Education and Information Technologies
https://doi.org/10.1007/s10639-019-10068-4

Predictive analytics in education: a comparison of deep learning frameworks

Tenzin Doleck[1] · David John Lemay[2] · Ram B. Basnet[3] · Paul Bazelais[2]

Received: 1 October 2019 / Accepted: 14 November 2019/Published online: 30 November 2019

Abstract
Large swaths of data are readily available in various fields, and education is no exception. In tandem, the impetus to derive meaningful insights from data gains urgency. Recent advances in deep learning, particularly in the area of voice and image recognition and so-called complete knowledge games like chess, go, and StarCraft, have resulted in a flurry of research. Using two educational datasets, we explore the utility and applicability of *deep learning* for educational data mining and learning analytics. We compare the predictive accuracy of popular deep learning frameworks/ libraries, including, Keras, Theano, Tensorflow, fast.ai, and Pytorch. Experimental results reveal that performance, as assessed by predictive accuracy, varies depending on the optimizer used. Further, findings from additional experiments by tuning network parameters yield similar results. Moreover, we find that deep learning displays comparable performance to other machine learning algorithms such as support vector machines, k-nearest neighbors, naive Bayes classifier, and logistic regression. We argue that statistical learning techniques should be selected to maximize interpretability and should contribute to our understanding of educational and learning phenomena; hence, in most cases, educational data mining and learning analytics researchers should aim for explanation over prediction.

Keywords Machine learning · Deep learning · Educational data mining · Learning analytics · Classification · Predictive analytics

1 Introduction

The ability and need to record, track, aggregate, and analyze data has trained focus on analytics in many fields. As more data about learners and learning becomes available, it

✉ Tenzin Doleck
doleck@usc.edu

Extended author information available on the last page of the article

is critical for educational researchers to better understand and use this data to gain insights into the education system and into teaching and learning activities (Aldowah et al. 2019; Doleck et al. 2016). However, it is equally important and pertinent to examine the analytical tools and techniques that enable us to make sense of the data and to build useful knowledge. Educational researchers have recognized both the value of data as a resource for knowledge discovery and on the affordances of computational models and theories for educational applications and research. Recent advances in computational methods, such as machine learning, have allowed educational researchers to generate data-driven insights about learning and learning outcomes (Avella et al. 2016; Doleck et al. 2016; Lemay & Doleck, 2019; Papamitsiou and Economides 2014; Romero and Ventura 2016). Machine learning methods, such as support vector machines and Naïve Bayes, have been frequently applied to educational datasets. Indeed, Jordan and Mitchell (2015) note that two trends have supported the progress and increased use of machine learning: proliferation of data and new learning algorithms.

From this context, the application of machine learning to educational data, two subfields have developed, Learning Analytics (LA) and Educational Data Mining (EDM) (Baker and Inventado 2014; Papamitsiou and Economides 2014; Siemens and Baker 2012). Research in this stream highlights the important potential of machine learning for educational research. Indeed, machine learning methods (Kotsiantis 2007) are increasingly being used to model student behaviors in learning environments (Baker and Inventado 2014). Despite wide application of machine learning techniques in educational research, relatively little attention has been paid to the development and application of deep learning techniques in the LA/EDM literature.

Deep learning is a form of machine learning, inspired by biological neural networks, which "allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction" (LeCun et al. 2015, p. 436). There are two key aspects of deep learning: "(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 and Yu 2014, p. 201). Interest in deep learning has increased substantially as deep learning has made important progress across a range of complex tasks including voice and image recognition, and so-called complete knowledge games such as chess, go, and StarCraft (Batmaz et al. 2018; Ismail Fawaz et al. 2019; Nguyen et al. 2019; Zhang et al. 2018). And indeed, deep learning has gained favor with researchers in many fields, and has already spurred a great deal of research (Ismail Fawaz et al. 2019; LeCun et al. 2015; Zhang et al. 2018).

