British Journal of Educational Technology Vol 0 No 0 2019 1–24
doi:10.1111/bjet.12836

The potential for student performance prediction in small cohorts with minimal available attributes

Edward Wakelam, **Amanda Jefferies**, **Neil Davey and Yi Sun**

Edward Wakelam is a PhD student and lecturer at the University of Hertfordshire researching the application of data mining techniques to learning analytics. Amanda Jefferies is professor of Technology-Enhanced Learning at the University of Hertfordshire. Neil Davey is a principal lecturer and research fellow at the University of Hertfordshire in the area of applied machine learning. Yi Sun is a senior research fellow at the University of Hertfordshire in the areas of applied machine learning and data visualisation. Address for correspondence: Edward Wakelam, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK. Email: e.wakelam@herts.ac.uk

Abstract

The measurement of student performance during their progress through university study provides academic leadership with critical information on each student's likelihood of success. Academics have traditionally used their interactions with individual students through class activities and interim assessments to identify those "at risk" of failure/withdrawal. However, modern university environments, offering easy on-line availability of course material, may see reduced lecture/tutorial attendance, making such identification more challenging. Modern data mining and machine learning techniques provide increasingly accurate predictions of student examination assessment marks, although these approaches have focussed upon large student populations and wide ranges of data attributes per student. However, many university modules comprise relatively small student cohorts, with institutional protocols limiting the student attributes available for analysis. It appears that very little research attention has been devoted to this area of analysis and prediction. We describe an experiment conducted on a final-year university module student cohort of 23, where individual student data are limited to lecture/tutorial attendance, virtual learning environment accesses and intermediate assessments. We found potential for predicting individual student interim and final assessment marks in small student cohorts with very limited attributes and that these predictions could be useful to support module leaders in identifying students potentially "at risk."

Introduction and motivation for study

An ability to predict individual student performance at appropriate points during a module, in particular their likely intermediate and final assessment marks, may provide module leadership with useful guidance on which individuals are "at risk" of failure or withdrawal. This information may then give lecturers and tutors an opportunity to make timely supportive interventions designed to increase the student's likelihood of success. The identification of students "at risk" of failure or withdrawal has become increasingly important to academics, tutors, support staff and institutions, for a variety of reasons. For the students themselves, the failure to achieve

Practitioner Notes

What is already known about this topic

- Learning analytics is a powerful tool in analysing student progress and predicting student outcomes.
- The majority of learning analytics research has focussed upon large data sets comprising of large student cohorts with a significant number of student attributes.
- The analysis of learning analytics data provides course leadership with the ability to identify students "at risk" of failure or withdrawal to allow the opportunity to make positive and timely interventions.
- The aggregation of learning analytics allows course leadership to potentially identify opportunities to improve course presentation and execution.

What this paper adds

Exploration of the potential for predicting student performance in small student cohorts where student data are limited by availability and/or institutional regulation.

- There is some potential for predicting student performance where the student cohort is small and student data are limited to attendance, virtual learning environment accesses and interim assessments. Prediction accuracy is similar to that achieved with large data sets.
- The analyses performed supported module leadership in identifying the need for timely student interventions.
- Random Forest and K-Nearest Neighbours machine learning techniques produced the most accurate prediction results.

Implications for practice and/or policy

Learning analytics can provide institutions with useful supporting data in small student cohort settings, where the availability of individual student data is restricted.

- Machine learning analyses can be provided alongside traditional institutional student performance measures currently made available to module leadership.
- Institutional restrictions often placed on student data availability and privacy are not necessarily a barrier to the deployment of learning analytics.
- The adoption of these methods requires appropriate additions to documented institutional intervention policy.

their potential is a waste, as are the consequent limitations on their future career development. Sometimes worse is the personal stress and trauma they consequently face, alongside the financial impact and the potential consequential effect on their families. In the UK for example, in academic year 2015/16, 6.4% of UK domiciled full-time entrants did not continue in their studies after their first year (Higher Education Statistics Agency, 2018). In Australia and the US, these figures are worse with attrition rates of over 21% (Australian Government Department of Education and Training, 2016) and over 25% (National Center for Education Statistics, 2017). For institutions, the financial impacts can be very significant, compounded by the consequential effects of published statistical measures of student drop-out rates and student satisfaction scores. Universities operate a sliding scale of refund levels to be applied should a student leave the course. In the case of the author's own university, the cost of refunds of full time UK and EU undergraduate student withdrawals can be as high as £27,750, and over 20% higher for non UK/EU students. This is based upon the recognition that in the vast majority of cases the university place cannot be filled by a suitable replacement and is based upon current annual fees of £9,250. Universities operate in a very competitive environment, and pay considerable attention to their place in league tables and how they may improve their position. The student satisfaction score is an integral part of each institution's overall score and is therefore an area of focus for university management and policies. In the modern HE/University system student non-attendance at lectures and tutorials remains high (Marburger, 2001; Mearman, Pacheco, Webber, Ivlevs, & Rahman, 2014) as course material has increasingly become available on-line and accessible to students 24/7. This reduction in face time between educators and students makes it increasingly difficult for tutors to identify students "at risk" who are struggling with the material or failing to engage. It has always been the case that students are able to request additional lecture/tutorial and face to face time with their tutors.

The application of machine learning techniques to predict student outcomes has made significant progress in recent years (Ashraf, Anwer, & Khan, 2018), providing academic leadership with useful information upon which to consider positive and timely interventions. These techniques have been applied in academic environments where so called "big data" is available (Daniel, 2015). We understand big data to be large student populations and a wide selection of data points (attributes) per student. For example, in the case of the OU, the machine learning analysis operates upon over 32,000 students and 27 attributes per student. However, many university courses/modules are comprised of relatively small student cohorts, often less than 30. A recent study, based upon 67 UK universities, found average class sizes of approximately 20 students (Huxley, Mayo, Peacey, & Richardson, 2018). In addition, academic institutions have legal and ethical obligations to maintain the privacy of individual student data (Corrin *et al.*, 2019), and therefore restrict its availability to prediction algorithms. Opportunities to make useful predictions based upon relatively small student cohorts combined with limited student attributes could provide educators with the capability to identify students "at risk" and make timely supportive interventions.

Research questions and problem formulation

Our study focusses upon two research questions:

Small student cohorts and limited student attributes

Is it possible and useful to predict student performance on courses comprising of relatively small student cohorts, where a very limited set of student attributes are readily available for analysis?

While there is evidence to show that predictions based upon large cohorts can provide educators with useful support in identifying students "at risk" (Heuer & Breiter, 2018), there is little evidence of the value that can be derived where cohorts are small, in this case 23 students. Given that most institutions have a significant number of courses which comprise of these smaller cohorts (Huxley *et al.*, 2018), more research could prove of value. In this case, the data are limited to lecture/tutorial attendance, virtual learning environment (VLE) accesses and five formal interim assessments. This question is critical given institutional protocols and concerns regarding the privacy of student data (see Section "The opportunity to make interventions"), coupled with the ethics of analysing and then taking subsequent action from the results. It is also the case that universities are confused as to whether in providing this data students are in fact giving prior (and legally supportable) approval for their inclusion in learning analytics (LA), and furthermore whether this entitles the institutions to categorise students and to be the catalyst/basis for interventions (Sclater & Bailey, 2015). This is equally true in the case of analytics based upon large student cohorts. Our research question addresses the combination of both small student cohorts and limited attributes.

The opportunity to make interventions

How useful would these analyses be in order to provide course leaders with the opportunity to make timely supportive interventions at appropriate points during the module? In this example, there is a relatively even spread of formal assessments throughout the duration of the course, including two at an early stage.

