eSTEeM project FINAL REPORT

Disproved predictions of at-risk students: Some students fail despite doing well, others succeed despite predicted as at-risk

Martin Hlosta, martin.hlosta@open.ac.uk

Christothea Herodotou, christothea.herodotou@open.ac.uk

Keywords: Students at-risk, OUAnalyse, Predictive Learning Analytics, Error Analysis, Interviews

Report submission date: 30/09/2020

Key staff associated with the project: Martin Hlosta, Christothea Herodotou, Tina Papathoma, Anna Gillespie, Vaclav Bayer, Zdenek Zdrahal

Suggested citation: Hlosta M., Herodotou C., Gillespie A., Papathoma T. (2020): Disproved predictions of at-risk students: Some students fail despite doing well, others succeed despite predicted as at-risk - final report of eSTEeM project, The Open University, Milton Keynes.

Executive Summary

Most of the research around the identification of at-risk students and the prediction of their performance using Machine Learning focuses on developing the most accurate model. Despite recognising the importance of transparency and understanding of the models, little effort has been made to investigate the errors made by these models. In this project, we address this gap by analysing large errors of predicting students at-risk of not submitting their assignments, when even the sophisticated machine learning model was confident about a student outcome, yet the result was different.

The underlying predictions are part of OUAnalyse and are available in most of the undergraduate OU modules to all tutors via the Early Alert Indicators (EAI) Dashboard. The models are updated every week to capture the dynamic changes in student learning behaviour.

We analysed both groups of errors: students predicted to submit their assignment, yet they did not (False Negative) and students predicted not to submit yet they did (False Positive). We conducted mixed-method analysis, combining quantitative analysis of predictions of more than 25,000 students with follow-up online interviews with 27 of them and thematic analysis. We focused on undergraduate level 1 modules on STEM faculty and analysed the predictions for the 1st Tutor Marked Assignment (TMA).

The quantitative analysis revealed that the most prevalent factor in False Positives was immediate growth of student activity after the predictions were generated. Interviews revealed that amongst those students the most prevalent themes were students that were working last minute and were able to overcome last-minute problems, students that had high study workload and dropped some of their other modules, or students who had either the knowledge required for the TMA or studied outside the VLE. In False Negatives, non-submission of assignments was explained mostly by financial reasons, family responsibilities or deferring the module because of high study workload.

Overall, the factors explaining the different outcomes were not related to any of the student data currently captured by the model. As a result of this study, data related to student finance will be part of the OUAnalyse model. We proposed that the absence of missing data can be handled by either giving students an initial questionnaire or letting tutors know so they are able to capture this before the module starts. Intervention strategies based on student recommendations are suggested as well as considerations that we will make available to tutors in the OUA training materials, which might lead to better understanding of the capabilities of Predictive Learning Analytics and subsequently its better usage.

1 Aims and scope of your project

The project investigates the reasons why predictions provided by even sophisticated machine learning models sometimes do not correspond to reality. More specifically, we examined the predictions of students that are at risk of not submitting or failing their first Tutor Marked Assignment (TMAs). These predictions are visualised in the Early Alert Indicators (EAI) Dashboard and are available to Associate Lecturers (ALs) in undergraduate modules at the OU (more than 300 modules per academic year) who make use of them in order to help students succeed with their studies. Several pilots in the previous years identified that (1) OUAnalyse usage by tutors is one of the two significant predictors of students' completing and passing a module (along with previous best score); (2) teachers accessing OUAnalyse at least 41% of the weeks a module runs had better student retention rates (56%) compared to 48% retention of the low engagement group (<=8% weekly usage) and (3) teachers had better student outcomes the year they were accessing OUAnalyse than the previous years when they had no access to it. These facts are evidenced in peer-reviewed manuscripts [2-5].

The goal of predictions is to help ALs with planning a potential intervention and ultimately retain students. From our experience, there are cases where data influencing students' performance occur irregularly, and it is not easy to be captured. For example, a student gets sick or a tutor intervenes with an at-risk student using a private phone.

Our goal was to provide more rigorous explanations of these cases. We aimed to first analyse quantitative patterns in the data followed by interviews. Our hypothesis is that we will not be able to identify all the reasons in the quantitative analysis and hence the need for interviews.