2 Predictive analytics in education

Predictive analytics is a commonly used approach in LA/EDM research (Peña-Ayala 2014). Predicting learning and learning outcomes from educational data is an important objective, bearing the potential to generate new insights for education and practice. Indeed, the issues of the use of predictive analytics and the implications thereof in shaping critical issues is highlighted by Rajni and Malaya (2015), who note that predictive analytics can "help improve the quality of education by letting decision makers address critical issues such as enrollment management and curriculum development" (p. 24). Whereas there is considerable educational research literature examining the utility of machine learning in predicting learner behaviors (Costa et al. 2017; Lykourentzou et al. 2009), comparatively fewer research effort has been directed at examining the suitability of deep learning for predictive analytics with educational data (Botelho et al. 2017; Doleck et al. 2019). In fact, some even note that "deep neural networks for student modeling are not yet well understood" (Jiang et al. 2018, p. 199).

It is therefore incumbent upon educational researchers to study deep learning techniques and their use in educational sciences to assess the potential of deep learning for LA/EDM (Pang et al. 2019). Thus, the objective of the study reported here is to assess the utility of deep learning by comparing various deep learning frameworks for classification tasks using two educational datasets. In doing so, we advance understanding of the potential scope, complexity, and applicability of deep learning methods for classification tasks applied to the specific context of education. While we ground this study in the context of LA/EDM research, it will no doubt be of interest to researchers in other fields.

3 Literature review: Deep learning in education

The application of deep learning (Do, Prasad, Maag, & Alsadoon, 2019; Ismail Fawaz et al. 2019; LeCun et al. 2015; Zhang et al. 2018) is an area of increasing interest to researchers and practitioners in education (Botelho et al. 2017; Doleck et al. 2019; Jiang et al. 2018); yet there exist relatively few research studies examining and assessing its use. We identify and review the available relevant literature on the use of deep learning models in the context of education. Contributions have, among others, been devoted to predicting learner behaviors and outcomes. Moreover, the bulk of the previous work has tended to rely on datasets from computer-based learning environments. While there are studies focusing on the application of deep learning models, we focus on studies that highlight the use of deep learning models in comparison to other related approaches. In the following, we give a brief overview on relevant work.

Using three different kinds of data (simulated data, Khan Academy data, and the Assistments benchmark dataset), Piech et al. (2015) compared Deep Knowledge Tracing (DKT), "flexible recurrent neural networks that are 'deep' in time to the task of knowledge tracing" (p. 1) to standard Bayesian Knowledge Tracing (BKT), "approach for building temporal models of student learning" (p. 2) in predicting student performance. The authors provide initial evidence on the improvements yielded by Deep Knowledge Tracing. They reported that, on all three datasets, Deep Knowledge Tracing outperformed Bayesian Knowledge Tracing. While this finding provides initial evidence of the efficacy of Deep Knowledge Tracing, however, it should be noted that other researchers have highlighted the potential weaknesses and shortcomings of the study, with some studies not able to reproduce the results after proper data formatting (e.g., Xiong et al. 2016).

Another related study is by Wilson et al. (2016), who study learner-system interaction data (three datasets) from computer-based learning systems. The authors compared item response theory (IRT) based proficiency estimation methods, which estimate "latent quantities corresponding to student ability and assessment properties such as difficulty" to Deep Knowledge Tracing (recurrent neural network model) to predict a student's future response given previous responses. Across all three datasets, the results revealed that IRT-based models do as well or better than Deep Knowledge Tracing.

Botelho et al. (2017) studied the use of deep learning model by applying deep learning models (recurrent neural networks, Gated Recurrent Unit networks, and Long-Short Term Memory networks) to the problem of sensor-free affect detection. They compared the results of deep learning models to past results obtained using traditional machine learning algorithms. The experimental results revealed that while the deep learning models achieved better AUC (Area under the ROC Curve), they do not, however, find any improvement in Cohen's kappa values.

