**International Journal of Technology, Science and Engineering** **ISSN: 2457-1016**

Vol. 7, Issue II, Apr-Jun 2024 http://www.bharatpublication.com/journal-detail.php?jID=25/IJTSE

# An Introduction To Machine Learning Technologies And How They Are Used In Online Education

**Rishik Gannavarapu**

*Arizona State University*

[1]*Received: 29 January 2024; Accepted: 23 March 2024; Published: 17 April 2024*

## ABSTRACT

We generate an incredible amount of data because of new technology, the internet, and connected objects. It is crucial to organise and contextualise these data so that they can be seen, comprehended, and reflected. Data analysis has historically been done by humans. But as data volumes increase, people are increasingly turning to automated systems that can mimic them. Machine learning refers to those systems that can solve issues by learning from data and data changes. Research on e-learning is significantly impacted by artificial intelligence, and technology enhanced learning environments (TELE) can be improved by implementing machine learning-based techniques. An overview of current discoveries in this field of study is provided in this paper. We begin by outlining the fundamental ideas of machine learning. Next, we showcase a few current projects that use machine learning in an e-learning setting.

## INTRODUCTION

Nowadays, practically everything we do creates a digital trail that details our whereabouts, explains our activity, and offers a wealth of additional information about our words, purchases, and other actions. The majority of computers, gadgets, and everything we use generate data because of both the capacity for data storage and the digitisation of society. For instance, we may retrieve data from parking lots, pay stations, smartphones, social media, images, videos, and more. Finding value and significance in all of the gathered data is essential.

Understanding phenomena, modelling behaviours, and making predictions are all made feasible by data analysis. In the past, machines used algorithms that people had created and analysed data to solve issues. These days, people provide data, which enables the machine to learn from it without explicit programming. We discuss the value of data. This is the machine learning tenet.

In actuality, people are conscious of the value of data and its potential richness. In fact, a number of scientific study fields, including medical [1] [2], e-commerce [3], industry [4][5], education [6][7], social networks [8][9], economics and finance [10], and others, have entered a significant age with the use of machine learning techniques to analyse complicated data.

Relationships between machine learning and other data science and AI ideas are depicted in Figure 1. Actually, data mining uses statistics to find patterns in raw data that may contain hidden information [11]. However, machine learning, a branch of artificial intelligence and computer science, makes predictions by learning from patterns. One of the key components of artificial intelligence and machine learning is deep learning. It is possible to characterise this new generation of machine learning as learning by layer, where the system must learn a little bit more on each layer.

[1] *How to cite the article:* Gannavarapu R. (April, 2024); An Introduction To Machine Learning Technologies And How They Are Used In Online Education; *International Journal of Technology, Science and Engineering*; Vol 7 Issue 2; 6-11

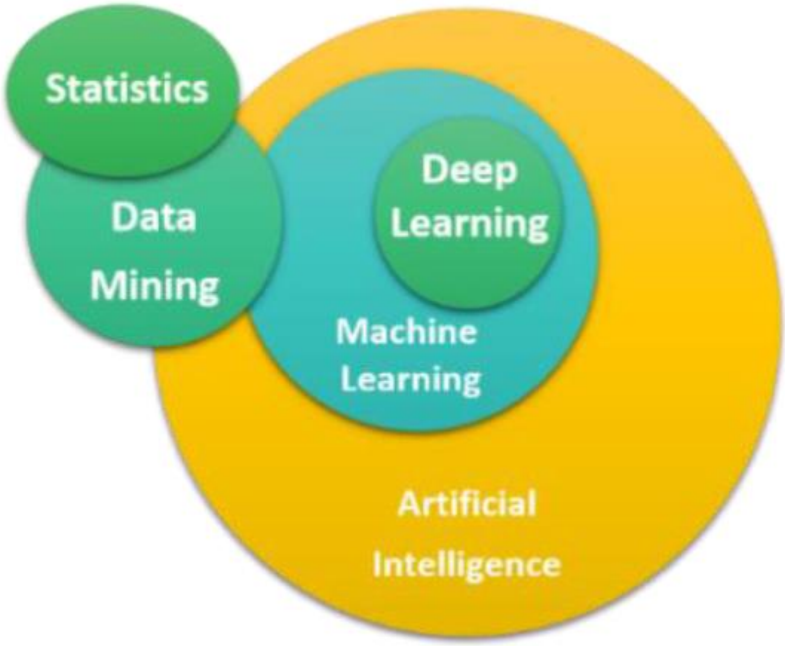


**Fig. 1 Machine Learning relationships to other related fields**

## MACHINE LEARNING

Machine learning is the process by which a computer learns how to do tasks given example data. We know that a machine's performance (P) increases when we offer it more encounters (E) with a specific task (T) [12]. Let's take the scenario where we want an email client to determine whether an email is spam or not. In this instance, experience E ought to consist of a collection of emails that have already been categorised as spam or not. Task T's job is to automatically categorise fresh emails. The machine's categorisation accuracy rate on a set of fresh emails is the performance P that ought to rise.