We focused on level one STEM modules and predictions for the first assignment only (A1). The first assignment in the first year of studies is when dropout is more likely to happen [7] hence it was selected as the focus of this study. Moreover, greater similarity in the learning design of modules could be achieved by selecting modules from a single faculty, hence the selection of modules from STEM.

We envisioned three types of results at the start of the project.

1) The results that will be made available to ALs who can support the application of predictive analytics to the teaching practice.
2) For students who are predicted as at-risk and there was a change in their behaviour that led to success, strategies used to retain at-risk students will be identified. These will be communicated to both ALs and students at the start of the module or during the intervention.
3) If features related to this change can be captured from the existing data, these will be used to feed to the learning algorithms.

2 Activities

The overall approach was to use quantitative and qualitative analysis to explain why the results of confident predictions differ. This schema of the methodology is depicted in Figure 1.

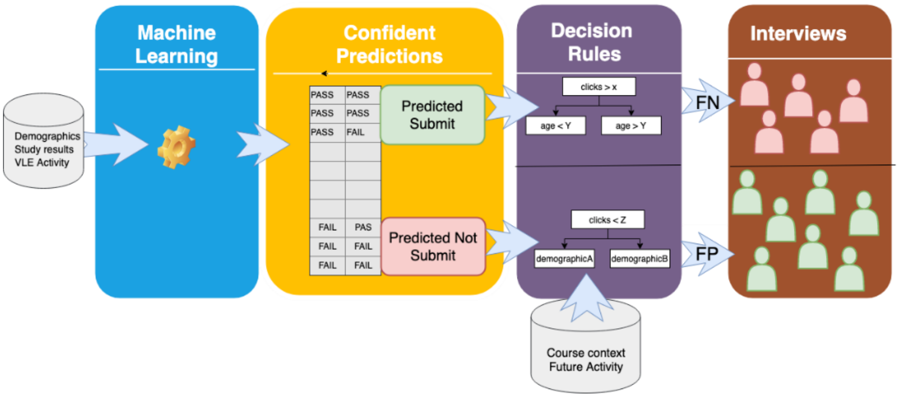

Figure 1 The schema of the methodology - quantitative analysis followed by the interviews.

2.1 Quantitative Analysis

The underlying predictive model uses (a) static data known at the start of a module and (b) VLE interactions to predict which students are at-risk of not submitting their next assignment. Static data include previous study results, number of modules studying, student demographics and indicators of whether the module is repeated. The previous presentation of the same modules used to train the data and the predictions that are generated weekly. This means that there are many models generated each week, one for each module. Gradient Boosted Machine (GBM) optimised for the highest ROC AUC[1] is used to generate the predictions, as well as the confidence of the prediction in interval [0;1]. The optimisation selects the parameters of the model that has the highest ROC AUC metric on the validation data, which is different than the training data. We selected the predictions two weeks before the first assignment (A1) deadline, where the predictions were accurate, yet there was still time for a possible intervention that could result in a change in student's behaviour such as incentivising students to submit their assignment.

Data For Analysis - data from all predictions were put in one table and extended by information not present in the prediction data:

1) **module context** - the length of the module, the number of assignments in the module, the weight of the assignment towards the final grade and
2) **future student activities** - data from the weeks following the predictions, that were unknown during the prediction's generation.

Confident Predictions selection - predictive models provide estimates trying to resolve the uncertainty about the external world, often with the existence of cases where the prediction is close to more classes. In education, we often see "swinging cases" where there are chances that a given student may eventually flip to a positive or negative outcome. We aim to analyse only clear errors and therefore, we excluded such predictions and selected only cases we were confident with. We applied a threshold 0.85 and selected only predictions with confidence >=0.85 for both classes, i.e. Submit and Not Submit. **The threshold 0.85 was selected based on our analysis of the Prediction Calibration Curve. The prediction**

[1] ROC AUC = Area Under the Curve of the Receiver Operator Characteristic - a commonly used metric in Machine Learning, which estimates the quality of the predictions under the different values of predictions threshold (usually 0.5). The metric works well also in the case of imbalanced classes, where Accuracy typically fails.

values were calibrated between 0.85 and 1.00. There was a drop in precision when in the 0.8 - 0.85 interval and that's why we selected the 0.85 threshold.[2]

Pattern Identification - separately for both of the error types, a Decision Tree was constructed to distinguish between (1) FP and True Positive (TP) and (2) FN and True Negative (TN) (minimum leaf size=30, depth=4). We chose the Decision Tree due to its ability to identify more complex patterns with dependencies between features. The resulting trees were then used to generate the decision rules, counting the number of covered students and the rule's confidence, with the minimum confidence set to 0.8[3].