Experiment design

We apply and compare three machine learning techniques, Decision Tree (DT), K-Nearest Neighbours (KNN) and Random Forest (RF) analyses to analyse and predict student performance, applied at appropriate points during module delivery. These points were selected to coincide with intermediate assessments. DT, KNN and RF methods were selected given their ability to perform well when some values are missing (Quinlan, 2014) and their widespread core use in LA research (Ashraf *et al*, 2018). Given that our experiment is designed to analyse student performance breakdown, missing values may be expected. In the case of our experiment, missing values occur where a student chooses not to take part in an interim assessment. For example, only the highest two of the three multiple choice assessments (see Table 5) count towards the student's final mark and in some cases students who scored highly in the first two of these assessments chose to not sit the third. After module completion, we applied RF analysis retrospectively at each intermediate assessment point to make overall module score predictions and evaluate their accuracy.

Literature review

We have structured our literature review to address each of the two research questions in turn.

Small student cohorts and limited student attributes

An inability to identify and consequently successfully support students "at risk" of failure or withdrawal presents two serious threats to universities. Firstly, the consequences of already budgeted student fees disappearing from university revenues are significant as can be seen by the percentages of student withdrawals. For example, the UK Higher Education Statistics Agency (HESA, 2018) performance indicators show that the percentage of full-time students not continuing after one year of study who started in 2015/16 was 6.4%. In the case of part-time students, the figure was 34.2%. In the case of American University students, Lin, Yu and Chen (Lin, Yu, & Chen, 2012) noted that predicted retention probability decreases from around 70% for a representative full-time student to 57% for a part-time student. In the case of open, distance environments retention and progression has been established to be a greater issue than for traditional full-time campus-based students Simpson (2006, 2013). Secondly, student satisfaction scores are an integral part of the scoring mechanism that determines a university's place in national and global rankings. The impact of these scores on rankings has been shown to be greater for more able students, for universities with entry standards in the upper-middle tier, and for subject departments facing more competition from other universities (Gibbons, Neumayer, & Perkins, 2015).

The prediction of student outcomes is a core component of the field of LA. LA is defined as " the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Ferguson, 2012). The overwhelming focus on LA in higher education has been devoted to the analysis of "big data" (Ashraf *et al.*, 2018) where the data comprises very large student cohorts and a large number of student data items.

In our experiment, we are interested in the application of LA for the prediction of intermediate and final student assessment marks, where the student cohort is small and with limited attributes. In order to provide ourselves with appropriate benchmarks for comparison we now discuss published comparative prediction accuracies across a variety of techniques, applied to large student cohorts with multiple student attributes.

In the case of large data sets, a variety of student attributes are used in the analyses summarised, including personal and admission data as well as previous educational records (Table 1) cited from Ashraf *et al.*, 2018.

A comparison of various data mining techniques (Ashraf *et al*, 2018) to predict student module marks using regression methods demonstrates achieved student prediction accuracy levels ranging from 50% to 97%. Accuracy is measured as the percentage accuracy of the prediction versus the actual student result. Accuracy levels are shown by algorithm (Table 2) and by summary attributes and algorithm (Table 3) cited from Ashraf *et al.*, 2018. These analyses included student numbers in excess of 10,000 and 77 attributes in some cases.

In their analysis of LA and interventions publications between 2007 and 2018, Wong and Li selected 23 case studies highlighting the measured benefits of LA in distance learning institutions (Wong & Li, 2018).

Table 1: Student attributes

Criteria	*Details*
Student demographic information	Age, gender, region, residence, guardian info
Previous results	Cleared certificates, scholarships and results
Grades	Recent assignment results, quizzes, final exam, CGPA, attendance
Social network details	Interaction with social media websites
Extra-curricular activities	Games partitions, sports, hobbies
Psychometric factor	Behaviour, absence, remarks

Table 2: Prediction accuracy by algorithm

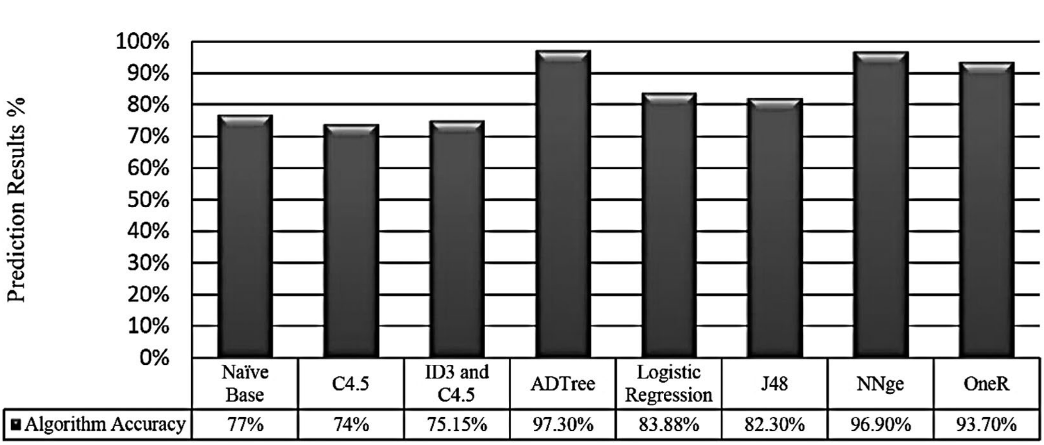
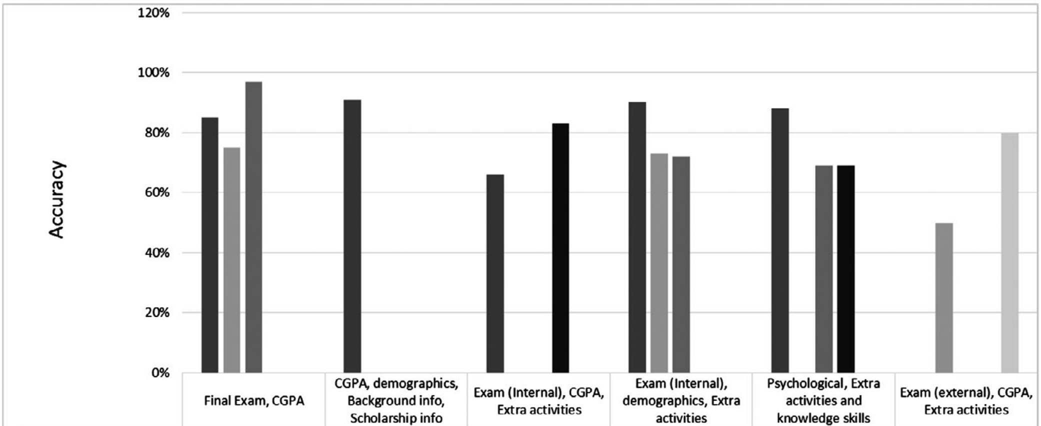


Table 3: Prediction accuracy by summary attributes and algorithm



	Final Exam, CGPA	CGPA, demographics, Background info, Scholarship info	Exam (Internal), CGPA, Extra activities	Exam (Internal), demographics, Extra activities	Psychological, Extra activities and knowledge skills	Exam (external), CGPA, Extra activities
Decision Tree	85%	91%	66%	90%	88%	
Naïve Bayes	75%			73%		50%
Neural Network	97%			72%	69%	
K-Nearest			83%		69%	
Support Vector Machine						80%

There is some evidence that interim assessment as part of the overall course assessment is a strong predictor of student success (Sclater, Peasgood, & Mullan, 2016). Case studies included in this report also identify a student's VLE accesses as a better predictor of success than their historical or demographic data. As with the majority of research conducted, these case studies measured very positive impacts from resulting interventions. A recent study (Heuer & Breiter, 2018) analysing student VLE activity across 22 courses and 32,593 OU students found student VLE accesses to be an important indicator of student performance. An experiment conducted on 200 students over a two-year period at Manchester Metropolitan University made extensive use of VLE usage to determine how to improve the design of learning environments (Stubbs, Martin, & Endlar, 2006).

