Students at Risk: Detection and Remediation

Irena Koprinska, Joshua Stretton and Kalina Yacef
School of Information Technologies, University of Sydney, Sydney, NSW 2006, Australia
{irena.koprinska, joshua.stretton, kalina.yacef}@sydney.edu.au

ABSTRACT

Detecting students at risk of failing is particularly useful and desirable when it is done in a timely manner and accompanied with practical information that can help with remediation. In this paper we investigate ways to detect students at risk of failing early in the semester for timely intervention. The context of our study is a first year computer programming course. We explore whether the use of several student data sources can improve the process: submission steps and outcomes in an automatic marking system that provides instant feedback, student activity in the discussion forum Piazza and assessment marks during the semester. We built a decision tree classifier that is able to predict whether students will pass or fail their final exam with an accuracy of 87% mid semester. The obtained rules are useful and actionable for teachers and students, and can be used to drive remediation.

Keywords

Student performance prediction; classification of failing and passing students; automatic grading system; discussion board; assessment and feedback.

1. INTRODUCTION

Computer programming is an essential skill for software engineers and computer scientists, and also an increasingly required skill for graduates of many other disciplines, such as science, medicine, economics and business. Key factors in how well a person will learn programming include regular practice, as well as quick and efficient correction of mistakes and misconceptions. This means that students must be provided with tools that allow them not only to practice their programming skills but also to receive timely and useful feedback, which can be challenging, especially for large introductory computer programming courses. Lack of regular practice and sufficient feedback, often leads to students becoming uninterested or disheartened, and giving up learning to program.

Innovative technology-enhanced teaching and learning tools can help to solve this problem. In our introductory programming courses, we use a combination of an automatic marking and instant feedback system (PASTA) and a sophisticated discussion board (Piazza). These tools not only provide a semi-independent platform for students to build and test their knowledge, but also the opportunity for useful data collection and analysis, that can be used to improve teaching and learning.

In this paper we describe how data collected from these two sources, together with data from assessment marks, can be used to identify students who are at risk of failing and need more careful guidance, early enough so that remediation is possible. To illustrate this we use data from a large first year programming course. Specifically, the goal of this study is three-fold:

(i) to investigate whether students at risk of failing can be identified early enough in the semester for timely intervention, using machine learning prediction methods and information from three different sources: automatic marking system (PASTA), discussion board (Piazza) and assessment marks;

(ii) to investigate whether the information from the automatic marking system and discussion board helps improve the predictive accuracy, in comparison to just using the assessment marks;

(iii) to investigate how useful and actionable the produced rules are for remediation.

2. DATA SOURCES

An important characteristic of our study is that it triangulates data from three different sources that contain information not only about student performance, but also about student activities. Each source offers useful perspectives on student learning: progression in code writing and diagnostic (PASTA), interaction and engagement (Piazza), student performance (assessment marks).

PASTA is an automatic marking and feedback system developed in our school. It allows students to submit their solution for an assessment task online, checks this solution against public and hidden tests set by the teacher and provides immediate feedback to the student about which public tests were passed and failed. Students can then correct their mistakes and resubmit until all these public tests are passed. Feedback about the hidden tests is released when marking is completed, along with manual feedback.

The use of PASTA has resulted in better student engagement and improved learning, because of the instant feedback and multiple submissions. The PASTA data contains, for each task and student, all sequences of assessment submissions, the tests that were passed and failed (and why), the time stamps and mark obtained.

Piazza (www.piazza.com) is a mix of discussion board and wiki, allowing students and teachers to post notes, ask and answer questions individually or collaboratively. It was developed with the aim to connect students and promote classroom engagement. The Piazza data contains, for each student, the number of questions asked, answered and viewed, and the time and content of the posts.

The third data source includes all assessment marks during the semester and the final exam mark and is described in Sec.4.

3. PREVIOUS WORK

Previous work on predicting failure rate of students has been performed, normally by predicting exam grade just before the exam. Kotsiantis et al. [1] predicted final exam performance based on assignment marks throughout the semester in a distance education environment. This prediction was performed only at the end of the semester, and the attributes used would not allow for a mid-semester prediction. They achieved an accuracy of 79% in predicting the final exam grade using an ensemble classifier.

