Analysing performace of first year engineering students

Zdenek Zdrahal
KMi, The Open University
Milton Keynes, UK
CIIRC, CTU in Prague
Prague, CZ
zdenek.zdrahal@open.ac.uk

Martin Hlosta
KMi, The Open University
Milton Keynes, UK
martin.hlosta@open.ac.uk

Jakub Kuzilek
KMi, The Open University
Milton Keynes, UK
CIIRC, CTU in Prague
Prague, CZ
jakub.kuzilek@open.ac.uk

Abstract

Too many students in the engineering disciplines drop out before the end of their higher education degree. This problem is grave, especially for first year university students. In this paper we analyse how students earn credits required for the successful completion of their first study year. In a case study of a European technical university with traditional classroom based education, we identify three groups of students: those who pass, those who earn only enough credits for staying in the program and those who fail. Important patterns can be found at start of the first exam period. We present a simple prediction model that identifies students who may benefit from early additional support, which would increase their chances to successfully complete the first study year and improves the retention rate of the university. The results are evaluated in two consecutive academic years.

1 Introduction

In some EU countries between 20% and 54% of students fail to complete their degrees [Qui13]. In distance education, the percentage of students who fail to complete the degree is about 78% [Sim10]. This paper analyses patterns of behaviour exhibited by a cohort of first year students, with the aim of discovering 'students at risk of failing' as early as possible, so that they can receive suitable support [AAC+14]. An anonymised dataset has been taken from a European technical university, which offers a three-year bachelor program, followed by two-year masters program in mechanical engineering. The education is organised in a traditional classroom-based manner involving lectures and tutorials. The selected dataset consists of all 943 students registered in the first year of the bachelors program in the academic year 2013/2014. This data was used for building the prediction model and for verification. Then, the model constructed from 2013/2014 data was applied on a new cohort of 1,077 first year students in the 2014/2015 academic year.

2 Study Organisation

2.1 Academic year

The academic year is divided into a winter and summer term each 13 weeks long and each followed by six week long examination periods. Between the end of the winter term and the start of the winter exam period there is

In: A. Wolff, Z. Zdrahal (eds.): Proceedings of the Data literacy for Learning Analytics Workshop, LAK 2016 Conference, Edinburgh, UK, 26-Apr-2016, published at http://ceur-ws.org

one week Christmas break. Prior to the start of winter and summer exam periods, still within the corresponding term, there are early exam periods of about three weeks. The summer exam period is extended by two more weeks after the summer holiday. Students can earn European Credit Transfer and Accumulation System (ECTS) credits in 21 courses, nine of which belong to the Science, Technology, Engineering and Mathematics (STEM) group. These are the most important courses for the qualification and therefore are rewarded by more credits [Che13]. In the winter term, students can earn up to 35 ECTS credits of which 21 are from STEM courses, in summer term it is 38 credits of which 26 are from STEM courses.

2.2 Progression rules

Based on their performance, the student can achieve one of three results:

- Pass – if the student earns at least 30 credits for the winter term subjects and 60 credits in total during the academic year. Out of the 30 credits for the winter term, the student must earn at least 15 credit before the end of the winter term exam period. Such students successfully complete the academic year. Students who pass have to earn all available credits from STEM subjects.
- Continuing – if the student earns more than 15 credits in the winter term and between 30 and 59 credits in total during the academic year. In order to successfully complete the first year, they must take and pass the failed or missing courses to earn the required credits.
- Fail – if the student earns fewer than 15 credits in the winter term, or fewer than 30 credits in total during the academic year, they are deregistered from the program.

In the analysed academic year 2013/14, the cohort of first year students achieved the following results:

- Pass 238 students, i.e. 25% of the registered students
- Continuing 205 students, i.e. 22% of the registered students
- Fail 500 students, i.e. 53% of the registered students

This progression rate is unsatisfactory both for students and for the institution. In order to propose possible improvements we need a deeper insight into the behaviour of these groups.

3 Analysing study patterns

First of all, we have to realize that the assignment of each student to the *pass*, *continue* or *fail* group is done at the end of the academic year. Though some students may irreversible fail early in the year (e.g. if they do not earn 15 credits in the winter term), many students fluctuate between *fail*, *continue* and *pass* groups. The dynamics of these groups classified at the end of academic year is shown in Figure 1. On the horizontal axis are dates in the study year with important milestones highlighted by a vertical line. The vertical axis shows the average number of credits for each class.

