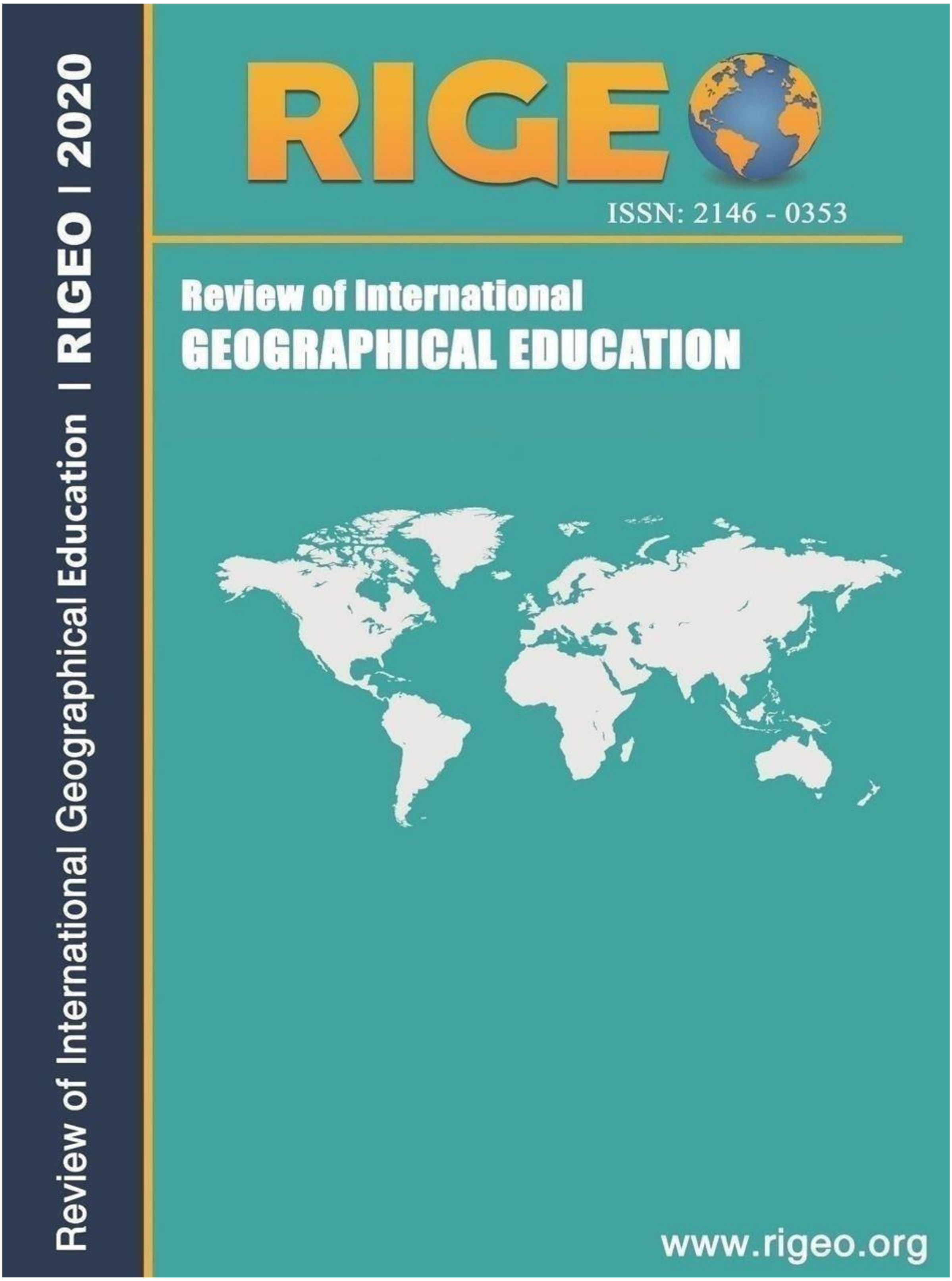
Review of International Geographical Education | RIGEO | 2020
RIGEO
ISSN: 2146 - 0353
Review of International
GEOGRAPHICAL EDUCATION
www.rigeo.org

1

A Comprehensive Overview of Machine Learning: Theory and Algorithmic Insights

Ms. FARHEEN SULTANA[1], Mrs. TAHERA ABID[2], Mrs. NAZIA AMREEN[3]
Department of Information technology
Nawab Shah Alam Khan College of Engineering and Technology

Abstract
A new wave of technology called SMAC (Social, Mobile, Analytic, Cloud) is making it possible for smart tools, networked processes, and big data to work together in the future. A huge amount of data has been created by this virtual world, which is speeding up the use of machine learning solutions and methods. Machine Learning lets computers behave like humans and change how they do things. When machine learning is used, every contact and action is turned into something the system can learn from and use the next time. This work gives an outline of this data analytics method that lets computers learn and do what people naturally do: discover things on their own. The basics of machine learning are covered, along with its meaning, terminology, and uses, explaining what it is, how it works, and why it is important. The machine learning technology plan is talked about to figure out and confirm its promise as a business and industry practice. The main goal of this work is to show why machine learning is the way of the future.

Keywords: *Machine Learning, Supervised Learning, Unsupervised Learning, Algorithms, Instant Learning, Ensemble Learning*

1. Introduction

In a broad sense, learning means picking up new behaviours, beliefs, information, skills, or tastes or changing ones that you already have. Personal learning theory, or how people learn, is made up of behaviourism, cognitivism, constructivism, experiential materialism, and social learning. Machines depend on data, which goes against what people do naturally, which is to learn from experience. To put it simply, machine learning (ML) is a type of artificial intelligence that lets computers think and learn on their own. To make computers more accurate, people have to change the things they do so they do them more often. How accurate something is is measured by how many times the actions chosen actually work.