The opportunity to make interventions

The objective of LA in this instance is to offer tutors the opportunity to identify and support the need to make timely interventions where a student's success is potentially "at risk." The LA cycle is shown in Figure 1 below (Ferguson & Clow, 2017).

In the UK the Open University (OU) is a world leader in the collection, intelligent analysis and use of large scale student analytics. It provides academic staff with systematic and high quality actionable analytics for student, academic and institutional benefit (Rienties, Nguyen, Holmes, Reedy, 2017). Rienties and Toetenel's, 2016 study (Rienties & Toetenel, 2016) identifies the importance of the linkage between LA outcomes, student satisfaction, retention and module learning design. These analytics are often provided through dashboards tailored for each of academics and students (Schwendimann *et al.*, 2017).

The OU's world-class Analytics4Action initiative (Rienties, Boroowa, Cross, Farrington-Flint *et al.*, 2016) supports the university-wide approach to LA. In particular, the initiative provided valuable insights into the identification of students and modules where interventions would be beneficial, analysing over 90 large-scale modules over a two-year period. Analytics4Action identifies

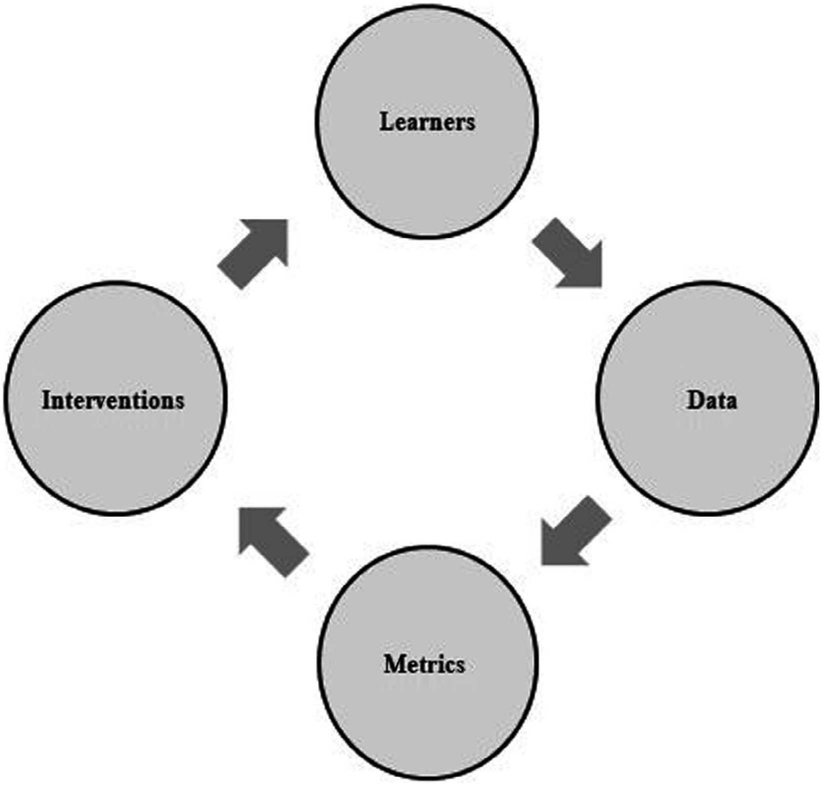


Figure 1: The learning analytics cycle

six phases for teachers and institutions to follow to successfully convert LA outcomes into actionable and impact-measurable interventions (Rienties, Boroowa, Cross, Kubiak *et al.*, 2016). The deployment of LA establishes the need and opportunity for student and module interventions (Clow, 2012). The study concludes that the faster the feedback loop to students, the more effective the outcomes. This is often an iterative process allowing institutions to understand and address systematic issues. Choi and colleagues (Choi, Lam, Li, & Wong, 2018) summarise the pros and cons of alternative intervention methods (Table 4), their study highlighting the benefits to staff faced with limited time and resources.

It is important to note that LA also provide institutions with the opportunity to address systematic issues with individual modules, as referenced above "module interventions" (Clow, 2012). Legal, ethical and moral considerations in the deployment of LA and interventions are key challenges to institutions. They include informed consent, transparency to students, the right to challenge the accuracy of data and resulting analyses and prior consent to intervention processes and their execution (Slade & Tait, 2019). These are well documented in a number of research papers including Pardo and Siemens (2014) and de Freitas *et al.* (2015). In addition, a comprehensive literature review of 86 publications was commissioned by Jisc (formerly the Joint Information Systems Committee, who provide UK universities and colleges with shared digital infrastructure and services including LA), to discuss the challenges faced by institutions and provide the background for a future code of practice (Sclater & Bailey, 2015). A discussion on ethical and data privacy issues in LA based on three studies in higher education and primary school contexts

Table 4: Pros and cons for the commonly-used intervention methods

Method	*Pros*	*Cons*
Email	• Least expensive • Allows personalisation via mail merge	• Students may easily overlook the message due to too many spam emails
Phone call	• Good for emergency matters – two-way synchronous communications	• Students may not be available and sometimes feel offended
Instant messaging	• Preferred communication channel for many students	• More costly than email as it requires one-to-one communications
LMS post & news	• Facilitates many-to-many asynchronous communications	• Requires students to login to the LMS and may overlook the posts and news
Group consultation	• Effective communication • Good for timid students	• Usually needs making appointments in advance and expensive for instructors
Face-to-face consultation	• Effective communication • One-to-one consultation	• Most expensive and usually needs to make appointments in advance
Video recording	• Effective instruction • Not restricted by time	• Substantial initial effort to record the instructions
Peer review	• Encourages critical evaluation • Students can learn from each other	• Requires good question design • Often conducted in class
E-tutorial	• Supplementary instructions available 24/7 (e.g. MyMathLab and MyStatLab developed by Pearson Publishing) • Suitable for highly motivated students	• May incur a price for students or instructors

(Rodríguez-Triana, Martínez-Monés, & Villagrá-Sobrino, 2016), specifically focusses on tutor-led approaches. Legislation has been in place for over two decades, specifically the European Data Protection Directive 1995 and the UK Data Protection Act 1998. More recently, General Data Protection Regulation (Guide to the General Data Protection Regulation, 2018) sets out the legal data protection principles which institutions and organisations are responsible for adhering to. In addition, despite their algorithmic accuracy intentions, there is growing research into the potential for machine learning approaches to introduce bias, such as class, gender and ethnicity (Wilson *et al.*, 2017).

