Computers and Education Open 2 (2021) 100060

Contents lists available at ScienceDirect

Computers and Education Open

journal homepage: www.sciencedirect.com/journal/computers-and-education-open

Perspectives on the challenges of generalizability, transparency and ethics in predictive learning analytics

Anuradha Mathrani [a,*], Teo Susnjak [b], Gomathy Ramaswami [c], Andre Barczak [d]

[a] *School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand*
[b] *School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand*
[c] *School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand*
[d] *School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand*

ARTICLE INFO

Keywords:
Learning analytics
Generalizability
Interpretability
Feature extraction
Transparency
Ethics protocol

ABSTRACT

Educational institutions need to formulate a well-established data-driven plan to get long-term value from their learning analytics (LA) strategy. By tracking learners' digital traces and measuring learners' performance, institutions can discern consequential learning trends via use of predictive models to enhance their instructional services. However, questions remain on how the proposed LA system is suitable, meaningful, and justifiable. In this concept paper, we examine generalizability and transparency of the internals of predictive models, alongside the ethical challenges in using learners' data for building predictive capabilities. Model generalizability or transferability is hindered by inadequate feature representation, small and imbalanced datasets, concept drift, and contextually un-related domains. Additional challenges relate to trustworthiness and social acceptance of these models since algorithmic-driven models are difficult to interpret by themselves. Further, ethical dilemmas are faced in engaging with learners' data while developing and deploying LA systems at an institutional level. We propose methodologies for apprehending these challenges by establishing efforts for managing transferability and transparency, and further assessing the ethical standing on justifiable use of the LA strategy. This study showcases underlying relationships that exist between constructs pertaining to learners' data and the predictive model. We suggest the use of appropriate evaluation techniques and setting up research ethics protocols, since without proper controls in place, the model outcome would not be portable, transferable, trustworthy, or admissible as a responsible outcome. This concept paper has theoretical and practical implications for future inquiry in the burgeoning field of learning analytics.

1. Introduction

The educational landscape has evolved with online learning management systems (LMSs) having facilitated any-time, any-place, and any-pace learning. Learners can interact with the different e-learning activities embedded in their institutional LMS; however, in doing so, they leave their digital traces or their digital footprints. Learner activities are captured via clickstream events associated with learners when they browse course content, navigate between course modules, download course material, participate over discussion forums, or upload/ submit assignments (e.g., uploading a file for assessment or submitting an online quiz for marking). Learners' clickstream data annotated with the background LMS data are stored in log files [46, 50] that provide digital footprint awareness to their educational institution. In other words, the digital footprint reveals experiences concerning "every article sound, image and information left, shared and clicked by the person [or leaner] in the digital environment either consciously or unconsciously" ([49], p.50).

Learning analytics (LA) is concerned with sense-making of learners' digital footprints with the aim of understanding learner engagement patterns, such as, how learners traverse course structures and access course content at their own pace and time (e.g., How often do learners watch a video uploaded on the LMS by the teacher? Or, how often do they speed up, pause or rewind the video? [20]). By using such bottom-up approaches, LA can assist in creating new insights for further optimizing online learning experiences. LA employs educational data

* Corresponding author at: School of Mathematical and Computational Sciences, Massey University, East Precinct Gate 1, Albany Expressway (SH17), Albany 0632, Auckland, New Zealand.
E-mail address: a.s.mathrani@massey.ac.nz (A. Mathrani).

https://doi.org/10.1016/j.caeo.2021.100060
Received 15 February 2021; Received in revised form 11 November 2021; Accepted 14 November 2021
Available online 20 November 2021
2666-5573/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

mining (EDM) methods (or reductionist techniques) for generating actionable insights which enable optimization of learning experiences. These methods consist of gathering clickstream data and combining them with other available learner data in order to generate meaningful features that describe learners' unique properties. Some of the commonly generated features include a number of learner logins into a LMS, their quiz completion times, grade averages and deviations from the cohort mean, number of course material downloads or the number of forum posts created/viewed amongst others. Once a rich set of features describing learners has been engineered on both current and past students, machine learning algorithms can then be used to generate predictive models. Subsequent steps frequently involve applying human judgement to the derived models so as to draw out insights which can facilitate enhancements to existing instructional services, and also inform institutions on future-oriented educational delivery strategies [47].

The field of LA is growing and there is ample evidence supporting its extensive applications in higher education [2,58]. However, many challenges still exist in conducting learning analytics and in achieving desired efficiencies. [50], p. 157] are of the view that "most institutions may not be ready to exploit the variety of available datasets for learning and teaching", since building a universal predictive model from the log files extracted from the LMS is not straightforward. The LMS data is scattered at different hierarchal levels that may be difficult to corelate [59]; moreover, the data have to be supplemented with additional data retrieved from other sources (e.g., student admission system, past study records). Once the combined data are pre-processed to a proper format, a predictive model is developed by inputting training datasets (or what is commonly referred to as *seen data*) to machine learning algorithms, whose outputs are then re-applied to new target data points (or to the *unseen data in respect to the algorithm*). Models which display high accuracy on unseen data are deemed to have generalized, or successfully transferred relevant knowledge from seen training datasets to the target data. In practice however, the training data and target data could differ in their composition and also the extent to which the target data represents the problem that is being posed which impacts the transferability or generalizability of the model. Further, ensuring interpretability of the model internals by providing explanations on the scope and rationale of the algorithmic functions used to generate them, is crucial for social acceptance of the models [52]. Watcher et al. suggest conveying human-understandable non-technical explanations to the intended learners on influences of specific features on the overall model predictions.

Alongside these issues also exist the ethical issues related to data privacy and data ownership. While institutions are privy to learners' course-related data, they also have access to learners' personal data (e.g., ethnicity, age, gender, prior study details, etc.), all of which must be used respectfully. The collection and usage of learner data by an institution has broader implications, such as, an increased power over the learner by their institution, or learners receiving little information on what data is being collected, or profiling learners based on race, socioeconomic status, ethnicity or gender, all of which raise moral questions pertaining to intrusion on students' rights and privacy [27, 40]; hence, any data policy used for LA implementation must align with an institution's core principles [39] before it can be used for developing any form of institutional capability.

This section has briefly introduced some challenges commonly faced by a learning analytics enterprise. In the following section, we present four research questions that drive this study. Next we provide an overview of the state of science regarding current endeavors in establishing LA systems; since concept papers are about "what do we do, where have we come from, and what are the areas yet to be examined" rather than covering extensive literature reviews ([19], p. 128). In developing convincing arguments and providing theoretical explanations, concept papers assimilate and combine selected pieces of literary and empirical evidence to form a logical chain of argumentation [24]. This study examines published literature that articulate implementation issues faced in LA, and in doing so, we direct the readers to key pieces of published literature that provide a deeper coverage of the major issues identified. Against this backdrop of previous studies and recent literature which explore general issues with the implementation of LA initiatives [31], we discuss operational challenges frequently encountered in building predictive models with educational datasets across multiple learner environments. Specifically, we discuss the generalizability, model transparency (which covers model interpretability and the explanation of predictions) as well as ethical concerns, and we suggest guiding frameworks to address them. Accordingly, a model generalizability framework is presented. The tensions faced in maintaining accuracy and effectiveness across low and high interpretability models are examined, and trade-offs around model transparency are identified. We further outline an institutional ethics protocol that can provide a regulatory structure for avoiding ongoing conflicts between having an algorithmic-driven strategy and maintaining learner privacy in a LA context. Finally, in the last section, we consolidate key points that emerged from our discussion to answer the research questions that can inform predictive model building activities for future analytics practice.

