

A Systematic Review of Techniques, Tools and Applications of Machine Learning

Abhishek, Student MCA
Manav Rachna International Institute
of Research and Studies, Faridabad,
India
abhisheknkumar97te@gmail.com

Abhishek Dhankar, Student MCA
Manav Rachna International Institute
of Research and Studies, Faridabad,
India
abhishekdhankhar11@gmail.com

Neha Gupta
Manav Rachna International Institute
of Research and Studies, Faridabad,
India
neha.fca@mriu.edu.in

Abstract— Machine Learning is a part of Artificial Intelligence. A branch of artificial Intelligence(AI), that offers the capability to the system by learning on their own and work better from experience without human intervention. By using machine learning (ML), software applications become more precise for outcome prediction with no requirement of explicit programmed. Machine learning has a fundamental concept i.e. building of algorithms for accepting input data and with the help of statistical analysis they can predict the output. Nowadays, machine learning has gained vast popularity, and its algorithms are used in every field like object detection, pattern recognition, text interpretation, and different research areas. One of the basic goals of machine learning is to teach computers to use the data to solve a particular problem. A fair range of machine-learning applications include e-mail classification training to distinguish spam from non-spam messages, fraud detection, etc. This paper aims at varied machine learning methods, algorithms, and tools required to run the machine learning projects. It also focuses on the advancements that are administered in order so that the current researchers can be benefited out of it.

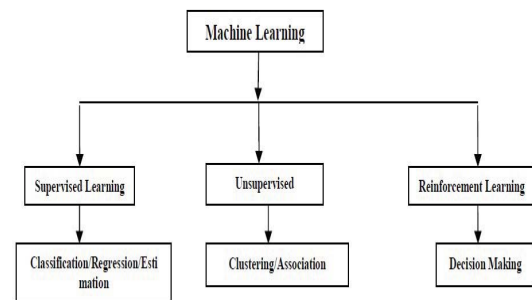
Keywords—Machine Learning, Clustering, Supervised Learning, Reinforcement Learning, Unsupervised Learning

I. INTRODUCTION

Machine Learning (ML) can be defined as a division of artificial intelligence which is based on the concept that machines can gain knowledge from data by itself, they can recognize already defined patterns. On the basis of this, machines can make decisions with negligible human involvement. It is the learning of a computer on its own from experience. As a perception of scientific development and improvement in technical AI systems, previous experiences without specific assistance code now allow them to learn and improve whenever exposed to new data. Machine learning is gaining attention and importance because of its meaningful impact in decision making. The main focus of machine learning is to apply various methods to find and extract hidden patterns from large stored databases. Machine learning and data mining algorithms are used in various fields' like cluster analysis, telecommunication industry, web search, tourism industry, financial forecasting, spam filter, networking, travel, credit scoring, fraud detection, computer design, electric load forecasting and many more. There are three categories of machine learning as shown in Fig.1

II. METHODS USED IN MACHINE LEARNING

A large number of ML algorithms have been implemented in recent years. Few may solve the problem in order to replace it with another. Common machine learning is based on unsupervised learning, reinforcement learning and supervised learning.



a) Supervised learning: The data is made up of a series of pre-labeled, target data input variables (training data). It produces a mapping function to map inputs to the needed outputs with the input variables. The adjustment of parameters continues until the device acquires an appropriate degree of accuracy for the teaching data. In this type of learning, predictions can be made on the basis of a known dataset (training dataset). An input variables (X) and output/response variables(Y) can be included in the above mentioned training dataset. These variables are helpful for building a model to predict the value of response variable(Y) for testing data. In supervised learning, to learn the mapping function with the help of input variables (X) and an output/response variable (Y), an algorithm can be created. Thus the mapping function would be:

$$Y = f(X)$$

b) Unsupervised learning: We only have training data in this type of learning, rather than outcome data. This data is not classified beforehand. In classifiers, existing patterns or clusters are recognized within data sets. It is a technique of machine learning in which there is no requirement of supervision for a model. Model will perform output by

discovering information on its own. This type of learning mainly handles unlabeled data. With the help of unsupervised learning algorithms, more difficult tasks can be performed easily than supervised learning algorithms. The most common unsupervised learning method is cluster analysis which is used to find a structure or pattern from the pool of uncategorized data.

