

An Empirical Research on Machine Learning Algorithms and Its Applications

Poorva Sanjay Sabnis

Assistant Professor, School of Management Sciences, Varanasi

ABSTRACT

In the today's era of globalization, machine learning has gained a lot of attention and popularity due to its vital play in the life of human being. The word "Machine Learning" can be defined as a set of computer algorithms that enables different systems to learn autonomously and produces outputs and additionally improve the results from various analysis and outputs. For that purpose, data will be provided to these algorithms, by which they automatically get trained to perform a certain task and then get a certain output. By analyzing the data, process and output, we can train the system and apply that for different real-life scenarios. There are various machine learning algorithms are available which can be used in various fields such as pattern recognition, object detection, text interpretation and different research areas. Basically, Machine learning is a part of AI (artificial intelligence) and it is used in the designing of algorithms based on the recent trends of data. This paper aims to introduce different machine learning algorithms, types of algorithms and highlighting the uses of machine learning algorithms in different sectors. It also focuses on how machine learning algorithm actually works. Overall this paper produces the work done by the authors in the area of machine learning and its algorithms.

poorva@smsvaranasi.com

Keywords : Algorithm, Machine learning, Data, Artificial intelligence

Computing Trendz (2020). DOI: 10.21844/cttjetit.v10.i01.17095

Introduction

Machine learning is playing a vital role in the study of computer science and it is considered as a part of Artificial Intelligence. In machine learning, the training is given to the machines to perform some specific task by its own. Basically, the origin of machine learning is from computational learning and pattern recognition which are under AI (artificial intelligence) area. For training which is given to the machine is based on some data sets, and based on these values, various algorithms are created and then applied [1]. For that purpose, the machine can make predictions and learn on the provided data sets, respectively. A model is programmed, then procedures are operated upon it, and inputs are taken to get data-driven predictions.

Corresponding Author: Poorva Sanjay Sabnis, Assistant Professor, School of Management Sciences, Varanasi, poorva@smsvaranasi.com

How to cite this article: Sabnis, P.S.. (2020). An Empirical Research on Machine Learning Algorithms and Its Applications, *Computing Trendz* 10(1&2): 34-42

Source of support: Nil

Conflict of interest: None

This procedure is follow to avoid firm static program instructions. Each instance in the data set which is used by the machine learning algorithms is depicted by features belonging to the same set [1]. These features can be categorical, binary or continuous. In the supervised learning, data is labeled; otherwise, it is unsupervised learning. Various applications are based on supervised learning.

Consider an example of recognizing spoken speech

in which conversion of the acoustic speech signal to ASCII text is taken place. There may be differences in pronunciation of words because of variations in gender, age or pronunciation. In this way, the idea in machine learning is to collect an extensive collection of data from people and plot those towards. The algorithms of machine learning are non-interactive because the model is built from the results gathered by the data set [2]. The past observations are studied to produce predictions. Machine learning is very much helpful to differentiate which emails are spammed and non-spammed. This could be achieved by gathering examples of non-spammed and spammed emails. Further, we can apply some algorithms to these samples to differentiate between spammed and non-spammed email [2].

Supervised Learning:

The basic motto is to map the function approximately which is mentioned above; so that whenever there is a new input data (x) then using the corresponding output variable, prediction is performed. It is called as supervised learning because it is the process of learning (from the training dataset) which may be thought like a teacher is supervising the entire learning process. In this way, we have to prepare “learning algorithm” which can iteratively makes predictions on the data provided for the training and is corrected by the “teacher”. This learning may stop when the algorithm achieves a level which is having acceptable performance or the desired accuracy [3].

Example of Supervised Learning can be given as following-

Consider a basket filled with some fresh fruits like apples, bananas, cherry, grapes etc.; now there is a task to arrange the same type of fruits in one dish.

Here, the earlier work is called as “training data” which learns the things from the training. It will happen because of a response variable which initiates y, if some particular fruit is there. This type of information is deducted from the data which can be used to train the model. This type of learning is called Supervised Learning. These problems are categorized under Classical Classification Tasks.

Unsupervised Learning:

The motto of unsupervised learning is to model the distributed data to learn more about the data. It is called as unsupervised learning because there is no particular answer for particular question and there is no such teacher present to train the data (unlike supervised learning). Here, algorithms are left to choose their own ways to discover and present the excellent structure in the data.

Example of Unsupervised Learning can be given as follows-

As similar example, consider a basket with some fresh fruits. Again the task is to arrange the same type of fruits in one dish [4].