2. Research questions

While LA manifests as an innovative data-driven capability that can personalize learning based on individual learner needs, researchers need to evaluate the theoretical and methodological stance pertaining to the conduct of their analytics strategy. Researchers encounter non-trivial challenges at all stages of developing LA systems. These can be of a technical nature such as developing and selecting relevant features for predictive modeling, as well as making design choices about which type of predictive model to use given positives and negatives associated with different types. The difficulties can also be of a non-technical nature and concern how the predictive models are used and how their usage is communicated to the learners. This study therefore reflects on challenges associated with the development and deployment of LA systems to enable meaningful transformation of learners' data into relevant features that can lead to improved instructional services.

Following questions are posed.

1 What are the key challenges in effectively deploying LA systems?
2 What difficulties are still encountered in producing generalizable predictive models?
3 What are the next frontiers in being able to extract more value from predictive models, rather than just predictions?
4 Which ethical dilemmas still remain in the deployment and operationalization of LA systems?

3. Current studies on learning analytics

Learning analytics is envisioned by educational institutions as a powerful force that can lead to more personalized learner experiences. It is considered as a way to "track individual student engagement, attainment and progression in near-real time, flagging any potential issues to tutors or support staff" ([42], p. 6.). With the use of predictive models built from historical student datasets, many educational institutions have implemented strategies to boost student retention rates, maintain quality assurance practices, reveal key determinants for academic achievement, bring about self-regulated learning (by predicting individual learning needs) and enhance learners' experience [56]. But when a disconnect emerges between training datasets and target (live) data, the utility of the predictive models can be degraded [45]. Generalizability (or transferability) of the derived model [6] is constrained when the training and target data are extracted from different distributions that exhibit different learner scenarios. Researchers from educational fields, such as the EDM and LA community, are thus to some degree restricted to the use of data belonging to similar courses when predicting students' performances.

In one study, the authors [9] built a universal predictive model from different MOOC (Massive Online Open Course) offerings. Data from three most recently finished MOOC course offerings, and also data from the initial weeks of an on-going course were used for building the predictive model. Using naïve algorithm and importance sampling approaches, they concluded that machine learning techniques should consider model performance on successive offerings of the same courses. The authors concluded that transferability can be improved when important sampling-based approach parameters are tuned by formulating a moving window size on longitudinal variables.

Successful transferability can take place in multiple ways, such as the reuse of some or all of the training data sets, or features extracted from those datasets. The transfer can also consist of reusing some model-specific settings extracted from a trained model to iteratively evaluate classifications in the target domain [23]. Hunt et al. put forward the transfer learning method for predicting students' graduation rates in undergraduate programmes. TrAdaBoost, an extended AdaBoost algorithm, was used to examine the effectiveness. That is instead of assuming all the training dataset (comprising a set of academic and demographic features of students belonging to different departments) came from the same distribution, the authors conducted two separate experiments each time using specific data for training. In the first experiment, the training set comprised all students apart from those studying engineering, while in the second one, the training set comprised all students that were suspended or on academic warnings. The experimental results showed that TrAdaBoost improved the accuracy of predictive models and recorded smallest error in both cases. Generally, TrAdaBoost helps improve the accuracy of predictive models by using the target set as a guide to select related data from the source set. However, when the target sample size is too small to be representative, TrAdaBoost does not improve performance because its selective process will be biased by the target samples and causes over-fitting to the target set. Moreover, when there is a variance in data distribution between training and target data, the predictive capability is compromised.

López-Zambrano, Lara, and Romero [32] proposed generic methods to check the feasibility of predictive models by grouping similar courses by degree or by similar level of usage of activities provided by LMS logs. Experimental results from a well-known classification algorithm (namely J48 from the Weka [55] software) showed that it is feasible to directly generate accurate models with an acceptable accuracy; however, the limitation is that the obtained models might result in low accuracy values with other courses that use different activities or actions compared to the course used for training. In such situations, the log files of the unseen data would show different events and then models efficacy would become compromised. [6]

Baesens, Ravi, Marsden, Vanthienen, and Zhao [5] add that deep analytic techniques (e.g., neural networks, support vector machines, ensemble methods) for building predictive capabilities rely on information (data) and trust, in which trust has not been given proper attention. Trustworthiness of any data-driven algorithmic decision relies on data quality and on fitness of data used that in turn inform specific learner features (e.g., number of forum posts or number of quiz attempts) on which the predictive capabilities are built. Researchers must have proper domain knowledge of the complex implementations of the underlying LMS and understanding of data characteristics in the digital footprint trail. This in turn will inform the internal design and improve predictability; and, position the LA enterprise in better providing human-understandable counterfactual explanations on the significance of any learner-related feature that has been considered by the model to impact the learners' performance [52]. With simple explanations, institutions can send the message across that they do not consider their learners as passive recipients. Explanations provide new grounds for conducting meaningful exchanges leading to ongoing interactions that further builds more trust in the operationalised predictive model. "Building trust is essential to increase social acceptance of algorithmic decision-making" (p. 4); however, explaining the rationale and functionality of the algorithms that together computationally process the raw data with different rule-sets to provide predictor values is not an easy task. To appreciate the reductionist power of analytics and make sense of the predictor variables, learners must be able to comprehend how the predictions align with their digital footprint; therefore, interpretability of the models at a high level by its intended audience is crucial.

Rubel and Jones [40] raise further questions regarding the conflicting positions between student privacy and learning analytics. While they recognize the benefits of LA, they caution on usage of other forms of student personal data, such as the students' socioeconomic status, their demographic profile, academic history, or their financial aid package. Classifying data categories statistically based on socioeconomic status, race, or gender without proper thought could perpetuate "old prejudices" and "have a stifling effect on individuals and society" ([51]; p. 254). Proper controls that allow for differential access based on the merits of the purpose of data usage in learning analytics should be formed. That is, collecting learner information based upon their religious observances or politics amongst others would be impermissible under these controls; however, information based on learning patterns so as to nudge students for enhancing their learning outcomes would be endorsed. Tene and Polonetsky suggest disclosure statements be made by institutions on their usage of individual data that has been harvested in log files, but without disclosing the internal logic of their proprietary algorithms (which constitute their trade secrets). Further, some meaningful explanations should be offered on algorithmic interpretability so as to increase societal acceptance and build trust in the automated decisions from the predictive models [52].

This section has discussed some of the recent research works on LA and highlighted both the opportunities and limitations. In particular, we have emphasized on the generalizability, transparency of predictive models and ethical challenges faced as predictions are tailored across diverse course offerings and learner groups. Next, we propose some key points to address these issues.

4. Challenges in learning analytics

This section provides more perspectives to the challenges that have been identified in learning analytics literature. Concept papers, in particular, do not have data; rather their focus is on integration of domain concepts to offer propositions that can serve as a bridge between validation and usefulness [19, 53]. We highlight generalizability, model transparency and ethical domain challenges as some of the key areas from literature. Model generalizability refers to transfer learning issues with regard to the relevance of patterns extracted from the training datasets for use on new data. Educational institutions are accountable for ensuring that the model's predictive abilities are reliable, besides also holding a social responsibility of explaining the significance of model's predictions to their intended audience (or to their current learners), referred to as the model interpretability domain or explainable artificial intelligence. Moreover, the ethical domain related to the proper collection and usage of learner-related data is an important consideration for the LA enterprise. We discuss these challenges in more detail in the following three subsections.