2.2 Qualitative Analysis, Online Interviews

Semi structured interviews were conducted (and recorded) using Skype or telephone with 27 students who met the FP and FN criteria - i.e. they were predicted as confident Submit or Not Submit in their first TMA. We selected the students with errors in the confident predictions for the cohort of students enrolled in currently running modules at the time that the data collection was about to start, i.e. 19J presentation. Emails were sent to 527 students in 16 modules including FP and FN groups. Out of these, 38 students replied positively (6.83%) but 25% of them cancelled their interview (n=9). The 27 students represent a 5.12% response rate. Interviewees were offered gift vouchers for their time. Interviews lasted between 20 and 40 minutes. The interview questions are included in the Attachment C.

To analyse the interviews, we grouped participants to (a) students predicted to submit yet they did not submit (FN; n=13) and (b) students predicted not to submit yet they submitted (FP; n=14). We anonymised participants and coded according to the group they belonged to (i.e. FN or FP followed by a number to ensure confidentiality).

2.3 Reporting results/findings

The results from the quantitative and qualitative analysis were planned to be submitted to renowned educational conferences (LAK and AIED). Moreover, a midterm report was finalised around the end of 2019 and the final report finished in September 2020. The results in this final report are planned to be submitted to an educational journal.

2.4 Ethical clearance

Before contacting students, we gained ethical approval from the University Student Research Project Panel (SRPP) and the Human Research Ethics Committee (HREC) to conduct interviews with students. We explained the use of predictive analytics to select candidates for interviewing. The approvals included sending the planned interview schedule with students and the consent form template for approval to the HREC. The comments that needed to be addressed were mostly concerning using Skype as a tool for online interviews and necessity to contact HREC again in case we decided to go for the questionnaire rather than the intended interview. The project was registered on the University Asset Register for GDPR compliance.

2.5 Changes against the plan

During the project, there were only minor changes against the plan we proposed. We had to deal with a lower response of students, but this is a common problem in OU studies, and it was expected. We submitted a reminder for the initial email, which helped us with recruiting more students and we reached the number we were aiming for. We did not need to distribute the questionnaire to students.

[2] https://scikit-learn.org/stable/auto_examples/calibration/plot_calibration_curve.html
[3] The minimum confidence 0.8 is a default value for the library that we used for the analysis (http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/) and commonly used value used in association rules analysis.

The lockdown that was caused by the COVID-19 virus did not affect our plans significantly as we aimed to conduct all the interviews online. Both LAK’20 and the AIED’20 conference were moved online which has led to less expenses. We sent more invitations to participants and ended up with 27 instead of 20 interviews. Compared to the original plan, we were able to recruit an associated lecturer, A.Gillespie to the interviews and contribute to the analysis.

3 Findings

3.1 Quantitative analysis

Considering only the predictions two weeks before the deadline of the first assignment, we analysed 50,442 predictions in 17 modules in 62 presentations between 2017-2019, having 36,352 unique students. After applying the threshold of 0.85, this resulted in 38,073 confident predictions of 29,247 students (383 FP; 1,507 FN; 2,671 TP and 33,512 TN. The ROC AUC computed over all predictions was 0.8897, and for the confident predictions 0.8958 which demonstrates high predictive accuracy of the predictive model. Taking Not Submit as a target class and setting up the prediction threshold on 0.5, the F1=0.6320 with Precision=0.6841 and Recall=0.5873.

For confident Not Submit predictions, the decision tree was able to explain 50.91% of the FP errors with 75.29% precision (195 students). The strongest attribute proved to be the number of clicks one week before the deadline of the first assignment *(clicks_1week_before_A1 > 0.66)* with confidence 0.77 (164/212 students with confident prediction submitted). Further filtering only to students who were younger than 37 years increased the confidence to 0.82 (138/168 students), and for students with A Levels or higher to confidence 0.88 (101/115 students).

