A Study on Machine Learning Technologies and their use in E-learning

C.Yosepu[1] , Dr.Jawaharalel Nehru[2] , Dr.N.Satheesh[3] , Dr.B.Rajalingam[4]
[1]Assistant Professor, Department of Computer Science and Engineering,
[2,3,4]Associate Professor, Department of Computer Science and Engineering,
St.Martin's Engineering College,TS,India.

Abstract—We produce an enormous amount of data as a result of new technologies, the internet, and connected objects. It is critical to place these data in context and organise them so that they can be perceived, understood, and reflected. Humans have traditionally analysed data. As the amount of data grows larger, humans are increasingly turning to automated systems that can replicate them. Machine learning refers to systems that can solve issues by learning from both data and changes in data. Artificial intelligence (AI) has a significant impact on e-learning research, and machine learning-based methodologies can be used to improve Technology Enhanced Learning Environments (TELEs) (TELE). This paper provides an outline of recent research findings in this topic. First, we'll go over some basic machine learning ideas. Then, we'll go over some recent work in the field of e-learning that uses machine learning.

Keywords:E-learning, Technology Enhanced Learning Environments, Data, Learners' Traces, Machine Learning, Deep Learning are some of the terms used to describe e-learning.

I.INTRODUCTION

Almost everything we do these days leaves a digital trail that characterises our actions, pinpoints our location, and offers a wealth of other data about what we say, buy, and so on. Most devices, machines, and everything we use produce data, thanks to both data storage capacity and societal digitalization. We can harvest data from pay stations, parking lots, smart phones, social media sites, films, and images, for example. All of this gathered data must be utilised and given meaning.

Understanding phenomena, modelling behaviours, and making predictions are all feasible with data analysis. Humans used to evaluate data, write algorithms, and then have the computer apply those methods to solve issues. Humans now introduce data, which allows machines to learn on their own without being expressly programmed. We discuss the importance of data. Machine learning works on this concept.

In actuality, there is an understanding of the value of data and its potential for richness. In fact, machine learning approaches for analysing complicated data have emerged as a significant trend in various scientific study disciplines, including medicine [1] [2], e-commerce [3], and industry .

Figure 1 depicts the connections between machine learning and various data science and artificial intelligence ideas. Statistics are used in data mining to uncover hidden information (patterns) from raw data [11]. Machine learning, on the other hand, is a branch of computer science and artificial intelligence that learns from patterns to make predictions. Deep Learning is one of the most important machine learning and artificial intelligence technologies. We can call it a new generation of machine learning because it is characterised by learning by layer, with each layer requiring the machine to learn a little bit more.

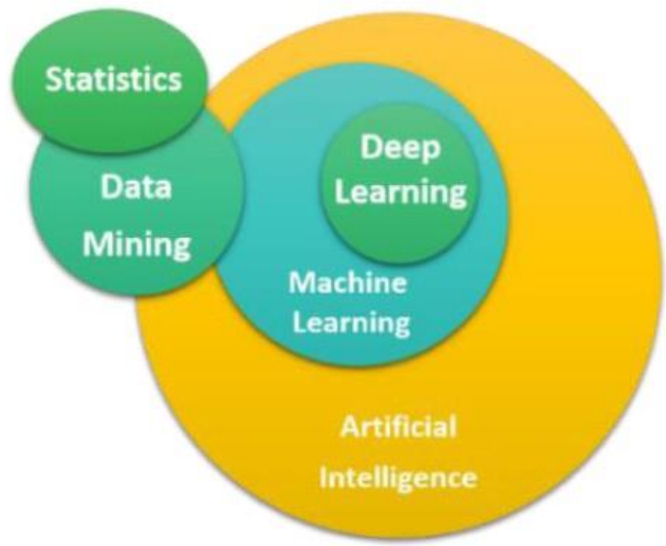

Figure 1: Relationships between Machine Learning and Other Related Fields

II. MACHINE LEARNING

A computer learns how to do tasks using example data in machine learning. We know that giving a machine additional encounters (E) with a specific task (T) improves its performance (P) [12]. Let's say we want an email client that can determine whether or not an email is spam. In this situation, the experience E should be a set of emails that have already been identified as spam or not. T's job entails automatically classifying fresh emails. The accuracy rate of the machine's classification on a set of new emails is a performance P that should improve.