4.1. Generalizability challenges

The intent of learning analytics is to uncover underlying relationship between predictors (e.g., assessment grades, participation via forum posts) and possible outcomes (e.g., final grades, course engagement level). A machine learning algorithm aids in computationally exploring data patterns in historic datasets and inferring rule-sets that can map predictor variables with outcomes being modelled. The result of this inductive process is a predictive model that can then be deployed to make predictions on live data. However, achieving high predictive accuracies on real-world application data using these inductive methods is one of the key challenges. The failure of predictive models to generalize may happen for numerous reasons, and these will differ in respect to the unique challenges with which each application domain is associated. Machine learning challenges that are most relevant to the LA domain are “the curse of dimensionality” ,[1] concept drift and class-imbalance ([41]. Moreover, the non-deterministic nature of the LA domain adds to the complexities around making accurate predictions about human behavior. While the dynamic nature of educational contexts can also compromise the predictive power induced from historic (training) data when applied to live (target) datasets. We observe this in dynamic environments when the training data used for deriving a predictive model ceases to correlate with the live data onto which the predictive model is being applied. In dynamic contexts such as LA, the underlying data may change frequently thus rendering the trained models using historic data, inaccurate. An example of this are models which have been developed for predicting learner outcomes for specific courses with strong dependencies on features representing different assessments. As the courses and the assessments evolve, or where the assessment syllabus and evaluation styles have changed, the historic training data risks losing relevance on the current live data. Such course-specific and highly tailored features are more powerful, but they have the potential to decrease the generalizability of the predictive models when their usage changes in subsequent deliveries [9, 16] .

Poor model generalization may also occur if the size of the training sample is not large enough for the machine learning algorithm to effectively create decision boundaries. In this instance high-dimensionality data are the culprit owing to the fact that it is tempting to exceed the number of features used in modeling in proportion to the size of the training datasets. Machine learning algorithms always uncover patterns. Many however are phantom patterns and do not correlate with reality. The more features an algorithm has access to, the higher the proclivity to discover meaningless patterns and overfit the model to irrelevant idiosyncrasies of the underlying dataset.

In the LA context, access to datasets used for machine learning may be limited to a few courses due to privacy and legal constraints in different jurisdictions, as well as to policies requiring opt-in consents from learners. Consequently, this may result in training datasets that are insufficient in size and therefore more prone to the negative effects arising from high-dimensionality data. More diverse and representative samples are critical for the field of LA research [16]. With small-sized datasets, the resulting models are even more likely to overfit by capturing residual noise rather than provide useful patterns. Alternatively, a model may underfit and thereby be unable to learn a complex decision boundary when the data volume is not rich enough to support this sufficiently. In either of the above scenarios, the accuracy obtained with the given training data may not match that of the models that have been deployed into a production environment, thereby limiting the usefulness of the derived models. Suggested ways to solve these issues are to make use of more general features, eliminating redundant or less-discriminatory features, incorporating more recent data points (while omitting some older data) and reducing the complexity of the models [11, 35]

It is well accepted within the machine learning community [12, 33] that the quality of features is more important than the choice of algorithms, or even the size of the training datasets. It has already been discussed that as courses evolve, often handcrafted features developed for an earlier course delivery may not express correlations with an ongoing course. Or they may not even exist in a subsequent course delivery.

The problem of generalization is considerable. But this is further confounded by the phenomenon of *concept drift*. Concept drift refers to what happens to a predictive model over time as the training data and the current real-life data become disconnected. For example, prior to 2000s, virtual learning environments in LA were rare. These days, they are ubiquitous. This shift in technology also represents a gradual shift in the manner that students have come to learn and generate their digital footprints – and this shift continues today. The consequence is that using historic educational data that reaches too far back in time risks producing models which are not relevant for making decisions about current cohorts of students. However, using too little of the historic data for training also risks producing models that overfit. A difficult challenge arises in this domain where frequent concept drift needs to be accounted for and detected. How this can be accomplished, is still an active area of research [33].

In addition to the above, educational datasets tend to also be highly class-imbalanced. Class-imbalanced dataset domains possess an unequal representation in the number of samples for the different dependent variables. Classes with proportionally much fewer samples than the larger classes tend to experience a degradation in accuracy due to the fact that machine learning algorithms generally focus more on majority classes. As a result, predictive models may behave differently in terms of generalizability on majority and minority classes [44]. For instance, if the training dataset consists of very few students labelled as *at-risk* students (or students facing learning challenges), then there is an increased chance of misclassification in detecting these students on live data streams. The overall accuracy would likely be biased towards students not at-risk. The proposed solutions to handle class imbalances are pre-processing the dataset in order to construct a more balanced training dataset. This may be done by under sampling, over sampling or synthetic sampling methods [14]. Under sampling implies removing samples from the majority class (i.e., the not at-risk students) to balance with the minority class (i.e., the at-risk students); over sampling involves creating copies of the existing minority class samples (i.e., the at-risk students) to match the majority class (i.e., the not at-risk students); while synthetic sampling involves increasing the minority class with synthetic samples using feature space similarity. While strategies for mitigating class imbalances exist, it is not always clear which strategy should be used for a particular dataset and the overall challenge of machine learning on this type of data is also an active area of research with many open questions [26].

Another concern raised by [18] is that currently predictive models are very focused on transferring knowledge from the source domain (training dataset) to a target domain (live data) irrespective of whether these domains are related; thus potentially resulting in low generalizability. For example, consider an example of two sets of undergraduate students enrolled in a university. In this scenario, one set comprises final-year undergraduate students while the second set comprises first-year undergraduate students who have just entered tertiary study. These two sets represent different domains. The domains are obviously related (i.e., both belong to tertiary education), but their activity log files (extracted from the LMS) are likely to represent different learning patterns. The first-year undergraduates will interact with different activities within LMSs in a specific pattern due to having no previous experience with them. This will differ to the navigation patterns of the final year students who are more experienced. Moreover, the final year student cohort would have more background knowledge of their chosen area of study, hence their approach towards using online study resources would differ. These two domains are different and deriving a universal model between them could result in poor generalizability, or negative knowledge transfer. Fig. 1 below outlines these generalizability challenges as identified from literature.

The above challenges are highly relevant for the LA domain in respect to machine learning and generalizability. More broadly however, generalizability of machine learning models in the context of Big Data are also complicated by the presence of noisy data and the necessity

[1] The curse of dimensionality is a phrase coined by Bellman [7], that refers to high-dimensional data, which in a LA context refers to learners’ data that has a very large number of features or attributes describing each student. .