### A. The process of machine learning

The following describes the seven steps that make up the generic machine learning method [13]. Data collection is the initial step. Because it will define how good the prediction model can be, it is a very crucial task. However, the majority of the time, the data we collect is unstructured, noisy, or requires additional formats in order to be utilised for machine learning. Therefore, data must be pre-processed and cleansed.

We may then start developing our machine learning model. To do this, we begin with feature engineering, which involves selecting the most pertinent characteristics from the data. Next, we attempt to determine which machine learning algorithm is most suited for the given situation. It is essential to obtaining the best result.

Training is the next task. This process involves using a portion of our data to gradually enhance machine learning's predictive capabilities. After training is finished, the model should be tested to determine how it would fare in comparison to the other, unseen data. Several metrics, including accuracy, precision, and recall, are used to gauge the performance evaluation. It is occasionally feasible to go back and refine training before retesting. The outcome of the machine learning process is the final stage. It may be an inference or a forecast.

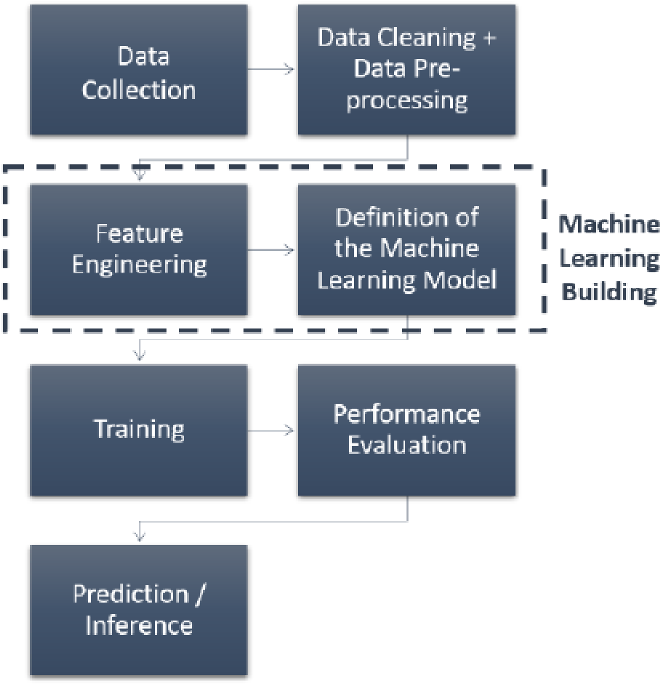


**Fig. 2 Components of a Generic machine learning model**

**Paradigms for machine learning**

Depending on the method employed for the learning process, machine learning can be categorised. Supervised, unsupervised, semi-supervised, and reinforcement learning were the four primary classifications [12].

A set of training data, also known as labelled data, in which the structure and result are known, is used in supervised learning. In order for the machine learning model to recognise patterns in the data, we need this data to train it. The model can be used to forecast data with unknown outcomes once it has been trained [14].

On the other hand, unsupervised learning techniques do not require prior labelling; instead, they learn structure from the data itself [15]. In other words, we can use unsupervised machine learning to identify patterns in labelled data.

Full label information isn't always accessible, though. When labels are few or costly to acquire, semi-supervised learning offers a strong foundation for utilising unlabelled data [16].

When we know what we want but don't know how to acquire it, the final machine learning method comes in handy. The idea is to try a number of ideas and determine which ones allow for the achievement of the intended outcome. An agent that must make decisions in a given environment can be used to formalise the reinforcement learning problem. The agent picks up a positive behaviour. This indicates that it gradually changes or picks up new behaviours and abilities. Therefore, the reinforcement agent just needs to be able to interact with the environment and gather information; it does not need to have total knowledge or control of the environment [17].

