Education and Information Technologies
https://doi.org/10.1007/s10639-022-11068-7

Dropout prediction in Moocs using deep learning and machine learning

Ram B. Basnet[1] · Clayton Johnson[1] · Tenzin Doleck[2]

Received: 19 December 2021 / Accepted: 21 April 2022

Abstract
The nature of teaching and learning has evolved over the years, especially as technology has evolved. Innovative application of educational analytics has gained momentum. Indeed, predictive analytics have become increasingly salient in education. Considering the prevalence of learner-system interaction data and the potential value of such data, it is not surprising that significant scholarly attention has been directed at understanding ways of drawing insights from educational data. Although prior literature on educational big data recognizes the utility of deep learning and machine learning methods, little research examines both deep learning and machine learning together, and the differences in predictive performance have been relatively understudied. This paper aims to present a comprehensive comparison of predictive performance using deep learning and machine learning. Specifically, we use educational big data in the context of predicting dropout in MOOCs. We find that machine learning classifiers can predict equally well as deep learning classifiers. This research advances our understanding of the use of deep learning and machine learning in optimizing dropout prediction performance models.

Keywords Educational big data · Machine learning · Deep learning · Predictive analytics · Learning analytics

✉ Tenzin Doleck
tdoleck@sfu.ca

Ram B. Basnet
rbasnet@coloradomesa.edu

Clayton Johnson
claypjay@gmail.com

[1] Colorado Mesa University, Grand Junction, CO, USA

[2] Simon Fraser University, Burnaby, BC, Canada

Published online: 13 May 2022

1 Introduction

Improvements in computational power and the expansion of data analysis opportunities have spawned the use of knowledge-based discovery processes in varied fields. The field of education is no exception to this trend (Doleck et al. 2020). Recent literature (Ang et al., 2020) stresses the need for educational researchers to take advantage of advanced analytical tools and techniques such as *deep learning* (LeCun et al., 2015) and the opportunities afforded by big data to derive enhanced insights (Baig et al., 2020; Sorensen, 2018). In fact, it should come as no surprise that researchers in education are finding ways to use diverse data as significant and relevant sources of educational intelligence in the advancement of educational decision making (Aldowah et al., 2019; Ang et al., 2020; Doleck et al. 2021).

The contemporary educational landscape is characterized by an expanding use of technology. Technology-based learning environments have expanded the contexts—either in an enabling or supporting role—within which people teach and learn (Doleck et al. 2017). A key affordance of technology-based learning environments is the facility to acquire, collate, and provide fine-grained data (Romero & Ventura, 2020). This is important because learners generate a sea of data through their interactions with such environments. Rapid advances in analytical tools and techniques have paved the way to provide enhanced data analytics pertinent to educational researchers to make learning decisions. Indeed, mining education data is of emerging interest to educational research, reflecting the growing influence of analytics in the field (Baek and Doleck 2020; Charitopoulos et al., 2020). In fact, emerging fields—educational data mining and learning analytics—of inquiry have developed to tap into this interest (Aldowah et al., 2019; Baek and Doleck 2021, 2022; Romero & Ventura, 2020).

2 Background: Deep learning and machine learning

Deep learning and machine learning algorithms are increasingly utilized as they enable the creation of new knowledge and predictions from data (Doleck et al. 2015, 2019; Sorensen, 2018). Advances in deep learning and machine learning have been especially useful in generating insights from *big* data in various fields and contexts (Basnet et al. 2019; Mohammadi et al., 2018; Obermeyer & Emanuel, 2016; Zhang et al., 2018; Zhou et al., 2017); and education is no exception (Charitopoulos et al., 2020). Although machine learning and deep learning have been around for some time now, there has been a surge of interest in recent years from education researchers. As should be the case, researchers have examined the use of deep learning and machine learning. Concomitantly, questions arise regarding their suitability for use with educational big data. Yet there is a relatively little research examining the differences in use and usefulness of deep learning and machine learning. Several scholars have stressed the need for additional studies of machine learning and deep learning, as such comparative analysis with big education data may be indispensable for predictive analytics efforts (e.g., Doleck et al. 2020).