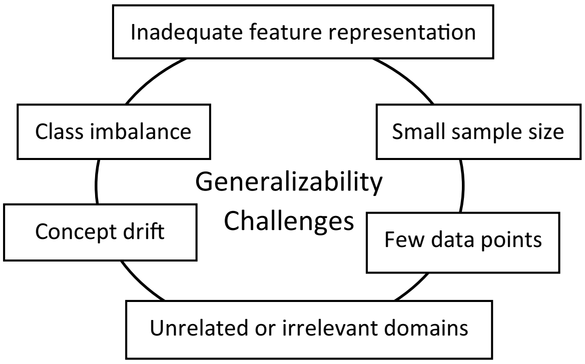
Inadequate feature representation
Class imbalance
Small sample size
Generalizability Challenges
Concept drift
Few data points
Unrelated or irrelevant domains

Fig. 1. Generalizability Challenges.

to learn with unreliable or contradicting data. This can also be compounded by the need to use sub-optimal algorithms due to the fact that the training data may be too large and cannot be processed and held entirely in a processing machine's working memory, which many algorithms require. These, as well as further challenges are reviewed in [28] and [57].

4.2. Model transparency (Interpretability and explainability) challenges

The rapid proliferation of predictive models into areas where previously human decision making was exclusive, has highlighted the need to be able to interrogate the mechanisms behind the models that drive their decisions. The goal is to ultimately generate glass-box models which provide transparency to the human-in-the-loop. There are a number of reasons why this is becoming important. Trust in black-box models is generally low, and trust in these systems can be forged through higher transparency. It is important to be able to verify the inner mechanics of the outputs of these algorithms in order to ensure that they are robust, reliable, and fair. Increasingly legal requirements are beginning to mandate that the predictive models account for their decisions and that the reasoning behind any automated decisions be clearly articulated to those affected by them. In addition, [52] point out the importance of those affected to have the ability to contest adverse decisions made by automated systems, and interestingly, to also have the ability to understand what would need to change in order to receive a desired result in the future, based on the current decision-making model.

The high interest in seeking new approaches to better understand the predictive modeling in real-life contexts such as education, has given rise to relatively new research fields such as Interpretable Machine Learning and Explainable Artificial Intelligence (XAI). The main goals of research in these spaces revolves around how *global model interpretability* and *model prediction explainability* can be achieved. Helpful literature surveys on these topics have emerged recently [1, 4, 13], together with some examples of some early work that is specific for the LA domain [3, 36]

Technically, *global model interpretability* deals with the challenge of making sense of the internals of a predictive model once a model has been trained by a machine learning algorithm. While *global model interpretability* highlights the behavior of the entire model at an abstract level, *model prediction explainability* on the other hand relates to the ability of a model to explain how it has arrived at a given prediction for a *specific* student. *Model interpretability* enables an institution to communicate to all students how a predictive model works using broad brushstrokes. M*odel prediction explainability* enables an institution to respond to a specific student query about how and why they might have been identified as an at-risk student given this student's unique data.

Some algorithms produce models whose internals are in the form of decision trees or rule-sets which are highly interpretable at a global level. With these algorithms, it is easy to see the decision points and threshold values for various features. However, higher accuracies are usually attained by algorithms that produce black-box models. Difficult trade-offs need to be made since some degree of accuracy or model interpretability will be sacrificed when choosing an algorithm. However, new suites of tools are emerging which are able to expose the internal logic of opaque models and induce them with adequate global interpretability, often through visualisations. Various approaches can be used such as generating proxy or surrogate models which approximate the underlying black-box model and generate interpretable models like decision trees (Trepan; [10]), rule-sets (BETA; [29]) or linear models (LIME; [37]). Apart from standard feature importance plots, more effective insights about the inner workings of models can be gleaned using tools that generate Partial Dependence Plots (PDP; [15]), which show how each feature affects the model's predictions across a range of values. While Individual Conditional Expectation (ICE; [21]) plots extend the PDPs with the ability to display the mean predicted outcomes for a range of values of a selected feature, meanwhile holding the values of other feature values constant. The challenge remains of matching the suitable tool for the particular educational dataset at-hand and performing extensive experimentation in order to identify the right tool.

In respect to explainability of predictions, global models with a high degree of interpretability can usually explain their individual decisions by highlighting the path through the decision tree that a single data point traverses, or in the case of rule-sets, listing the selected rules which were triggered by given predicates being met. In the instance of *k*-Nearest Neighbour models, *k* number of most similar students to the target student can be returned for inspection and comparison.

With opaque algorithms, once again additional tools are required in order to explain the model's reasoning. Recently, Shapley Additive Explanations (SHAP; [34]) have gained popularity in their effectiveness to visually explain the drivers of a model's decision-making process. Anchors [38] have also been recently developed as a tool that imparts a high degree of explainability. Anchors extend LIME by creating proxy models which are able to approximate non-linear functions and output a most succinct decision rule that "anchors" the prediction for a given data point for a given precision requirement. This means that rule anchors a prediction (the prediction will not change) with a given decision rule even if values change in other feature values, thus highlighting the key features for a given student. Using an opposite approach, counterfactuals [52] search out the smallest required change to a student's values which would result in a change of prediction. A counterfactual explanation in a case of a student, offers insights in terms of a minimum shift in key features that would need to take place in order to achieve a different outcome to what is currently predicted.

In summary, achieving full transparency and interpretability of operationalised predictive models in educational settings is challenging for a number of reasons already outlined, and presents delicate tradeoffs that need to be made (refer Fig. 2). It can be tempting to use machine learning models that come with a high degree of intrinsic interpretability, but they produce less accurate models. The reverse is true with black-box algorithms. However, the trade-off is that additional tools need to be used in order to unpack both the internals of the models at an abstract level, as well as a suite of tools that provide explanations at an individual level of each student. The second scenario places additional burdens on educational providers to have larger and more skilled teams of data scientists who are able to work with a wide range of tools.

4.3. Ethical challenges

There is considerable evidence that confirms the value of learning analytics in the enhancement of institutional teaching environments. However, many institutions worldwide are still at early stages in their adoption of LA and in their practice of using data-informed approaches for improving instructional services and supporting learners [43], as they deal with associated ethical challenges. From an ethical standpoint, the field of learning analytics sits in contrast with other big data

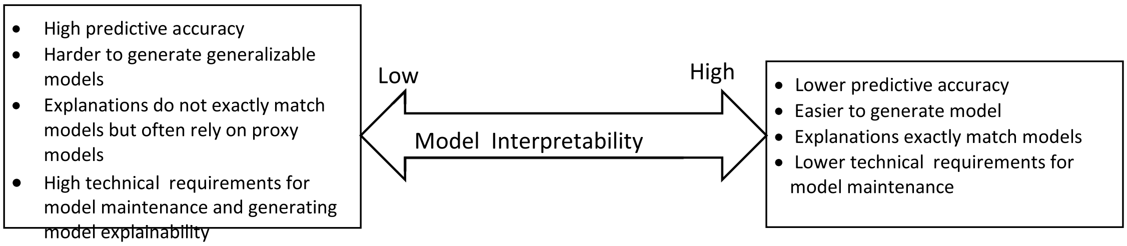
High predictive accuracy
Harder to generate generalizable models
Explanations do not exactly match models but often rely on proxy models
High technical requirements for model maintenance and generating model explainability
Low
High
Model Interpretability
Lower predictive accuracy
Easier to generate model
Explanations exactly match models
Lower technical requirements for model maintenance

Fig. 2. Trade-offs between Low and High Model Transparency.

analytics (e.g., marketing analytics), since the digital footprints used are directly linked with individual students who can be identified via unique identifiers [40]. This raises questions about what constitutes acceptable or ethical analytics activities; that is, to what degree should the learners be informed about the details of how their data is used and whether explicit consent should be sought. Moreover, taking actions on automated predictions or recommendations from predictive models introduces levels of uncertainty, as future possibilities are conceived based upon their alignment with historic data. Institutions that use LA must confront lack of predictive certitude in deciding the effectiveness of predictive outcomes. Therefore, when flagging particular students, such as those who may be facing learning difficulties with intent to provide them with appropriate pedagogical interventions, they must provide learners with explanations on why such interventions are being actioned.