Researchers have officially described ML in the relevant journals. Arthur Samuel came up with the word "machine learning" (ML) in 1959. He said that ML is the study of making machines smart enough to learn without being directly coded [1]."A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E [2]." This is a more recent phrasing that has been more useful in engineering.

It is important to note that machine learning is a multidisciplinary field with many study areas that support its survival. These are shown in Figure 1.

A lot of what ML models do is linked to computational statistics, whose main goal is to use computers to make predictions. As well as that, it is linked to Mathematical Optimisation, which connects statistical models, uses, and tools.

The problems in the real world are very complicated, which makes them great for using ML. Machine learning can be used in many areas of computing to create and program clear algorithms that work very well. Some examples are filtering spam emails, finding fraud on social networks, trading stocks online, finding faces and shapes, making medical diagnoses, predicting traffic, recognising characters, and suggesting products. In the real world, machine learning is used for things like Google cars that drive themselves, Netflix showing people movies and TV shows they might like, online recommendation engines like "more items to consider" and "get yourself a little something" on Amazon, and finding credit card fraud.

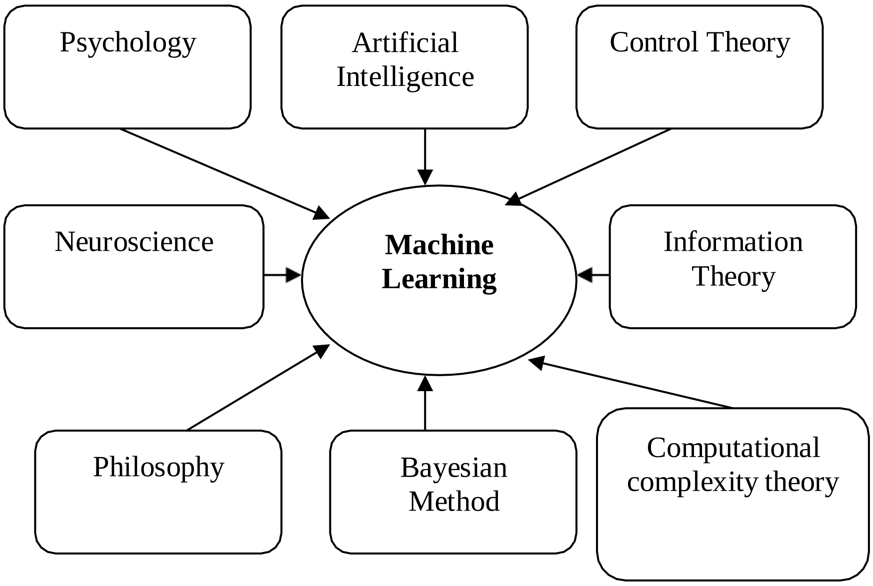

Fig.1. The Multi-disciplinary ML

The main goal of this work is to give an outline of how Machine Learning has changed over the years, including different methods, uses, and problems. This essay gives a general outline of machine learning, including where it came from and how it has changed over the years. This is how the study is set up: In Section 2, the issues in data science that ML methods can help with are talked about. In Section 3, you can read about how ML has changed over time. In Section 4, we talk about the general model of machine learning. In Section 5, we explain how machine learning works. In sections 6 and 7, the different machine learning models and methods are explained. In section 8, the problems that machine learning faces and its future prospects are briefly discussed. Finally, in section 9, the study is concluded.

2. Data Science problems and Machine Learning

Machine learning is required to make the computers sophisticatedly perform the task without any intervention of human beings on the basis of learning and constantly increasing experience to understand the problem complexity and need for adaptability.

- *Tasks performed by Human Beings:* There are lots of tasks performed on day to day basis by human beings, but the main concern is that to perform the tasks perfectly and to perform under well-defined program. Examples: Cooking, Driving, Speech Recognition.
- *Tasks beyond Human Capabilities:* Another category of tasks includes which can be done by machine learning in effective manner is analysis of large and complex data sets like Remote Sensing, Weather forecasting, Ecommerce, Web search etc. With large

amounts of data, it becomes really complex for human beings to predict meaningful data.

Machine learning has proven capabilities to inherently solve the problems of data science. Hayashi and Chikio [3] define data science as, “a concept to unify statistics, data analysis, machine learning and their related methods in order to understand and analyze actual phenomena" with data”. Before taking to problem solving, the problem must be categorized suitably so that the most appropriate machine learning algorithm can be applied to it. Any problem in data science can be grouped in one of the following five categories as shown in Figure 2.

