An Introduction To Machine Learning Technologies And How E-Learning Uses Them

Likhita Akkina

Arizona State University, Arizona

DOI:10.37648/ijrst.v14i01.005

[1]*Received: 09 December 2023; Accepted: 31 January 2024; Published: 19 February 2024*

ABSTRACT

We generate a staggering amount of data because of modern technologies, the internet, and connected objects. It is crucial to arrange and contextualize this data so that they can be seen, comprehended, and reflected. Humans have traditionally analyzed data. But as data volumes rise, people are turning more and more to automated systems that can mimic them. Machine learning refers to those systems that can learn from data as well as changes in data to solve problems. Technology Enhanced Learning Environments (TELE) can be improved by implementing machine learning-based techniques, and artificial intelligence has a significant influence on e-learning research. An overview of current discoveries in this field of study is presented in this paper. Firstly, we outline the main ideas behind machine learning. Next, we showcase a few new projects that use machine learning in an online learning environment.

INTRODUCTION

These days, nearly everything we do leaves a digital trail that reveals several more details about our words, purchases, and other actions, in addition to describing our activity and location. The majority of the computers, gadgets, and things we use now generate data because of advances in data storage and the digitization of society. For instance, we are able to retrieve data from pay stations, parking lots, smartphones, social media, films, images, and more. Gaining insight and significance from all of this gathered data is essential.

Making predictions, modeling behaviors, and comprehending phenomena are all made feasible by data analysis. In the past, people-built algorithms, analyzed data, and then machines used those algorithms to solve issues. In the modern world, humans provide data and let the machine learn from it without explicit programming. We discuss the power of information. This is how machine learning works.

In actuality, people are aware of the value that data has and the richness it can contain. In fact, a number of scientific study fields, including medicine [1] [2], e-commerce [3], industry [4][5], education [6][7], social networks [8][9], economics and finance [10], etc., have identified the analysis of complicated data using machine learning techniques as a significant era.

Figure 1 illustrates how several data science and artificial intelligence topics relate to machine learning. Actually, data mining uses statistics to identify patterns in raw data that 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 technologies in artificial intelligence and machine learning is deep learning. It is possible to refer to this new wave of machine learning as layer-by-layer learning, where the machine must learn a little bit more at each level.

[1] How to cite the article: Akkina L.; February 2024; An Introduction To Machine Learning Technologies And How E-Learning Uses Them; *International Journal of Research in Science and Technology*, Vol 14, Issue 1, 33-40, DOI: http://doi.org/10.37648/ijrst.v14i01.005

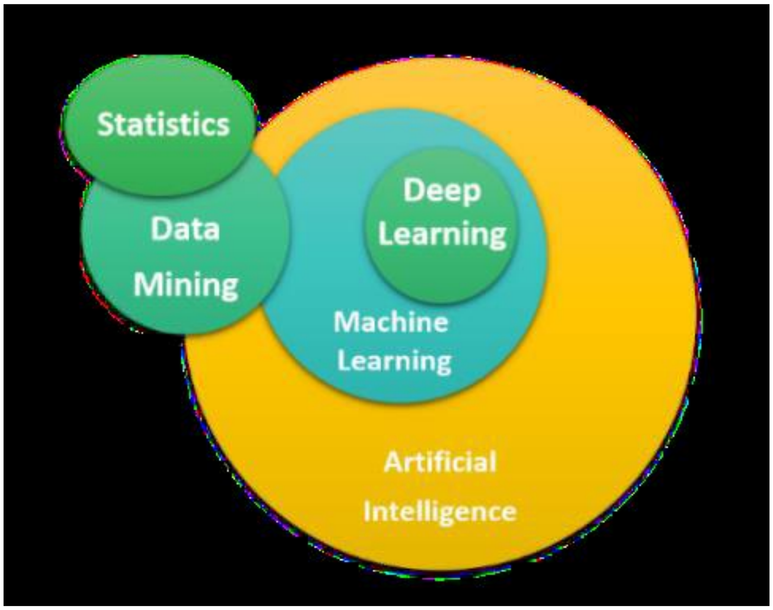

Fig. 1: Relationships between Machine Learning and Related Fields

In machine learning, a computer picks up task-specific knowledge from sample data. It is well known that a machine's performance (P) increases with increasing experience (E) with a given job (T) [12]. Let's say, for illustration, that we want an email client to identify whether or not an email is spam. Here, experience E should be a collection of emails that have already been flagged as spam or not. The task that T completes is automatically classifying fresh emails. The accuracy rate of the machine's classification on a set of fresh emails is the performance P that ought to rise.

A. The process of machine learning

The next section outlines the seven steps that make up the general machine learning method [13]. Gathering data is the first step. This work is crucial since it will establish how excellent predictive model is conceivable. However, the majority of the data we collected were either unstructured, noisy, or required additional formats in order to be meaningful for our machine learning. Thus, pre-processing and data cleaning are required.

We may then start developing our machine learning model. To do this, we first perform feature engineering, which involves extracting the most pertinent features from the data. Next, we attempt to determine which machine learning algorithm is most appropriate for the given scenario. Getting the finest outcomes possible requires it.