A. Process of machine learning

The generic machine learning process is broken down into seven steps [13]. The first step is to gather information. It is a crucial assignment because it will define how good you are.It is possible to create a prediction model. However, the data we gathered is frequently unstructured, has a lot of noise, or must be transformed into different formats in order to be helpful for our machine learning. As a result, data must be cleansed and pre-processed.

We can then start working on our machine learning model. To do so, we first perform feature engineering, in which we identify the most relevant data features, and then we strive to find the optimal machine learning algorithm for the situation at hand. It is critical to achieving the greatest potential outcomes.

The next step is to begin training. In this step, we use a portion of our data to increase machine learning's prediction skill incrementally. After the model has been trained, it is time to put it to the test and evaluate how it performs against data that has yet to be seen. Various metrics such as accuracy, precision, and recall are used to evaluate performance. It is sometimes feasible to go back and improve training before retesting. The result of the machine learning is the final phase. It could be a guess or an inference.

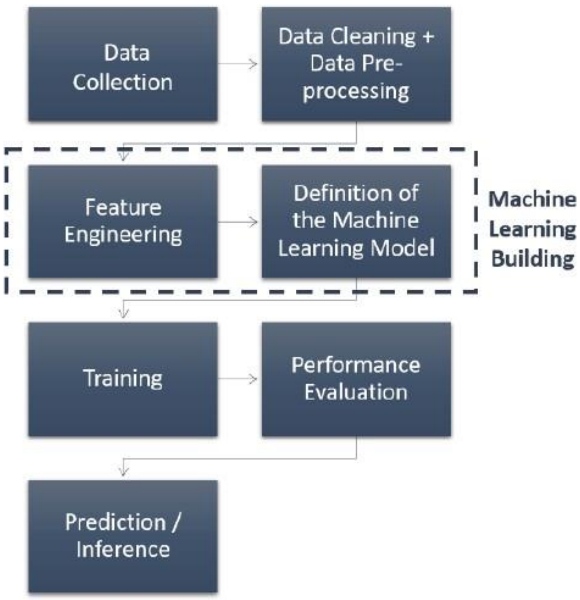

Figure 2: Generic Machine Learning Model Components

B. Models of machine learning

The approach employed for the learning process can be utilised to classify machine learning. supervised, unsupervised, semi-supervised, and reinforcement learning were defined as the four primary types [12].

We have a collection of training data or labelled data in supervised learning, and we know the structure and outcome. We use this information to create a machine learning model that can recognise patterns in the data. We may use the model to predict results of data with uncertain outcomes once it has been trained [14].

Unsupervised learning approaches, on the other hand, learn structure from the data without the necessity for prior labelling [15]. That is to say, we can use unsupervised machine learning to discover patterns in labelled data.

However, complete label information is not always available. When labels are few or expensive to collect, semi-supervised learning provides a powerful framework for utilising unlabeled data [16].The last machine learning method is useful when we know what we want but don't know how to acquire it. The idea is to try out a few different options and discover which ones allow us to get the desired result. The problem of reinforcement learning can be expressed as a decision-making agent in a given environment. The agent develops a positive attitude. This means it gradually changes or gains new habits and talents. As a result, the reinforcement agent does not need total knowledge or control of the environment, merely the ability to interact with it and gather data [17].