In line with Botelho et al. (2017), but focusing on comparing deep neural networks and feature engineering, Jiang et al. (2018), using data from middle school students learning in an open-ended learning environment (Betty's Brain), compared deep neural network approaches with a feature engineering approach for predicting affective states and behavior. The two approaches were compared using cross-validated performance (kappa and A′ values). The results revealed that in general deep learning model displayed similar or higher A′ values, while the feature engineering approach resulted in higher kappa values.

Work by Mao et al. (2018) provides a comparison of the effectiveness of Bayesian Knowledge Tracing (BKT) and Intervention-BKT (IBKT) to deep learning based model (Long Short Term Memory (LSTM)) using data from two intelligent tutoring systems. According to the authors, BKT is essentially a two-state Hidden Markov Model, IBKT incorporates different types of instructional interventions into BKT, and LSTM is a special type of recurrent neural network. In addition to testing the three models (BKT, IBKT, and LSTM), the authors also tested additional variants incorporating skill discovery method (SK); thus, testing six models (BKT, IBKT, LSTM, BKT + SK, IBKT+SK, and LSTM+SK). The comparison exercise was conducted for two different student modeling tasks: post-test scores and learning gains. On the first task, predicting post-test scores, BKT and BKT + S yielded better results in comparison to other models. In contrast, for the second task, predicting learning gains, LSTM and LSTM+SK outperformed other models.

Thus, past studies exhibit mixed findings regarding the performance of deep learning models in the context of education. Against this backdrop, it is worthwhile to scrutinize the use of deep learning. As such, there is a need to reconcile the inconsistent findings about the use of deep learning models in LA/EDM research. This study aims to advance the LA/EDM literature by further illuminating the use and applicability of deep learning in education.

4 Purpose

We are particularly interested in understanding the use of various deep learning frameworks/libraries for modeling educational data. We attempt to meet this goal by doing the following: we test the performance of several deep learning frameworks/libraries across two educational datasets. Among the various methods employed in LA/EDM research, the most popular method is classification (Papamitsiou and Economides 2014), which is what we focus on in this study. Regarding the deep learning frameworks/libraries evaluated, we use the following: Keras, Tensorflow, Theano, fast.ai, and Pytorch (for a review, see Nguyen et al. 2019).

5 Datasets

We use two datasets from educational contexts for our experiments: MOOC dataset (Lemay and Doleck 2019) and CEGEP Academic Performance dataset (Bazelais et al. 2018). We use two datasets to account for particularities in datasets, and because prediction analyses may yield different results depending on the characteristics of the dataset. The details of the data are provided below.

The MOOC dataset, which includes 6241 instances, consists of ten video-viewing features (i.e., independent variables) tabulated on a weekly basis to accord with weekly assignments and an outcome (i.e. dependent) variable, that is, performance on the weekly assignment. The ten video-viewing features were calculated from EdX log files and consist in the number of videos viewed per week, number of stops, pauses, rewinds, fastforwards, average fraction played, average time spent watching, average fraction completed, average playback rate, and standard deviation of playback rate. These video-viewing features are developed at length in Brinton and Chiang (2015).

The CEGEP academic performance dataset, which includes 309 instances, consists of information about students age, gender, prior academic performance (high school performance), and an outcome variable (i.e. enrollment in honors science courses).

6 Analyses and results

The first analysis evaluates how accurately the ten video-viewing features predict performance. The second analysis evaluates how accurately the three features (age, gender, and prior academic performance) predict enrollment in science courses at the college level.