Romero et al. [2] predicted final student marks based on Moodle usage data - the number of: quizzes passed and failed, assignments done, messages sent and read on the discussion board, and also the time spent on the assignments, quizzes and discussion board. They measured the geometric mean of the accuracies per class, which is an appropriate measure for imbalanced datasets as theirs, and achieved 67% with decisions trees. More recently, in [3] the same group investigated predicting the student grade (pass or fail) based on the student participation in a discussion forum, achieving accuracy of 75% using data collected in the middle of the semester and 90% using data collected at the end of the semester

Similar to student failure rates, student dropout rates have been studied, using a variety of assessment and non-assessment attributes. Agnihotri and Ott [4] predicted the likelihood of students dropping out of university after their first semester based on data provided such as admission information, placement tests and financial information. They were able to predict the retention of students with recall of 73% and precision 54%. Lykourentzou et al. [5] predicted dropout rate of students in an e-learning course environment, using the learning management system's extensive logs. They use machine learning techniques to achieve a 75-85% accuracy in the early sections of the course, and 97-100% accuracy in the final sections.

In this paper we extend previous work on predicting the students at risk of failing by using data from an automatic marking system and an advanced discussion board, in addition to assessment marks, from a computer programming course. We show how to define useful attributes from each data source, investigate if the student traces on the automatic marking system and discussion board help to improve predictive accuracy, and analyse how useful the prediction rules are for driving remediation.

4. CONTEXT OF THE STUDY

The study was conducted in the context of a large first year computer programming course with 223 students.

4.1 Assessment components

The six assessment components are summarised in Table 1.

Table 1. Assessment components

Homeworks	10	Weekly	Marks, Piazza
Task 1	2	Week 4	Marks, PASTA, Piazza
Task 2	6	Week 6	Marks, PASTA, Piazza
Practical test	16	Week 7	Marks, Piazza
Assignment	16	Week 12	Marks, PASTA, Piazza
Exam	50	Exam period	Marks, Piazza

The weekly homeworks were due before the computer lab and included multiple choice questions mainly requiring reading and understanding code. Their goal was to prepare students for the lab. The two tasks and assignment were programming assessments, with increasing level of difficulty, submitted via PASTA. Students were provided with skeleton code and required to complete the missing parts. The practical test involved writing code to solve five tasks with increasing difficulty levels in front of the computer. The exam, conducted at the end of the semester, was paper-based and required mainly writing code for solving problems. All assessment components were individual except for the assignment, where students had the choice of working individually or in pairs; 57% of students worked individually and 43% worked in pairs.

4.2 Predicted Variable

We predict the exam grade based on the marks of the other assessment components during the semester and the student activities on PASTA and Piazza. The two grades are defined as F (exam mark below 50, N=76), notF (exam mark of 50 and above, N=147). We chose the exam grade as a performance index because the exam: (i) is the major and most comprehensive assessment component, (ii) is conducted under strict conditions which minimises cheating, (iii) is independent of the other assessment components. The exam mark is highly correlated with the final mark (r=0.937).

4.3 Attributes

Table 2 summarises the student attributes that we defined to characterise student performance and activity.