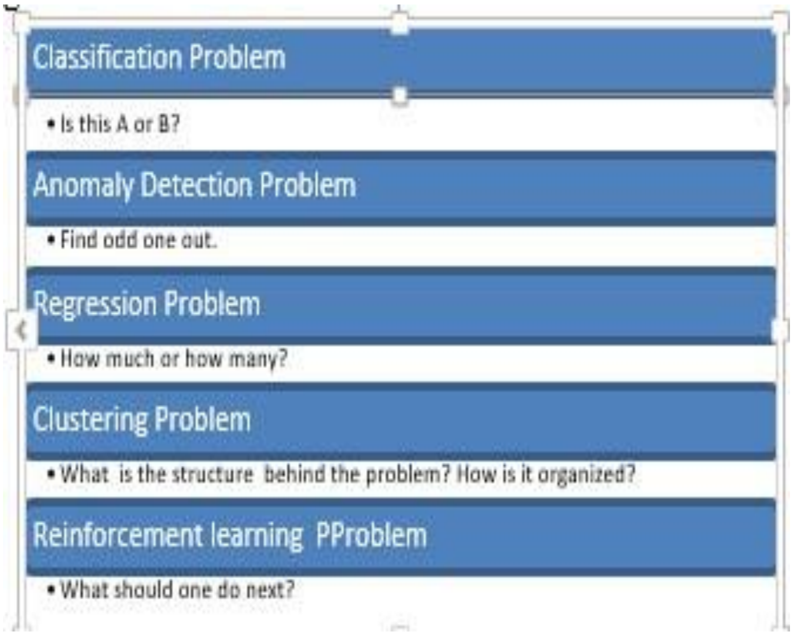

Fig 2. Types of Problems

Thus depending on the type of problem, an appropriate machine learning approach can be applied. The various categories are explained below:

- *Classification Problem-* A problem in which the output can be only one of a fixed number of output classes known apriori like Yes/No, True/False, is called a classification problem. Depending on the number of output classes, the problem can be a binary or multi-class classification problem.
- *Anomaly Detection Problem-* Problems which analyze a certain pattern and detects changes or anomalies in the pattern fall under this category. For example, credit card companies use anomaly detection algorithms to find deviation from the usual transaction behavior of their client and raises alerts whenever there is an unusual transaction. Such problems deal with finding out the outliers.

- *Regression Problem-* Regression algorithms are used to deal with problems with continuous and numeric output. These are usually used for problems that deal with questions like, ‘how much’ or ‘how many’.
- *Clustering Problem-* Clustering falls under the category of unsupervised learning algorithms. These algorithms try to learn structures within the data and attempt to make

clusters based on the similarity in the structure of data. The different classes or clusters are then labeled. The algorithm, when trained, puts new unseen data in one of the clusters.

- *Reinforcement Problem*-Reinforcement algorithms are used when a decision is to be made based on past experiences of learning. The machine agent learns the behaviour using trial and error sort of interaction with the continuously changing environment. It provides a way to program agents using the concept of rewards and penalties without specifying how the task is to be accomplished. Game playing programs and programs for temperature control are some popular examples using reinforcement learning.

3. Development of Machine Learning

The words, Artificial Intelligence and Machine Learning are not new. They have been researched, utilized, applied and re-invented by computer scientists, engineers, researchers, students and industry professionals for more than 60 years. The mathematical foundation of machine learning lies in algebra, statistics, and probability. Serious development of Machine Learning and Artificial Intelligence began in 1950's and 1960's with the contributions of researchers like Alan Turing, John McCarthy, Arthur Samuels, Alan Newell and Frank Rosenblatt. Samuel proposed the first working machine learning model on Optimizing Checkers Program. Rosenblatt created Perceptron, a popular machine learning algorithm based on biological neurons which laid the foundation of Artificial Neural Network [4, 5, 6]. The following table 1 depicts the illustrious, expansive and practical development of machine learning.

Table 1. Development of ML

1950	Alan Turning created "Turning Test" to check a machine's intelligence. In order to pass the Turning Test, the machine should be able to convince humans that there they are actually talking to a human and not a machine.
1952	Samuel created a highly capable learning algorithm than can play the game of Checkers with itself and get self-trained.
1956	Martin Minsky and John McCarty with Claude Shannon and Nathan Rochester organized a conference in Dartmouth in 1956 where actually Artificial Intelligence was born.
1958	Frank Rosenblatt created Perceptron, which laid the foundation stone for the development of Artificial Neural Network (ANN).
1967	The Nearest Neighbor Algorithm was proposed which could be used for "Pattern Recognition".
1979	Stanford University students developed "Stanford Cart", a sophisticated robot that could navigate around a room and avoid obstacles in its path.
1981	Explanation Based Learning (EBL) was proposed by Gerald Dejong, whereby, a computer can analyze the training data and create rules for discarding useless data [7]
1985	NetTalk was invented by Terry Sejnowski, [8] which learnt to pronounce English words in the same manner that children learn.
1990s	The focus of Machine Learning shifted from Knowledge-driven to Data Driven. Machine Learning was implemented to analyze large chunks of data and derive conclusions from it [9]
1997:	IBM invented the Deep Blue computer which was able to beat World Chess Champion Gary Kasparov.

2006	The term "Deep Learning" was coined by Geoffery Hinton which referred to a new architecture of neural networks that used multiple layers of neurons for learning.
2011	IBM's Watson, built to answer questions posed in a natural language, defeats a Human Competitor at Jeopardy Game.
2012	Jeff Dean from Google, developed GoogleBrain, which isa Deep Neural Network to detect patterns in Videos and Images.
2014	Facebook invented the "DeepFace" algorithm based on Deep Neural Networks capable of
	recognizing human faces in photos.
2015	Amazon proposed its own Machine Learning Platform. Microsoft created "Distributed Machine Learning Toolkit" for efficient distribution of machine learning problems to multiple computers to work parallel to find a solution [10,11]. Elon Musk and Sam Altman, created a non-profit organization- OoeaAI, with the objective of using Artificial Intelligence to serve human beings.
2016	Google proposed DeepMind which is regarded as the most complex Board Game. Google AlphaGo program becomes the first Computer Go program to beat a professional human player. It is based on the combination of machine learning and tree searching techniques [12].
2017	Google proposed Google Lens, Google Clicks, Google Home Mini and Google Nexus based phones which use Machine Learning and Deep Learning Algorithms. Nvidia proposed NVIDIA GPUs- The Engine of Deep Learning. Apple proposed Home Pod which is a Machine Learning Interactive device.