III.E-LEARNING APPLICATIONS FOR MACHINE LEARNING

Nowadays, everyone, including students, employers, and others, wants to study and expand their knowledge in a variety of sectors. Education systems are undergoing major transformation as a result of the widespread adoption of lifelong learning, and e-learning is becoming increasingly popular. As a result, the number of Technology Enhanced Learning Environments (TELE) offering open or private online courses and other sorts of services has skyrocketed. Machine learning approaches have been developed to analyse the vast amount of data supplied by TELE. It's a good idea to look at ways to leverage this strong new technology to improve e-learning.

A. Sentiment Analysis

The extent of student happiness with a Massive Open Online Course (MOOC) has recently been defined as success [18]. Sentiment analysis can be used to predict learner satisfaction by identifying complicated emotions [19]. Researchers intend to determine the polarity of learners' attitudes, positive and negative sentiments, using forum messages in MOOCs [19]. They compare five supervised machine learning algorithms: Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, and Nave Bayes, which have been employed more frequently in contributions linked to prediction in MOOCs. The Random Forest approach was shown to be the most reliable.

It's critical to comprehend the function of emotions in MOOC students' learning experiences. On the one hand, regulation of achievement emotions, according to [20], may help to promote learner engagement. [20] built a supervised machine learning model based on SVM to categorise achievement emotions automatically. SVM was chosen because it outperforms Nave Bayes, Logistic Regression, and Decision Tree in terms of performance. On the other side, [21] uses huge data from assignment completion, comments, and forums to examine learners' emotional inclinations in order to analyse course acceptability. [21] study the link between emotional dispositions and learning effects using semantic analysis and machine learning.

B.Predicting student behaviour

The application of machine learning in forecasting student conduct was the subject of an interesting literature study [22]. Student classification and dropout prediction were specified as two research objectives.

- Classification of students:

Personalities, histories, knowledge, abilities, and preferences undoubtedly play an important role in the learning process. Recommender systems help people find the best material for them.

each student Learner profiling and classification is essential not only for personalising learning but also for identifying abandonment causes and a variety of other purposes. In table 1, we outline some recent machine learning-based student classification studies.

STUDENT CLASSIFICATION TABLE I

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

- Prediction of dropouts:

To assess interactive behaviour traces left over TELE, a variety of machine learning approaches were used. Logistic regression (LR) has been the most commonly utilised technique to predict student dropout in the MOOC setting, according to [27], who focuses on learners' clickstream data, with an accuracy of 89 percent. SVM and Decision Tree are ranked second and third, respectively. Natural Language Processing Technique is ranked third.

C.Self-Controlled Learning

Because the majority of TELE has little external instructor monitoring, learners are expected to make decisions about their own activities [28]. Individuals with good self-regulated learning

(SRL) skills, which are defined as the ability to plan, manage, and control one's own learning process, can learn faster and better in this circumstance [29]. MOOC aims learners to self-evaluate the quality or progress of their work, make objectives and plans, and provide them the option to reread notes, logs, tests, or learning materials to prepare for testing, among other things, as it is one of the e-learning platforms enabling SRL techniques [30]. Despite all of these advantages, several researchers believe it is still vital to improve student SRL using a machine learning technique.

Learners' log traces and survey responses were used to create a model.

In an asynchronous online course at a women's institution in South Korea, [31] contribute to a better understanding of how students learn and how education should be organised to facilitate SRL. Researchers in this study look at student profiles and the student SRL process through time. Initially, they proposed three fundamental SRL characteristics:

SRL analytics were directed by time investment in content learning, study regularity, and help-seeking in asynchronous online courses, which served as the foundation for SRL analytics and guided the selection of log variables. Second, they used the silhouette method to identify student subpopulations using the K-medoids clustering algorithm. After you've discovered existing clusters and their learning patterns, you can move on to the next step.

By referring to each week's log data, [31] employ random forest classification as a decision tree-based machine learning technique to predict cluster membership.