c) Semi-Supervised learning: It consists of a middle space between supervised and unsupervised algorithms. It contains a combination of little amount of labelled data with huge amount of unlabelled data.

d) Reinforcement learning: The machine is conditioned to formulate action on a particular decision using this learning process, which results in reward or feedback. The machine was programmed to look through past experience for the most meaningful activities. Under this method, suitable actions can be taken in a particular condition to get the reward maximization. Learning is possible only when there is an interaction with the environment by the agents or learners. Rewards will be given to agents for their right performance and they will be penalized for wrong performance. No human intervention is required for learning by the agent. Reward maximization and penalty minimization is the way by which an agent can learn. There is a system of reward and punishment for training of algorithms.

III. ALGORITHMS OF MACHINE LEARNING

Massive numbers of algorithms are employed to erect models of machine learning. All algorithms are grouped by their learning methodology, as follows:

a) Regression algorithms - Regression algorithms are making predictions by modeling the relationship between variables using an error measure. The Regression methodology predicts continuously varying values. It is a statistical tool with the help of which dependent and independent variables' relationship could be found. It is also used for modelling a target value that is completely dependent on independent predictors. To find correlation between two variables, regression analysis is used. Cause and effect relationship between variables cannot be found with regression. Regression and classification differ to each other in terms of output variable as shown in fig. 2

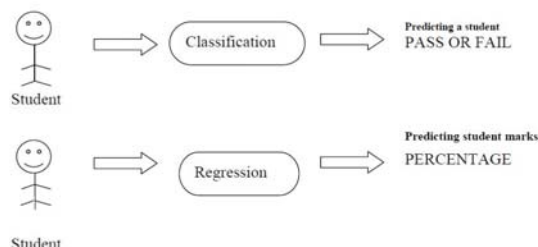


Fig.2 Classification vs. Regression Output

b) Algorithms using Decision Tree - Decision trees algorithms are primarily used for classification problems. By sorting them by using their values, they divide attributes into two or more classes. There are nodes and branches in every tree. Each node is the attribute of the group and each value is represented by one branch.

c) Bayesian algorithms - Machine learning, including statistical data and algorithm, is the multidisciplinary field of computer sciences. This model is easy to construct and does not have a complex iterative parameter estimation making it very useful for large datasets.

d) Modelling using Artificial Neural Network - Artificial neural networks based models use configuration of the biological neuron and work as per the supervised approach of machine learning. It is made up of artificial neurons that have weighed connections between units. Parallel distributed computing networks are also well known to them.

e) Data clustering algorithms- This algorithm separates objects into various batches. It groups the item into clusters with some resemblance between each sub-set. It is an unsupervised method of learning and its techniques are known as clustering and partitioning of the hierarchy or network.

f) Algorithms for Association Rule Mining - It is a technique to find out rules between different data objects within a large database. e.g. people that are going to watch a movie, most likely to buy popcorn..

g) Ensemble Algorithms- These Algorithms are based on unsupervised learning techniques. It separates the data into several different data groups. These groups develop self-supporting learning models. These learning models are mixed together to establish the right hypothesis.

h) Algorithms using Dimensionality Reduction- In case of enormous number of dimensions, large space volumes are typically used to minimize dimension. The statistical importance is needed. Methods for reducing dimensionality used to restrict the number of dimensions represented by the object and delete unrelated and unessential data that minimize computer cost. Classification and regression are based on two methods.

IV. BACKGROUND / RELATED PAPERS

A mathematical approach was introduced by Authors Foulloy et al. (2018)[3] and was employed using a machine learning model to analyze and apply it to the prevision of the working time of solar irradiation. The high, low, or medium weather variance was employed in this strategy, such as Ajacio, Odeillo and Tilos. The model was contrasted with the average and multi-layer automated regression control.

Chen et al. (2011) [4] discussed a procedure to forecast the demand of tourism that is SVR built using chaotic genetic algorithm (CGA), like SVRCGA, which overcomes the premature local optimum problem. This paper suggested a

SVRCGA model that has outclassed other methodologies reviewed in the paper.