For the confident Submit predictions, the model distinguished 18.73% of the FN errors with precision 68.73% (1,036 students). The only significant pattern discovered was:

module_avg_clicks > 69 & clicks_1week_before_A1 <= 0.01 & trend_clicks_to_next_week <= -0.86

with confidence 0.83 (242/292 students with confident prediction did not submit). This relates to a dramatic decrease in students' activity in the last week before the first assignment in modules with high activity.

Both types of errors were associated with a change of student activity after the predictions were generated, either a drop for students that seem on track or a steep increase for at-risk students. The presence of younger and more educated students in FP might suggest that for these groups it can be easier to exert last minute effort and succeed, while for the older and with lower education ones it might be hard (e.g. limited time available, other obligations) to do that. Results showed that only data driven analysis is not adequate and prompts for investigating further these results through exploring students' experiences that may explain why confident predictions were disproved. Therefore, in the second stage of this study, we conducted in depth online interviews with students with errors in predictions.

3.2 Qualitative Analysis – False Positives (Submit despite predicted as Not Submit)

The reasons why the predictions were so much off in the 14 False Positives, were grouped into 4 themes – 1) Last Minute Catching Up, 2) Previous Knowledge, 3) Using external resources and 4) Decrease of High Workload.

1. Last minute catching up

Aligned with the quantitative analysis results, six students had to catch up because of Late Study Loan approval, Health problems, Family issues, Technical problems with the computer, Low study motivation or a combination of these issues. The predictions reacted to their very low activity, yet they were able to escape from the issue and submit for several reasons. One student (P22FP) did not study but guessed correctly the answers to the TMA because that was the only possibility given their health issues. Two students were contacted by a tutor (P08FP; P27FP). For one of them the tutor provided help on how to catch up with the printed materials. Note that this student also had health issues despite passing the initial "Are you prepared" test the student didn't believe they have the necessary background to pass the module. This student, despite completing the module, failed the exam.
Two other students could be described as 'self-organised and self-motivated' and these characteristics helped them overcome their study problems (P03FP; P04FP). One of them appreciated the possibility to

download the tutorial that they could not attend. Three of the students also admitted using offline or External resources, rather than material on VLE, which influenced their prediction of being flagged as at-risk.

'My student finance was late to come through. So I sat there and prepared not to start the year as I thought I was not going to get it in time…. I was in Ireland as well, so I didn't have my laptop for some of the time.'

'I had to actually go to the library and I had to book it in and I could only book it in for an hour a day because of the demand so it kind of messed up how I work. It is so much easier for me when I am working in a comfortable environment usually at home and I am working on my own laptop. I don't know why but it messes with me when I am not working on my own laptop.'

One student didn't want to contact the tutor despite knowing that it was late half due to their pride and possibly had some knowledge from the previous module failure. Despite the student's perception of being on track after the TMA1, this student later failed the module.

2. Previous Knowledge

Four students stated that they already had the knowledge required for the first TMA, which means they did not need to use the VLE materials or used them in a lower study frequency. Two students (P23FP; P09FP), mentioned previous knowledge related to their experience and for one of them, the module was considered easy. For the remaining two students (P05FP; P18FP), the previous knowledge was due to repeating the same module they had failed in the past. Interestingly, these two submitted the first TMA and even reached the completion but they failed the module in the end. This might tell us that despite the predictions being wrong because the students submitted, they were not good enough to pass the whole module. Flagging these students to their tutors can be beneficial. Moreover, one of the students declared their low motivation and one of them faced difficult personal experiences during the first TMA.

"I just went through the assessment. Like I said, most of it I have kind of done before but when I needed to do specific pieces of work, like pieces of study, I just kind of find what I needed. "

3. Using external resources

Because five students did most of their work offline or using external materials, the model considered them not engaged and at risk. The increased activity observed after the predictions can be explained by using VLE only to prepare and submit their TMA at the last minute.

'... I studied from the books and then I only used the websites when preparing my TMAs.'

For two of them (P24FP; P02FP), using offline materials was a preferred way of studying - one of them declared that this was necessary due to travelling.

'I imagine the thing with the system, the biggest thing is that I downloaded the materials, so I am doing it offline. I am not interacting with the website as much as other people.'