Various deep learning frameworks/libraries (Keras, Theano, Tensorflow, Pytorch, and fast.ai,) were applied on the two datasets and compared for predictive accuracy. The Keras interface is a high-level neural networks API, written in Python, usable for TensorFlow, Theano, and CNTK as a back-end (Home-Keras Documentation 2019). Theano, a Python-based deep learning framework, is used to optimize and evaluate mathematical expressions (Theano 1.0.0 documentation 2019)). TensorFlow, an open-source platform developed by Google, to develop and train

Table 1 Keras (Using TensorFlow as backend)- Classification Accuracies

Optimizer	Accuracy	*SD*
adadelta	58.29%	4.80%
adagrad	61.80%	1.75%
adam	69.19%	1.68%
adamax	68.81%	1.69%
nadam	68.95%	2.25%

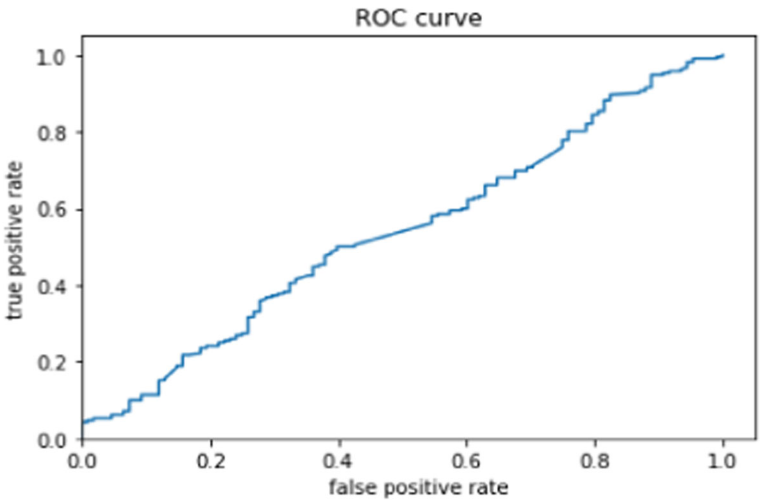

Fig. 1 TensorFlow (BoostedTrees)- ROC Curve

machine learning models using high-level APIs like Keras (TensorFlow 2019). PyTorch is an open source machine learning framework that can be used with popular libraries and packages such as Cython and Numba (PyTorch 2019). fast.ai is an open-source Python-based library that uses PyTorch (Fast.ai 2019).

We also sought to expand the scope of our analysis by considering the effects of adjusting network parameters. Specifically, we evaluate the effects of tuning a network as certain hyperparameters can be adjusted to improve predictive accuracy (Nguyen et al. 2019). We assess the results of changing the size of the network by making the network smaller and larger. Furthermore, as deep learning models are prone to overfitting, we also assess the dropout technique (Srivastava et al. 2014) developed to mitigate this issue.

Finally, we also report the results of popular machine learning algorithms (Support Vector Machines, Naïve Bayes, Logistic Regression, and K-Nearest Neighbors) on the two classification tasks. We do so to illustrate how machine learning algorithms perform relative to deep learning. All the analyses were carried out in Python, using the popular Scikit-learn library (Pedregosa et al. 2011). In this study, for all the experiments, we use accuracy as the measure of model performance.

Table 2 Keras (Using Theano as backend)- Classification Accuracies

Optimizer	Accuracy	*SD*
adadelta	68.25%	1.49%
adagrad	66.37%	2.50%
adam	68.01%	1.99%
nadam	69.00%	2.08%
sgd	66.82%	2.58%

Table 3 Pytorch- Classification Accuracies

Optimizer	Accuracy	*SD*
adadelta	67.37%	27.14%
adagrad	67.94%	32.79%
adam	68.25%	46.55%
rmsprop	68.13%	25.97%
sgd	67.26%	26.80%

7 Results: MOOC dataset

For the first dataset, Keras (using TensorFlow as backend) resulted in accuracies ranging from 58.29% to 69.19% (see Table 1). TensorFlow (BoostedTrees) resulted in an accuracy of 63.13%; see Fig. 1 for the ROC curve. Keras (using Theano as backend) resulted in accuracies ranging from 66.37% to 69.00% (see Table 2). Pytorch resulted in accuracies ranging from 67.26% to 68.25% (see Table 3). Since we ran grid search to find the best accuracy, standard deviations are not reported for these experiments. And finally, fast.ai Framework (using Pytorch as backend) resulted in an accuracy of 68.19% (*SD*: 2.76%). Overall, Keras (using TensorFlow as backend) displayed the best performance.