The total number of ECTS credits for each group is shown in full lines, the STEM credits are depicted by dashed lines. Periods of the academic year are denoted as follows: (a) winter early exam period as a part of the winter term, (b) Christmas break, (c) winter exam period, (d)+(e) summer term, (e) summer early exam period, (f) summer exam period, (g) summer holiday, (h) end of the academic year. The beginning of the winter term is not shown as it is not important for our analysis.

3.1 Predicting the final result from the result of the winter term

When the winter exam period is over and the winter term results are known, it is possible to provide a crude estimate of the overall progression. The students who have not earned 15 credits in the winter term have definitely failed and are deregistered. In the presented case there are 446 such students. The rest of the cohort continues studying to finish at the end of the academic year in one of the groups: *pass*, *continue* or *fail*.

The probability of a different final result, depending on the number of credits earned at the end of the winter exam period is shown in Figure 2. According to the rules, all students with fewer credits than 15 have failed. Arguably, some students who failed could have been saved if action was taken earlier. It is obvious in Figure 1 that the differentiation into *pass*, *continue* and *fail* groups exists already at the start of the winter exam period.

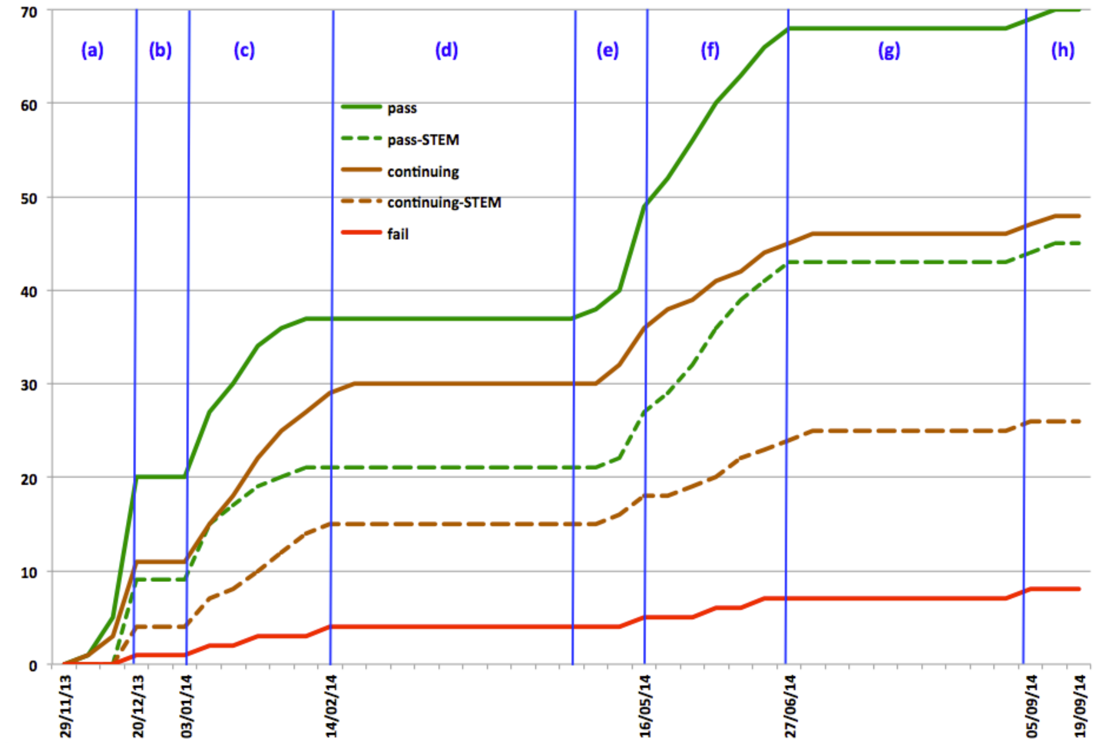

Figure 1: Average number of credits earned by different groups in time

The credits earned at the start of the winter exam period are awarded both for early exams but also for activities, assignments, and test carried out during winter term.