## E-LEARNING APPLICATIONS FOR MACHINE LEARNING

Everyone wants to study and expand their knowledge in a variety of industries these days, including employers, students, and others. Education systems are facing significant modernisation challenges as lifelong learning gains traction, and e-learning is gaining popularity. The number of Technology Enhanced Learning Environments (TELE) providing open or private online courses and other services has skyrocketed as a result of all of this. Machine learning techniques for analysing the vast volume of data generated by TELE have surfaced. Studying how to take advantage of this potent new technology to improve e-learning is beneficial.

### A. Analysis of sentiment

Success in Massive Open Online Courses (MOOCs) is now measured by how satisfied students are with the course [18]. Sentiment analysis can be used to predict learner satisfaction by identifying complicated emotions [19]. In [19], researchers aim to determine the polarity of learners' sentiments—both positive and negative—through forum messages in MOOCs. They contrast five supervised machine learning algorithms—Logistic Regression, Support

Vector Machine, Decision Tree, Random Forest, and Naïve Bayes—that have been utilised more often in MOOCs that deal with prediction. The most dependable method, according to the results, was Random Forest.

It's critical to comprehend how emotions play a part in MOOC participants' educational experiences. Controlling achievement emotions could, on the one hand, increase learner engagement, claims [20]. Create a supervised machine learning model based on SVM [20] to automatically classify achievement emotions. Because SVM outperforms Naïve Bayes, Logistic Regression, and Decision Trees, it was adopted. Conversely, [21] use large data from forums, comments, and completed homework to follow students' emotional inclinations in order to analyse their acceptance of the courses. [21] examine the connection between learning effects and emotional inclinations using machine learning and semantic analysis.

### B. Predicting student behaviour

The topic of using machine learning to predict student behaviour has been covered in an intriguing literature study [22]. Two research objectives were determined: dropout prediction and student classification.

Classification of students:

Undoubtedly, learning is greatly influenced by personalities, backgrounds, information, abilities, and preferences. The purpose of recommender systems is to provide each student with the best content. Classifying and profiling students is essential for a variety of reasons, including identifying desertion issues and personalising learning. Table 1 provides a summary of several recent studies that use machine learning to classify students.

**TABLE I. STUDENT CLASSIFICATION**

| Paper | Machine Learning Algorithm | Classification goal | Results |
|---|---|---|---|
| [23] | k-means Support Vector Machine (SVM) Naïve Bayes | Classification of engaged and disengaged faces of students with dyslexia | accuracy with 97–97.8% |
| [24] | Backpropagation (BP), Support Vector Machine (SVM), Gradient Boosting Classifier (GBC) | classification of student performance | Accuracy: BP = 87.78%, 83.20%= 83.20%, GBC= 82.44% |
| [25] | Decision Tree, Logistic regression, k-nn, SVM, random forest algorithms | Classification of successful and unsuccessful students | K-nn gives the higher accuracy = 85% |
| [26] | K-modes clustering algorithm Naive Bayes classifier | Classification of learner's learning style | Accuracy = 89% |

**Dropout prediction:** Interactive behaviour traces left throughout TELE have been analysed using a variety of machine learning techniques. The most popular method for predicting student dropout in MOOC environments is logistic regression (LR), which has an accuracy of 89%, according to [27], who focusses on learners' clickstream data. Second prize goes to SVM and Decision Trees, while third place goes to Natural Language Processing Techniques.

### C. Self-Regulated Learning

In the majority of TELE, students must make decisions about their own activities with minimal external teacher supervision [28]. In such scenario, people who possess high self-regulated learning (SRL) skills—which are defined as the capacity to organise, direct, and control their learning process—are able to learn more quickly and more effectively than people who lack these abilities [29]. Since MOOC is one of the e-learning platforms that supports SRL techniques [30], it encourages students to assess the quality or progress of their own work, create objectives, make plans, and have the opportunity to review notes, logs, tests, or learning materials in order to get ready for exams, among other things. Despite all of those aspects, many researchers still believe that improving student SRL using a machine learning technique is crucial. [31] help to improve knowledge of how students learn and how instruction should be planned to facilitate SRL in an asynchronous online course at a women's institution in South Korea based on learners' log traces and survey responses. Researchers in this study go on to identify student profiles and analyse the student SRL process over time. Initially, they proposed three essential SRL characteristics;

The selection of log variables was informed by the amount of time spent on content learning, study regularity, and help-seeking, all of which are applicable to asynchronous online courses and form the foundation of SRL analytics. Second, they used the silhouette method and the K-medoids clustering algorithm to identify student subpopulations. [31] employ random forest classification, a decision tree-based machine learning technique, to predict cluster membership by referring to each week's log variable after identifying existing clusters and their learning patterns.