Module description

The selected course instance is a Level 6 (Final Year undergraduate) Computer Science module, duration 15 weeks (including a 3 week vacation period and 2 weeks allocated for submission and review of each of the two final assessments) comprising five intermediate summative assessments and no final examination. Each week students are expected to attend a two-hour lecture and one-hour tutorial. During the course of the module, there are 10 lectures and 9 tutorials. Three EVS (Electronic Voting System) in-class tests are included, with the best two results counting towards the final overall module assessment (see Table 5). The module has a profile of early

Table 5: Module assessments

Week no.	*Name*	*Description*	*Number of weeks to complete assessment*	*Submit on week no.*	*Result publication week no.*	*Percentage contribution to final result*
1	EVS1	Multiple choice	Immediate	4	4	5%
2	EVS2	Multiple choice	Immediate	6	6	5%
3	EVS3	Multiple choice	Immediate	10	10	5%
4	Group Presentation	Group work and presentation	6	11	12	40%
5	Individual Report	Technical Report	8	15	18	50%

"low stakes" assessments with "higher stakes" assessments later in the module. The module VLE comprises of eight sections, including the course guide for example, however student focus was overwhelmingly on the News and Teaching sections.

Note that only the highest two scores of the three EVS results contribute to the final result.

Dataset description

The student cohort is 23. For each student the attributes collected comprise attendance at lectures/tutorials, VLE accesses and intermediate assessment results spread throughout the module (Table 6). Ethics approval limited analysis to dynamic data collected during course execution. Static attributes such as gender, age, prior academic results were not included.

Methodology

We applied three machine learning techniques, DT (regression), KNN and RF to predict student assessment marks, using only their attendance, VLE accesses, and intermediate summative assessments results. The aim of these techniques is to create a model that takes these input values to predict the value of a target variable, in this case the students' assessment marks.

Summary of Machine learning techniques

Decision Tree
DTs are a tree-like model of successive decisions, where each leaf in the tree is a decision, with its corresponding probability, followed by a consequential branch leading to the next leaf, ultimately leading to a prediction (Horning, 2013).

K-Nearest Neighbours
KNN iteratively searches for the most similar (nearest) data points in a given data set, allowing classification of the target data point and consequently prediction (Zhang, 2016).

Random Forest
RF analysis is an ensemble prediction method which uses multiple DTs and averaging their individual results in order to predict a target variable (Horning, 2013).

Design of experiments to meet research questions

Commencing at module registration, each student's attendance at lectures and tutorials was recorded, both as a simple count and as a percentage of overall module tutorials/lectures to date. As well as cumulative attendance, we recorded the delta increases between the measurement points, which were selected to coincide with intermediate assessments. A continuous count of

Table 6: Student attributes

Attribute	*Data range*
Lecture/tutorial attendance	1–19
Delta increase in attendance from prior period	1%–100%
Cumulative VLE News section accesses	0–unlimited
Cumulative VLE Teaching section accesses	0–unlimited
Cumulative VLE accesses	0–unlimited
EVS1 result	0%–100%
EVS2 result	0%–100%
EVS3 result	0%–100%
Group presentation result	0%–100%
Individual report result	0%–100%

individual student "accesses" on items in the VLE was maintained. Of the 8 sections of the VLE, 99% of student accesses were in only 2 sections, News and Teaching. The News section included all module announcements and weekly reminders of tasks to complete. The Teaching section included all course material. For the purposes of the experiment we included each of these two section accesses in our analyses. Intermediate and final assessment results were recorded for each student. This resulted in the data set shown in Table 6. For each analysis point, each of DT, KNN and RF analyses were carried out and the resultant predictions compared with actual student results and the level of accuracy measured. These analyses included the overall module result at module completion. Regression methods were selected to enable the prediction of an actual assessment mark, as opposed to classification methods which would simply predict a pass or fail. This data mining method is often used in the construction of predictive models (Daniel, 2015). The measurement methods used were percentage relative error/accuracy, Mean Squared Error (MSE) and Correlation Coefficient (CC). Prediction accuracies between the analysis methods were compared. We then repeated our analyses combining the two VLE section accesses (see Table 6) into one total in order to determine sensitivity. The progressive prediction results at each assessment point were shared with the module leader for consideration of potential interventions during module delivery. To provide module leadership with data which could potentially support their choice of intervention approach, tabular and graphical comparative analyses of attendance, VLE accesses and intermediate assessment results were also provided. Additionally, we repeated the prediction analyses at each assessment point, based upon the assessment results data alone, excluding attendance and VLE "accesses" in order to compare results. Upon availability of the overall module result after module completion, we were able to revisit our collected data at each assessment point and perform overall module result prediction analyses at each point. We selected RF for these analyses given that it delivered the most accurate predictions in our earlier analyses. Upon module completion, the correlation between all assessments, including overall module results was investigated.

Performance measurement

Percentage relative accuracy is measured as the percentage accuracy of the prediction compared to the actual student result. This permitted a direct comparison with the measurement method used by Ashraf *et al.*, 2018 which compared the results of various data mining techniques, as described in Section "The opportunity to make interventions." MSE measures how close a prediction (regression) line is to the set of actual data points, by calculating the distances from the points to the prediction line (distances are the "errors"), squaring them and calculating their average (mean). The squaring removes any negative signs as well as giving more weight to the larger differences. CC measures how strongly variables are related to each other by dividing their covariance by the product of their standard deviations. A CC of +1 indicates a perfect positive correlation, which means that as variable X increases, variable Y increases and while variable X decreases, variable Y decreases. A CC of −1 indicates a perfect negative correlation. For the purposes of identifying the strongest overall correlations for each analysis technique we calculate the average using absolute CC values.

Experimental results

Research question 1

The value and usefulness of prediction based upon small student cohorts (in this case <30) and where organisational barriers limit the availability of student data. In this case, our data only includes Attendance, VLE accesses and assessment marks. We summarise our results under each of machine learning analyses and traditional statistical methods.

Machine learning analyses

For each of three prediction accuracy measures, Relative % Accuracy, MSE and CC, we present the results of each of DT, KNN and RF analyses, carried out at each assessment point (Tables 7–9). In each case, this includes both the analyses results where VLE News and Teaching accesses are included as separate attributes and where they are combined as one attribute. Prediction accuracy is calculated as 100%—Absolute value of (Actual assessment result—predicted result)/100%. The results of each technique are then discussed.

The overall module result is an arithmetic combination of the intermediate assessment results (see Table 5) and, therefore, we would expect all the prediction methods at the module result assessment point to deliver the most accurate results. This is clearly the case with accuracy between 81% and 91%, averaging 86%. The less than 100% accuracy in each case may be explainable by a combination of inaccuracies in the prediction techniques used and the influence of attendance and VLE access data. RF and KNN (K = 3) with VLE accesses combined delivered the highest average prediction each with accuracies of 75%. Importantly for potential intervention opportunities, predictions at each of the intermediate assessment points using these analysis techniques, although mixed (between 56% and 88%) were promising in several cases, with accuracies at 70% or above at 9 of the 12 points. The least accurate results were delivered by DT Regression and KNN (K = 1) with VLE accesses combined, averaging 65% and 66%, respectively.

As with our Relative % Error measure, the most accurate prediction results (in the case of MSE these are the closest results to zero) are as expected at the overall module result assessment point. At this point, MSE values are between 0.01 and 0.03. Similarly to Relative % Error measure, RF and KNN (K = 3) with VLE accesses combined delivered the most accurate prediction results, excluding the overall module result predictions, with average MSE values of 0.046 and 0.047 respectively. The least accurate results were delivered by KNN (K = 1) with VLE accesses combined, DT and KNN (K = 1) with average MSE values of 0.08, 0.07 and 0.07 respectively.