This time there is no information about those fruits beforehand, it's the first time that the fruits are being seen or discovered. Now there is challenge how to group similar fruits without any prior knowledge about those. For that purpose, firstly any physical characteristic of a particular fruit is selected, for example color.

Then the fruits are arranged on the basis of the color. Now the groups will be created as follows:

RED COLOR GROUP: apples & cherry fruits.

GREEN COLOR GROUP: bananas & grapes.

So now, take another physical character say, size, so now the groups will be something like this.

RED COLOR AND BIG SIZE: apple.

RED COLOR AND SMALL SIZE: cherry fruits.

GREEN COLOR AND BIG SIZE: bananas.

GREEN COLOR AND SMALL SIZE: grapes.

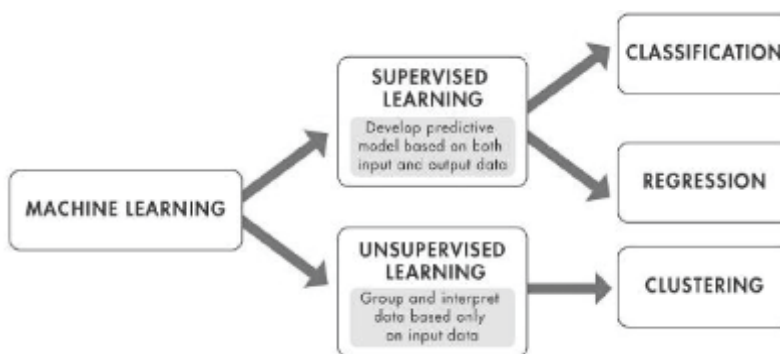
And the job is done. Here, there is no need to know or learn anything beforehand. That means no train data and no response variable. This type of learning is known as Unsupervised Learning.

Table 1 Comparison between supervised and unsupervised learning

Supervised Learning	Unsupervised Learning
Uses trained labeled data.	Uses trained unlabeled data.
Takes direct feedback to check whether prediction about output is correct or not.	No any feedback.
Prediction about the output.	Finding the hidden patterns in data.
Provision of input data as well as output to the model.	Provision of only input data to the model.
The main aim is to train the model so that it can predict the output when it is given new data.	The main aim is to find the hidden patterns and useful insights from the unknown dataset.
Requires supervision to train the model.	No requirement of any supervision to train the model.
Further division- Classification and Regression problems.	Further division- Clustering and Associations problems.
Generates more accurate result.	Less accurate result as compared to supervised learning.
It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc.	It includes various algorithms such as Clustering, KNN, and Apriori algorithm.

We can define Machine Learning is a method of data analysis that automates analytical model building [5]. It is a sub-branch of artificial

intelligence which mainly focuses on systems which can learn from data, identify patterns and make decisions by its own thinking



Artificial intelligence is a technology which impact on user interaction and affected by the Internet. In the upcoming days, impact of AI is likely to grow enormously. AI has the potential to change the way of human interaction, not only with the digital world, but also with each other, through their work and through other socioeconomic institutions – for better or for worse.

How Do Machine Learning Algorithms Work?

As the machine learning algorithm is associated with data, the working of Machine learning algorithms is like a loop. The machine learning technique which we are using, final result depends on the type of ML algorithms used. The data can be continuously improving by going through the cycles, such as these:

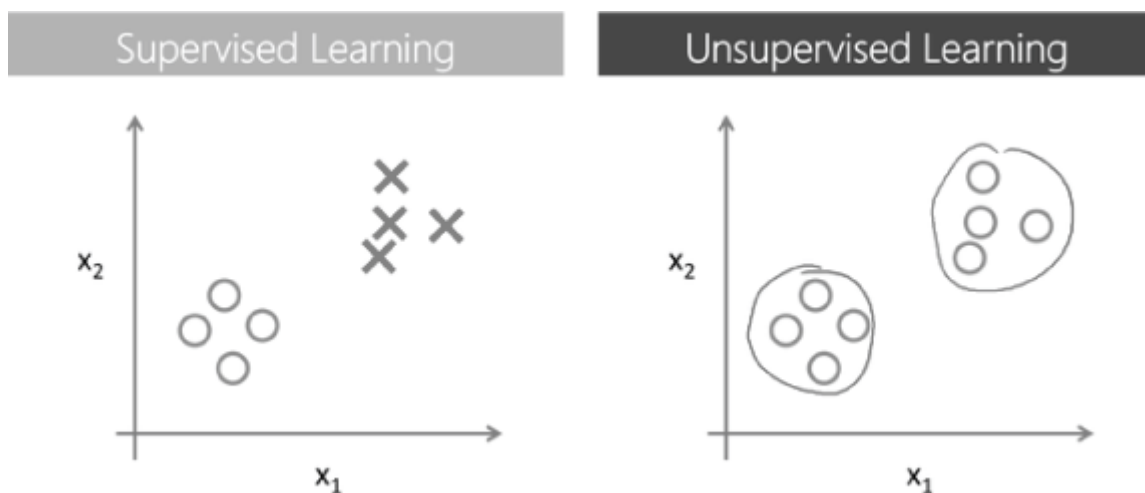
- Unlabeled data comes from various sources into one storage.
- Label the data as per requirement.
- The main task of ML algorithms is to sort that data.
- Decide the patterns for the data.