4. The Generic Model of ML

ML is used to solve various problems that require learning on the part of the machine. A learning problem has three features:

- Task classes (The task to be learnt)
- Performance measure to be improved
- The process of gaining experience

For example, in a game of checkers, the learning problem can be defined as:

- Task T: Playing the game
- Performance Measure P: number of games won against the opponent.
- Experience E: practicing via playing games against itself and consistently improving the performance.

The generic model of machine learning consists of six components independent of the algorithm adopted. The following figure 3 depicts these primary components.

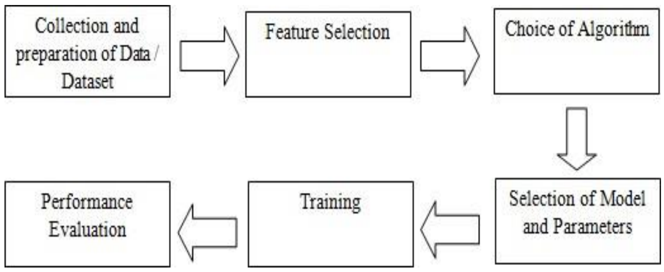

Fig 3. Components of a Generic ML model

Each component of the model has a specific task to accomplish as described next.

i. **Collection and Preparation of Data**: The primary task of in the machine learning process is to collect and prepare data in a format that can be given as input to the algorithm. A large amount may be available for any problem. Web data is usually unstructured and contains a lot of noise, i.e., irrelevant data as well as redundant data. Hence the data needs to be cleaned and pre-processed to a structured format.

ii. **Feature Selection**: The data obtained from the above step may contain numerous features, not all of which would be relevant to the learning process. These features need to be removed and a subset of the most important features needs to be obtained.

iii. **Choice of Algorithm**: Not all machine learning algorithms are meant for all problems. Certain algorithms are more suited to a particular class problem as explained in the previous section. Selecting the best machine learning algorithm for the problem at hand is imperative in getting the best possible results. The various ML algorithms are discussed in Section 6. iv. **Selection of Models and Parameters**: Most of machine learning algorithms require some initial manual intervention for setting the most appropriate values of various parameters.

v. **Training**: After selecting the appropriate algorithm and suitable parameter values, the model needs to be trained using a part of the dataset as training data.

vi. **Performance Evaluation**: Before real-time implementation of the system, the model must be tested against unseen data to evaluate how much has been learnt using various performance parameters like accuracy, precision and recall.

5. Machine Learning Paradigms

Depending on how an algorithm is being trained and on the basis of availability of the output while training, machine learning paradigms can be classified into ten categories. These include: supervised learning, semi-supervised learning, unsupervised learning, reinforcement learning, evolutionary learning, ensemble learning, artificial neural network, Instance-based learning, dimensionality reduction algorithms and hybrid learning [4, 5, 6, 13,14,15,16 17]. Each of these paradigms is explained in the following sub-sections.

5.1. Supervised Learning

Under supervised learning, a set of examples or training modules are provided with the correct outputs and on the basis of these training sets, the algorithm learns to respond more accurately by comparing its output with those that are given as input. Supervised learning is also known as learning via examples or learning from exemplars. The following figure 4 explains the concept.

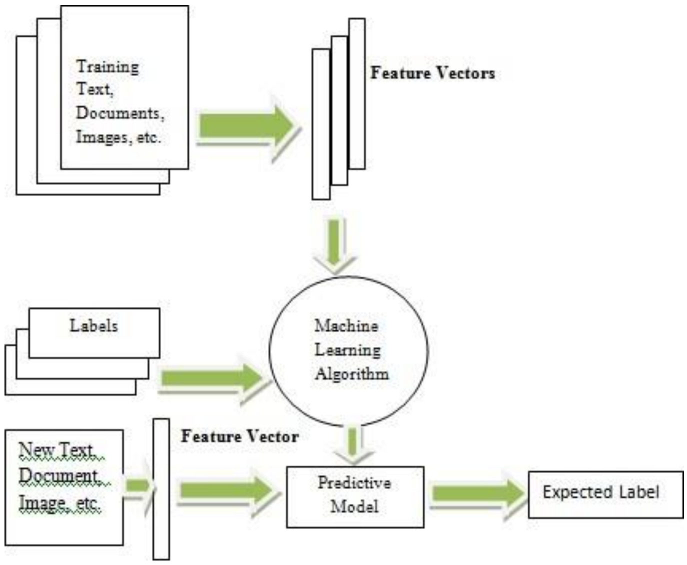

Fig. 4. Supervised Learning [4]

Supervised learning finds applications in prediction based on historical data. For example: to predict the Iris species given a set of its flower measurements or a recognition system that determines whether an object is a galaxy, a quasar or a star given a colored image of an object through a telescope, or given an e-commerce surfing history of a person, recommendation of the products by e-commerce websites [4]. Supervised learning tasks can be further categorized as classification tasks and regression tasks. In case of classification, the output labels are discrete whereas they are continuous in case of regression.

5.2. Unsupervised Learning

The unsupervised learning approach is all about recognizing unidentified existing patterns from the data in order to derive rules from them. This technique is appropriate in a situation when the categories of data are unknown. Here, the training data is not labeled. Unsupervised learning is regarded as a statistic based approach for learning and thus refers to the problem of finding hidden structure in unlabeled data. Figure 5 explicates the concept.