8 Results: CEGEP dataset

For the second dataset, Keras (using TensorFlow as backend) resulted in accuracies ranging from 62.85% to 85.13% (see Table 4). TensorFlow (BoostedTrees) resulted in an accuracy of 90.32%; see Fig. 2 for the ROC curve. Keras (using Theano as backend) resulted in accuracies ranging from 62.20% to 84.82% (see Table 5). PyTorch resulted in accuracies ranging from 76.05% to 84.79% (see Table 6). Since we ran grid search to find the best accuracy, standard deviations are not reported for these experiments. And finally, fast.ai Framework (using PyTorch as backend) resulted in an accuracy of 88.55% (*SD*: 3.71%). Overall, TensorFlow (BoostedTrees) displayed the best performance; see Fig. 2 for the ROC curve.

Additional experiments on the network We go a step further, to consider the effects of tuning network parameters and features. Along with testing the effects of network size (Nguyen et al. 2019), we also train a dropout network on both the hidden and visible layers (Srivastava et al. 2014). For this set of experiments, we employed Keras using TensorFlow as backend. The results for the MOOC data and CEGEP data are presented

Table 4 Keras (Using TensorFlow as backend)- Classification Accuracies

Optimizer	Accuracy	*SD*
adadelta	62.85%	31.36%
adagrad	82.55%	4.06%
adam	84.82%	3.15%
rmsprop	85.13%	3.23%
sgd	84.16%	1.60%

Table 5 Keras (Using Theano as backend)- Classification Accuracies

Optimizer	Accuracy	*SD*
adadelta	62.20%	30.44%
adagrad	82.87%	3.76%
adam	84.82%	3.15%
rmsprop	83.84%	1.92%
sgd	84.16%	1.60%

in Tables 7 and 8 respectively. For the MOOC data, accuracies ranged from 67.81% to 69.13% (see Table 7). For the CEGEP data, accuracies ranged from 84.15% to 86.06% (see Table 8). Overall, the findings show minimal effects for tuning these networks.

Machine learning algorithms We now present the results of machine learning algorithms as a comparative evaluation of deep learning and machine learning. The classification accuracies for the two datasets are presented in Tables 9 and 10. For the MOOC dataset, we found predictive accuracy for the various machine learning algorithms ranging from 63.04% to 69.31% predictive accuracy. For the CEGEP dataset, we found predictive accuracy for the various machine learning algorithms ranging from 84.16% to 90.60% predictive accuracy. Overall, we find that machine learning algorithms yield similar prediction performance as deep learning.

9 Discussion

Interest in deep learning has picked up dramatically in recent years; however, the study of deep learning techniques applied to educational data is emergent. Therefore, to assess the utility and relevance of deep learning in LA/EDM research, we undertook a comparative study of various deep learning approaches and other machine learning techniques available as open-source packages in the Python programming environment using two different educational data sets. The first dataset employed student video-viewing behaviors to predict performance in a MOOC. The second dataset employed student characteristics to predict enrollment in college-level honors science program.

For the first dataset, we found predictive accuracy for the various deep learning techniques ranging from 58.29% to 69.19%. For the second dataset, we found predictive accuracy for the various deep learning techniques ranging from 62.20% to 90.32%. Moreover, we found negligible improvements from using network tuning techniques. (testing the effects of network size (Nguyen et al. 2019) and training a dropout network

Table 6 PyTorch- Classification Accuracies

Optimizer	Accuracy
adadelta	76.05%
adagrad	84.79%
adam	84.46%
rmsprop	84.46%
sgd	84.46%