- Focus is to enhance and optimize business processes
- Automate the learning processes.

Requirement of good machine learning algorithm:

- Data preparation: Collect hardware and software information necessary to collect, process, analyze, integrate, and store this data.
- ML algorithms: We have to choose ML algorithm compatible with given data set.
- Automation and iterative processes: These are required for automation and repetitive process.
- Scalability: Prepare your system for future growth.
- Ensemble modeling: The idea here is when you use two or more algorithms at the same time to sync their results and get more precise results.

There is a lot of specific terms when it comes to machine learning. Let's stop for a moment to take a look at the terms and understand what they mean[6].



Machine getting to know (ML) is a place of laptop technological know-how and a subset of synthetic intelligence which presents the computer systems with the functionality to suppose and examine with out being programmed explicitly. Machine getting to know is utilized in diverse computational duties and its root goal is to teach the gadget with the assist of records provided, the records may be categorised in case of supervised getting to know and unlabeled in case of unsupervised getting to know as a way to produce higher consequences for the desired problem. The fundamental cognizance is to make computer systems examine from beyond experience [8].

Machine Learning and Big Data Challenges

Data Collection & Usage

If you truly accumulate facts, not anything will happen (besides for the reality that your garage could be full.) To remodel incoming facts into value-introduced enterprise insights, you want to apprehend what type of facts you want and the way you intend to apply it.

Therefore, earlier than you begin to teach your device mastering set of rules on a dataset, write down the desires for the set of rules and what type of facts may be beneficial to attain that goal. For example, if you're looking to determine which eating place to pick, you don't want climate forecast for the day (except you need to take a seat down at the terrace) or the call of the sous-chef (except you're monitoring down a Michelin one)[9].

If you turn out to be getting that facts as well, it won't modify the results, however the system of wading thru the numbers could take extra attempt and time. When we're speaking approximately massive facts projects, this difficulty turns into

even extra important.

Security

Cyber protection is one of the most up to date subjects those days. With such a lot of human beings sharing their private statistics and hundreds of thousands of bots producing even greater on-line facts, it's distinctly smooth to sway public opinion closer to one or some other decision. Think of Brexit and Cambridge Analytical scandal (the Brexit film indicates it pretty well). Or Equifax leaks - it's all approximately illicit use of amassed facts.

Therefore, whilst you are gathering statistics, you want to make certain you're now no longer violating the regulation. This is mainly touchy because of the General Data Protection Regulation that got here out in May of 2018. So far, GDPR works most effective in the bounds of the European Union. Nevertheless, you want to keep in mind that if, for example, you're placed with inside the United States but you've got got human beings from the EU come on your internet site or use your application, you want to make certain you get consent from them for facts collection[10].

When cyber protection meets device learning, you want to make certain the facts you accumulate is regulated (clean permissions for exclusive degrees of users), informed (legitimate consent agreement), monitored, and transparent. The facts amassed on this manner may be used accurately for schooling your ML algorithms.

Data Validation

When you've got got incoming facts from numerous reassets, you want to be aware of the credibility of your facts. Is there any facts that may purpose harm? Semi-supervised device gaining

knowledge of algorithms assist with automating the labeling process. They can undergo unlabeled facts and examine it to the to be had examples of already-categorized facts.

Right Algorithms

Every device gaining knowledge of version has a motive and is designed to carry out particular tasks. Therefore, one of the demanding situations in ML is selecting the proper set of rules that might convey you the insights you're seeking out. For instance, an unmonitored set of rules won't be capable of label your facts successfully due to the fact its number one motive is to search for patterns, and vice versa. Before you enforce ML, as soon as again, you want to jot down down your expectancies from this generation and what's the give up goal.