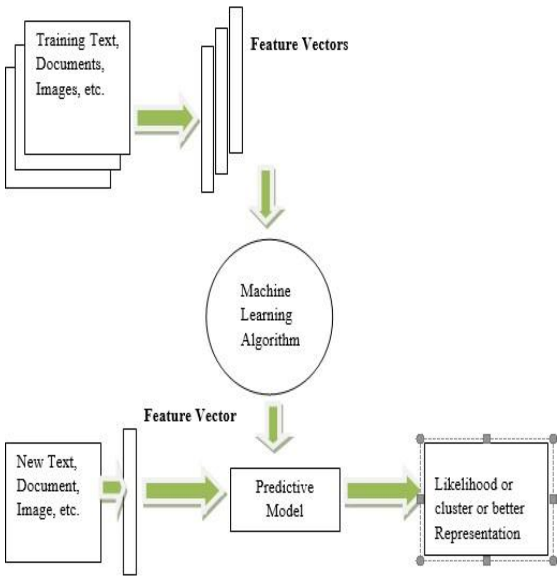

Fig 5. Unsupervised Learning

5.3. Reinforcement Learning

Reinforcement learning is regarded as an intermediate type of learning as the algorithm is only provided with a response that tells whether the output is correct or not. The algorithm has to explore and rule out various possibilities to get the correct output. It is regarded as learning with a Critic as the algorithm doesn't propose any sort of suggestions or solutions to the problem.

5.4. Evolutionary Learning

It is inspired by biological organisms who adapt to their environment. The algorithm understands the behavior and adapts to the inputs and rules out unlikely solutions. It is based on the idea of fitness to propose the best solution to the problem.

5.5. Semi-Supervised Learning

These algorithms provide a technique that harnesses the power of both - supervised learning and unsupervised learning. In the previous two types output labels areeither provided for all the observations or no labels are provided. There might be situations when some observations are provided with labels but majority of observations are unlabeled due to high cost of labeling and lack of skilled human expertise. In such situations, semi-supervised algorithms are best suited for model building. Semi supervised learning can be used with problems like classification, regression and prediction [4, 13,18].

It may further be categorized as Generative Models, Self-Training and Transductive SVM.

5.6. Ensemble Learning

It is a machine learning model in which numerous learners (individual models) are trained to solve a common problem. Unlike other machine learning techniques which learn a single hypothesis from the training data, ensemble learning tries to learn by constructing a set of hypotheses from the training data and by combining them to make a prediction model [15,19] in order to decrease bias(boosting), variance (bagging), or improve predictions(stacking). Ensemble learning can be further divided into two groups:

- Sequential ensemble approaches- These are the methods in which the base learners are constructed sequentially (AdaBoost). This method exploits the dependence between the base learners.
- Parallel ensemble approaches- In these, the base learners are independent of each other, so this relationship is exploited by constructing the base learners in parallel (e.g. Random Forest) ➢ *Bagging:* It stands for bootstrap aggregation. It implements homogenous learners on sample populations and takes the mean of all predictions [4, 15,16]. For example, M different trees can be trained on dissimilar subsets of data and compute the ensemble as:

$$f(x) = 1/M \sum_{m=1}^{M} f_m(x)$$

➢ *Boosting:* It is an iterative technique that adjusts the observation's weight on the basis of last classification. It tries to fit a sequence of weak learner models that performs a little better than just random predicting e.g. small decision trees [4,16]. AdaBoost stands for adaptive boosting and is the most widely used boosting algorithm.

5.7. Artificial Neural Network

The organic neural network helps artificial neural networks (ANNs) work better. A neural network is a group of neurone cells that are linked together to help electrical signals move through the brain. A neurone, which is a nerve cell, is the basic building block of a neural network to learn. Each neurone has four parts: the dendrites, which are receptors; the soma, which processes electrical signals; the nucleus, which is the neuron's centre; and the axon, which sends signals. An ANN works in a way that is similar to a biological neural network. It has three layers: the input layer, the hidden layer, and the output layer. The links in this kind of network are weighted, and the network learns by changing the weights of the connections so that processing can happen in parallel. Some well-known algorithms are the Radial Basis Function Network (RBFN), the Backpropagation algorithm, the Perceptron learning algorithm, and the Hopfield Networks. Based on learning behavior, ANN can be further classified as:

- *Supervised Neural Network* – The inputs and the outputs are presented to the network as training data. The network is trained with this data by adjusting the weights to get accurate results. When it is fully trained, it is presented with unseen data to predict the output.
- *Unsupervised Neural Network*- In unsupervised neural network, the network is not provided with any output. It tries to find some structure or correlation among the input data and group those data together in a group or class. When new data is presented to it as input, it identifies its features and classifies it in one of the groups based on similarities.
- *Reinforcement Neural Network*- As humans interact with their environment and learn from mistakes, a reinforcement neural network also learns from its past decisions by way of penalties for wrong decisions and rewards for good decisions. The connection weights producing correct output are strengthened, while those producing incorrect responses are weakened.

5.8. Instance based learning

This learning method doesn't define the target function at the start like other machine learning methods do. Instead, the training data gives a clear picture of what the target function is. Instead, it just saves the training instance and waits to generalise until a new instance is identified. That's why it's also called "lazy learner." These kinds of methods collect examples of past problems and use a similarity measure to find the closest match between new data and examples in the database. Then, when new data is given, the system makes an estimate based on that closest match [4]. The lazy learner guesses the target function locally and differently for each new instance that needs to be classified instead of widely for the whole instance space. This means that it is faster to train but takes longer to make predictions [16]. A few well-known instance-based algorithms are k-means, k-medians, hierarchical clustering, and expectation maximisation.