Training is the next assignment. In this step, we employ a portion of our data to gradually enhance the predictive capacity of machine learning. forecast. After training is finished, the model should be tested to determine how it would fare against the remaining, unseen data. The appraisal of performance is gauged by a number of factors, including recall, accuracy, and precision. It is occasionally possible to go back and refine instruction before retaking the test. The outcome of the machine learning process is the final stage. It might be an inference or forecast.

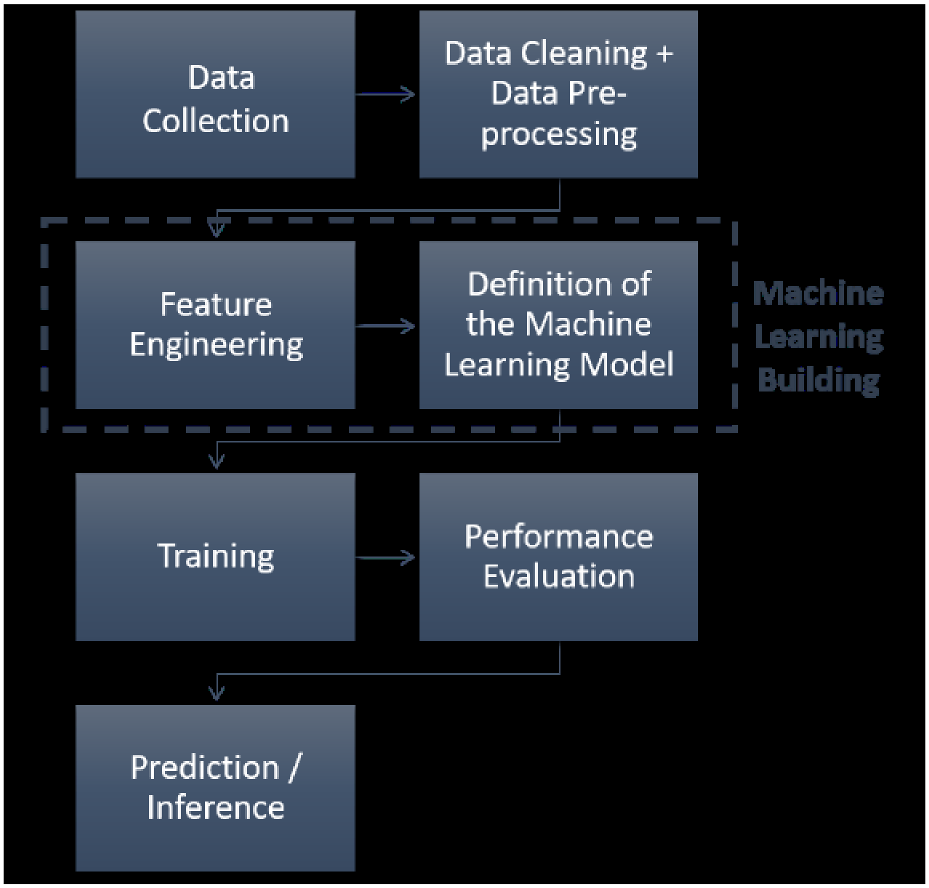

Fig. 2: Generic machine learning model components

B. Paradigms of machine learning

Based on the methodology employed for the learning process, machine learning can be categorized. supervised, unsupervised, semi-supervised, and reinforcement learning were found to be the four primary groups [12].

In supervised learning, the structure and result of a collection of training data, also known as labelled data, are known to us. With the help of this data, we train a machine learning model to recognize patterns in the dataset. After the model is trained, it can be used to forecast data sets with uncertain outcomes [14].

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

Complete label information isn't always accessible, though. In situations when obtaining labels is difficult or costly, semi-supervised learning offers a potent framework for utilizing unlabeled data [16].

The final machine learning strategy comes in handy when we know what we want but are unsure how to acquire it. The idea is to try out multiple ideas and determine which ones allow us to get the intended outcome. An agent that must make decisions in a given environment can be used to define the reinforcement learning problem. The agent picks up a positive habit. This indicates that it gradually changes or picks up new behaviors and abilities. Therefore, the reinforcement agent just needs to be able to interact with the environment and gather data—it does not need to have total control or understanding of it [17].

E-LEARNING APPLICATIONS FOR MACHINE LEARNING

These days, learners and employers alike are eager to expand their expertise in a variety of sectors. Lifelong learning is causing education systems to face significant modernization challenges, and e-learning is growing in popularity. All of this causes a sharp increase in the quantity of Technology Enhanced Learning Environments (TELE) that provide various services including private or public online courses. It is now possible to analyze the vast amounts of data generated by TELE using machine learning techniques. Studying the best ways to utilize 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]. Researchers aim to determine the polarity of learners' sentiments—positive and negative sentiments—through forum messages in MOOCs [19]. Five supervised machine learning algorithms—Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, and Naïve Bayes—that have been utilized more frequently in MOOC contributions relevant to prediction are compared. The most dependable method, according to the results, was Random Forest.