As with average % accuracy and MSE, CC prediction results are strongest at the overall module result assessment point, with CC values between 0.05 and 0.74. However, in the case of CC, it is DT with VLE accesses combined that delivers our strongest prediction results with an average CC of 0.4, followed by RF with VLE accesses combined and KNN, K = 3 each with an average CC of 0.29. The least accurate results were delivered by KNN, K = 1 and K = 2, with VLE accesses combined giving us CC values of 0.13 and 0.17 respectively. The remaining analysis techniques delivered promising prediction results with CC values between 0.22 and 0.28. In order to investigate the corresponding effect of attendance and VLE access data, we repeated the analyses using only the assessments and excluding all other data. The results were mixed with only very small variations leading us to believe that inaccuracies in the prediction techniques themselves are the major contributor. We present an illustrative subset of the results (Table 10).

Results including all attributes are shown first and results using the assessment results only (ie excluding attendance and VLE accesses) are shown second. We can see that the comparative results are mixed. Recommendations for further work include investigating the predictive effect of cumulative multi-year analyses on the inclusion of attendance and VLE accesses data. After module completion, we performed an overall module result prediction analysis at each assessment point, using RF analysis (Table 11).

We obtained average student final result prediction accuracies of between 82% and 86% using RF analyses. However, the variance between individual student predictions and their actual final result at each assessment point was high, with accuracies ranging from 11% to 99% (Table 12). MSE and CC accuracies performed in line with relative % accuracy analyses.

Table 7: Prediction accuracy measured by relative % accuracy

Relative % accuracy	*EVS1*	*EVS2*	*EVS3*	*Group presentation*	*Individual report*	*Module result*	*Average % Accuracy*	*Ave % Accuracy (Excl. Module result)*
Decision Tree Regression	72%	33%	57%	74%	64%	90%	65%	60%
Decision Tree Regression (Combined VLE Clicks)	77%	32%	57%	96%	69%	88%	70%	66%
K Nearest Neighbour, K = 1	71%	54%	52%	90%	70%	88%	71%	67%
K Nearest Neighbour, K = 1 (Combined VLE Clicks)	73%	46%	52%	89%	57%	81%	66%	63%
K Nearest Neighbour, K = 2	74%	49%	66%	86%	74%	89%	73%	70%
K Nearest Neighbour, K = 2 (Combined VLE Clicks)	74%	58%	63%	89%	69%	81%	72%	71%
K Nearest Neighbour, K = 3	74%	55%	74%	74%	72%	89%	73%	70%
K Nearest Neighbour, K = 3 (Combined VLE Clicks)	76%	60%	68%	88%	73%	82%	75%	73%
Random Forest	80%	56%	70%	81%	71%	91%	75%	72%
Random Forest (Combined VLE Clicks)	80%	50%	65%	90%	71%	86%	74%	71%

Table 8: Prediction accuracy measured by mean squared error

Mean Squared Error	*EVS1*	*EVS2*	*EVS3*	*Group presen-tation*	*Individual report*	*Module result*	*Ave MSE*	*Ave MSE (Excl. Module result)*
Decision Tree Regression	0.0767	0.1489	0.1051	0.0411	0.0603	0.0137	0.0743	0.0743
Decision Tree Regression (Combined VLE Clicks)	0.0459	0.1435	0.1019	0.0127	0.0603	0.0158	0.0634	0.0634
K Nearest Neighbour, K = 1	0.0806	0.0969	0.1464	0.0216	0.0611	0.0213	0.0713	0.0713
K Nearest Neighbour, K = 1 (Combined VLE Clicks)	0.0736	0.1101	0.1426	0.0247	0.0838	0.0315	0.0777	0.0777
K Nearest Neighbour, K = 2	0.0527	0.0982	0.0781	0.0261	0.046	0.0217	0.0538	0.0538
K Nearest Neighbour, K = 2 (Combined VLE Clicks)	0.0634	0.0755	0.0841	0.0229	0.0586	0.032	0.0561	0.0561
K Nearest Neighbour, K = 3	0.0527	0.0842	0.0591	0.0334	0.0532	0.0181	0.0501	0.0501
K Nearest Neighbour, K = 3 (Combined VLE Clicks)	0.0613	0.0669	0.0692	0.0028	0.0526	0.0289	0.0470	0.0470
Random Forest	0.0341	0.0657	0.0756	0.0359	0.0461	0.0191	0.0461	0.0461
Random Forest (Combined VLE Clicks)	0.0465	0.0922	0.0726	0.0189	0.0542	0.0196	0.0507	0.0507

Table 9: Prediction accuracy measured by correlation coefficient

Correlation Coefficient	*EVS1*	*EVS2*	*EVS3*	*Group presentation*	*Individual report*	*Module result*	*Ave CC*	*Ave CC (Excl. Module result)*
Decision Tree Regression	−0.0912	−0.4518	0.0706	−0.0224	0.1732	0.7386	0.2580	0.1618
Decision Tree Regression (Combined VLE Clicks)	0.2754	−0.5090	−0.0426	0.7853	0.1732	0.6942	0.4133	0.3571
K Nearest Neighbour, K = 1	0.042	0.0843	0.2329	0.558	0.0262	0.5363	0.2466	0.1887
K Nearest Neighbour, K = 1 (Combined VLE Clicks)	−0.0295	−0.0083	−0.1433	0.4638	0.2651	0.1394	0.1749	0.1820
K Nearest Neighbour, K = 2	−0.1536	−0.34	0.0701	0.3899	0.1541	0.5424	0.2750	0.2215
K Nearest Neighbour, K = 2 (Combined VLE Clicks)	−0.0295	0.1093	0.0683	0.5106	−0.0404	0.0455	0.1339	0.1516
K Nearest Neighbour, K = 3	−0.0876	−0.3019	0.2973	0.146	−0.1536	0.7391	0.2876	0.1973
K Nearest Neighbour, K = 3 (Combined VLE Clicks)	−0.3137	0.0535	0.1928	0.5069	−0.0402	0.2075	0.2191	0.2214
Random Forest	0.4165	0.1443	0.0648	0.1352	0.1711	0.5985	0.2551	0.1864
Random Forest (Combined VLE Clicks)	0.0438	−0.2986	0.1289	0.62	0.0732	0.579	0.2906	0.2329

Table 10: Comparison of analyses including all attributes against those using assessment results only

Analysis Technique	*Prediction accuracy measure*	*EVS3*	*Group presentation*	*Individual report*
K Nearest Neighbour, K = 3 (Combined VLE Clicks)	Relative % Accuracy	74%/67%	74%/82%	72%/73%
	Mean squared error	0.0591/0.0553	0.0334/0.0427	0.0532/0.0498
	Correlation coefficient	0.2973/0.4164	0.146/−0.2093	−0.1536/0.1254

Correlations between assessments

An analysis of the cross-correlation between each of the interim assessments and the overall module result (Table 13) shows moderate, high and very high correlations with the overall module result. Of these five interim assessments, we found high and very high correlations between the two major interim assessments (Group Presentation and Individual Report) and the overall module result. The initial three interim assessments were all moderately correlated with the overall module result.

Graphical analyses to support potential interventions

Example graphical analyses performed at EVS3 and individual report assessment points are discussed and shown below (Figures 2–7). In each figure, the student identification number (1 to 23) is labelled on the x axis. Note that student 10 withdrew from the module prior to assessment commencement.

Machine learning predictions for students 12 and 14 highlighted 62% and 97% negative disparities with their actual and expected progress raising concerns with module leadership. We can see form this table that in both cases their attendance records are very high and therefore not a cause for leadership concern. Student 22 had scored well in EVS1 and EVS2 assessments and given that the best two of the three assessments only are included chose not to take EVS3.

A glance at this chart shows that both student 12 and student 14 are registering average VLE accesses and this could be an area for concern and potential intervention.