5.9. Dimensionality reduction algorithms

In the last few decades, clever machine learning models have been used in many complicated and data-heavy fields, such as biology, science, medicine, the economy, and finance. But the MLbased tools that are already out there aren't fast or flexible enough to handle huge amounts of data. The fact that data has a lot of dimensions has made it hard to handle. Another problem is that there isn't enough data. It's pricey to find the global optimal for this kind of data. By cutting down on the amount of dimensions in the data, a dimensionality reduction method helps cuts down on the cost of computing. To do this, it gets rid of unnecessary and duplicate data and cleans the data to make the results more accurate. To find and use the underlying structure in the data [4, 5], dimension reduction works without being watched. One way to reduce the number of dimensions is to use multidimensional scaling (MDS), principal component analysis (PCA), linear discriminant analysis (LDA), principal component regression (PCR), and linear discriminant analysis (LDA). These can all be combined with classification and regression algorithms.

5.10. Hybrid Learning

Though ensemble learning appeared as a relief to researchers dealing with the common problems of computational complexity, over fitting and sticking to local minima in classification algorithms, researchers have found problems with ensemble learning. Complicated ensemble of multiple classifiers makes it difficult to implement and difficult to analyze the results. Instead of improving accuracy of the model, ensembles may tend to increase error at the level of individual base learner. Ensembles may result in poor accuracy as a result of selection of poor classifiers in combination. Recent approach to deal with such problems is hybridization i.e. creating ensemble of heterogeneous models. In this, more than one method is combined for example, combining clustering and decision tree or clustering and association mining etc.

Out of all the above-mentioned learning paradigms, the supervised learning is by far the most popular with researchers and practitioners.

6. Machine Learning Algorithms

In this section, we focus on some popular machine learning algorithms from the different paradigms [4, 5, 6, 16] explained in the preceding section. Although the number of algorithms falling within each paradigm are numerous and reported across pertinent literature, in this study we consider only few of these. The following table 2 briefly explains few of these algorithms.

These algorithms have a wide domain of practical applications, some of which are described in the next section.

Table 2. ML Algorithms

Paradigm	Algorithm	Description

Supervised Learning	Decision Tree	Decision Tree is a technique for approximating discrete valued target function which represents the learnt function in the form of a decision tree [10]. A decision tree classifies instances by sorting them from root to some leaf nodes on the basis of feature values. Each node represents some decision (test condition) on attribute of the instance whereas every branch represents a possible value for that feature. Classification of an instance starts at the root node called the decision node. Based on the value of node, the tree traverse down along the edge which corresponds to the value of the output of feature test. This process continues in the sub-tree headed by the new node at the end of the previous edge. Finally, the leaf node signifies the classification categories or the final decision. While using a decision tree, focus is on how to decide which attribute is the best classifier at each node level. Statistical measure like information gain, Gini index, Chi-square and entropy are calculated for each node to calculate the worth of that node [10]. Several algorithms are used to implement decision trees. The most popular ones are: Classification and Regression Tree (CART), Iterative Dichotomiser 3 (ID3), Automatic Interaction Detection (CHAID), Chi-Squared C4.5 and C5.0 and M5

	Naïve Bayes	Naïve Bayes classifies using Bayes' Theorem of probability. Bayes' theorem calculates the posterior probability of an event (A) given some prior probability of event B represented by P(A/B) as follows: $$P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$ Where, • A and B are events. • P(A) and P(B) are the probabilities of observing A and B independent of each other. • P(A/B) is the conditional probability, i.e. Probability of observing A, given B is true. • P(B/A) is the probability of observing B, given A is True. Naïve Bayes' classifiers fall under the category of simple probabilistic classifiers based on the concept of Bayes' Theorem having strong independence assumptions among the features. It is particularly suited when the dimensionality of the inputs is high [4,5].

	Support Vector Machines	SVMs can be used for classification as well as regression problems. It is a supervised learning algorithm. It works on the concept of margin calculation. In this algorithm, each data item is plotted as a point in n-dimensional space (where n is the number of features we have in our dataset). The value of each feature is the value of the corresponding coordinate. It classifies the data into different classes by finding a line (hyper plane) which separates the training datasets into classes. It works by maximizing the distances between the nearest data point (in both classes) and the hyper plane that we can call as margin.

	Regression Analysis	Regression analysis is a predictive modelling technique which investigates the relationship between a dependent (target) and independent variable(s) (predictor). It is an important tool for analysing and modelling of data. In this method, we try to fit the line/curve to the data points so as to minimize the differences between distances of data points from the curve or line. There are various kinds of regression analysis like linear, logistic and polynomial.
Unsupervised Learning	K-Means Clustering	K-means is a popular unsupervised machine learning algorithm for cluster analysis. Its goal is to partition ‘n’ observations into ‘k’ clusters in which each observation belongs to the cluster having the nearest mean, serving as a prototype of the cluster. The mean of the observations in a particular cluster defines the center of the cluster.
Instance based Learning	K-nearest Neighbours	It is a non-parametric method used for classification and regression. Given N training vectors, KNN algorithm identifies the k-nearest neighbours of an unknown feature vector whose class is to be identified.
Ensemble Learning	Random Forest	It is an ensemble learning method used in classification and regression. It uses bagging approach to create a bunch of decision trees with random subset of data. The output of all decision trees in the random forest is combined to make the final decision trees. There are two stages in Random Forest Algorithm, one is to create random forest, and the other is to make a prediction from the random forest classifier created in the first stage.
Dimensionality Reduction	Principal Component Algorithm	It is primarily used for reducing dimensionality of data set. It helps in reducing the number of features of the data set or the number of independent variables in the data set. It uses orthogonal transformation to convert correlated

		variables into a set of linearly uncorrelated variables called principal components.