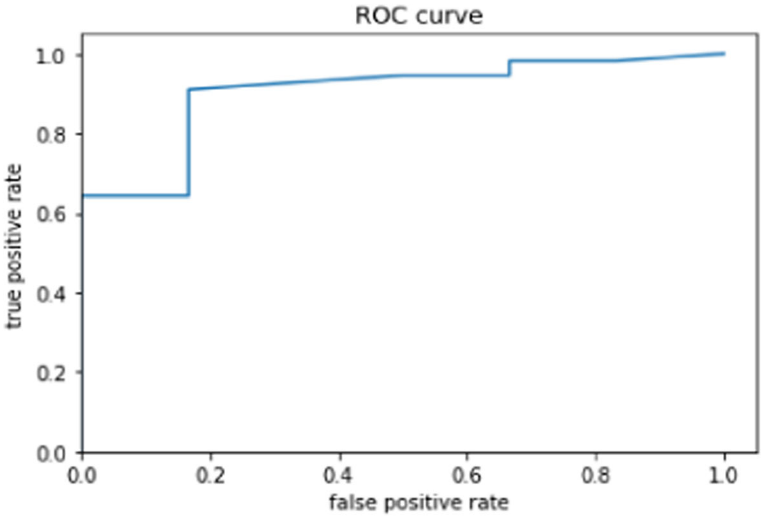

Fig. 2 TensorFlow (BoostedTrees)- ROC Curve

on both the hidden and visible layers (Srivastava et al. 2014)): for the MOOC data, accuracies ranged from 67.81% to 69.13% and for the CEGEP data, accuracies ranged from 84.15% to 86.06%. We conducted additional analysis to assess the predictive accuracy of machine learning algorithms on the two datasets. We found that other machine learning techniques performed as well and in some cases better than the deep learning approaches we tested on our two educational datasets. For the first dataset, we found predictive accuracy for the various machine learning algorithms ranging from 63.04% to 69.31%. For the second dataset, we found predictive accuracy for the various machine learning algorithms ranging from 84.16% to 90.60%. Our findings align with previous work that highlight that deep learning algorithms do not necessarily outperform traditional machine learning algorithms on educational data (e.g., Doleck et al. 2019; Xiong et al. 2016).

Given these results, and similar convergent findings, we are given to question the wider applicability of deep learning methods to LA/EDM research. Deep learning neural networks need to be trained on large datasets, they are sensitive to new data, and prone to overfitting (Batmaz et al. 2018; Marcus 2018; Sünderhauf et al. 2018; Xiao et al. 2018). In fact, Xiong et al. (2016) highlight the need to properly prepare datasets for better understanding and utilizing deep learning techniques. While certain important gains have been registered in some educational research, mainly with respect to user trace data for affect detection and gaze tracking, the shortcomings listed above

Table 7 MOOC data- Classification Accuracies

	Accuracy	*SD*
Dropout on hidden layer	69.13%	2.89%
Dropout on visible layer	67.81%	2.76%
Larger Network	69.07%	1.59%
Smaller Network	69.07%	2.09%

Table 8 CEGEP data- Classification Accuracies

	Accuracy	*SD*
Dropout on hidden layer	84.15%	0.87%
Dropout on visible layer	84.17%	6.49%
Larger Network	86.06%	4.08%
Smaller Network	85.45%	2.91%

significantly limit the overall applicability of deep learning techniques, and reveal the need to further scrutinize the use of deep learning.

We argue that selection of appropriate statistical learning techniques for educational data mining and learning analytics research should be theoretically or methodologically motivated. In educational research, we are often less concerned with prediction than we are with explaining or understanding the phenomenon under study. Hence, statistical learning techniques should be selected to maximize interpretability and should contribute to our understanding of educational and learning phenomena (Lemay and Doleck 2019). Whereas deep learning has demonstrable merits, the requirements for building robust models and their limited interpretability at present constrain their usefulness for the study of education, teaching, and learning. Finally, to realize the potential benefits of machine learning and deep learning depends in large part on the data, as such, it will be crucial to continue to recognize the challenges of predictive analytics, especially in relation to the quality and quantity of data (Rajni and Malaya 2015).