Interventions could comprise automated notifications that can nudge students by recommending relevant learning resources or by making some other provision in the form of personalized human assistance to help students overcome their learning difficulties [56]. Even though the intent of flagging students is to improve their learning environment, the fact remains that LA can also be considered to be a form of surveillance [25]. This tension between surveillance concerns and getting the true value out of LA has made it difficult to devise concrete ethical guidelines. Having said this, all scholarly research studies must follow high ethical standards. This involves conducting a proper ethical scrutiny by the concerned institution and by the analytics team to ensure *appropriate* protocols have been used in data collection and analysis. The word *appropriate* in the context of educational dataset for LA implies that institutional codes of conduct should cover elements of informed consent, privacy and de-identification, clearly state the scope and motive of learner data tracking, define the boundaries on data usage and have measures to prevent unauthorised access and disclosure of learner data [48]. However, getting ‘informed consent’ for participation would not be possible from past students (whose data has been used for training the model), or be feasible in covering large scale LA projects [27]; therefore, LA should be considered as “development or improvement of technological resources within an ethical framework” (p. 2862). Kitto and Knight further caution ethics committees, asking them to acknowledge the diversity across applied research disciplines from traditional education research, adding that “informed consent” may not always be possible within LA projects. Hence, we suggest definition of an institutional research ethics protocol that lays out detailed guidelines with respect to their technology deployment strategy that recognizes the purpose of the learner-generated data before leveraging any benefits from learning analytics.

Fortunately, the digital age has broadened everyone’s perspective on how digital footprints left on online public platforms can be leveraged by online agencies (e.g., advertising and marketing agencies). Online data traces can be linked to our persona such as to our social media profile, physical appearance, current location and to other personal interests, which can then be assessed by commercial agencies for their competitive advantage. Learners too are somewhat aware that they leave their digital footprints when interacting with the institutional LMS over the course of their study. However, if an institution intends to use learners’ digital footprints for LA, they hold the responsibility of conveying their intention explicitly to the enrolled learners [48]. That is, they must reiterate to the learners that their online interactions are being recorded in the log files of a LMS; further, that the data from the user generated log files may be used by their institution for analytics. Therefore, as a first step, the research protocol should account for managing regulatory practices to ensure that learner privacy and confidentiality are not compromised when LA approaches are deployed for institutional advantage. In other words, institutions must acknowledge to all enrolled students that their digital footprints (captured via online interactions on LMSs) would serve as proxy data for analyzing their online behaviors. The proxy data would be mined and subsequently analyzed for gathering insights on learner behaviors that would in turn be used for improving overall instructional services. These services include creating models on user behavior, user experience, user profiles, trend analysis or for modeling various learner knowledge domains [8]. The benefits and limitations of these services must be explicitly stated in simple and non-technical language for ease in comprehension by the learners.

Moreover, in the case where historical learner datasets are to be used for developing instructional services, the provision of ‘informed consent’ from students no longer holds since these students are not currently enrolled at the institution. Therefore, institutions must ensure proper research protocols are followed to preserve the privacy and confidentiality of their past learner cohorts. First and foremost, instead of using actual unique student identifiers that can identify past students, the institution should follow a proper data management plan, such as to apply pseudonymization. Pseudonymization differs from anonymization where the “data subject is not or no longer identifiable”, since here the specific data subject cannot be identified without the use of “additional information” [17, 54]. The Article 4(5) adds that “such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person”. In the context of LA implementation, pseudonymization techniques would imply replacing the real identifier of each student with another unique identifier that in no way can be connected to the actual student. Further, this procedure must be conducted in a fool-proof manner by institutionally approved data custodians to protect the re-identification of learners in compliance with Article 25. Hash algorithms (e.g., SHA-256) can be used to convert the unique student identifier into a fixed-length unique value that is used instead. The data stewards are therefore responsible that all personal identifiable information (e.g., name, address, and contact information) are removed and stored separately before using the pseudonymized data for model development. In this manner, the training data used for model development cannot be linked to any particular individual.

The key actionable output of LA systems are interventions. Since the aim is to essentially develop early warning systems that identify students who are at risk of underperformance or discontinuation, and to subsequently activate appropriate interventions in order to avoid these probable outcomes, clear understanding of what constitutes effective interventions must be known. However, the existing research into efficacy of various intervention types based on outputs of predictive models is unclear [30]. Further research needs to be conducted into effective strategies for segmenting different learner types into groups which represent distinctive profiles, which ultimately have different needs and responsiveness to various types of intervention strategies. The instructional services henceforth produced from education data mining techniques will further inform on subsequent intervention strategies. Evidence-based strategies would relate to how current students who have been flagged as being at-risk or those facing learning difficulties are to be supported by tutoring staff in overcoming their learning challenges. However, there is danger of oversimplifying the intervention support strategy as an outcome of the model. We advise caution in simply setting up any intervention strategy and suggest that institutions consider multiple socio-pedagogical approaches for assisting students in overcoming their learning difficulties. Rubel and Jones [40] state that intervention strategies must stay clear of the student's personal choices that are central to their conception of well-being or social tolerance (e.g., religion or politics). Instead, the strategies should be via tailoring of teaching practice or via interventions such as allocating relevant course resources conducive to improving student learning. Fig. 3 gives an overview of the ethical process that has just been described.

Meanwhile, more stringent legislation around data privacy like GDPR, require high levels of openness about operationalized analytics systems, and particularly the ability to explain to affected learners how certain automated decisions were formed, together with a list of all the contributing factors. While not all internals of a predictive model need to be explainable, there does however need to exist a mechanism that retraces prediction outputs for learners on-demand [22]. This is both a technical (Wachter et al., 2017) and a capability challenge which represents an ethical dilemma if predictive models are rushed into deployment without the ability to satisfy these requirements.

Finally, LA is a burgeoning field, and there is a dearth of educational datasets for the emergent researcher community to practice and hone their EDM skills. Another ethical concern faced by educational institutions is related to the sharing of student data with third parties in the current global environment [40]. Educational institutions are legally bound in ensuring privacy of student data thereby limiting reproducibility and replication studies in learning sciences. The advent of MOOCs run by global providers (e.g., Coursera, edX, FutureLearn) offer another view of the emerging learning environments; although sharing of the learner data here too is restricted by strict privacy regulations [16]. Recognizing these restrictions, wherein full anonymity of each individual has to be maintained, many MOOC providers (e.g., HarvardX, Coursera) have released limited non-identifiable data via MORF (MOOC Replication Framework), a platform that allows researchers to deposit anonymized data (e.g., assessment details, grades, time stamps of student interactions, demographic information), that can be used by researchers in controlled environments while maintaining full privacy of student data. Further, researchers adhere to a global level ethics instrument that has been in place by the MOOC provider for responsible use of the anonymized data.