One other student (P07FP) expressed that because English is not their first language, they preferred to use other sources – external audio materials. Two students (P02FP; P09FP) used WhatsApp to communicate with other students rather than using the module forum as this led to instantaneous responses whereas the students forums were slow in replying. For two other students (P09FP; P27FP) external resources were accompanied with other contributing factors, such as tutor intervention and last-minute studying of P27FP and previous knowledge of P09FP.

4. Decrease High Workload

Eight out of 14 students studied at least one other module by the time of the module, for which they have been interviewed, started. Three students studied two other modules. Two of them (P01FP; P06FP) dropped one of the modules, which might be a reason why they were able to not only submit but pass their module despite having two modules, which is still high study intensity. The third one (P23FP) continued with all three modules (each 30 credits) and passed all of them. One of the reasons for submitting and passing the module might be their previous knowledge (see theme 2). The remaining five students had two modules in total consecutively, but the different prediction was explained by other themes.

Aligned with the result of Calvert et al. [1], for the FP cohort, two of our participants also highlighted the importance of starting preparations for the assignment early on. In three cases, submission of the assignment may be attributed to students' prior knowledge.

3.3 Qualitative Analysis – False Negatives (Not Submit despite predicted as Submit)

False Negatives (N=13) revealed four different factors explaining why students did not submit despite their high chances of submitting including 1) Finance Related Problems, 2) Deferral for another module, 3) Deferral for Family Responsibilities and 4) Other issues.

1. Finance related problems

Four students reported issues with finance (P14FN; P12FN; P19FN; P21FN). P12FN reported that they were unable to submit their assignment as they had not received their student finance on time and was subsequently deregistered from the module. When asked if they approached the university to re-register, they reported that they did not think that would have been possible so they did not pursue it and they are no longer studying. A further issue was the difficulty experienced in trying to get an assessment for disability funding.

.' …. on Dec 12th I was deregistered by the (university) and on Dec 17th SFE wrote to me and said we will pay your course tuition fees.'

P14FN also stated that they did not think they could get their finance in place in time for the cut-off date at which they would be deregistered so also deferred. As a high achiever, that student continued to appear as active on the VLE while the transfer to their later module date was completed hence the prediction they would submit. P20FN stated that the loan was delayed because the student filled a wrong form when applying for it.

For P21FN, the loan got through, but with a great delay. The student was also affected by family issues, submitted the first TMA with a great delay, but withdrew in the end.

2. Deferral for another module

Of this group, four students (P25FN; P13FN;P11FN; P15FN) were studying 120 credits and chose to defer their registration until a later intake with two (P11FN; P15FN) stating that studying 2 modules whilst also working and having family commitments became unmanageable, particularly as they also found the module content difficult. They both remained registered on their other module and reported that prior to the submission for the first assignment they had expected to continue studying both modules and had therefore engaged in all their module activities. P13FN also stated that they quickly realised that they could not manage to study two modules and hence their decision to defer until a later intake.

3. Deferral – Family responsibilities

Apart from P25FN, two other students (P21FN; P26FN) reported their study was affected by family responsibilities and despite only studying one module they decided to withdraw. P26FN contacted their tutor but did not receive a reply. There was an industrial strike planned in December 2019 and one of the factors why some tutors did not get back to students might be their participation in the strike. We did not investigate this connection; however, we suggest that any further planned strike should be considered as a potential risk for students and their engagement with tutors. For example, module chairs might consider prompting tutors to put an automatic answer in their email. This will not leave students wondering whether their tutor has forgotten to answer or is simply not working at that moment. P21FN submitted the TMA with a delay but then withdrew anyway.

4. Other

The rest of the students (N=4) showed specific reasons why the prediction considered them to submit but they did not. One student (P17FN) stated that as the weighting for the first assignment was only 7% of the overall course work assessment mark, they did not feel it was worth submitting, whereas P20FN found it too difficult and struggled with technical issues such as zipping files for submission. As an old student, it would be beneficial to increase their digital skills to prevent similar problems in the future.

Two students in the end passed the module. P10FN stated that they struggled with language issues as English was not their first language. Despite having done the relevant study, due to illness they decided not to submit their first assignment as they did not think it was good enough and they were not granted an extension due to lateness of the request. However, they remained registered and passed the module. P16FN had issues accessing the internet which might account for the prediction they would not submit as their online activity would be curtailed. The extension allowed the student to submit later and finally also pass the module.