IV. CONCLUSION

To improve the learning experience, e-learning experts have invested a lot of time evaluating learners' data using machine learning approaches. This appears to be prudent, given that the student is the most important component in the e-learning world. However, to the best of our knowledge, no research has been done on using learning data to measure content quality in order to enhance it.

As a result, we will concentrate our future efforts on evaluating e-learning content using machine learning. The major goal is to assist course creators in the educational reengineering process using machine learning and a variety of parameters, including previous learner interactions.

REFERENCES

1. S. B. Rakhmetulyeva, K. S. Duisebekova, A. M. Mamyrbekov, D. K. Kozhamzharova, G. N. Astaubayeva, and K. Stamkulova (2018). For determining the efficiency of tuberculosis treatment, a classification method based on SVM was used. 130, 231-238 in Procedia Computer Science.
2. M. V. Kabyshev and S. V. Kovalchuk (2019). Diabetes mellitus case study for the development of a customised mobile assistant for chronic disease patients. 123-133 in Procedia Computer Science, vol. 156.
3. G. Zhu, Z. Wu, Y. Wang, S. Cao, and J. Cao (2019). Tourism E-commerce Purchase Decisions on the Internet. 100887, Electronic Commerce Research and Applications
4. B. Brik, B. Bettayeb, M. H. Sahnoun, and F. Duval (2019). Machine Learning-Based Approach to Predicting System Disruption in Industry 4.0 667-674 in Procedia Computer Science.
5. Y. Han, Q. Zeng, Z. Geng, and Q. Zhu (2018). A case study of complex petrochemical industries using a unique fuzzy extreme learning machine for energy management and optimization. 165, 163-171 in Energy Conversion and Management.
6. K. F. Hew, X. Hu, C. Qiao, and Y. Tang (2019). What factors influence students' satisfaction with MOOCs? A supervised machine learning and sentiment analysis approach based on gradient boosting trees. 103724, Computers & Education.
7. B. Hmedna, A. El Mezouary, and O. Baz (2019). In MOOCs, how do learners choose to process information? An investigation based on data. 371-379 in Procedia Computer Science, 148.
8. M. Birjali, A. Beni-Hssane, M. Erritali, M. Beni-Hssane, M. Beni-Hssane, M. Beni-Hssane, M. Beni-H (2017). Suicide sentiment prediction in social networks using machine learning and semantic sentiment analysis techniques. 65-72 in Procedia Computer Science.
9. K. V. Kumari and C. R. Kavitha (2019). Machine Learning for Spam Detection in R Computer Networks and Communication Technologies International Conference (pp. 55-64). Singapore: Springer.
10. H. Ghoddusi, G. G. Creamer, and N. Rafizadeh (2019). A survey of machine learning in energy economics and finance. 709-727 in Energy Economics.
11. J. Liu, X. Kong, X. Zhou, X. Wang, L. Zhang, D. Lee, I. Lee, X. Xia, F. Xia, F. Xia, F. Xia, F. Xia, F. Xia, F. Xia, F. Xia, F. Xia (2019). A bibliographic review of data mining and information retrieval in the twenty-first century. Computer Science Review, vol. 34, no. 100193, p. 100193.
12. Portugal, P. Alencar, and D. Cowan (2018). A systematic review of the usage of machine learning techniques in recommender systems. 97, 205-227, Expert Systems with Applications
13. J. Alzubi, A. Nayyar, and A. Kumar (2018, November). An overview of machine learning, from theory to algorithms. Conference Series of the Journal of Physics (Vol. 1142, No. 1, p. 012012). IOP Publishing is a publishing house based in the United Kingdom.
14. D. R. Schrider and A. D. Kern (2018). A new paradigm in population genetics: supervised machine learning. Trends in Genetics, vol. 34, no. 4, pp. 301-312.
15. J. F. Rodriguez-Nieva and M. S. Scheurer (2019). Unsupervised machine learning is used to determine topological order. 1. Nature Physics
16. Oliver, A. Odena, C. A. Raffel, E. D. Cubuk, and I. Goodfellow (2018). Deep semi-supervised learning techniques are realistically evaluated. Advances in Neural Information Processing Systems (Advances in Neural Information Processing Systems) (pp. 3235-3246).
17. V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, V. François-Lavet, P. Henderson, P. Henderson (2018). This is a primer on deep reinforcement learning. 219-354 in Foundations and Trends® in Machine Learning, 11(3-4).
18. K. F. Hew, X. Hu, C. Qiao, and Y. Tang (2019). What factors influence students' satisfaction with MOOCs? A supervised machine learning and sentiment analysis approach based on gradient boosting trees. 103724, Computers & Education.