Training Dataset

When you're education your device gaining knowledge of set of rules, you want an excellent and big education dataset, in order that the set of rules can become aware of the fundamental patterns, facts, and insights. If your dataset is small, the consequences is probably very biased. The reassets for facts can come from facts collecting services (for instance, your business' Google Analytics account), pattern datasets (fake facts), or it could be bought from third-party. Whatever you do, don't forget that your facts desires to be credible and validated, it have to be applicable for the set of rules's motive, and it have to be cleaned. It additionally have to be legally received facts (for apparent reasons.)

In a few cases, you may use device gaining knowledge of algorithms for small facts sets, however maximum of the time, it's now no longer cost-effective. Therefore, both makes certain

you've were given sufficient facts or don't put money into ML algorithms at this point[11].

Data Noise

According to IBM researchers, "Pareto's Principle applies: 80% of a facts scientist's precious time is spent surely finding, cleansing, and organizing facts, leaving simplest 20% to carry out analysis." Data noise is any facts that isn't applicable to the ML set of rules's motive. Like we've referred to above, in our instance with the eating place. If you're seeking out a eating place with particular cordon bleu or a dish, getting facts approximately sous-cooks at eating places or their menus might be vital on your motive. If, however, you're seeking out an excellent area to devour with inside the vicinity, the call of the cordon bleu might be facts noise, due to the fact it'd suggest without a doubt not anything to you. Data noise may be incomplete facts, inconsequential facts, anomalous bits, and facts that may not be identified.

Use of Machine Learning

Banking & Financial Services

Various monetary offerings and banks address a number of numerical data, and that is one of the pleasant makes use of of gadget mastering algorithms for anomaly detection and fraud prevention. The insights, except shielding from high-threat customers and symptoms and symptoms of manipulations, additionally assist to become aware of funding possibilities or use gadget mastering algorithms for trading.

Government

Public Safety and software companies can also enjoy the insights supplied via way of means of information mining and system getting to know.

One of the implementation regions is electricity efficiency, which allows to decrease the costs and the payload. Machine getting to know algorithms for face popularity assist with surveillance and safety from identification theft.

Healthcare

Machine getting to know for healthcare predictions is a totally fast-developing fashion because of wearable gadgets and sensors. Thanks to them, the patient's information may be supplied for the gadget getting to know algorithms in real-time, assisting to keep lives.

Big information analytics, in mixture with gadget getting to know algorithms, also can assist in studying the traits or figuring out crimson flags in phrases of prognosis and treatment.

Machine getting to know algorithms for photo processing and gadget getting to know algorithms for photo type are the technology in the back of the capacity to discover atypical formations in numerous human organs and assist early most cancers detection, amongst different causes.

HUSPI had a risk to offer IT consulting offerings to at least one such undertaking known as Homeopath. Using gadget getting to know algorithms for sample recognition, gadget getting to know algorithms for prediction, and gadget getting to know algorithms for regression, the system, as soon as launched, might constantly replace its information with more recent findings, making the destiny patients' remedies extra precise.

Retail & Ecommerce

With the assist of device mastering algorithms for advice systems, retail and Ecommerce companies

can experience better client acquisition. Retailers also can seize data, examine it, and use it to offer a customized purchasing experience, put in force advertising campaigns, optimize prices, control deliver planning, and get client insights.

Oil, Gas, And Energy Sector

With the assist of system getting to know algorithms for advice systems, retail and Ecommerce corporations can experience better consumer acquisition. Retailers also can seize data, examine it, and use it to offer a customized purchasing experience, put into effect advertising campaigns, optimize prices, control deliver planning, and get consumer insights.

Transportation & Automotive

What is the maximum green direction? How to boom visitors capacity? How to optimize the lighting fixtures device with inside the city? All of those questions may be responded with the assist of insights from gadget studying algorithms. Have you ever used Waze and loved its cap potential to evolve the direction at the fly to the high-quality viable one given the set of filters? Yep, that is gadget studying at work. Data evaluation and modeling also are useful gear for agencies along with transport providers, public transportation, and different agencies that address transport. Driverless vehicles also are powered via way of means of gadget studying algorithms, specifically their automated power reaction systems. Besides real-lifestyles examples, it's also used with inside the Gaming industry. For example, Grand Theft Auto makes use of a collision detection gadget studying set of rules for shifting humans and vehicles.

Advertising Technologies (Adtech)

AdTech companies closely lean on system

mastering algorithms. One of the motives for it's miles the reality that AdTech is one of the proper Big Data industries.

Conclusion:

This paper explains machine learning algorithms and lists all the recent applications in the research area. In today's world, every person is making use of machine learning techniques whether unknowingly or knowingly. The paper illustrates the supervised and unsupervised techniques in detail and explains where machine learning is applied. This paper also explains how machine algorithm actually works on different data set. As we are facing various challenges of big data now-a-days, this paper also explains the different challenges on big data with machine leaning. And finally this paper explains the uses of machine learning in different sectors.