Note about mental health

Despite not reporting as the primary factor of non-submission, it is possible that for three students the knowledge of specific disability (Mental health) might increase their chance of being flagged (P12FN; P15FN; P20FN). One of the students had issues with finance loan approval and the two others deferred this module and continued in another one. Whilst recognising that this is not the case for all students with mental health issues, it is possible that for students and at certain periods of their lives, it might be challenging for them to manage a high workload and a prediction that knows there is a risk of this specific flag can identify this in the data.

Note about extension

Granted extension for present only for one FP (this student had to guess the answer in the end anyway) and four FN, one more FN also refused the extension - i.e. for 5 out of 13 students the extension did not support them towards submission. Since the predictions considered them on track, it is possible that these students got into an issue too late and could not recover from their problem. More work regarding extension and its influence should be conducted to better understand it's impact. Moreover, for 3 students the extension was not recorded in the data that is available for OUAnalyse and it should be investigated whether this is a) due to some modules not having this recorded or b) because the tutor did not record the extension in the system or c) for any other reason.

3.4 Limitations

There are some limitations that can be addressed in future research. The low response rate of students is rather common and this may be explained by the fact that most students have full or part time occupations while they are also studying that may have prevented them from taking part in any research activities. Recruiting more participants in the future, including new to the university students, students from non 1st year STEM courses and examining errors in TMA2 onwards can enrich our understanding of why errors in predictions occur.

4 Impact

We aimed to include different types of suggestions on how to deal with the limitations of the predictions in the future:

1. Enhancing the predictive models with data that is already collected at the OU yet not captured in the predictions.
2. Recommendation to collect new data, which might be relevant for predictions.
3. Possible intervention strategies for students to escape from being at-risk to succeed. These can be also fed into the developed personal study recommender.
4. Set of possible issues to pay attention, which can be made available to tutors and students at the beginning of the module.

The set of possible issues to pay attention to as well as the link to the report is now being incorporated in the new training materials for ALs and will be available in the Tutors' induction information in TutorHome and referenced from OUAnalyse dashboard directly.
Despite the short time that the research has been available so far, it has attracted attention. It has been referenced by Simon Buckingham Shum in his blog post Should predictive models of student outcome be "colour-blind"? [6], where he stresses the importance of models and errors understanding.

4.1. Improved prediction capabilities

The results from the interviews revealed one type of information that should be included in the predictive model of OUAnalyse – **data about the student funding**. This includes several individual attributes:

- Flag whether a student applied for a loan for the module
- Status of the application for a loan – Eligible - Payment Pending - Changes - Attended/Decline
- Price area (ENG/SCO/WAL)

This information is being included in OUAnalyse and evaluated for the 2020J presentations. Initial results and the results of a separate project dealing with qualification progression suggest that this is important, especially at the start of the module. Despite the student being active in VLE, not obtaining the student loan can mean they are at risk of deferral. This can allow tutors to identify that the student does not need study help but rather some financial advice. In addition, we investigate other factors including specific disability types as some disabilities may pose higher risk to complete and pass the module than others.

Information from other studied modules - OUAnalyse already takes into account student's credit workload but it does not take into account the situation in the other modules. The interviews revealed that some students, despite being enrolled in more modules, handle their studies well. Knowing that the student is both active and submitting their TMAs in both modules might suggest that the workload itself is not an issue and might resolve some of the potential False Positives. This assumption needs to be, however, properly evaluated.

Extension for the next TMA - despite the extension being available to tutors in OUAnalyse dashboard, this information is taken into predictions only for future TMAs. For example, for predicting submission of TMA2, the predictions take into account possible extension of TMA01. Until recently, we only had this information once TMA was submitted.
Taking extension for the currently predicted TMA is our next planned step. The extensions need to be examined closely as we found out that not all students with extensions had their extensions recorded in the data. The analysis should include which system is used for the submission and whether this has changed since the last presentation as this might include errors in the predictions. It is possible that taking late submission from the previous TMA might be a better indicator than the extension itself.

4.2 New data collection

Apart from using the finance data new types of data might be possible to collect.

1. Tutorial related data – despite the information captured within the OU systems, currently this is not being used because of its complexity. Some data can be extracted from VLE data exports from

Moodle but some of them can only be found in other systems, e.g. VOICE. This should help to have an integrated view of whether students booked their tutorial, attended them or whether they downloaded the materials related to the tutorial afterwards.