19. P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Muoz-Merino, I. Estévez-Ayres, and C. D. Kloos (2018, April). A case study of sentiment analysis in MOOCs. The IEEE Global Engineering Education Conference (EDUCON) was held in 2018. (pp. 1489-1496). IEEE.
20. W. Xing, H. Tang, and B. Pei (2019). Looking at the role of success emotions in MOOC discussion forums beyond positive and negative emotions. 100690, The Internet and Higher Education.
21. L. Wang, G. Hu, and T. Zhou (2018). On-line MOOC education: a semantic study of learners' emotional inclinations. Sustainability, vol. 10, no. 6, p. 1921.
22. V. F. de Souza and G. Perry (2019). Machine learning is being used to identify student behaviour in MOOCs. 30–39 in International Journal of Innovation Education and Research.
23. S. S. A. Hamid, N. Admodisastro, N. Manshor, A. Kamaruddin, and A. A. A. Ghani (2018, February). Student involvement prediction utilising a machine learning method in a dyslexia adaptive learning model. In Soft Computing and Data Mining: An International Conference (pp. 372-384). Cham. Springer.
24. B. Sekeroglu, K. Dimililer, and K. Tuncal (2019, March). Machine learning algorithms are used to predict and classify student performance. In the 8th International Conference on Educational and Information Technology, published in 2019. (pp. 7-11). ACM.
25. S. Fedushko and T. Ustyianovych (2019, January). Using machine learning techniques and quantitative modelling methodologies, predict student success factors. Computer Science, Engineering, and Education Applications International Conference (pp. 625- 636). Cham. Springer.
26. O. EL AISSAOUI, Y. E. A. EL MADANI, L. OUGHDIR, and Y. E. A. EL ALLIOUI (2019). To anticipate the learners' learning styles, a combination of supervised and unsupervised machine learning techniques is used. 87-96 in Procedia Computer Science, vol. 148.
27. F. Dalipi, A. S. Imran, and Z. Kastrati (2018, April). Review of research difficulties for MOOC dropout prediction using machine learning approaches. The IEEE Global Engineering Education Conference (EDUCON) was held in 2018. (pp. 1007-1014). IEEE.
28. J. Wong, M. Baars, D. Davis, T. Van Der Zee, G. J. Houben, and J. Paas a. (2019). A thorough review of supporting self-regulated learning in online learning settings and MOOCs. 35(4-5), 356-373, International Journal of Human–Computer Interaction.
29. R. F. Kizilcec, M. Pérez-Sanagustn, and J. J. Maldonado (2017). In Massive Open Online Courses, self-regulated learning strategies predict learner behaviour and goal attainment. Computers & Education, vol. 104, no. 1, pp. 18-33.
30. R. Garcia, K. Falkner, and R. Vivian (2018). Self-Regulated Learning Strategies for Computer Science Using E-Learning Tools: A Systematic Literature Review 123, 150-163 in Computers & Education.
31. D. Kim, M. Yoon, I. H. Jo, and R. M. Branch (2018). A case study at a women's institution in South Korea used learning analytics to enhance self-regulated learning in asynchronous online courses. Computers & Education, vol. 127, no. 2, pp. 233-251.