Limitations The findings of the current study have to be reflected against the exploratory nature of the study. This study is limited in that we only compared two educational datasets which we chose for convenience. While we believe these are typical of educational datasets, we do not claim our findings are generalizable to all kinds of educational data in the field of LA/EDM. Indeed, deep learning techniques do find their uses for analyzing massive data streams generated through online behavioral trace data.

Future directions As mentioned, the findings of the current study are a function of the nature of the data used in the study. Future studies using big data are needed to ascertain the value of deep learning in educational research. In the current work, we focused on supervised learning using labeled data. Future work ought to test the use of deep learning for semi-supervised and unsupervised learning as well. Further, more work can also be pursued to test deep learning models of different complexity. Finally, a key line of inquiry

Table 9 MOOC data; Machine Learning Algorithms- Classification Accuracies

Machine Learning Algorithms	Accuracy	*SD*
Support Vector Machines	68.37%	6.95%
Naïve Bayes	64.17%	5.25%
Logistic Regression	69.31%	7.20%
K-Nearest Neighbors	63.04%	7.37%

Table 10 CEGEP data; Machine Learning Algorithms- Classification Accuracies

Machine Learning Algorithms	Accuracy	*SD*
Support Vector Machines	90.60%	6.05%
Naïve Bayes	84.16%	12.0%
Logistic Regression	84.17%	10.9%
K-Nearest Neighbors	90.58%	5.79%

for future research will be to demonstrate the generalizability of deep learning models using different kinds of data (Botelho et al. 2017). More generally, additional research is required as the nature of data and the methods/techniques to analyze the data coevolve (Poitras et al. 2016).

10 Conclusion

We compared deep learning techniques against other machine learning approaches on two typical educational datasets from different contexts to explore the suitability of deep learning for LA/EDM research. We find that performance, as assessed by predictive accuracy, varies depending on the optimizer used in the deep learning libraries. Moreover, we find that deep learning displays comparable performance to other machine learning algorithms. In fact, machine learning algorithms perform as well or better, with fewer and less stringent data requirements. Educational researchers are advised to favor interpretability and explanation over accuracy as criteria for selecting computational techniques for LA/EDM research.

References

Aldowah, H., Al-Samarraie, H., & Fauzy, W. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics, 37*, 13–49. https://doi.org/10.1016/j.tele.2019.01.007.

Avella, J., Kebritchi, M., Nunn, S., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning, 20*(2), 13–29.

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics* (pp. 61–75). New York, NY: Springer.

Batmaz, Z., Yurekli, A., Bilge, A., & Kaleli, C. (2018). A review on deep learning for recommender systems: Challenges and remedies. *Artificial Intelligence Review, 52*(1), 1–37. https://doi.org/10.1007/s10462-018-9654-y.

Bazelais, P., Lemay, D. J., Doleck, T., Hu, X. S., Vu, A., & Yao, J. (2018). Grit, mindset, and academic performance: A study of pre-University science students. *Eurasia Journal of Mathematics, Science and Technology Education, 14*(12), 1–10. https://doi.org/10.29333/ejmste/94570.

Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017). Improving sensor-free affect detection using deep learning. In E. André, R. Baker, X. Hu, M. M. T. Rodrigo, & B. Boulay (Eds.), *Proceedings of the 18th international conference on artificial intelligence in education* (pp. 40–51). London, UK: Springer International Publishing.

Brinton, C. G., & Chiang, M. (2015). MOOC performance prediction via clickstream data and social learning networks. *IEEE Conference on Computer Communications (INFOCOM)*, 2299–2307.

Costa, E., Fonseca, B., Santana, M., de Araújo, F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior, 73*, 247–256. https://doi.org/10.1016/j.chb.2017.01.047.

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., Jarrell, A., Poitras, E. G., Chaouachi, M., & Lajoie, S. P. (2016). A tale of three cases: Examining accuracy, efficiency, and process differences in diagnosing virtual patient cases. *Australasian Journal of Educational Technology, 36*(5), 61–76. https://doi.org/10.14742/ajet.2759.

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.