2. Student questionnaire - To eliminate cold-start of not knowing some of the student preferences, students might be given a short questionnaire that might identify information currently not captured:
   - o Student preference towards online vs printed materials – some students know in advance that they will do most of their studies in printed materials. Not having this data will potentially keep some students being flagged as at-risk, which might not be a problem, but it might slightly increase the time tutor spends with them even if they do not need it.
   - o Previous student knowledge – modules can introduce the "are you ready for this module" test that could be easily integrated in the existing infrastructure. One student that the test was not really challenging and did not reflect what the module is really about. Other students were complaining that the module was too easy for them.
   - a. Student Learning Maturity - this might also be used to capture learning habits in a more generic way to discover student's learning habits. Some students, despite faced with difficult life events, showed a good level of determination and organisation and were able to handle them better than others. Such skill was not related only to students with a HE or Postgraduate degree therefore using their previous qualification data might not be enough.
   - b. Digital skills – Despite only one student reporting that they didn't know how to zip the TMA for the submission, discovering students' digital skills might prevent similar problems in the future, especially for older students, where the technology barrier might exist. Some modules employ the strategy of "dummy TMA" but some of the students in the interview expressed their frustration of doing this.

4.3 Possible intervention strategies

(AL, SST) Financial help – after including the finance information, OUAnalyse can play a role in triggering potential problems related to not obtaining the loan or when the loan is delayed.

(AL) Student Workload – already highlighted to students yet it is reasonable to remind them that high workload can lead to their potential problems in the future. They should weigh their options and the commitment needed to study and complete the module.

(AL) Escaping from problems – the number of interviewed students that faced unexpected problems and got into time pressure was not high. But some existing strategies included
- reminding the possibility to download the tutorial if they were not able to attend it online
- send the important materials for the next TMA in the email
- prevention - developing good organisation and time management habits before starting to study the module. When faced with an issue under time, students with such skills were more resistant and likely to submit the TMA and pass the module. This included students who are used to doing work under deadline pressure.

(MC) Strike related issues – module chairs can consider sending a message to all tutors suggesting putting information about a potential strike in case the strike occurs around the TMA submission. This should not be mandatory, but it might help students to realise that the tutor might not respond. The student might have been informed by The OU about the strike, but it might be that students did not read this message and are unaware of the situation. It is up to the university to consider how to approach this issue.

4.4. ALs - topics and issues about predictive modelling to pay attention

Tutors should be aware that not only predictive analytics cannot be 100% correct, there are issues that the models will possibly not be able to address as some of the data are not collected. Even after enhancing the predictions with the additional data, tutors should keep in mind that:

When student is flagged as highly at risk:

- Students might study outside the VLE – this might be due to their preferences in reading printer materials or necessities related to travelling while working. A possible solution can be to ask the students in the group before the module starts about student preferences and their plans.
- They might already have the necessary knowledge – students do not study because they don't to or because they think they do not need to. This research showed that even when students think they have the necessary knowledge, especially from the previous run of the course, and they might decide not to prepare for the first TMA, it might negatively influence their subsequent progress.

When student is predicted as confidently Submit/Pass:

- Students might study more than one module and may decide to defer a module in favour of another one.
- Students might defer the module because their loan did not get through. Although this should now be included in the predictive model, it is better to be vigilant.
- An unexpected event might happen or a student might realise that their family responsibilities are too high, and they have to defer the module.

5 Conclusions

The project focused on the mixed-method analysis of machine learning errors of Predictive Learning Analytics models used in OUAnalyse for the first TMA in Level 1 STEM modules. As expected, not all the errors could be explained by the existing data (i.e. the quantitative analysis). The most explaining attribute turned out to be students' future change in VLE behaviour, after the predictions were generated. interviews with 27 students helped us to better understand the causes of this change and why the models failed. We identified data that are now captured to improve the predictions, suggested new data that might be collected and also proposed some intervention strategies that we can learn from these errors. Findings show that there are still unexpected circumstances which occur during study such as changes in family and/or work responsibilities, unexpected health issues and computer problems which cannot be accounted for through using PLA. To address these issues, what is needed is good communication with tutors; clear guidance; tutorial participation; peer support and advice from student support services. Future research should focus on validating the findings across all faculties and also in later TMAs, taking into account the improvements in data collections proposed in this research.