It's critical to comprehend how emotions play a part in MOOC students' educational experiences. On the one hand, [20] suggests that managing accomplishment emotions could enhance student involvement in the classroom. Create a supervised machine learning model based on SVM [20] to automatically classify achievement emotions. Since SVM outperforms Naïve Bayes, Logistic Regression, and Decision Tree in terms of performance, it was accepted. On the other hand, [21] use large data from homework completion, comments, and forums to track learners' emotional inclinations and analyze how well the courses are received. [21] examine the connection between emotional inclinations and learning effects using machine learning and semantic analysis.

B. Predicting the behavior of students

A fascinating survey of the literature [22] has tackled the topic of using machine learning to forecast student behavior. Two research objectives were established: dropout prediction and student classification.

Categorization of students:

Undoubtedly, a person's personality, experience, education, preferences, and talents all have a significant role in their ability to learn. Recommender systems are designed to provide the most relevant content to every student. Creating learner profiles and classifications is essential for several reasons, including identifying characteristics that contribute to student abandonment and personalizing the learning experience. Table 1 provides an overview of some recent studies that use machine learning to classify students.

Table I: Classification of Students

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%

Predicting dropouts:

A range of machine learning methodologies have been employed to examine the traces of interactive behavior left by users on TELE. Logistic regression (LR) has been the most often used technique to predict student dropout in MOOC environments, with 89% accuracy, according to [27], who focuses on learners' clickstream data. Second rank goes to SVM and Decision Trees, while Natural Language Processing Technique comes in third.

C. Self-Controlled Education

In most TELE, students are expected to make decisions about their own activities because there is minimal external teacher supervision [28]. Then, people who possess high self-regulated learning (SRL) abilities—which are defined by the capacity to organize, direct, and oversee one's own learning process—are able to learn more quickly and effectively than people who lack these abilities [29]. One of the e-learning platforms that supports SRL techniques [30], MOOC encourages students to assess the caliber or progress of their own work, create objectives, make plans, and have the option to review notes, logs, exams, or other learning resources to get ready for an exam, among other things. Despite all of those advantages, many researchers still believe that improving student SRL through machine learning techniques is crucial.

[31] help to improving understanding of how students learn and how instruction should be planned to encourage 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 work move on to the identification of student profiles and the analysis of the student SRL process over time. Initially, they proposed three fundamental SRL characteristics—time invested in learning content, study frequency, and help-seeking—that apply to asynchronous online courses and led the choice of log variables. Second, they used the silhouette method and the K-medoids clustering algorithm to identify student subpopulations. [31] employ random forest classification as 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 very hard to analyze learner data using machine learning techniques. Given that the student is seen as the primary component in the e-learning space, this seems sensible. To the best of our knowledge, no research has been done to date on the use of learning data for content quality measurement and improvement.

Therefore, the evaluation of e-learning content by machine learning will be the main emphasis of our future study. The primary goal is to support course designers in the process of reengineering education using machine learning findings and a variety of other elements, particularly historical student interactions.

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] Wang, L., Hu, G., & Zhou, T. (2018). Semantic analysis of learners' emotional tendencies on online MOOC education. Sustainability, 10(6), 1921.

[22] de Souza, V. F., & Perry, G. (2019). Identifying student behavior in MOOCs using Machine Learning. International Journal of Innovation Education and Research, 7(3), 30-39.

[23] Hamid, S. S. A., Admodisastro, N., Manshor, N., Kamaruddin, A., & Ghani, A. A. A. (2018, February). Dyslexia adaptive learning model: student engagement prediction using machine learning approach. In International Conference on Soft Computing and Data Mining (pp. 372-384). Springer, Cham.

[24] Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019, March). Student performance prediction and classification using machine learning algorithms. In Proceedings of the 2019 8th International Conference on Educational and Information Technology (pp. 7-11). ACM.

[25] Fedushko, S., & 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.

[26] EL AISSAOUI, O., EL MADANI, Y. E. A., OUGHDIR, L., & EL ALLIOUI, Y. (2019). Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles. Procedia computer science, 148, 87-96.

[27] Dalipi, F., Imran, A. S., & Kastrati, Z. (2018, April). MOOC dropout prediction using machine learning techniques: Review and research challenges. In 2018 IEEE Global Engineering Education Conference (EDUCON) (pp. 1007-1014). IEEE.

[28] Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2019). Supporting self-regulated learning in online learning environments and MOOCs: A systematic review. International Journal of Human–Computer Interaction, 35(4-5), 356-373.

[29] Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. Computers & education, 104, 18-33.

[30] Garcia, R., Falkner, K., & Vivian, R. (2018). Systematic literature review: Self-Regulated Learning strategies using e-learning tools for Computer Science. Computers & Education, 123, 150-163.

[31] Kim, D., Yoon, M., Jo, I. H., & Branch, R. M. (2018). Learning analytics to support self-regulated learning in asynchronous online courses: A case study at a women's university in South Korea. Computers & Education, 127, 233-251.