There is inconsistency and ambiguity in the use of the terms, deep learning and machine learning. Therefore, it is necessary to make clear the differences between machine learning and deep learning. The following subsections provide a primer on both machine learning and deep learning. Before doing so, we situate machine learning and deep learning in the context of artificial intelligence (AI). AI has been portrayed as transformational in both the academic literature and press, with the potential to be the "most important general-purpose technology of our era" (Brynjolfsson & McAfee, 2017, p. 2). Instigated by the popularity of AI, the adoption of AI and the suite of associated applications has become widespread in both practice and research. AI and associated algorithms are widely touted for their ability to render big noisy data meaningful.

It is important to note here that there is much confusion around the terms AI, machine learning, and deep learning (Chah, 2019); in fact, they are frequently used interchangeably (Jakhar & Kaur, 2019), raising the need to clarify the definitions of the three terms. A definition that may be useful in understanding AI was offered by Nguyen et al. (2019), who stated that AI is: "any technique that aims to enable computers to mimic human behaviour, including machine learning, natural language processing (NLP), language synthesis, computer vision, robotics, sensor analysis, optimization and simulation" (p. 78). Importantly, a key clarification provided in the literature frames AI as encompassing *machine learning* and *deep learning* (Jakhar & Kaur, 2019; Nguyen et al., 2019).

2.1 Machine learning

Machine learning represents "a subset of AI techniques that enables computer systems to learn from previous experience (i.e. data observations) and improve their behavior for a given task" (Nguyen et al., 2019, p. 78). Machine learning generally spans three subdomains: supervised learning, unsupervised learning, and reinforcement learning (Nguyen et al., 2019; Qiu et al., 2016). At the same time, Nguyen et al. (2019) note that machine learning algorithms have no strict categorization; that is, they can fit in more categories. A recent review (Cui et al., 2019) of predictive learning analytics documented the following machine learning techniques as the most frequently used: decision tree, logistic regression, naïve Bayes classifier, random forest, and support vector machine.

2.2 Deep learning

Deep learning is "a subset of NNs that makes the computational multi-layer NN feasible" (Nguyen et al., 2019, p. 79). Deep learning has grown in popularity as a solution for predictive analytics using large-scale datasets (Qiu et al., 2016). In contrast to machine learning, deep learning algorithms can extract their own features from the raw data (Bini, 2018). Deep learning has two distinct features: "(1) models consisting of multiple layers or stages of nonlinear information processing; and (2) methods for supervised or unsupervised learning of feature representation at successively higher, more abstract layers" (Deng & Yu, 2014, p. 201). Although deep learning has started to gain the attention of education researchers (Baker et al., 2017) some have noted that the results of deep learning applied to educational data are mixed, sometimes conflicting (Wilson et al., 2016).

With an understanding of machine learning and deep learning now in place, we turn our attention to the purpose of the study.

3 Purpose of the study

This paper is motivated by the need to more comprehensively understand the use of machine learning and deep learning with *big* educational data which encompasses digital traces of students' learning behaviors. Despite rapidly growing interest in deep learning and machine learning, the joint examination of deep learning and machine learning has not received commensurate research attention. Moreover, more research has been called for to better understand the realization of expected benefits with big data (Sorensen, 2018). As a step toward addressing this, we use *big* educational data, specifically, data from massive open online courses (MOOCs) to compare the predictive capacities of machine learning and deep learning.

Studies of machine learning and deep learning in education share an emphasis on prediction, which is therefore the focus in this paper. Specifically, we frame our study in the context of predicting dropout in MOOCs, that is, classification of students into dropouts and non-dropouts (see for a review, Dalipi et al., 2018). Dropout in MOOCs is considered a major concern in the MOOC literature (Aldowah et al., 2019). Predictive modeling investigations are particularly critical for optimizing dropout prediction model performance and permits better decision making with respect to providing feedback and supporting students. Jin (2020) notes that the research on predictive analytics with MOOCs generally falls into two categories: dropout prediction as a binary classification task and dropout prediction as a time series classification. The current study belongs to the former category.