As above, using students 12 and 14 as our examples, we can see that their high average EVS1 and EVS2 results indicate why machine learning prediction disparities were evident.

Machine learning predictions for students 19 and 20 highlighted 159% and 179% negative disparities with their actual and expected progress raising concerns with module leadership. We can see from this table that both students are maintaining average attendance.

A consideration of this chart shows that both student 19 and student 20 are registering above average VLE accesses but may still be an area for potential intervention.

As above, using students 19 and 20 as our examples, we can see that both have good average assessment results to date. In this case, module leadership considered intervention unnecessary.

Research question 2

We now consider the value and usefulness of prediction analyses for intervention opportunities. For these analyses to be of value for interventions they must be available to module leadership while sufficient time is left for successful interventions to be made and any consequent positive effects to be achieved by the student. The early and mid-timed assessments in the

Table 11: Module result prediction at each assessment point

Analysis technique	*Prediction Accuracy Measure*	*EVS1*	*EVS2*	*EVS3*	*Group presentation*	*Individual report*	*Average accuracy*
Random Forest	Relative % accuracy	82%	82%	86%	83%	85%	84%
	Mean squared error	0.0325	0.0334	0.0253	0.0323	0.0206	
	Correlation coefficient	−0.0207	0.1114	0.3763	0.1866	0.5483	

Table 12: Range of individual student final result percentage prediction accuracies at assessment points

Prediction accuracy	*EVS1*	*EVS2*	*EVS3*	*Group presentation*	*Individual report*
Lowest	38%	11%	52%	28%	35%
Highest	98%	98%	100%	99%	99%

Table 13: Assessments correlation matrix

	EVS1	*EVS2*	*EVS3*	*Group presentation*	*Individual report*	*Overall module result*
EVS1	1.00	0.53	0.63	0.47	0.44	0.55
EVS2		1.00	0.59	0.60	0.60	0.66
EVS3			1.00	0.42	0.42	0.51
Group Presentation				1.00	0.73	0.90
Individual report					1.00	0.95
Overall module result						1.00

Key		
Very highly correlated	(0.9 to 1.0)	
Highly correlated	(0.7 to 0,89)	
Moderately correlated	(0.5 to 0.69)	
Low correlation	(0.3 to 0.49)	

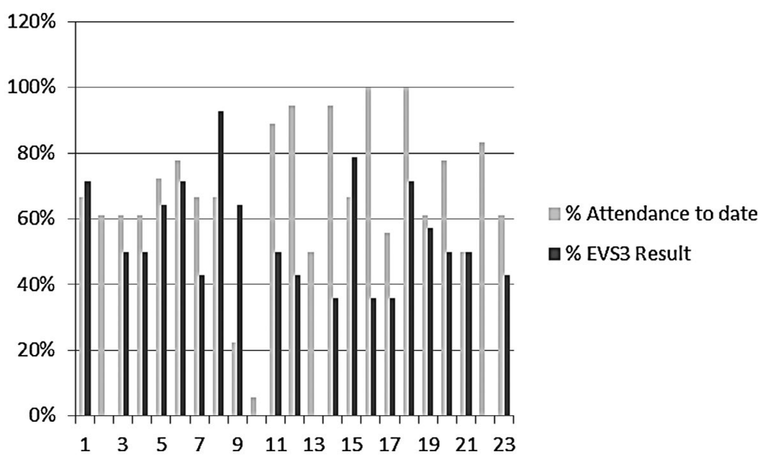


Figure 2: Attendance to date v EVS3 result

selected module provided this opportunity. The progressive prediction analyses conducted may also provide module leadership with useful data in respect of module and assessment design. For example, if our predictions on individual assessments were consistently accurate, it may be that these assessments are adding little value in their current form and require revision. Adaptive learning systems dynamically adjust the number of questions upwards and downwards and dynamically adjust student learning paths depending upon student performance (Wakelam *et al.*, 2015). The graphical analyses (Section "Graphical analyses to support potential interventions") proved useful for module leadership to perform "at a glance" assessments of

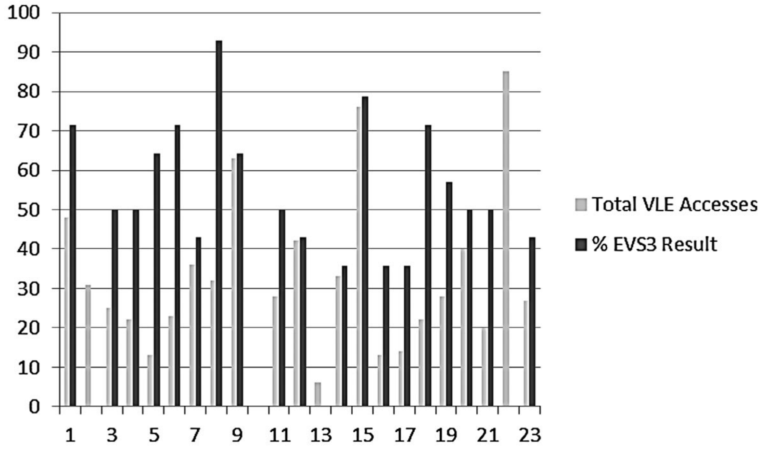


Figure 3: Total VLE accesses v EVS3 result

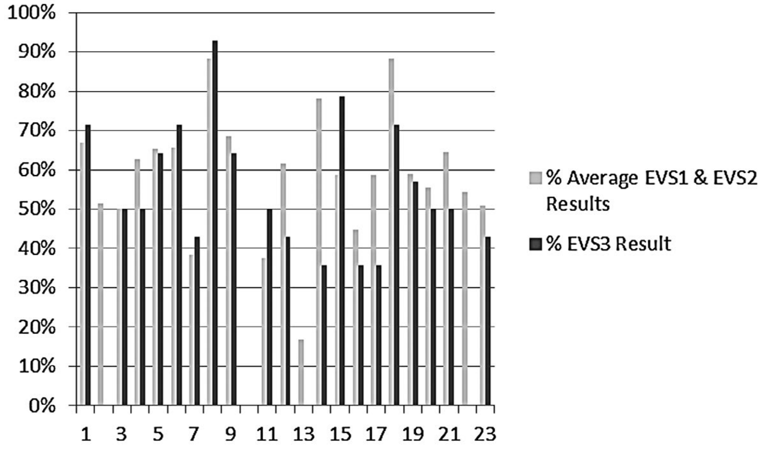


Figure 4: Average of EVS1 and EVS2 results v EVS3 result

student activities. For example, where a student prediction suggests a performance risk, module leadership were able to quickly view their attendance and VLE usage in support of personal experience of the student. This in itself may suggest intervention methods, ranging from encouraging improved attendance or more usage of VLE material. In the case of this module, we were able to review students where machine learning predictions identified potential poor outcomes, supported by "at a glance" comparisons of their attendance, VLE accesses and prior assessment marks. This information coupled by module leadership knowledge of each student through face-to-face lectures and tutorials supported direct interventions, including coaching and the provision of additional teaching material. These interventions may be grouped under the heading of providing additional scaffolding to students. Research conducted by Stubbs *et al.* at Manchester Metropolitan University discusses how a metaframework for assisting the design of learning frameworks to educational designers to support improved learning outcomes (Stubbs *et al.*, 2006).