The pattern found without any exception in the data is that all students without any credit at the start of the winter exam period fail. This rule applies, though there are plenty of opportunities to earn the sufficient number of credits during the winter exam period to move the student from the *fail* status at least to the *continue* one.

If students at risk can be identified prior to the start of the regular exam period, it might be possible to provide them with additional support. For example, one of the important challenges for the first year students is how to organize their preparation for the difficult STEM exams with high credit values to maximise the total number of credits earned in the winter term. The university is prepared to assign a tutor to each small student group to support them through this period, if there is a chance that this effort is efficient and meaningful.

3.2 Selecting students for interventions

The predictive model for selecting students who may need additional support is based on the assumption that the pattern of student behavior persists across different school years; therefore the results calculated from historical data are applicable in the future. Our goal is to select the group of at-risk students based on the knowledge of the number of credits the students earned at the start of the winter exam period (or any other fixed date) and the final classification of students into the above mentioned three groups.

Let $n(x)$ be the number of students who earned x credits, and N the total number of students. Similarly, we denote $n(pass \cap x)$ the number of students who have earned (at the start of the winter exam period) exactly x credits and at the end of academic year belong to the *pass* group. We can define probability $P(x) = \frac{n(x)}{N}$, the joint probability $P(pass, x) = \frac{n(pass \cap x)}{N}$ and the conditional probability $P(pass|x) = \frac{P(pass,x)}{P(x)}$.

Similar probabilities can be defined for the *continue* and *fail* groups. $P(pass|x) + P(continue|x) + P(fail|x) = 1$ holds for any x. The conditional probabilities for the three groups and different number of credits at the start of exam period are shown in Figure 3. The goal is to estimate the allocation of students to the groups based on the early information about the credit distribution. Let us define the error caused by not assigning students with x credits to the *pass* group as $e(pass|x) = P(continue|x) + P(fail|x)$. Similarly, we can define the error function for the other two groups.

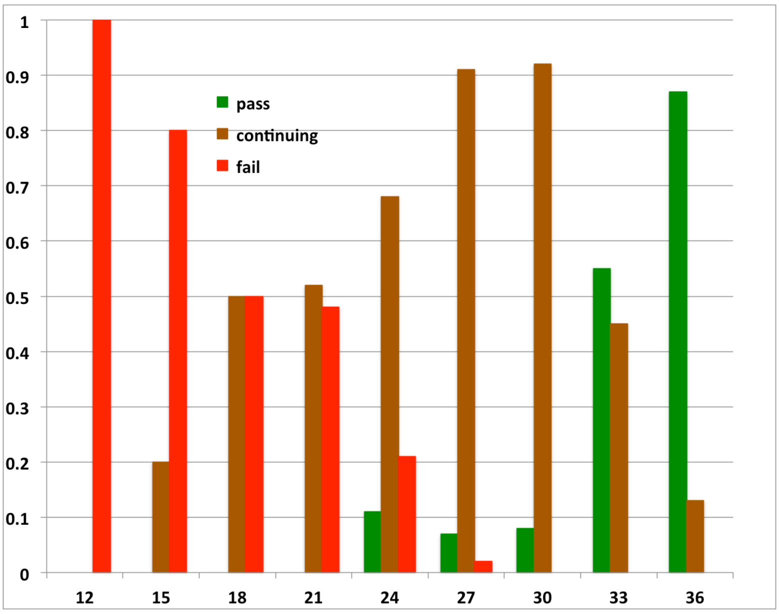

Figure 2: Probabilities of the final result based on the number of ECTS credits earned at the end of the winter term

It is reasonable to assume monotonicity: up to some x_1 the students will be classified as *fail*, then from x_1 to x_2 as *continue* and above x_2 as *pass*. The values x_1 andx_2 are selected to minimize the expression for error:

$$E(x_1, x_2) = e_{fail}(x_1) + e_{cont}(x_1, x_2) + e_{pass}(x_2),$$

where:

$$e_{fail}(x_1) = \sum_{0}^{x_1} e(fail|x),$$

$$e_{cont}(x_1, x_2) = \sum_{x_1}^{x_2} e(continue|x),$$

$$e_{pass}(x_2) = \sum_{x_2}^{72} e(pass|x).$$

In Figure 3, the values x_1 andx_2 are shown as blue vertical lines. The value 72 in the last sum is the highest number of credits that can be earned. The terms measure misclassification of the *fail*, *continue* and *pass* classes, respectively. In Figure 3, some values of x are missing. There is no course awarded by 1 credit and therefore value $x = 1$ is not in the histogram. The combinations 26, 31 and 33 could be achieved but do not exist in the analysed dataset. The minimum of $E(x_1, x_2)$ has been achieved for $x_1 = 6$ and $x_2 = 20$. The error $e_{cont}(x_1, x_2)$ of the *continue* class contributes by about 50% to the value $E(x_1, x_2)$.