This paper contributes to the predictive modeling research in MOOCs and seeks to answer the following research question: What are the differences in predictive power of deep learning and machine learning for dropout prediction in MOOCs?

4 Method

In response to a growing emphasis on predictive analytics in education, researchers seek to measure and evaluate performance accurately. We showcase our work with data from two MOOCs. We compare the performance of each model at predicting dropout. The analyses in this paper is based on open datasets obtained from the following sources: XuetangX (Feng et al., 2019) and KDD Cup Dataset (KDDCup15, 2015). The datasets and data processing procedures are described in greater detail in the following subsections.

4.1 Data sources

4.1.1 Data set 1

We downloaded the XuetangX dataset (Feng et al., 2019) in csv format. The csv files were preprocessed and the feature vectors were generated. The following preprocessing steps were conducted:

- Users' age was missing in some instances. We converted missing age to 0 and age values less than 10 or greater than 70 were also converted to 0.
- For gender coding: we used 1 for male, 2 for female and 0 for missing.
- We used integer encoding (1 to 7) for categorical values of the user's educational levels ("Bachelor's", "High", "Master's", "Primary", "Middle", "Associate", "Doctorate") and used 0 for missing value.
- We also used integer encoding to convert categorical feature for major degrees ('math', 'physics', 'electrical', 'computer', 'foreign language', 'business', 'economics', 'biology', 'medicine', 'literature', 'philosophy', 'history', 'social science', 'art', 'engineering', 'education', 'environment', 'chemistry').
- We performed Standard Scalar Transformation (StandardScaler — scikit-learn 0.24.1 documentation, 2021) on the numeric features.
- The features generated are listed below:

– action_count
– seek_video_count
– play_video_count
– pause_video_count
– stop_video_count
– load_video_count
– problem_get_count
– problem_check_count
– problem_save_count
– reset_problem_count
– problem_check_correct_count
– create_thread_count
– create_comment_count
– delete_thread_count
– delete_comment_count
– click_info_count
– click_courseware_count
– click_about_count
– click_forum_count
– click_progress_count
– close_courseware_count
– age
– gender

- education
- user_enroll_count
- course_enroll_count
- course_category.

- After preprocessing the data, we ended up with 225,642 samples.

4.1.2 Dataset 2

KDD Cup Dataset was downloaded (KDDCup15, 2015). The zip file contains train and test log files. We combined the train and test dataset resulting in 200,904 samples to run ten-fold cross-validation. Unlike the XuetangX dataset, this dataset does not provide certain demographic features such as age, sex, education, etc. Therefore, we conducted two experimental evaluations using different feature sets—smaller and larger feature set.

For the first experiment, we used a smaller feature set. We generated 15 numeric features and performed Standard Scalar Transformation—Standard Scalar Transformation (StandardScaler — scikit-learn 0.24.1 documentation, 2021) on those features.

The features included the following:

- action_count
- server_navigate_count
- server_access_count
- server_problem_count
- server_page_close_count
- server_video_count
- server_discussion_count
- server_wiki_count
- browser_navigate_count
- browser_access_count
- browser_problem_count
- browser_page_close_count
- browser_video_count
- browser_discussion_count
- browser_wiki_count

For the second experiment, we used a larger feature set (140 features) generated by Peng and Aggarwal (2015). This was done to compare with the smaller set of 15 features that we generated.

5 Libraries and algorithms

In the literature, various frameworks/libraries are used across the areas of interest. In this paper we engage with the various machine learning and deep learning frameworks/libraries.

5.1 Machine learning

To conduct our baseline experiments we used Scikit-learn (Pedregosa et al., 2011) library's supervised learning classifiers. Among the different families of classifiers available, we picked the following that are typically used in the literature.