List of deliverables

[1] Hlosta, M.; Zdrahal, Z.; Bayer, V. and Herodotou, C. (2020). **Why Predictions of At-Risk Students Are Not 100% Accurate? Showing Patterns in False Positive and False Negative Predictions.** In: Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK20), 23-27 Mar 2020, Frankfurt am Main, Germany.
Accepted in Learning Analytics and Knowledge conference and presented virtually in March 2020 – the paper is focused around the first part of the project, the quantitative explanation of the errors.

[2] Hlosta, M.; Papathoma, T. Herodotou, C. (2020). **Explaining Errors in Predictions of At-Risk Students in Distance Learning Education.** In: Artificial Intelligence in Education, Lecture Notes in Computer Science (LNCS), Springer, pp. 119–123.

Accepted in AIED'20 conference and presented virtually in July 2020 – the paper mostly tackles the second part of the project, covering 12 out of 27 interviews that were available during the submission.

Figures and tables

Figure 1 - The schema of the methodology - quantitative analysis followed by the interviews.

References

1. Calvert C. (2017): Succeeding Against the Odds: eSTEeM Final Report.

2. Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., Hlosta, M. (2019): A large-scale implementation of Predictive Learning Analytics in Higher Education: The teachers' role and perspective. Educational Technology Research and Development.

3. Herodotou, C., Rienties, B., Hlosta, M., Boroowa, A., Mangafa, C. & Zdrahal, Z. (2020). The scalable implementation of predictive learning analytics at a distance learning university: Insights from a longitudinal case study. Internet and Higher Education. URL: http://oro.open.ac.uk/68953/

4. Herodotou, C., Hlosta, M., Boroowa, A., Rienties, B., Zdrahal, Z. & Mangafa, C. (2019). Empowering online teachers through predictive learning analytics, In: British Journal of Educational Technology (BJET), 50(6) pp. 3064–3079. URL: http://oro.open.ac.uk/62192/

5. Herodotou, C., et al (2017). Implementing predictive learning analytics on a large scale: the teacher's perspective. in Proceedings of the seventh international learning analytics & knowledge conference.

6. Shum S.B. (2020, July 14): Should predictive models of student outcome be "colour-blind"?, URL: http://simon.buckinghamshum.net/2020/07/should-predictive-models-of-student-outcome-be-colour-blind (Accessed: 01 Oct 2020).

University approval processes

- *SRPP/SSPP – Approval from the Student Research Project Panel/Staff Survey Project Panel was obtained according to the Open University's code of practice and procedures before embarking on this project. Application number 2019/089*
- *Ethical review – An ethical review was obtained according to the Open University's code of practice and procedures before embarking on this project. Reference number HREC/3335/Hlosta*
- *Data Protection Impact Assessment/Compliance Check – A Data Protection Impact Assessment/Compliance Check was obtained according to the Open University's code of practice and procedures before embarking on this project. Data Protection registration number* 28 04 012

Appendices

Appendix A – Metrics for your project

Appendix B – Confidential Commentary (attached separately)

Appendix C – Interview Schedule

Appendix A – Metrics for your project

Project staff	
Number of academic, academic-related staff who contributed to the project	5
Number of days spent working on the project for all staff involved, including the project lead(s)	38 days *Anna 5days, Tina 6 days,* *Martin 20 days, Zdenek 1 day, Vaclav 1 day*
Number of ALs and number of days contribution to the project	1
Number of students involved as co-researchers/co-collaborators on the project and any student incentives provided	1 (the involved PhD student is also an AL)
Student survey data (if applicable)	
Number of students surveyed	N/A
Number of student respondents	N/A
Student interview data (if applicable)	
Number of students interviewed	27
Student focus group data (if applicable)	
Number of students involved either as interviewers or interviewees	N/A (they were not focus groups)
AL survey data (if applicable)	
Number of ALs surveyed	N/A
Number of AL respondents	N/A
AL interview data (if applicable)	
Number of ALs interviewed	N/A
AL focus group data	
Number of ALs involved either as interviewers or interviewees	N/A (1 tutor was involved in online interviews)

Appendix B – Confidential Commentary

attached in a separate file