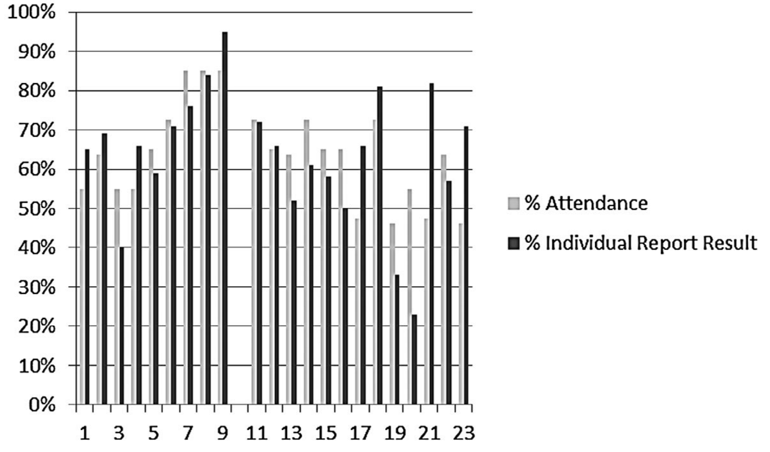


Figure 5: Attendance to date v individual report result

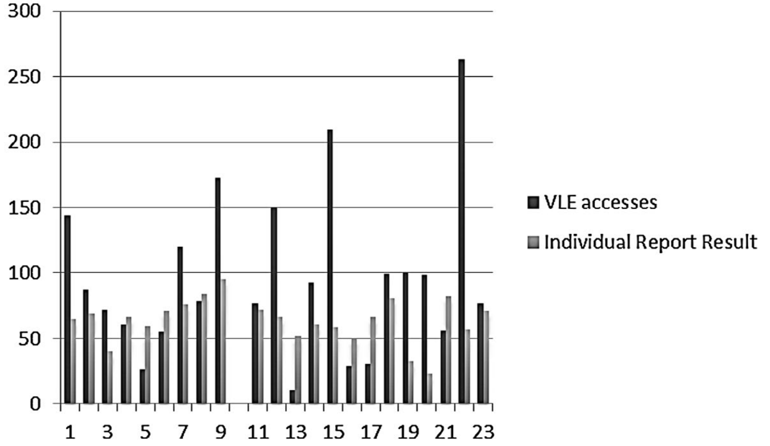


Figure 6: Total VLE accesses v individual report result

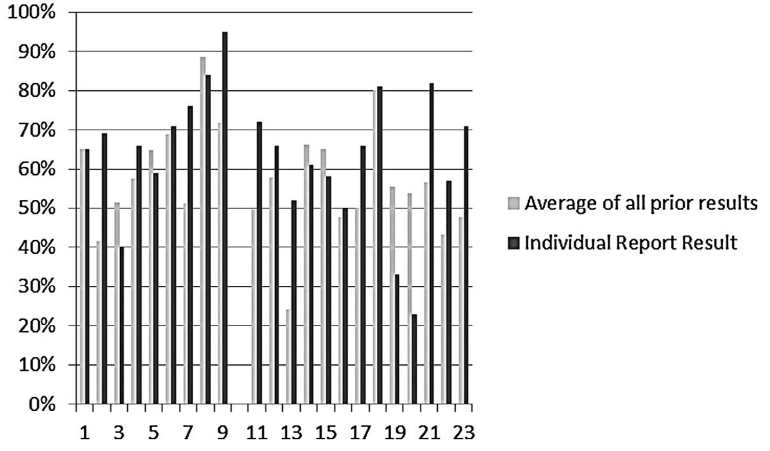


Figure 7: Average of EVS1, EVS2, EVS3 and group presentation results v individual report result

Discussion and conclusions

Research question 1

Is it possible and useful to predict student performance on courses comprising of relatively small student cohorts, where a very limited set of student data is readily available for analysis?

Experimental results show some potential for analysing and predicting student assessment marks on courses comprising relatively small student cohorts, and where only very limited set of student data is readily available for analysis. The average prediction accuracy across all machine learning techniques used was 67%, with KNN and RF prediction accuracy between 66% and 75%. This compares favourably with student prediction accuracy levels achieved across a variety of machine learning techniques applied to large student cohorts with significantly more student attributes (Ashraf *et al.*, 2018). The results in Ashraf and colleagues' study ranged from 50% to 97% (Tables 2 and 3). Importantly for potential intervention opportunities, we obtained some promising results at the point of the third assessment, approximately two thirds of the way through the module, with prediction accuracies of 74% and 70% for KNN and RF Analyses respectively. Reducing the attributes used in our analyses gave us mixed results. Combining VLE News and Teaching accesses into one total had very little effect upon prediction accuracy, in some cases giving a 1% improvement and in others the reverse. Reducing the attributes to only the intermediate assessment results gave us mixed results in comparison with prediction accuracy using all available attributes, hence we could not reliably consider student interventions. Similarly, this provided us with little opportunity to determine the effect of including attendance and VLE accesses on prediction accuracy. We believe that the inclusion of all available attributes may be considered as at least benign to our analyses. There is some evidence (Heuer & Breiter, 2018) that the analysis of VLE accesses alone can be a useful predictor of student performance. Future work accumulating year-on-year module data to investigate the effects on prediction accuracy of multi-year data may provide further insight. As we might expect, the final assessment, the student's Individual Report which is submitted in week 15 of 18, contributing 50% to their overall mark, correlated very highly (CC 0.95) with their overall module result. Additionally, the penultimate assessment, the Group Presentation, submitted in week 11 of 15, correlated highly (CC 0.9) with the overall module result. Usefully, for the potential of earlier intervention opportunities, given their early assessment points of weeks 4, 6, 10 of 15, we found moderate correlations (CCs of 0.55, 0.66 and 0.51 respectively) between EVS1, EVS2 and EVS3 and the overall module result. In particular, student usage of VLE material and correlations between attendance and VLE usage on assessment marks provided valuable insights.

Research question 2

How useful would these analyses be in order to provide course leaders with the opportunity to make timely supportive interventions at appropriate points during the module?

The analyses demonstrated three opportunities for module leadership to identify potentially "at risk" students and to consider appropriate timely interventions. These were machine learning analyses at intermediate assessment points, and the identification, post module completion, of which intermediate assessments provided the likeliest indicators of overall module success. Student performance in their third assessment, week 10 of 15, appears to be a useful measure of individual progress. In this experiment, module leadership were then able to review attendance and VLE access patterns for students whose performance was of concern. Alongside personal experience of the student in question an intervention decision could then be made. In the case of the module, our analyses led to module leadership identifying two specific opportunities for direct, interventions, both following the third assessment, EVS3. In each case, a student's predicted performance showed a likelihood of failing their next assessment. In case 1, further analysis showed a reduction in tutorial attendance. In case 2, analysis showed a combination of reduced lecture/tutorial attendance coupled with minimal activity in the VLE. This enabled leadership to engage in positive discussions with each student and provide specific guidance on their future studies. A variety of possible interventions are described in Section "The opportunity to make interventions", but could be as simple as evidence based discussions drawing a student's attention to their attendance, arranging additional individual or group lectures/tutorials or the availability of further and focussed supporting material on the VLE. Graphical analyses allowing the visualisation of relationships between attributes provides module leadership with further opportunities to identify any interesting correlations which could support positive interventions. These graphical presentations compared different combinations of attendance, VLE usage and assessment results providing easily referenceable "at a glance" supporting material to machine learning results for module leadership. In the case of the module in this experiment, we found these representations supported intervention decisions. Given their significant mark contribution to the overall module result this was to be expected. Additionally, promising results at the earlier third assessment point gave module leadership the opportunity to consider interventions in time for their effects to be useful.