- Random Forest (RF): A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting (RandomForestClassifier—scikit-learn 0.24.1 documentation, 2021). We used the default sub-sample size of 100 to build each tree. The sub-sample size can be controlled with the max_samples parameter when the boostrap parameter is True which is the default parameter in the random forest api in Scikit Learn framework.
- XGBoost (XGB): XGBoost is a scalable tree boosting system that is widely used by data scientists and provides state-of-the-art results on many problems. XGBoost algorithm handles sparse data and its cache access patterns, data compression, and sharding features enable the algorithm to solve large-scale problems using a minimal amount of resources (Chen & Guestrin 2016).
- AdaBoost Classifier (AB): AdaBoost classifier is a meta-estimator that begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases (AdaBoostClassifier—scikit-learn 0.24.1 documentation, 2021).
- Decision Trees Classifier (DT): Decision Trees classifier creates model that predicts the value of a target variable by learning simple decision rules inferred from the data features (Decision Trees—scikit-lean 0.24.1 documentation, 2021). Decisions trees white box models are easy to visualize, understand, and interpret.
- K-Neighbors Classifier (kNN): k-Neighbors classifier in Scikit learn implements k-nearest neighbors statistical algorithm where an sample object is classified by a plurality vote of its k nearest neighbors where k is a typically a small positive integer. If $k=1$, e.g., then the sample is simply assigned to the class of its single nearest neighbor.
- Support Vector Machines (SVM): A support vector machine constructs a hyper-plane or set of hyper-planes in a high or infinite dimensional space, which can be used for classification, regression, or other tasks. Intuitively,

a good separation is achieved by the hyper-plan that has the largest distance to the nearest training data points of any class (so-called functional margin) (Support Vector Machines—sckit-learn 0.24.1 documentation, 2021).

- Logistic Regression (LR): Logistic regression is a linear model for classification problems where the probabilities describing the possible outcomes of a single sample are modeled using a logistic function (Linear Models—scikit-learn 0.24.1 documentation, 2021).
- Linear Discriminant Analysis (LDA): LDA is a classifier with a linear decision boundary, generated by fitting class conditional densities to the data and using Baye's rule. The model fits a Gaussian density to each class, assuming that all classes share the same covariance matrix (Linear Discriminant Analysis— scikit-learn 0.24.1 documentation, 2021).

5.2 Deep learning

There are several deep learning frameworks that can be useful for learning with big data. The frameworks considered for deep learning include:

- Keras: The Keras interface is a portable, high-level neural networks API, written in Python, usable for TensorFlow, CNTK, and Theano as a back end, and was originally developed as part of the Open-ended Neuro-Electronic Intelligent Robot Operating System (ONEIROS). This interface has multiple advantages when it comes to research and development, primarily its portability. Since Keras is written to support three major deep learning frameworks and potentially more in the future, minimal alterations are required to switch the framework in use.
- TensorFlow: TensorFlow is a machine learning open-source platform developed by Google with extensive industry use. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in machine learning (ML) and developers easily build and deploy ML powered applications. This framework is available in Python, Java, JavaScript, and C + +, while also supporting the Internet of Things (IoT) devices.
- fast.ai: was designed with the purpose of making deep learning solutions more accessible to researchers and developers of diverse backgrounds. fast.ai is an open-source Python-based library that uses PyTorch—another popular deep learning framework—as backend. fast.ai optimizes PyTorch and simplifies the experimentation process for deep learning researchers.

6 Experiments and results

We ran our base experiments using Jupyter Notebooks [https://github.com/rambasnet/PredictingMOOCDropouts]. The accuracy of the various models (in percent) is employed to evaluate the performance of the models. We compare

Table 1 Machine learning results

Classifiers	Mean Accuracy	SD
RF	0.83894	0.00214
XGB	0.84582	0.00230
AB	0.83116	0.00218
DT	0.76300	0.00274
kNN	0.82757	0.00182
SVM	0.75853	9e-05
LR	0.81857	0.00214
LDA	0.81386	0.00184
NB	0.81077	0.00161

Table 2 Deep learning results

Model	Accuracy	Bal-acc	Rec	Prec	AUC	F1
Keras-Tensorflow	77.49 ± 6.87	63.43 ± 3.90	77.49 ± 6.87	78.24 ± 4.20	63.43 ± 3.90	85.55 ± 5.80
fast.ai	83.48 ± 0.16	70.07 ± 0.66	83.48 ± 0.16	82.82 ± 0.10	70.07 ± 0.66	89.82 ± 0.06

the prediction accuracy results of machine learning and deep learning models. To compare the machine learning and deep-learning classifiers, we separated each dataset 80/20 random split. 80% of the split was used for baseline comparison of the models using five-fold cross-validation. 20% was held out as the validation dataset.