Shahrabi et al. (2013) [6] have proposed a new combined intelligent model for the estimation of tourism demand. Modular Genetic-Fuzzy prediction system with the assistance of a genetic fuzzy expert system. It noticed, however, that the exactness of the power prediction of MGFFS is better than the approaches of the Classical Time Series model.

Guang-Bin Huang et al. (2015) [7] discusses the fundamental functions of ELMs, such as kernels, random features and random neurons, compares ELMs' output and tends to outline classification, vector supporting and regression applications.

The next day's stock price forecasting model, based on the four layer FMAS (Fuzzy Multi Agent System), is proposed by Turksen et al. (2012) [8]. The teamwork of smart agents was used for the artificial intelligence model. Authors investigate that FMAS is an effective method for problems with stock price prediction as it exceeds all pre-emptions.

Table 1 shows the comparative study of literature review related to various machine learning models

TABLE I. COMPARATIVE STUDY OF LITERATURE REVIEW RELATED TO MACHINE LEARNING

S. N.	Name of Author(s)	Method/Technique Used	Result
1	Diego Buenaño-Fernández, David Gil and Sergio Luján-Mora (2019)	Decision Tree/Classification	A methodology was built up to monitor and predict grades of students.
2	Boran Sekeroglu, Kamil Dimililer and Kubra Tuncal (2019)	To predict student performance, following machine learning algorithms have been used; Long-Short Term Memory (LSTM,) Support Vector Regression(SVR) Backpropagation (BP)	Highest Prediction rate was achieved by Support Vector Regression (SVR) while Backpropagation (BP) was able to find lowest Prediction rate
3.	E. T. Lau, L. Sun and Q. Yang(2019)	11 input variables, two layers of hidden neurons and one output layer was used in the composition of Artificial neural network model.	To evaluate and predict the student's CGPA, a model was built using the their socio-economic background and entrance exam result data
4.	Sahar Al-Sudani and Ramaswamy Palaniappan(2019)	Multilayered Neural Network was used to identify the student's degree	A model was built using Multilayered Neural Network and compared the

		class whether it is of a good or bad	same with other classifiers like K-Nearest Neighbor, Support Vector Machine and Decision Tree,
5.	Lubna Mahmoud and Abu Zohair(2019)	Support vector machine with radial kernel and Linear discriminant analysis classifier were used to predict students' performance.	A prediction model was created using the key indicators in the small dataset with the help of visualization and clustering algorithm
6.	Wenhao Xie, Gongqian Liang, Z honghui Dong, Baoyu Tan, and Baosheng Zhang(2019)	Sample Selection Strategy along with Random-SMOTE algorithm	Imbalanced classification was handled using Support Vector Machine (SVM) classification algorithm.
7.	N. Sai Sragvi Vibhushan and Vikas B. (2018)	feature selection algorithms namely Principal Component Analysis, ReliefAttribute Eval and CfsSubsetEval.	A classification model was built for the analysis of various classifiers namely Naïve Bayes, Sequential Minimal Optimization, J48 and Multilayer Perceptron with and without the ensemble method,
8.	Dhanashree Mane, Pranali Namdas and Pooja Gargade (2018)	C4.5, Artificial Neural Network and Q Learning (Reinforcement learning algorithms)	Generate prediction model using machine learning algorithms
9.	Mushtaq Hussain and Wenhao Zhu•Wu Zhang (2018)	Digital electronics education and design suite (DEEDS) was used by TEL system along with machine learning algorithms like Artificial Neural Network and Support Vector Machine	Generate Higher accuracy model using K-fold cross validation
10 .	Ali Daud, Naif Radi Aljohani and Rabeeh Ayaz Abbasi(2017)	Generative and discriminative classification models using Learning Analytics	A Significant model was generated to get the result by using machine learning algorithms
11 .	Sungar,V. A., Shinde P.D. and Rupnar M.V.(2017)	Incremental Learning algorithms	Model is built using ensemble technique with incremental algorithm for career selection path
12	Yasmeen	ID3 decision	A model was