4 Evaluation of predictions

Two types of evaluations were performed in order to test the performance of the models:

1. Applying the prediction model on the 2013/14 data used for its training to verify whether the models captures the key properties of the dataset.

2. Applying the prediction model built from 2013/14 data on the 2014/15 cohort to find out how the identified patterns transfer between consecutive years.

Precision of the prediction is calculated as the ratio of the number of students predicted correctly to the number of all students predicted as being member of the class. *Recall* is the ratio of the number of students correctly predicted to the number of students in the category.

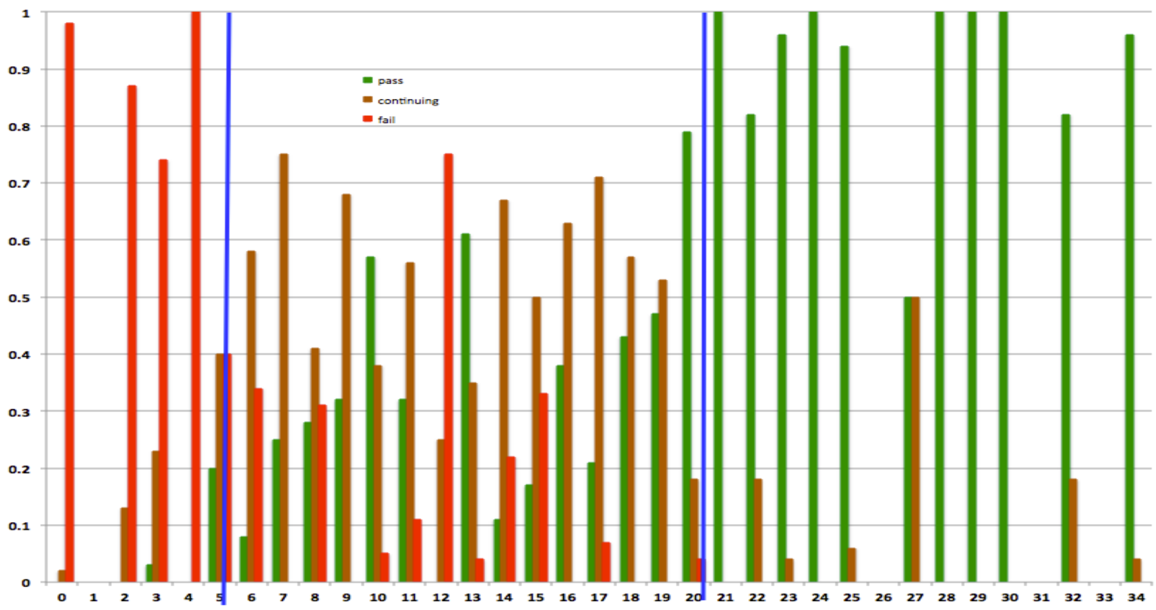

Figure 3: Probabilities of final results for different number of credits earned at the start of 2013/14 winter exam period

4.1 Evaluating 2013/14 cohort

The predictions for the 2013/14, cohort from which the predictive model was built, are as follows: In class predicted as *fail* ($x < 6$), 100% students were predicted correctly. In the class predicted as *continue* ($6 \leq x \leq 20$), 50% students were at the end of the academic year in the class *continue*, 31% were in *pass* and 19% students failed. In the class predicted as *pass* ($x > 20$), 92% students were in the class *pass*, 8% were continuing and no student in reality failed. This result is certainly encouraging because it indicates that this simple predictive model can in a very early stage of the exam period predict with good accuracy the future student outcome. However, we should be aware of the fact that the training and testing datasets were identical, and at the moment of predictions the study results were already known.