6.1 XuetengX dataset

We first show the performance of machine learning methods. The results are presented in Table 1. The experimental results revealed that XGB yields the best accuracy (84.582%). Whereas, in examining the deep learning results (Table 2), we find that fast.ai performs the best with an accuracy of 83.48%. When comparing machine learning and deep learning results, we observe that the best performing machine learning classifier, XGB, outperforms the best performing deep learning classifier. Of note is that deep learning does not show higher accuracy in the predictive task.

We then used the model generated by XGB to classify the validation set. Out of 10,813 samples representing drop-out label, 5785 (53.50% true positive) were correctly predicted as dropout and the rest 5,028 were falsely predicted as no-dropout (46.50% false negative). Similarly, out of 34,316 no-dropout labelled samples, 32,448 (95.56% true negative) and 1,868 (5.44% false positive) were predicted as dropout. Since the database is unbalanced, we get the weighted average precision of 0.84, recall of 0.85 and f1-score of 0.84 with the support of 45,129 samples.

Table 3 Machine learning results

Classifiers	Mean	SD
RF	0.85549	0.00185
XGB	0.85958	0.00205
AB	0.85824	0.00192
DT	0.80715	0.00225
kNN	0.84539	0.00250
SVM	0.85965	0.00180
LR	0.85435	0.00166
LDA	0.84992	0.00162
NB	0.84812	0.00201

6.2 KDD dataset: Smaller feature set

Our experiments with the KDD dataset are divided into two parts. We first used a smaller feature set. As seen in Table 3, the highest accuracy was realized by SVM (85.965%). In comparison, Table 4 which provides the results for the deep learning methods, reveals that the highest accuracy was realized by Keras-Tensorflow (86.02%). In comparing the machine learning results with deep learning methods, we find that the best performing deep learning classifier, Keras-Tensorflow (86.02%), barely outperforms the best performing machine learning classifier, SVM (85.965%).

6.3 KDD dataset: Larger feature set

In this section, we show the performance of machine learning and deep learning methods with a larger feature set. According to the results, among the machine learning classifiers, the highest accuracy was realized by AB (87.636%) (Table 5). As illustrated in Table 6, among the deep learning classifiers, fast.ai (87.52%) achieved the highest accuracy. When comparing machine learning and deep learning results, we observe that the best performing machine learning classifier, AB (87.636%), barely outperforms fast.ai (87.52%), the best performing deep learning classifier. It is interesting to note the differences in performance with a smaller and larger feature set; thus, highlighting that the feature selection step can influence the classification performance.

Table 4 Deep learning results

Model	Accuracy	Bal-acc	Rec	Prec	AUC	F1
Keras-Tensorflow	86.02 ± 0.06	72.67 ± 0.38	86.02 ± 0.06	85.09 ± 0.07	72.67 ± 0.38	91.54 ± 0.04
fast.ai	85.98 ± 0.08	71.71 ± 0.38	85.98 ± 0.08	85.10 ± 0.10	71.71 ± 0.38	91.57 ± 0.05

Table 5 Machine learning results

Classifiers	Mean	SD
RF	0.87552	0.00307
XGB	0.87560	0.00299
AB	0.87636	0.00298
DT	0.80623	0.00341
kNN	0.85430	0.00265
SVM	0.87199	0.00265
LR	0.87282	0.00271
LDA	0.86984	0.00313
NB	0.85063	0.00209

Table 6 Deep learning results

Model	Accuracy	Bal-acc	Rec	Prec	AUC
Keras-Tensorflow	85.03 ± 0.23	85.03 ± 0.23	84.08 ± 0.20	72.99 ± 0.03	90.84 ± 0.16
fast.ai	87.52 ± 0.13	87.52 ± 0.13	86.86 ± 0.13	75.50 ± 0.59	92.43 ± 0.06