5. Conclusions, limitations and future directions

This paper has provided a much-needed perspective on the challenges encountered in deploying LA systems. Literature-based evidence in response to the first research question – What are the key challenges in effectively deploying LA systems? – has identified three challenges, namely, transferability or generalizability, model transparency, and ethical challenges. The LA movement espouses the premise that computational exploration of learners' historical data could lead to relevant feature extraction and the development of predictive models that can profile currently enrolled students based on their learning needs. This can be further leveraged by the educational institution in facilitating intervention strategies for supporting learner communities in overcoming their learning difficulties. While it is tempting to have a general model solution that can be used across multiple courses and learner cohorts, in practice this is rather difficult to deliver. Moreover, model transparency concerns too need to be addressed for relevance, robustness, fairness, and social acceptance. Finally, the use of educational datasets has additional ethical concerns, such as responsible use of learners' data so that individual learner's privacy and confidentiality are

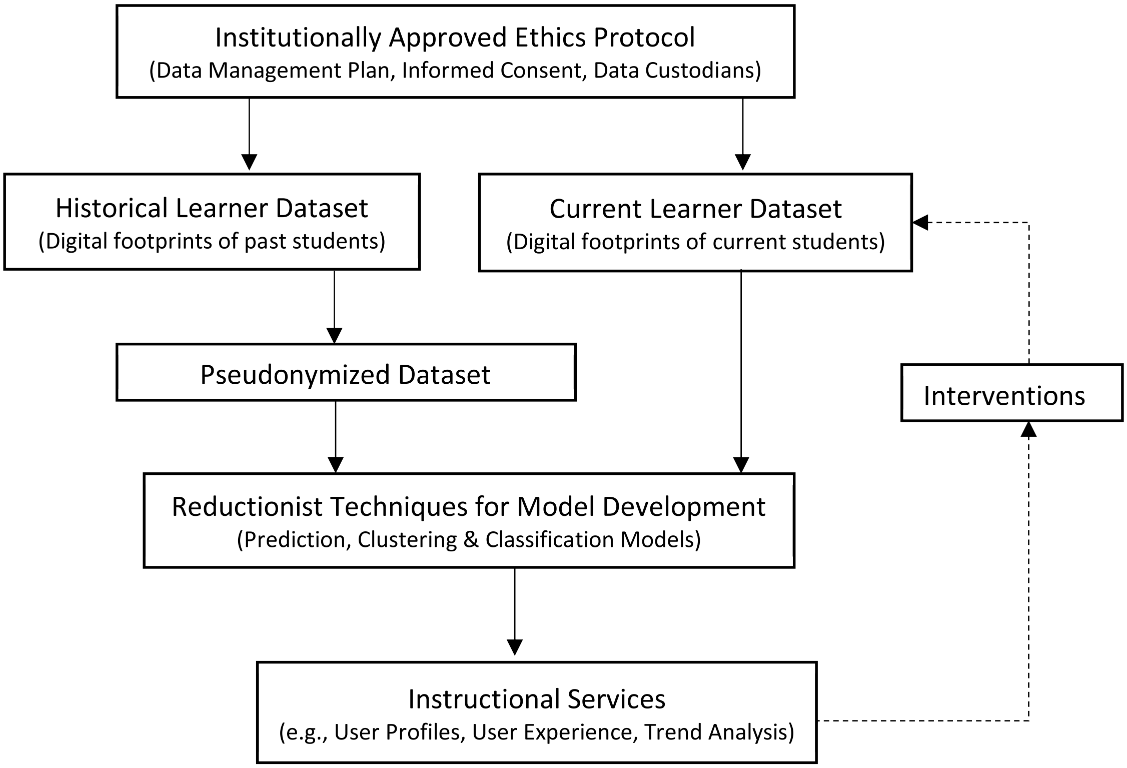
Institutionally Approved Ethics Protocol
(Data Management Plan, Informed Consent, Data Custodians)
Historical Learner Dataset
(Digital footprints of past students)
Current Learner Dataset
(Digital footprints of current students)
Pseudonymized Dataset
Interventions
Reductionist Techniques for Model Development
(Prediction, Clustering & Classification Models)
Instructional Services
(e.g., User Profiles, User Experience, Trend Analysis)

Fig. 3. Research Ethics Protocol.

not compromised.

Discussions pertaining to the second research question – What difficulties are still encountered in producing generalizable predictive models? – have revealed generalizability challenges associated with the dynamic nature of the domain, feature engineering and selection, small dataset sizes, unbalanced datasets, and concept drift amongst others (refer Section 4.1). For accurate transfer learning to take place by the prescribed models, educational institutions must first consider the constraints both in the use of learners' digital footprints and also the learning context. Exploratory analyses to acquire a basic understanding of the data and the learning context must precede any machine learning algorithmic analyses. Learning situations evolve with new course syllabus, changes to assessment structures, different learner cohorts and with diverse pedagogical approaches used by different tutoring staff. These lead to challenges in extracting relevant features that require much analysis and careful selection. However, a strong predictor in one learning scenario may become a weak predictor in another learning scenario. Therefore, we recommend re-evaluating the model design after regular intervals, such as after each successive course offering. Predictive models are driven by hindsight or historical data; it is crucial to ensure that the historical (or training) data aligns well with the target data to avoid generalization degradations brought on by concept drift. The analytics task is not a one-off task that concludes once a model is developed, rather it is an iterative empirical process of trial and error.

The third question – What are the next frontiers in being able to extract more value from predictive models, rather than just predictions? – has revealed gaps related to model transparency by the intended audience (refer Section 4.2). Algorithmic decisions as a consequence of predictive models are not easily interpretable. Hence, to achieve transparency, institutions must convey human-understandable explanations of the logic behind the internals of machine learning algorithms to their learners. In other words, simply informing a student about some algorithmic predictor value (e.g., AUC) as a measure of their learning behavior is by no means adequate; rather, both the high-level model behavior and the explanations of specific predictions must be made available in simple layman terms for non-technical people (i.e., to the currently enrolled students in this case). We have provided an overview of some strategies for explaining the model's reasoning by way of counterfactuals, proxy models and visuals (e.g., decision trees, feature importance plots, individual conditional expectation, etc.) as the next frontiers in extracting more value, rather than merely stating predictions.

The fourth question – Which ethical dilemmas still remain in the deployment and operationalisation of LA systems? – has further divulged ethical predicaments in the usage of learners' digital footprints being harvested within the institutional learning management platform. We acknowledge the crucial role learning analytics can play in transforming educational delivery with better flow of customized instructional services; however, we caution institutions on preserving the privacy and confidentiality of their learners. For any research to be recognized as a scholarly research outcome, all concerns related to its ethical conduct must be addressed first. However, we find that the ethical perspective in the deployment and operationalisation of LA systems is not explicitly stated in literature. We advise the use of an institutional research ethics protocol that clearly outlines the institutional strategy. Most importantly, disclosure statements on the use of learner data must be explicitly communicated to current learners, while historical data used for advancement/refinement of the model must be pseudonymized and all additional data that can lead to re-identification kept safe with authorized data custodians. The created model(s) is proprietary to the concerned institution; hence it is not required that institutions disclose their technical practices (e.g., ensemble of algorithms used).