References

- Singh, P.K., Kar, A.K., Singh, Y., Kolekar, M.H. and Tanwar, S. eds., 2019. *Proceedings of ICRIC 2019: Recent Innovations in Computing* (Vol. 597). Springer Nature.)
- Osisanwo, F.Y., Akinsola, J.E.T., Awodele, O., Hinmikaiye, J.O., Olakanmi, O. and Akinjobi, J., 2017. Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), pp.128-138.
- Tschandl, P., Codella, N., Akay, B.N., Argenziano, G., Braun, R.P., Cabo, H., Gutman, D., Halpern, A., Helba, B., Hofmann-Wellenhof, R. and Lallas, A., 2019. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *The Lancet Oncology*, 20(7), pp.938-947.
- Abdallah, I., Dertimanis, V., Mylonas, H., Tatsis, K., Chatzi, E., Dervilis, N., Worden, K. and Maguire, E., 2018, June. Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data. In *Proceedings of the European Safety and Reliability Conference* (pp. 3053-3061).
- Osisanwo, F.Y., Akinsola, J.E.T., Awodele, O., Hinmikaiye, J.O., Olakanmi, O. and Akinjobi, J., 2017. Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), pp.128-138.
- Liu, Y., Arunachalam, S. and Temme, K., 2020. A rigorous and robust quantum speed-up in supervised machine learning. *arXiv preprint arXiv:2010.02174*.
- Amruthnath, N. and Gupta, T., 2018, April. A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)* (pp. 355-361). IEEE.
- Cerioti, M., 2019. Unsupervised machine learning in atomistic simulations, between predictions and understanding. *The Journal of chemical physics*, 150(15), p.150901.
- Cerioti, M., 2019. Unsupervised machine learning in atomistic simulations, between predictions and understanding. *The Journal of chemical physics*, 150(15), p.150901.
- Lee, Y., Ragguett, R.M., Mansur, R.B., Boutilier, J.J., Rosenblat, J.D., Trevizol, A., Brietzke, E., Lin, K., Pan, Z., Subramaniapillai, M. and Chan, T.C., 2018. Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. *Journal of affective disorders*, 241, pp.519-532.
- Al-Garadi, M.A., Hussain, M.R., Khan, N., Murtaza, G., Nweke, H.F., Ali, I., Mujtaba, G., Chiroma, H., Khattak, H.A. and Gani, A., 2019. Predicting cyberbullying on social media in the big data era using machine learning algorithms: Review of literature and open challenges. *IEEE Access*, 7, pp.70701-70718.
- Sharma, D. and Kumar, N., 2017. A review on machine learning algorithms, tasks and applications. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 6(10), pp.2278-1323.
- Bannach-Brown, A., Przybyła, P., Thomas, J., Rice, A.S., Ananiadou, S., Liao, J. and Macleod, M.R., 2019. Machine learning algorithms for systematic review:

- reducing workload in a preclinical review of animal studies and reducing human screening error. *Systematic reviews*, 8(1), pp.1-12.
- Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A. and Aljaaf, A.J., 2020. A systematic review on supervised and unsupervised machine learning algorithms for data science. *Supervised and Unsupervised Learning for Data Science*, pp.3-21.
- Fedushko, S. and Ustyianovych, T., 2019, January. Predicting pupil's successfulness factors using machine learning algorithms and mathematical modelling methods. In *International Conference on Computer Science, Engineering and Education Applications* (pp. 625-636). Springer, Cham.
- Kumar, D.P., Amgoth, T. and Annavarapu, C.S.R., 2019. Machine learning algorithms for wireless sensor networks: A survey. *Information Fusion*, 49, pp.1-25.
- Manogaran, G. and Lopez, D., 2017. A survey of big data architectures and machine learning algorithms in healthcare. *International Journal of Biomedical Engineering and Technology*, 25(2-4), pp.182-211.
- Das, K. and Behera, R.N., 2017. A survey on machine learning: concept, algorithms and applications. *International Journal of Innovative Research in Computer and Communication Engineering*, 5(2), pp.1301-1309.
- Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M.P., Shyu, M.L., Chen, S.C. and Iyengar, S.S., 2018. A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*, 51(5), pp.1-36.
- Sen, P.C., Hajra, M. and Ghosh, M., 2020. Supervised classification algorithms in machine learning: A survey and review. In *Emerging technology in modelling and graphics* (pp. 99-111). Springer, Singapore.