7 Discussion

Researchers can harness the digital traces of students' learning behaviors to investigate the ways of improving learning. One key exercise has been to predict various student learning outcomes. This has been achieved by leveraging computational techniques. However, despite the popularity of deep learning and machine learning, only a small body of research has jointly explored deep learning and machine learning for prediction with educational big data. To address the lack of relevant research, the present study provides a comprehensive comparison of predictive performance using deep learning and machine learning models. To this end, we use educational *big* data in the context of predicting dropout (dropout/no dropout) in MOOCs. Development of predictive models for such purposes are crucial for unearthing salient factors that potentially influence learner outcomes. More specifically, such exercises are useful for identifying at-risk students and for efforts geared toward understanding learner needs.

We posed the following research question: What are the differences in predictive power of deep learning and machine learning for dropout prediction in MOOCs? To answer the research question, we trained and tested several machine learning and deep learning models for the dropout/no dropout prediction exercise using two datasets: XuetangX dataset (Feng et al., 2019) and KDD Cup dataset (KDDCup15, 2015). Through a set of experiments comparing the accuracy of the models on the data used to create them, both machine learning and deep learning methods showed good predictive capacity for the purpose of predicting dropout in MOOCs. For the XuetangX dataset, accuracy reached as high as 84.582%. And, for the KDD Cup dataset, accuracy reached as high as 87.636%. For the XuetangX dataset, XGB resulted in the highest predictive performance. For the KDD dataset, Keras-Tensorflow resulted in the highest predictive performance when using a smaller feature set, while AB resulted in the highest predictive performance when using a larger feature set. The experimental estimations produced by machine learning and deep learning methods were compared, which leads us to the central finding of this paper—the machine learning classifiers performed as well and, in some cases, better than deep learning. This is in congruence with previous work documenting that deep learning algorithms do not necessarily outperform traditional machine learning classifiers on educational data (e.g., Doleck et al. 2019; Xiong et al., 2016). This finding is important in that it tempers optimism regarding deep learning. Moreover, this raises the need to place greater importance on the context to which deep learning is applied. In fact, Wilson et al. (2016) note that: "deep learning has a promising future in educational data mining, but that future depends on data sets that have a much richer encoding of the exercises and learning context" (p. 7).

While deep learning and machine learning have been used and evaluated as distinct approaches, the findings of the present study suggest that there may be utility in using deep learning particularly when used in complement with traditional machine learning techniques; and/or using either approach depending on the research context. In summary, our findings contribute to a growing literature on predictive analytics in educational research.

7.1 Limitation and future directions

The focus of this work was on comparing deep learning and machine learning on predicting dropout in *two* MOOCs. The sample we considered in our experiments may not be representative of other populations of MOOC learners. As such, generalizability of the models is a concern. Therefore, further work should examine additional MOOCs and learner samples. Data issues such as the unbalanced nature of the datasets ought to be considered; future work can experiment with weighted average precision. It is important to acknowledge that the difference in mean performance between any two algorithms may not necessarily be significant. Thus, future research should examine the significance of differences in performance between algorithms. In the experiments, we did not use deep learning to automatically extract the features. Future work should explore the influence of direct use of deep learning for feature extraction and model training. In the present paper, we did not demonstrate feature importance ranking; future work can provide this additional information for stakeholders interested in the relative importance of the different features contributing to dropout. And, finally, it is important to consider the impact of fake learners on MOOCs, as recent research (Alexandron et al., 2019) suggests that fake learners can bias the findings of analytical experiments using MOOC learner-system data.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

AdaBoostClassifier—scikit-learn 0.24.1 documentation (2021). Retrieved 24 January 2021, from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html

Aldowah, H., Al-Samarraie, H., Alzahrani, A., & Alalwan, N. (2019). Factors affecting student dropout in MOOCs: A cause and effect decision-making model. *Journal of Computing in Higher Education, 32*(2), 429–454. https://doi.org/10.1007/s12528-019-09241-y

Alexandron, G., Yoo, L., Ruipérez-Valiente, J., Lee, S., & Pritchard, D. (2019). Are MOOC learning analytics results trustworthy? With fake learners, they might not be! *International Journal of Artificial Intelligence in Education, 29*(4), 484–506. https://doi.org/10.1007/s40593-019-00183-1