This concept paper takes a problem-centered approach with the main purpose of "developing logical and complete arguments for associations rather than testing them empirically" ([19], p. 127); hence, no experimental design or empirical data has been provided. Further, it does not cover analytic technicalities, such as choosing the right machine learning algorithm or tuning of the machine learning algorithms. Concept papers are meant to provide a tightly focused literature overview, since their objective is to put forth a bridge between validation and usefulness of constructs within some identified domain. Gilson and Goldberg advise the use of figures to clearly depict authors' views on how these constructs are related. Figs. 1, 2 and 3 showcase constructs for establishing and managing data-driven approaches related to the generalizability, model transparency, and ethical domains. While tracking and measuring learner performance can make education providers more aware of their instructional services, we encourage policy makers and institutional authorities to consider these constructs and question themselves on how their analytics approach is suitable, meaningful, and justifiable.

Declaration of Competing Interest

We have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

References

[1] Adadi A, Berrada M. Peeking Inside the Black-Box: a Survey on Explainable Artificial Intelligence (XAI). IEEE Access 2018;6:52138–60. https://doi.org/10.1109/ACCESS.2018.2870052.

[2] Aldowah H, Al-Samarraie H, Fauzy WM. Educational data mining and learning analytics for 21st century higher education: a review and synthesis. Telematics Inf 2019;37:13–49. https://doi.org/10.1016/j.tele.2019.01.007.

[3] Alonso JM, Casalino G. Explainable artificial intelligence for human-centric data analysis in virtual learning environments. 6-7 June 2019. Novedrate, Italy: Paper presented at the International Workshop on Higher Education Learning Methodologies and Technologies Online; 2019.

[4] Angelov PP, Soares EA, Jiang R, Arnold NI, Atkinson PM. Explainable artificial intelligence: an analytical review. WIREs Data Mining Knowl Discov 2021;11(5): e1424. https://doi.org/10.1002/widm.1424.

[5] Baesens B, Ravi B, Marsden JR, Vanthienen J, Zhao JL. Transformational issues of big data and analytics in networked business. MIS Q 2016;40(4):807–18. https://doi.org/10.25300/MISQ/2016/40:4.03.

[6] Baker RS. Challenges for the Future of Educational Data Mining: the Baker learning analytics prizes. Journal of Educational Data Mining 2019;11(1b):1–17. https://doi.org/10.5281/zenodo.3554745.

[7] Bellman R. Adaptive control processes: a guided tour. princeton legacy library. Princeton, NJ: Princeton University Press; 1961.

[8] Bienkowski MA, Feng M, Means B. Enhancing teaching and learning through educational data mining and learning analytics: an issue brief. Center for technology in learning. Washington, DC: U.S.: SRI International; 2012 (ED-04-CO-0040, https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf, Task 0010).

[9] Boyer, S., & Veeramachaneni, K. (2015, 2015//). Transfer learning for predictive models in massive open online courses. Paper presented at the Artificial Intelligence in Education, Cham.

[10] Craven M, Shavlik J. Extracting Tree-Structured Representations of Trained Networks. Adv Neural Inf Process Syst 1996;8:24030. https://proceedings.neurips.cc/paper/1995/file/45f31d16b1058d586fc3be7207b58053-Paper.pdf.

[11] Ding, M., Wang, Y., Hemberg, E., & O'Reilly, U. (2019). Transfer Learning using Representation Learning in Massive Open Online Courses. Paper presented at the 9th International Conference on Learning Analytics & Knowledge, Madrid, Spain. https://doi.org/10.1145/3303772.3303794.

[12] Domingos P. A few useful things to know about machine learning. Commun ACM 2012;55(10):78–87. https://doi.org/10.1145/2347736.2347755.

[13] Dosilovic FK, Br j M, Hlupic N. Explainable artificial intelligence: a survey. In: The 41st International Convention on Information and Communication Technology, Electronics and Microelectronics. MIPRO; 2018. p. 0210–5.

[14] Elrahman SMA, Abraham A. A Review of Class Imbalance Problem. Journal of Network and Innovative Computing 2013;1:332–40. https://www.mirlabs.net/jnic/secured/Volume1-Issue1/Paper31/JNIC_Paper31.pdf.

[15] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann. Statist. 2001;29(5):1189–232. https://doi.org/10.1214/aos/1013203451.

[16] Gardner J, Brooks C, Andres JM, Baker RS. MORF: a Framework for Predictive Modeling and Replication At Scale With Privacy-Restricted MOOC Data. Paper presented at. In: the 2018 IEEE International Conference on Big Data. Big Data; 2018. https://doi.org/10.1109/BigData.2018.8621874.

[17] GDPR. Regulation (EU) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (No. 02016R0679 - EN - 04.05.2016 - 000.002). Portal of the Publications Office of the EU; 2016. Retrieved from http://data.europa.eu/eli/reg/2016/679/2016-05-04 Retrieved from http://data.europa.eu/eli/reg/2016/679/2016-05-04.

[18] Ge L, Gao J, Ngo H, Li K, Zhang A. On handling negative transfer and imbalanced distributions in multiple source transfer learning. Statistical Analysis and Data Mining: The ASA Data Science Journal 2014;7(4):254–71. https://doi.org/10.1002/sam.11217.

[19] Gilson LL, Goldberg CB. Editors' Comment: so, What Is a Conceptual Paper? Group & Organization Management 2015;40(2):127–30. https://doi.org/10.1177/1059601115576425.

[20] Ginda M, Richey MC, Cousino M, Börner K. Visualizing learner engagement, performance, and trajectories to evaluate and optimize online course design. PLoS ONE 2019;14(5):e0215964. https://doi.org/10.1371/journal.pone.0215964.

[21] Goldstein A, Kapelner A, Bleich J, Pitkin E. Peeking Inside the Black Box: visualizing Statistical Learning With Plots of Individual Conditional Expectation. J Comput Graph Statist 2015;24(1):44–65. https://doi.org/10.1080/10618600.2014.907095.

[22] Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? ArXiv preprint 2017. https://arxiv.org/abs/1712.09923, (https://arxiv.org/abs/1712.09923).

[23] Hunt XJ, Kabul IK, Silva J. Transfer learning for education data. 13–17 August 2017. Halifax, NS, Canada: KDD Workshop; 2017. Paper presented at the, http://ml4ed.cc/attachments/HuntTransfer.pdf.

[24] Jaakkola E. Designing conceptual articles: four approaches. AMS Review 2020;10(1):18–26. https://doi.org/10.1007/s13162-020-00161-0.

[25] Jones, K., & Salo, D. (2017). Learning analytics and the academic library: professional ethics commitments at a crossroads (No. id 2955779). Rochester, NY. Retrieved from https://ssrn.com/abstract=2955779 Retrieved from https://ssrn.com/abstract=2955779.

[26] Kaur H, Pannu HS, Malhi AK. A Systematic Review on Imbalanced Data Challenges in Machine Learning: applications and Solutions. ACM Comput Surv 2019;52(4):79. https://doi.org/10.1145/3343440. Article.

[27] Kitto K, Knight S. Practical ethics for building learning analytics. British Journal of Educational Technology 2019;50(6):2855–70. https://doi.org/10.1111/bjet.12868.

[28] L'Heureux A, Grolinger K, ElYamany HF, Capretz MAM. Machine Learning With Big Data: challenges and Approaches. IEEE Access 2017;5:7776–97. https://doi.org/10.1109/ACCESS.2017.2696365.