.	Altujjar, Wejdan Altamimi, Isra Al-Turaiki and Muna Al-Razgan (2016)	tree induction algorithm	generated for the identification of critical courses which are a great indicator of student level of achievement.
13	Hana Bydžovská (2015)	Collaborative Filtering Method for student modeling	A model was built up to check that whether the students can be adequately characterized with the help of their previously passed course. The with the help of model built in results has shown that

V. SIGNIFICANT APPLICATIONS OF MACHINE LEARNING

Significant applications of Machine learning techniques are broadly classified into supervised and unsupervised learning applications. Classification techniques include applications related to face recognition, pattern recognition, character identification, medical diagnosis and web advertising are subject to supervised training. Clustering, association analysis, CRM, summarization, compression of images, bioinformatics can be used for unattended research. Enhanced game-play learning and robot control are widely used.

a) Application of ML in Drug Discovery:

Drug discovery and development pipelines are long, complex and varied. Machine Learning (ML) methods include a range of tools to enhance exploration and decision-making with ample and high quality data for well-specified questions. ML implementations occur at all drug development stages. Examples include target validation, prognostic biomarker identification and analysis of clinical trial digital pathology data.

b) Application of ML in cancer prediction and prognosis:

ML's capability to identify main characteristics from complicated datasets indicates its significance in detecting cancer [11]. A variety of ML techniques have been widely applied in cancer research to develop predictive modelling which lead to effective and accurate decision-making, including Artificial Neural Networks (ANNs), Bayesian Networks (BNs), Supporting Vector Machines (SVMs) and Decisional Trees (DTs). While ML methods can clearly improve understanding of the progression of cancer, an appropriate degree of validation is required to take these methods into account in daily clinical practice.

c) Applications of ML in Cyber Security:

Machine learning is used effectively for designing authentication systems, testing the functionality of the

protocol, assessing the safety of human interaction proofs, smart meter data profiling, etc. Machine learning helps ensure that various systems are secure. Some work has been done to improve the efficiency of algorithms for machine learning and to protect them against various attacks.

d) Application of ML in Healthcare:

At present there are abundant data on medical imaging stored electronically and Deep Learning algorithms can be fed with these datasets for trends and anomalies to be observed and discovered. Imaging data can be interpreted as much as highly qualified radiologists can by machines and algorithms identify suspected spots on the skin, lesions, tumors, and brain bleeds. Therefore it is important to extend exponential use of AI / ML tools / platforms to assist radiologists.

VI. CONCLUSION

Within this paper, machine learning methods and algorithms have been discussed along with their implementations. This paper also explored algorithms that explain the different types of machine learning techniques. Various machine learning applications and numerous processing tools are also being examined. We studied various algorithms for machine learning in different fields, in combination with conventional approaches, applied in recent years and studied the limit values of previous models in the literature review section.

REFERENCES

- [1] Law R. (1998), "Room occupancy rate forecasting: a neural network approach", International Journal of Contemporary Hospitality Management, vol. 10 Issue 6, pp 234 – 239
- [2] Chen K-Y, Wang H. (2007), "Support vector regression with genetic algorithms in forecasting tourism demand", Tourism Management 28, pp 215–226.
- [3] Makridakis S., Spiliotis E., Assimakopoulos A. (2018), "Statistical and Machine Learning forecasting methods: Concerns and ways forward", PLoS ONE 13, 2018.
- [4] Fazel Zarandi M. H, Hadavandi E., Turksen B. (2012), "A Hybrid Fuzzy Intelligent Agent-Based System for Stock Price Prediction", International Journal Of Intelligent Systems, Vol. 00, pp 1–23.
- [5] Barzegar R., Moghaddam A., Adamowski J., Ozga-Zielinski B. (2017), "Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model", Springer, ISBN 00477-017-1394.
- [6] Pai P.F., Hung K-C, Lin K-P. (2014), "Tourism demand forecasting using novel hybrid system", Expert Systems with Applications 41, pp 3691–3702
- [7] Sun X., Sun W., Wang J., Zhang X., Gao Y. (2016), "Using a Greye Markov model optimized by Cuckoo search algorithm to forecast the annual foreign tourist arrivals to China", Tourism Management 52
- [8] Shahrabi J., Hadavandi E., Asadi S. (2013), "Developing a hybrid intelligent model for forecasting problems: Case study of tourism demand time series", Knowledge-Based Systems 43, pp 112–122.
- [9] Hong W-C., Dong Y., Chen L-C, Wei S-Y. (2011), "SVR with hybrid chaotic genetic algorithms for tourism demand forecasting", Applied Soft Computing 11, pp 1881– 1890.