Ang, K., Ge, F., & Seng, K. (2020). Big educational data & analytics: Survey, architecture and challenges. *IEEE Access, 8*, 116392–116414. https://doi.org/10.1109/access.2020.2994561

Baek, C., & Doleck, T. (2020). A bibliometric analysis of the papers published in the journal of artificial intelligence in education from 2015–2019. *International Journal of Learning Analytics and Artificial Intelligence for Education, 2*(1), 67–84. https://doi.org/10.3991/ijai.v2i1.14481

Baek, C., & Doleck, T. (2021). Educational data mining versus learning analytics: A review of publications from 2015–2019. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2021.1943689

Baek, C., & Doleck, T. (2022). Educational data mining: A bibliometric analysis of an emerging field. *IEEE Access, 10*, 31289–31296. https://doi.org/10.1109/access.2022.3160457

Baker, R., Beck, J. E., Chi, M., Heffernan, N. T., & Mozer, M. (2017). Workshop on deep learning with educational data. In *Proceedings of the 10th International Conference on Educational Data Mining Workshops and Tutorials*.

Baig, M., Shuib, L., & Yadegaridehkordi, E. (2020). Big data in education: a state of the art, limitations, and future research directions. *International Journal of Educational Technology In Higher Education*, *17*(1). https://doi.org/10.1186/s41239-020-00223-0

Basnet, R. B., Shash, R., Johnson, C., Walgren, L. & Doleck, T. (2019). Towards detecting and classifying network intrusion traffic using deep learning frameworks. *Journal of Internet Services and Information Security, 9*(4), 1–17. https://doi.org/10.22667/JISIS.2019.11.30.001

Bini, S. A. (2018). Artificial intelligence, machine learning, deep learning, and cognitive computing: What do these terms mean and how will they impact health care? *The Journal of Arthroplasty, 33*(8), 2358–2361. https://doi.org/10.1016/j.arth.2018.02.067

Brynjolfsson, E., & McAfee, A. (2017). *The business of artificial intelligence*. Harvard Business Review. Retrieved from https://starlab-alliance.com/wp-content/uploads/2017/09/The-Business-of-Artificial-Intelligence.pdf

Chah, N. (2019). Down the deep rabbit hole: Untangling deep learning from machine learning and artificial intelligence. *First Monday*. https://doi.org/10.5210/fm.v24i2.8237

Charitopoulos, A., Rangoussi, M., & Koulouriotis, D. (2020). On the use of soft computing methods in educational data mining and learning analytics research: A review of years 2010–2018. *International Journal of Artificial Intelligence in Education, 30*(3), 371–430. https://doi.org/10.1007/s40593-020-00200-8

Chen, T & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. http://arxiv.org/abs/1603.02754

Cui, Y., Chen, F., Shiri, A., & Fan, Y. (2019). Predictive analytic models of student success in higher education. *Information and Learning Sciences, 120*(3/4), 208–227. https://doi.org/10.1108/ils-10-2018-0104

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

Decision Trees—scikit-lean 0.24.1 documentation (2021). Retrieved 24 January 2021, from https://scikit-learn.org/stable/modules/tree.html

Deng, L., & Yu, D. (2014). Deep learning: Methods and applications. *Foundations and Trends in Signal Processing, 7*(3–4), 197–387. https://doi.org/10.1561/2000000039

Doleck, T., Basnet, R. B., Poitras, E. G., & Lajoie, S. P. (2015). Mining learner-system interaction data: Implications for modeling learner behaviors and improving overlay models. *Journal of Computers in Education, 2*(4), 421–447. https://doi.org/10.1007/s40692-015-0040-3

Doleck, T., Bazelais, P., & Lemay, D. J. (2017). Examining CEGEP students' acceptance of CBLEs: A test of acceptance models. *Education and Information Technologies, 22*(5), 2523–2543. https://doi.org/10.1007/s10639-016-9559-9