7. Applications of Machine Learning

Machine learning problems range from game playing to self-driven vehicles. Table 3 enlists some popular real-life applications of ML.

Table 3. Applications of ML

Application	Description
Playing Checkers Game	A computer program learns to play checkers game, improvises its performance as determined by its ability to win at various class of tasks involving the game, through experience obtained by playing games against itself.
Speech Recognition	The most sophisticated speech recognition systems these days deploy machine learning algorithms in some forms. Example: SPHINX system [20] learns speaker-specific sounds and words from speech signals. Various Neural Network learning methodologies for interpreting hidden Markov Models are highly effective for automatically customizing speakers, dictionary, noise etc.
Autonomous Vehicles	Machine learning models are these days being applied to drive autonomous vehicles like Cars, Drones etc. Example: Google Driver Less Cars, Tesla Cars. Machine learning techniques are also highly effective in controlling sensor-based applications.
Filtering Emails (Spam Emails)	Machine learning can be applied to filter spam emails. The machine learning based model will simply memorize all the emails classified as spam emails by user. When new email arrives in inbox, the machine learning based model will search, compare and based on the previous spam emails. If new email matches any one of them, it will be marked as spam; else it will be moved to user's inbox.
Robotics and Artificial Intelligence	Machine learning is regarded as improved approach to problem solving. Using base knowledge and training data with machine learning models, learning can be improved which will take robotics and AI to next generation levels.
Web and Social Media	• Naïve Bayes classifiers have been successfully applied in the field of text mining, may it be spam filtering or classifying the web page, an email or any document. • Facebook uses Naïve Bayes' to analyze status update expressing positive and negative emotions. • Document Categorization: Google uses Naïve Bayes algorithm for document categorization. • K-means clustering is used by search engines like Google, Yahoo to cluster web pages by similarity. • Apriori is used by websites such as Amazon or Flipkart to recommend which items are purchased together frequently. • Another common application of Apriori is the Google auto-complete. When a person types a word, Google search engine looks for other associated words that go together with the word earlier typed word.

	• Sentiment analysis on social networking sites is a typical text classification problem solved using application of variety of ML algorithms [21, 22, 23,24]
Medical Field	TRISS: Trauma & Injury Severity Score, which is widely used to predict mortality in injured patients, was originally developed by Boyd et al. using logistic regression. Many other medical scales used to assess severity of a patient have been developed using logistic regression.
Bayesian Methods	Bayes' Theorem is one of the most popular methods for calculating conditional probabilities given the set of hypothesis. It can be used to solve complex data science and analytics problems by integrating with various Machine Learning models and algorithms. Some examples real world problems that can be solved using Bayesian methods are: • Suggestions on Netflix • Auto-correction • Credit card fraud detection • Google page ranking • Facial recognition • Climate modeling forecast • Stock trading systems

8. Analysis and Discussion

Given the wide range of applicability of ML, it faces a number of challenges. Some of them are as follows:

- The machine learning algorithms require large volumes of data to be accurate and efficient which is still not available to researchers. Tech giants like Facebook and Google have had access to such enormous data which is why they are leading in the field of Artificial Intelligence. It becomes even more difficult to get this data in the fields like banking and healthcare where sparse digital data is available making it tough to make accurate predictions.
- Spam Detection: Given email in an inbox, the intelligent systems developed so far are still not able to correctly detect the spam mail. It ends up in sending spam in inbox and non-spam mails to spam directory.
- Machine learning algorithms have not yet been successful in identifying the objects and images. This field is still an open research field for machine learning. Though we have mentioned a few challenges in machine learning, there are many more fields which still challenging for deep learning algorithms i.e. speech understanding, credit card fraud detection, face detection, digit recognition given a zip code, and product recommendation etc.

Machine learning algorithms are in continuous development and will definitely become more widespread in the years to come. They are useful in many different applications; there is significant brainpower and funding behind pushing the boundaries towards innovation [25]. Some open areas of application include:

- Deep learning, e.g. for predicting stock market trends, designing circuits, identifying illnesses, designing voice-controlled devices, and much more (save special attention for Generative Adversarial Neural Networks)
- Data mining and big data analytics, e.g. for predicting business market trends
- Natural language processing, e.g. in search engines
- Hardware accelerators for new AI architectures, e.g. from AMD and Intel
- Simulation environments for evaluation and testing, e.g. for self-driving cars and virtual reality
- Healthcare machine learning (medical imaging, working with clinical data, making sense of genomic data of huge populations)
- HCI (Human Computer Interaction), keep advancing better interfaces and usability between different devices with the rise of cloud computing and IoT

Given the current rate of advancement in the field, it has a bright future. Applications of machine learning have the potential to expand dramatically in the near future.

9. Machine Learning in Future

Machine learning is likely to be a part of almost all software programs in the near future. Here are some ideas about how machine learning will be used in the future: With Natural Language Processing and machine learning, computers can quickly learn to understand the meaning and context of words. Soon, computers will be able to talk like people. In the near future, machine learning tools and methods will be able to connect to the internet and keep track of the most important data. This will make it easier for programs to remember what they've learnt, so they won't have to be trained over and over again. It's possible to improve personalisation and suggestions, which would lead to better and more useful experiences.