- [10] Zhang Y., Wang Y., Zhou G., Jin J., Wang B., Wang X., Cichocki A. (2018), "Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces", *Expert Systems With Applications* 96, pp 302–310.
- [11] Buenaño-Fernández, D., Gil, D. & Lujan-Mora, S., 2019, "Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study", *International Journal of environmental, cultural, economic, and social sustainability of human beings*, MDPI, 11(10), pp.1-18.
- [12] Sekeroglu, B., Dimililer, K. & Tuncal, K., 2019, "Student Performance Prediction and Classification Using Machine Learning Algorithms", *ICEIT 2019: Proceedings of the 2019 8th International Conference on Educational and Information Technology*, March 2019, pp.7–11
- [13] Lau, E.T., Sun, L. & Yang, Q., 2019, "Modelling, prediction and classification of student academic performance using artificial neural networks", *International Journal of SN Applied Sciences*, 1(9), pp.1-10
- [14] Sudani, S. & Palaniappan, R., 2019, "Predicting students' final degree classification using an extended profile", *Education and Information Technologies (Springer Journal)*, 24, pp.2357-2369
- [15] Mahmoud, L. & Zohai, A., 2019, "Prediction of Student's performance by modelling small dataset size", *International Journal of Educational Technology in Higher Education*, 16, pp.1-18
- [16] Xie, W., Liang, G., Dong, Z., Tan, B. & Zhang, B., 2019, "An Improved Oversampling Algorithm Based on the Samples' Selection Strategy for Classifying Imbalanced Data", *Mathematical Problems in Engineering by Hindawi*, 2019, pp.1-13
- [17] Sragivi, S.N., Vibhushan & Vikas., 2018, "Evaluation Of Various Feature Selection algorithms In Educational Data Mining", *International Journal of Computational Engineering Research*, 8(6), pp.225-3005
- [18] Mane, D., Namdas, P., Gargade, P. & Rathi, S.S., "Predicting student performance using machine learning approach", *Vishwakarma Journal of Engineering Research*, 2(4), pp.13-18
- [19] Hussain, M., ZHU, W., Zhang, W. & Ali, S., 2018, "Using machine learning to predict student difficulties from learning session data", *Internal Journal of Artificial Intelligence Review*, 52(1), pp 22-134
- [20] Bakhshinategh, B., Zaiane, O.R., ElAtia, S. & Ipperciel, D., 2018, 'Educational data mining applications and tasks: a survey of the last 10 years', *Education and Information Technologies*, 23(1), pp.537–553.
- [21] Daud, A., Aljohani, N.R., Abbasi, R.A. & Abbas, F., 2017, "Predicting Student Performance using Advanced Learning Analytics", *Proceedings of the 26th International Conference on World Wide Web Companion*, pp.415-421
- [22] Sungar, V.A., Shinde, P.D. & Rupnar, M.V., 2017, "Predicting Student's performance using Machine Learning", 2017, *International journal of communications on applied electronics*, 7(11), pp.11-15
- [23] Chaudhari, K.P., Sharma, R.A., Jha, S.S., & Bari, R.J., 2017, "Student performance prediction system using data mining approach", *International Journal of Advance Research in Computer Science and Communication Engineering*, 6(3), pp.833–839.
- [24] Altujjar, Y., Altamimi, W., Turaiki, I. & Razgan, M., 2016, "Predicting critical courses affecting student performance: A case study", *Symposium on Data Mining Applications*, 82(2016), pp.65-71
- [25] Saa, A.A., 2016, 'Educational Data Mining & Students' Performance Prediction', *International Journal of Advanced Computer Science and Applications*, 7(5), pp.212-220.
- [26] Bydzovska, H., 2015, "Student Performance prediction using collaborative filtering methods", *International Conference on Artificial Intelligence in Education*, 5, pp.550-553. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (*references*)