4.2 Application on 2014/15 cohort

In order to test practical applicability, the predictive model constructed from 2013/14 data is applied to the 2014/15 cohort. In the analysed academic year 2014/15 the cohort of 1,077 first year students achieved the following results:

- Pass – 269 students, i.e. 25% of the registered students
- Continuing – 214 students, i.e. 20% of the registered students
- Fail – 594 students, i.e. 55% of the registered students out of which 60 students failed in the summer term.

The measures used for evaluation predictions are precision and recall. As an indicator of sensitivity with respect to the calculated border values x_1 and x_2 we have evaluated also the nearest lower and higher values. The values of precision and recall for the three predicted classes are shown in Table 1, Table 2 and Table 3.

Table 1: Predictions for the *fail* class

	$x_1 = 4$	$\mathbf{x_1 = 5}$	$x_1 = 6$
precision	0.915	**0.841**	0.826
recall	0.936	**0.956**	0.970

Table 2: Predictions for the *continue* class

$x_1 = 4$	$x_2 = 19$	$x_2 = 20$	$x_2 = 21$
precision	0.490	0.468	0.464
recall	0.822	0.855	0.859
$\mathbf{x_1 = 5}$	$x_2 = 19$	$\mathbf{x_2 = 20}$	$x_2 = 21$
precision	0.490	**0.470**	0.467
recall	0.822	**0.855**	0.860
$x_1 = 6$	$x_2 = 19$	$x_2 = 20$	$x_2 = 21$
precision	0.489	0.463	0.459
recall	0.677	0.710	0.714

Table 3: Predictions for the *pass* class

	$x_2 = 19$	$\mathbf{x_2 = 20}$	$x_2 = 21$
precision	0.900	**0.910**	0.940
recall	0.600	**0.590**	0.490

The predictions of the *fail* and *pass* classes are very accurate, the precision of the *continue* class is significantly lower. Moreover, the precision of the class predicted as *fail* is negatively affected by the 60 students who fail in the summer term but in the winter term behave as *continue* or even as *pass*.

5 Conclusions

The overall result is in accordance with our intuition when we look at Figure 3. For the original goal of our analysis the fact that the students belonged eventually to the *pass* class have been misclassified as *continue* is not a significant problem. These students, are close to the border between the *continue* and *pass* category, and an intervention provided by tutors may help them to successfully complete the first year, i.e. to *pass*. The important outcome is the high precision, and especially high recall, for students classified as *fail*. These are the most vulnerable students and the university wants to offer them a helping hand. Those who failed may need only a little help and they will be able to continue.

The most important finding is that it is possible to predict the final result of students of traditional classroom-based university from their performance at the beginning of the study year, prior to the first main exam period.

5.1 Future work

The described approach makes use only of the number of credits earned at the beginning of the winter exam period. There is a large group classified as potentially failing. By analysing data at the end of the winter exam period we have found that a significant number of these students do not earn any credits during the exam period. Since we have also their demographic data including their age, gender, the institution, type and the date of completion of their pre-university education, the future work will concentrate on mining interesting patterns from these data. The results may, for example, indicate that these students need a short course prior to their start of their university studies.

The results of the presented analysis have been assessed by the academics from the university that provided the data. In January 2016, the described method has been applied on the current cohort and the university intends to provide individual support for the at-risk students. At the time of writing this paper, the final impact of predictions and university interventions have not been fully evaluated, but the initial results suggest that the number of students who fail in the 2015/16 winter term decreased by about 24% compared to study year 2014/15.

References

[AAC+14] Everaldo Aguiar, G Alex Ambrose, Nitesh V Chawla, Victoria Goodrich, and Jay Brockman. Engagement vs performance: Using electronic portfolios to predict first semester engineering student persistence. volume 1, pages 7–33. UTS ePRESS, 2014.

[Che13] Xianglei Chen. Stem attrition: College students' paths into and out of stem fields. statistical analysis report. nces 2014-001. *National Center for Education Statistics*, 2013.

[Qui13] J. Quinn. Drop-out and completion in higher education in europe among students from under-represented groups. *European Commission*, 2013.

[Sim10] Ormond Simpson. 22% - can we do better? supporting students at a distance. *http://www.ormondsimpson.com/USERIMAGES/Retention%20literature%20review.pdf*, 2010.