10. Conclusion

Because of digitisation and the Internet change, there is more and more organised and random data that needs to be analysed. Machine learning is one of the most important trends in technology. It gives computers the smarts to get information from data. Researchers and practitioners alike are using machine learning to solve difficult problems in real life. This has made the field of study busy and dynamic, with people from many businesses and countries contributing. This essay gives an in-depth look at the steps and methods used in Machine Learning. The goal is to fully understand what Machine Learning is for, how it can help, and how far it can go as a technology-based answer.

References

[1] There are some works in machine learning that use the game of checkers [1]. The IBM Journal of Research and Development published in 1959, July, 3(3), 210–209.

[2] Mitchell, T. (1997). Page 2 of Machine Learning by McGraw Hill. ISBN 978-0-07-042807-

[3] Hayashi, Chikio (January 1, 1998). "What does Data Science mean? Some basic ideas and a heuristic example. Data science, sorting, and methods that are similar. Research into sorting things into groups, analysing data, and putting together knowledge. Springer Japan, pages 40–51. Find more at doi:10.1007/978-4-431-65950-1_3. ISBN 9784431702085

[4]"A Review on Machine Learning Techniques" by Sandhya N. Dhage and Charanjeet Kaur Raina was published on March 16, 2016, in the International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), ISSN: 2321-8169, Pages 395–399 .

[5]AyonDey, "Machine Learning Algorithms: A Review," International Journal of Computer Science and Information Technology (IJCSIT).

[6] A study from the Royal Society in April 2017 called "Machine learning: the power and promise of computers that learn by example" (ISBN: 978-1-78252-259-1) and Information Technologies, Vol. (3), 2016,1174–1179

[7] Minton S, Zweben M. How People Learn, Plan, and Schedule: A Brief Overview. The book Machine Learning Methods for Planning 1993 has pages 1–29.

[8] Sejnowski T. Net talk: A parallel network that learns to read out loud. Complex Systems 1 (1987): 145–68.

[9] Han J, Cai Y, & Cercone N. Finding numeric rules in relational databases through data. Journal of IEEE Transactions on Knowledge and Data Engineering, vol. 1, no. 1, 1993, pp. 29–40.

[10] Member inference attacks on machine learning models by Shokri R, Stronati M, Song C, and Shmatikov V. Safety and Privacy (SP), 2017 IEEE Symposium on May 22, 2017 (pp.3–18). IEEE.

[11] Li M, Jiang JW, Smola AJ, Ahmed A, Josifovski V, Long J, Shekita EJ, and Su BY. Usingthe Parameter Server to scale up distributed machine learning. On October 6, 2014, InOSDI published Volume 14, Pages 583–598.

[12] Chen JX. The history of computers: AlphaGo. Computing in Science & Engineering. 2016 Jul;18(4):4–7.

[13] A study by Al-Jarrah OY, Yoo PD, Muhaidat S, Karagiannidis GK, and Taha K. A look at how to make machine-learning work well with big data. 87–93. Big Data Res. 2015; 2(3): 87–93.

[14] The article "A Review of Machine Learning Algorithms for Text Documents Classification" by A. Khan, B. Baharudin, and L. Hong Lee was published in the Journal of Advances in Information Technology in February 2010.

[15] Zhou ZH. (2009). Ensemble Learning. Li S.Z. and Jain A. (eds.) Encyclopaedia of Biometrics.Springer, Boston,MA

[16] Tom M. Mitchell, "Machine Learning," WCB McGraw-Hill, 2013. ISBN-13: 978-1-25-909695-2, ISBN-10:1-25-909695-5.

[17] The use of machine learning to track physical exercise with shoe sensors (Tang W, Zhang T, & Sazonov E, 2017). Cyberphysical Systems: Integrated Computing and Engineering Design. September 26, 2013, 187 pages.

[18] Zhu X. A review of the literature on semi-supervised learning. Computer Science, University of Wisconsin–Madison; 2006;2(3):4.

[19] Wright IH, Frank E, Hall MA, and Pal CJ. Data mining: useful tools and methods for machine learning. Morgan Kaufmann; October 1, 2016.

[20] Lee KF, Hon HW, Hwang MY, Mahajan S, and Reddy R. “The SPHINX speech recognition system.” International Conference on Acoustics, Speech, and Signal Processing (ICASSP '89), Vol. 1, pp. 445–448; May 1989.

[21] Kumar A. and Sharma A. wrote a paper called “Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence” (21). Webology, December 2017,Vol.14, No. 2.

[22] An empirical study of Twitter and Tumblr for sentiment analysis using soft computing techniques was presented at the International Conference on Soft Computing and Applications (ICSCA 2017), which was part of the World Congress on Engineering and Computer Science 2017. The paper can be found in volume 1, pages 1–5, and has the ISBN number 978-988-14047-5-6.

[23] Kumar A. and Abraham A. wrote an article called “Opinion Mining to Assist User Acceptance Testing for Open-Beta Versions”. It was published in the Journal of Information Assurance and Security in 2017 and can be found on pages 146–153.

[24] Billsus D. and Pazzani MJ. Content-based suggestion systems. The flexible web 2007 (pp. 325– 341). Springer in Berlin and Heidelberg.

[25] Information Retrieval & Machine Learning: Supporting Technologies for Web Mining Research & Practice' by Bhatia MPS and Kumar A. was published in Webology in Volume 5, No. 2 in 2015.