[29] Lakkaraju H, Kamar E, Caruana R, leskovec J. Interpretable & Explorable Approximations of Black Box Models. ArXiv_preprint 2017. https://arxiv.org/abs/1707.01154.

[30] Larrabee Sønderlund A, Hughes E, Smith J. The efficacy of learning analytics interventions in higher education: a systematic review. British Journal of Educational Technology 2019;50(5):2594–618. https://doi.org/10.1111/bjet.12720.

[31] Leitner P, Ebner M, Ebner M. Learning Analytics Challenges to Overcome in Higher Education Institutions. In: Ifenthaler D, Mah D-K, Yau JY-K, editors. Utilizing learning analytics to support study success. Cham: Springer International Publishing; 2019. p. 91–104. https://doi.org/10.1007/978-3-319-64792-0_6. https://doi.org/10.1007/978-3-319-64792-0_6.

[32] López-Zambrano J, Lara JA, Romero C. Towards Portability of Models for Predicting Students' Final Performance in University Courses Starting from Moodle Logs. Appl Sci 2020;10(1):354. https://doi.org/10.3390/app10010354.

[33] Lu J, Liu A, Dong F, Gu F, Gama J, Zhang G. Learning under Concept Drift: a Review. IEEE Trans Knowl Data Eng 2018;31(12):2346–63. https://doi.org/10.1109/TKDE.2018.2876857.

[34] Lundberg S, Lee SI. A unified approach to interpreting model predictions. ArXiv preprint 2017. https://arxiv.org/pdf/1705.07874.pdf, (https://arxiv.org/pdf/1705.07874.pdf).

[35] Moubayed A, Injadat M, Nassif AB, Lutfiyya H, Shami A. E-Learning: challenges and Research Opportunities Using Machine Learning & Data Analytics. IEEE Access 2018;6:39117–38. https://doi.org/10.1109/ACCESS.2018.2851790.

[36] Putnam V, Conati C. Exploring the need for explainable artificial intelligence (XAI) in intelligent tutoring systems (ITS). IUI Workshops; 2019. Paper presented at the.

[37] Ribeiro, M.T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": explaining the Predictions of Any Classifier. Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA. https://doi.org/10.1145/2939672.2939778.

[38] Ribeiro, M.T., Singh, S., & Guestrin, C. (2018). Anchors: high-Precision Model-Agnostic Explanations. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1) Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/11491.

[39] Romero C, Ventura S. Educational data mining and learning analytics: an updated survey. WIREs Data Mining and Knowledge Discovery 2020;10(3):e1355. https://doi.org/10.1002/widm.1355.

[40] Rubel A, Jones KML. Student privacy in learning analytics: an information ethics perspective. Inf Soc 2016;32(2):143–59. https://doi.org/10.1080/01972243.2016.1130502.

[41] Sagi O, Rokach L. Ensemble learning: a survey. WIREs Data Mining Knowl Discov 2018;8(4):e1249. https://doi.org/10.1002/widm.1249.

[42] Sclater N, Mullan JS. Learning analytics and student success – assessing the evidence. JISC 2017. Retrieved 5/01/2021 Retrieved from https://repository.jisc.ac.uk/id/eprint/6560 from https://repository.jisc.ac.uk/id/eprint/6560.

[43] Sclater N, Peasgood A, Mullan JS. Learning analytics in higher education: a review of UK and international practice. VOCEDplus 2016. Retrieved from www.jisc.ac.uk/reports/learning-analytics-in-higher-education, 39. www.jisc.ac.uk/reports/learning-analytics-in-higher-education.

[44] Seah CW, Ong YS, Tsang IW. Combating negative transfer from predictive distribution differences. IEEE Trans Cybern 2013;43:1153–65. https://doi.org/10.1109/TSMCB.2012.2225102.

[45] Shimodaira H. Improving predictive inference under covariate shift by weighting the log-likelihood function. J Stat Plan Inference 2000:227–44.

[46] Siemens G. Learning Analytics: the Emergence of a Discipline. Am Behav Sci 2013;57(10):1380–400. https://doi.org/10.1177/0002764213498851.

[47] Siemens G, Baker RSJd. Learning analytics and educational data mining: towards communication and collaboration. Paper presented at the. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. British Columbia, Canada: Vancouver; 2012. https://doi.org/10.1145/2330601.2330661.

[48] Slade S, Prinsloo P. Learning Analytics: ethical Issues and Dilemmas. Am Behav Sci 2013;57(10):1510–29. https://doi.org/10.1177/0002764213479366.

[49] Surmelioglu Y, Seferoglu SS. An examination of digital footprint awareness and digital experiences of higher education students. World J Educ Technol 2019;11(1):48–64. https://eric.ed.gov/?id=EJ1205431.

[50] Tempelaar DT, Rienties B, Giesbers B. In search for the most informative data for feedback generation: learning analytics in a data-rich context. Comput Human Behav 2015;47:157–67. https://doi.org/10.1016/j.chb.2014.05.038.

[51] Tene O, Polonetsky J. Big data for all: privacy and user control in the age of analytics. Northwestern J Technol Intellectual Property 2013;11(5):239–73. https://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1.

[52] Wachter S, Mittelstadt B, Russell C. Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harv J Law Technol 2018;31(2):841–87. https://doi.org/10.2139/ssrn.3063289. https://ssrn.com/abstract=3063289 or https://doi.org/.

[53] Weick KE. Theory Construction as Disciplined Imagination. Acad Manag Rev 1989;14(4):516–31. https://doi.org/10.2307/258556.

[54] Wes M. Looking to comply with GDPR? Here's a primer on anonymization and pseudonymization. The Privacy Advisor 2017. Retrieved from https://iapp.org/news/a/looking-to-comply-with-gdpr-heres-a-primer-on-anonymization-and-pseudonymization/.

[55] Witten IH, Frank E, Trigg L, Hall M, Holmes G, Cunningham SJ. Weka: practical machine learning tools and techniques with java implementations (No. working paper 99/11). Hamilton, New Zealand: University of Waikato, Department of Computer Science. Retrieved from; 1999. https://hdl.handle.net/10289/1040 Retrieved from https://hdl.handle.net/10289/1040.

[56] Wong B, Li K, Choi SPM. Trends in learning analytics practices: a review of higher education institutions. Interact. Technol. Smart Educ. 2018;15:132–54. https://doi.org/10.1108/ITSE-12-2017-0065.

[57] Zhou L, Pan S, Wang J, Vasilakos AV. Machine learning on big data: opportunities and challenges. Neurocomputing 2017;237:350–61. https://doi.org/10.1016/j.neucom.2017.01.026.

[58] Ramaswami, G., Susnjak, T., Mathrani, A., Lim, J., & Garcia, P. (2019). Using educational data mining techniques to increase the prediction accuracy of student academic performance. *Information and Learning Sciences, 120*(7/8), 451-467. https://doi.org/10.1108/ILS-03-2019-0017.

[59] Umer, R., Mathrani, A., Susnjak, T., & Lim, S. (2019, 30, March - 1 April 2019). *Mining Activity Log Data to Predict Student's Outcome in a Course.* Paper presented at the International Conference of Big Data and Education, London, UK. https://doi.org/10.1145/3322134.3322140.