## CONCLUSION

To improve the learning experience, e-learning experts have worked hard to analyse student data using machine learning techniques. Given that the student is seen as the primary element in the e-learning domain, this appears to be a sensible move. To the best of our knowledge, no research has been done on using learning data to gauge the quality of material and make improvements.

Therefore, we will use machine learning to evaluate e-learning content in our future work. The primary goal is to assist course designers in the process of reengineering education based on machine learning findings and a variety of criteria, particularly the interactions of previous students.

## REFERENCES

[1] Rakhmetulayeva, S. B., Duisebekova, K. S., Mamyrbekov, A. M., Kozhamzharova, D. K., Astaubayeva, G. N., & Stamkulova, K. (2018). Application of classification algorithm based on SVM for determining the effectiveness of treatment of tuberculosis. Procedia computer science, 130, 231-238.

[2] Kabyshev, M. V., & Kovalchuk, S. V. (2019). Development of personalized mobile assistant for chronic disease patients: diabetes mellitus case study. Procedia Computer Science, 156, 123-133.

[3] Zhu, G., Wu, Z., Wang, Y., Cao, S., & Cao, J. (2019). Online Purchase Decisions for Tourism E-commerce. Electronic Commerce Research and Applications, 100887.

[4] Brik, B., Bettayeb, B., Sahnoun, M. H., & Duval, F. (2019). Towards Predicting System Disruption in Industry 4.0: Machine Learning-Based Approach. Procedia Computer Science, 151, 667-674.

[5] Han, Y., Zeng, Q., Geng, Z., & Zhu, Q. (2018). Energy management and optimization modeling based on a novel fuzzy extreme learning machine: Case study of complex petrochemical industries. Energy conversion and management, 165, 163-171.

[6] Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2019). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. Computers & Education, 103724.

[7] Hmedna, B., El Mezouary, A., & Baz, O. (2019). How Does Learners' Prefer to Process Information in MOOCs? A Data-driven Study. Procedia computer science, 148, 371-379.

[8] Birjali, M., Beni-Hssane, A., & Erritali, M. (2017). Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. Procedia Computer Science, 113, 65-72.

[9] Kumari, K. V., & Kavitha, C. R. (2019). Spam Detection Using Machine Learning in R. In International Conference on Computer Networks and Communication Technologies (pp. 55-64). Springer, Singapore.

[10] Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. Energy Economics, 81, 709-727.

[11] Liu, J., Kong, X., Zhou, X., Wang, L., Zhang, D., Lee, I., ... & Xia, F. (2019). Data Mining and Information Retrieval in the 21st century: A bibliographic review. Computer Science Review, 34, 100193.

[12] Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. Expert Systems with Applications, 97, 205-227.

[13] Alzubi, J., Nayyar, A., & Kumar, A. (2018, November). Machine learning from theory to algorithms: an overview. In Journal of Physics: Conference Series (Vol. 1142, No. 1, p. 012012). IOP Publishing.

[14] Schrider, D. R., & Kern, A. D. (2018). Supervised machine learning for population genetics: a new paradigm. Trends in Genetics, 34(4), 301-312.

[15] Rodriguez-Nieva, J. F., & Scheurer, M. S. (2019). Identifying topological order through unsupervised machine learning. Nature Physics, 1.

[16] Oliver, A., Odena, A., Raffel, C. A., Cubuk, E. D., & Goodfellow, I. (2018). Realistic evaluation of deep semi-supervised learning algorithms. In Advances in Neural Information Processing Systems (pp. 3235-3246).

[17] François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An introduction to deep reinforcement learning. Foundations and Trends® in Machine Learning, 11(3-4), 219-354.

[18] Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2019). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. Computers & Education, 103724.

[19] Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., Estévez-Ayres, I., & Kloos, C. D. (2018, April). Sentiment Analysis in MOOCs: A case study. In 2018 IEEE Global Engineering Education Conference (EDUCON) (pp. 1489-1496). IEEE.

[20] Xing, W., Tang, H., & Pei, B. (2019). Beyond positive and negative emotions: Looking into the role of achievement emotions in discussion forums of MOOCs. The Internet and Higher Education, 100690.

[21]. https://scholar.google.com/citations?user=GZZEOHkAAAAJ&hl=en

[22] https://scholar.google.com/citations?user=R94525UAAAAJ&hl=en

[23] https://scholar.google.com/citations?user=WYURT_IAAAAJ&hl=en

[24] https://scholar.google.com/citations?user=xh9HeqgAAAAJ&hl=en