Fast.ai. (2019). Retrieved from https://www.fast.ai/

Home-Keras Documentation. (2019). Retrieved from https://keras.io/

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. (2019). Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery, 33*(4), 917–963. https://doi.org/10.1007/s10618-019-00619-1.

Jiang, Y., Bosch, N., Baker, R., Paquette, L., Ocumpaugh, J., Andres, J. M. A. L., Moore, A. L., & Biswas, G. (2018). Expert feature-engineering vs. deep neural networks: Which is better for sensor-free affect detection? In *Proceedings of the 19th international conference on artificial intelligence in education* (pp. 198–211). London, UK: Springer.

Jordan, M., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. In I. Maglogiannis et al. (Eds.), *Emerging artificial intelligence applications in computer engineering* (pp. 3–24). Amsterdam, Netherlands: IOS Press.

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

Lemay, D. J., & Doleck, T. (2019). Grade prediction of weekly assignments in MOOCs: Mining Video-Viewing Behavior. Education and Information Technologies. https://doi.org/10.1007/s10639-019-10022-4.

Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education, 53*(3), 950–965. https://doi.org/10.1016/j.compedu.2009.05.010.

Mao, Y., Lin, C., & Chi, M. (2018). Deep learning vs. Bayesian knowledge tracing: Student models for interventions. *JEDM | Journal of Educational Data Mining, 10*(2), 28–54.

Marcus, G. (2018). Deep learning: A critical appraisal. arXiv preprint arXiv:1801.00631.

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.

Pang, B., Nijkamp, E., & Wu, Y. (2019). Deep learning with TensorFlow: A Review. *Journal of Educational and Behavioral Statistics*. https://doi.org/10.3102/1076998619872761.

Papamitsiou, Z., & Economides, A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society, 17*(4), 49–64.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications, 41*(4), 1432–1462. https://doi.org/10.1016/j.eswa.2013.08.042.

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in Neural Information Processing Systems* (pp. 505–513).

Poitras, E. G., Lajoie, S. P., Doleck, T., & Jarrell, A. (2016). Subgroup discovery with user interaction data: An empirically guided approach to improving intelligent tutoring systems. *Educational Technology & Society, 19*(2), 204–214.

PyTorch. (2019). Retrieved from https://pytorch.org/

Rajni, J., & Malaya, D. (2015). Predictive analytics in a higher education context. *IT Professional, 17*(4), 24–33. https://doi.org/10.1109/mitp.2015.68.

Romero, C., & Ventura, S. (2016). Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7*(1), e1187. https://doi.org/10.1002/widm.1187.

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252–254). ACM.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929–1958.
Sünderhauf, N., Brock, O., Scheirer, W., Hadsell, R., Fox, D., Leitner, J., et al. (2018). The limits and potentials of deep learning for robotics. *The International Journal of Robotics Research, 37*(4–5), 405–420. https://doi.org/10.1177/0278364918770733.
TensorFlow. (2019). Retrieved from https://www.tensorflow.org/
Theano 1.0.0 documentation. (2019). Retrieved from http://deeplearning.net/software/theano/
Wilson, K. H., Karklin, Y., Han, B., Ekanadham, C. (2016). Back to the basics: Bayesian extensions of IRT outperform neural networks for proficiency estimation. In *Proceedings of Educational Data Mining* (pp. 539–544).
Xiao, C., Choi, E., & Sun, J. (2018). Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review. *Journal of the American Medical Informatics Association, 25*(10), 1419–1428. https://doi.org/10.1093/jamia/ocy068.
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.

Affiliations

Tenzin Doleck[1] **· David John Lemay**[2] **· Ram B. Basnet**[3] **· Paul Bazelais**[2]

David John Lemay
david.lemay@mail.mcgill.ca

Ram B. Basnet
rbasnet@coloradomesa.edu

Paul Bazelais
paul.bazelais@mail.mcgill.ca

[1] University of Southern California, Los Angeles, CA, USA

[2] McGill University, Montreal, QC, Canada

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