Implications to practice and/or policy

University expectations are currently that the application of LA necessitates the availability of so-called "big data," in particular, modules with large student cohorts. Our results show that university practice can now usefully consider smaller scale deployments of LA. Where student attributes for analysis are limited to readily available data such as student attendance, VLE accesses and intermediate assessment results, with no inclusion of demographic/personal data, either none, or very limited modifications are necessary to university policies. It is good practice to provide students with a clear explanation of what data are being collected and how the analysis is being done, allowing them to individually opt in or opt out of LA implementations. In addition, alternative intervention methods should be documented and where possible students given the opportunity to express their preferences. For example, dashboard presentation of predictions, system generated emails, offers of face to face supportive meeting with course tutors.

Future work

We plan to perform DT, KNN and RF analyses using classification (i.e. binary prediction of pass or fail), instead of regression, and compare student marks prediction accuracy with the results of this experiment.

In order to investigate the predictive results of the respective scenarios of a wider range of student attributes and of a large cohort size, we would propose to conduct two related experiments. Firstly, where the student cohort is small, but where a wider selection of student attributes is available, for example, prior student module marks and examination results from previously attended institutions. Secondly, where the student cohort is much larger, but with the same student attributes as with this experiment.

It would be valuable to accumulate year-on-year module data to investigate the effects of the inclusion of multi-year data on prediction accuracy, using the same analyses techniques as in this experiment.

The module in our experiment is comprised of a relatively even spread of formal assessments, with two at an early stage. The effects upon prediction accuracy of applying the same experimental analyses to a module where either there are fewer intermediate assessments or where they are conducted later in the module may be of value.

Finally, a logical next step would be to design and conduct an experiment which tracks and measures any resulting changes in individual student attendance, VLE accesses and assessment scores resulting from academic staff interventions.

Statements on open data, ethics and conflict of interest

The data in this study can be accessed upon request.

Formal ethical approval for the conduct of the experiment described in this study was granted by the university.

The authors confirm that there is no conflict of interest in this study.

References

Ashraf, A., Anwer, S., & Khan, M. G. (2018). A Comparative study of predicting student's performance by use of data mining techniques. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, *44*(1), 122–136.

Australian Government Department of Education and Training. (2016). Attrition, Success and Retention. *Higher Education statistics, Appendix 4*.

Choi, S. P., Lam, S. S., Li, K. C., & Wong, B. T. (2018). Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. *Journal of Educational Technology & Society*, *21*(2), 273–290.

Clow, D. (2012, April). The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 134–138). New York: ACM.

Corrin, L., Kennedy, G., French, S., Shum, S. B., Kitto, K., Pardo, A., … Colvin, C. (2019). *The ethics of learning analytics in Australian higher education*. Melbourne: University of Melbourne.

Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, *46*(5), 904–920.

de Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., & Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*, *46*(6), 1175–1188.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304–317.

Ferguson, R., & Clow, D. (2017, March). Where is the evidence?: a call to action for learning analytics. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 56-65). Vancouver: ACM.

Gibbons, S., Neumayer, E., & Perkins, R. (2015). Student satisfaction, league tables and university applications: Evidence from Britain. *Economics of Education Review*, *48*, 148–164.

HESA. (2018). *Non-continuation summary: UK Performance Indicators 2016/17.* Retrieved from https://www.hesa.ac.uk/news/08-03-2018/non-continuation-summary

Heuer, H., & Breiter, A. (2018). Student success prediction and the trade-off between big data and data minimization. *DeLFI 2018-Die 16. E-Learning Fachtagung Informatik*.

Horning, N. (2013). Introduction to decision trees and random forests. *American Museum of Natural History*.

Huxley, G., Mayo, J., Peacey, M. W., & Richardson, M. (2018). Class size at university. *Fiscal Studies*, *39*(2), 241–264.

Lin, T. C., Yu, W. W. C., & Chen, Y. C. (2012). Determinants and probability prediction of college student retention: New evidence from the Probit model. *International Journal of Education Economics and Development*, *3*(3), 217–236.

Marburger, D. R. (2001). Absenteeism and undergraduate exam performance. *The Journal of Economic Education*, *32*(2), 99–109.

Mearman, A., Pacheco, G., Webber, D., Ivlevs, A., & Rahman, T. (2014). Understanding student attendance in business schools: An exploratory study. *International Review of Economics Education*, *17*, 120–136.

National Center for Education Statistics. (2017). Retention of first-time degree-seeking undergraduates at degree-granting postsecondary institutions, by attendance status, level and control of institution, and percentage of applications accepted: Selected years, 2006 to 2015. *Digest of Education Statistics*. Retrieved from https://nces.ed.gov/programs/digest/d16/tables/dt16_326.30.asp

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438–450.

Quinlan, J. R. (2014). *C4. 5: Programs for machine learning*. San Mateo, CA: Elsevier.

Rienties, B., Boroowa, A., Cross, S., Farrington-Flint, L., Herodotou, C., Prescott, L., & Woodthorpe, J. (2016). Reviewing three case-studies of learning analytics interventions at the Open University UK. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 534–535). Edinburgh: ACM.

Rienties, B., Boroowa, A., Cross, S., Kubiak, C., Mayles, K., & Murphy, S. (2016). Analytics4Action evaluation framework: A review of evidence-based learning analytics interventions at the open university UK. *Journal of Interactive Media in Education*, *2016*(1). 1–11.

Rienties, B., Nguyen, Q., Holmes, W., & Reedy, K. (2017). A review of ten years of implementation and research in aligning learning design with learning analytics at the Open University UK. *Interaction Design and Architecture (s)*, *33*, 134–154.

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior*, *60*, 333–341.

Rodríguez-Triana, M. J., Martínez-Monés, A., & Villagrá-Sobrino, S. (2016). Learning analytics in small-scale teacher-led innovations: Ethical and data privacy issues. *Journal of Learning Analytics*, *3*(1), 43–65.

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., … Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, *10*(1), 30–41.

Sclater, N., & Bailey, P. (2015). Code of practice for learning analytics. *Joint Information Systems Committee (JISC)*.

Sclater, N., Peasgood, A., & Mullan, J. (2016). *Learning analytics in higher education*. London: Jisc.

Simpson, O. (2006). Predicting student success in open and distance learning. *Open Learning: The Journal of Open, Distance and e-Learning*, *21*(2), 125–138.

Simpson, O. (2013). Student retention in distance education: Are we failing our students? *Open Learning: The Journal of Open, Distance and e-Learning*, *28*(2), 105–119.

Slade, S., & Tait, A. (2019). *Global guidelines: Ethics in learning analytics*. Oslo: International Council for Open and Distance Education (ICDE).

Stubbs, M., Martin, I., & Endlar, L. (2006). The structuration of blended learning: Putting holistic design principles into practice. *British Journal of Educational Technology*, *37*(2), 163–175.

UK Government. (2018). *Guide to the General Data Protection Regulation (GDPR)*. Retrieved from https://www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation

Wakelam, E., Jefferies, A., Davey, N., & Sun, Y. (2015). The potential for using artificial intelligence techniques to improve e-Learning systems.In *ECEL 2015 Conference Proceedings*. Hatfield.

Wilson, A., Watson, C., Thompson, T. L., Drew, V., & Doyle, S. (2017). Learning analytics: Challenges and limitations. *Teaching in Higher Education*, *22*(8), 991–1007.

Wong, B. T., & Li, K. C. (2018, July). Learning analytics intervention: A review of case studies. In *2018 International Symposium on Educational Technology (ISET)* (pp. 178–182). Osaka: IEEE.

Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine*, *4*(11), 218.