Doleck, T., Lemay, D. J., Basnet, R. B., & Bazelais, P. (2020). Predictive analytics in education: A comparison of deep learning frameworks. *Education and Information Technologies, 25*(3), 1951–1963. https://doi.org/10.1007/s10639-019-10068-4

Doleck, T., Lemay, D. J., & Brinton, C. (2021). Evaluating the efficiency of social learning networks: Perspectives for harnessing learning analytics to improve discussions. *Computers and Education*. https://doi.org/10.1016/j.compedu.2021.104124

Doleck, T., Poitras, E. & Lajoie, S. (2019). Assessing the utility of deep learning: Using learner-system interaction data from BioWorld. In J. Theo Bastiaens (Ed.), *Proceedings of EdMedia + Innovate Learning* (pp. 734–738). Amsterdam, Netherlands: AACE.

Feng, W., Tang, J., & Liu, T. X. (2019). Understanding dropouts in MOOCs. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 517–524).

Jakhar, D., & Kaur, I. (2019). Artificial intelligence, machine learning and deep learning: Definitions and differences. *Clinical and Experimental Dermatology, 45*(1), 131–132. https://doi.org/10.1111/ced.14029

Jin, C. (2020). MOOC student dropout prediction model based on learning behavior features and parameter optimization. *Interactive Learning Environments*, 1-19. https://doi.org/10.1080/10494820.2020.1802300

KDDCup15. (2015). KDDCUP Data. Retrieved from http://lfs.aminer.cn/misc/moocdata/data/kddcup15.zip

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Linear Discriminant Analysis—scikit-learn 0.24.1 documentation (2021). Retrieved 24 January, 2021, from https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html

Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018). Deep learning for IoT Big data and streaming analytics: A survey. *IEEE Communications Surveys & Tutorials, 20*(4), 2923–2960. https://doi.org/10.1109/comst.2018.2844341

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López García, Á., Heredia, I., et al. (2019). Machine learning and deep learning frameworks and libraries for large-scale data mining: A survey. *Artificial Intelligence Review, 52*(1), 77–124. https://doi.org/10.1007/s10462-018-09679-z

Obermeyer, Z., & Emanuel, E. (2016). Predicting the future — big data, machine learning, and clinical medicine. *New England Journal of Medicine, 375*(13), 1216–1219. https://doi.org/10.1056/nejmp1606181

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research, 12*, 2825-2830.

Peng, D., & Aggarwal, G. (2015). Modeling MOOC dropouts. *Entropy,10*(114), 1–5. http://cs229.stanford.edu/proj2015/235_report.pdf

Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, *2016*(1). https://doi.org/10.1186/s13634-016-0355-x

RandomForestClassifier—scikit-learn 0.24.1 documentation (2021). Retrieved 24 January 2021, from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wires Data Mining and Knowledge Discovery*, *10*(3). https://doi.org/10.1002/widm.1355

Sorensen, L. (2018). "Big Data" in educational administration: an application for predicting school dropout risk. *Educational Administration Quarterly, 55*(3), 404–446. https://doi.org/10.1177/0013161x18799439

StandardScaler — scikit-learn 0.24.1 documentation. (2021). Retrieved 24 January 2021, from https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

Support Vector Machines—sckit-learn 0.24.1 documentation (2021). Retrieved 24 January, 2021, from https://scikit-learn.org/stable/modules/svm.html

Wilson, K. H., Xiong, X., Khajah, M., Lindsey, R. V., Zhao, S., Karklin, Y., ... & Heffernan, N. (2016). Estimating student proficiency: Deep learning is not the panacea. In *Neural Information Processing Systems, Workshop on Machine Learning for Education*.

Xiong, X., Zhao, S., Van Inwegen, E. G., & Beck, J. E. (2016). Going deeper with deep knowledge tracing. In *Proceedings of 9th International Conference on Educational Data Mining* (pp. 545–550).
Zhang, Q., Yang, L., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion, 42*, 146–157. https://doi.org/10.1016/j.inffus.2017.10.006
Zhou, L., Pan, S., Wang, J., & Vasilakos, A. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing, 237*, 350–361. https://doi.org/10.1016/j.neucom.2017.01.026