Table 2. Attributes extracted from the three data sources

I. Assessment marks
homework_mark, task1_mark, task2_mark, prac_quiz_mark, assignment _mark (numeric) - mark (%) awarded for each assessment component
w7_homework_mark (numeric) – same as *homework_mark*, but only counting homeworks submitted before the end of week 7
II. PASTA activity – submission history
Starting and finishing times for assessments
task_start, task_finish, assignment_start, assignment_finish (numeric) – the average number of days before the due date that a student will start or finish the tasks or assignment
early_task, early_assignment (nominal, *yes/no*) - *yes* if the student starts the tasks faster than the average user; *no* otherwise
Multiple assignment submissions – improvement and consistency
marks_per_attempt_tasks, marks_per_attempt_assignments (numeric) – the average number of marks per PASTA submission of a task or assignment (including non-compiling submissions)
assignment_first_mark (numeric) - mark awarded for the student's first submission for the assignment
assignment_improvement (numeric) – the slope of the trendline of the student's assignment marks over each compiling submission; a larger number indicates rapid improvement
assignment_only_improvement (nominal, *yes/no*) - *yes* if the student's marks for compiling assignment submissions never decrease; *no* otherwise
assignment_consistency (nominal, multiple values) - goodness of fit (R^2) over each of the student's compiling submissions for the assignment, [-1, 1]; close to 1/-1 - linear increase/decrease in marks over submissions, close to 0 - random distribution of marks. Discretised as: *single* for single compiling submission, *none* for no assignment submission; *small/medium/high/very_ high* otherwise.
Pair work
pair_assignment (nominal, *yes/no*) - *yes* if the student worked in a pair for the assignment; *no* otherwise
Assignment submission statistics
single_submission (nominal, *yes/no/none*) - *yes* for one compiling assign. submission, *no* for more than one, *none* for no submission
assignment_total_submissions - total number of assignment submissions

assignment_compiling_submissions - number of compiling assignment submissions	
III. Piazza activity – views, questions and answers	
piazza_views, piazza_questions, piazza_answers (numeric) - number of posts viewed, questions asked or answered by the student on Piazza	
piazza_activity (numeric) calculated as: (*piazza_views* + 10*(*piazza_ questions* + *piazza_answers)* + 5*(*piazza_posts* - *piazza_answers*)) / *total_posts,* where *piazza_posts* is the total number of contributions made by the student (asking or answering a question, or posting a comment), and *total_posts* is the total number of question threads on Piazza	
piazza_active_viewer, piazza_active_questioner, piazza_active_answerer (nominal, *yes/no*) - *yes* if the student has an average or higher number of posts viewed, questions asked or questions answered; *no* otherwise	
w7_piazza_views, w7_piazza_questions, w7_piazza_answers, w7_piazza_activity (numeric) and *w7_piazza_active_viewer, w7_piazza_active_questioner, w7_piazza_active_answerer* (nominal, *yes/no*) - same as the respective attributes without prefix *w7,* but only counting Piazza's posts up until the end of week 7	

5. CAN WE PREDICT FAILING AND PASSING STUDENTS MID-SEMESTER?

We investigate whether we can predict accurately the students who will fail and pass the exam, based on the information available at two time points during the semester (and before the exam): in the middle of the semester (end of week 7) and at the end of the semester, just before the final exam (end of week 15). By the end of week 7, the students would have completed half of the homeworks, the two tasks and the practical test.

We built a Decision Tree (DT) classifier. One example in the data corresponds to one student and is described with the extracted attributes. An advantage of DTs is that the set of if-then rules they generate provides an explanation about the prediction which can be easily understood by teachers and students and directly applied.

Selecting appropriate attributes is very important for successful classification. Starting with the full set of attributes from Table 2, we used several methods for attribute subset selection [6] (manual and automatic such as correlation-based and wrapper, and combinations of them), before applying the DT algorithm. Although DTs have an inbuilt mechanism for attribute selection (only a subset of the attributes appear in the tree), their performance benefits from prior attribute subset selection. We report the best results. In all experiments, we used 10-fold stratified cross validation as an evaluation procedure.

Table 3 shows the accuracy results using data from all three sources and Figure 1 shows the generated DTs. The numbers in the brackets next to a leaf node in the trees give information about the coverage and correctness of the rule, e.g. (51/3) means that the rule covered 51 examples from the data, 3 of them we classified incorrectly and the remaining 48 were classified correctly.

Our results show that it is possible to predict the failing and passing students mid semester equally well as at the end of the semester – the two trees achieved the same accuracy, 87%. This accuracy is high enough to be useful in practical applications.

An examination of the confusion matrix shows that for the mid-semester tree the misclassifications are due to more failing students being classified as non-failing than the opposite. For the end of semester tree, there is no dominant misclassification type.

Table 3. Accuracy and number of rules using all three sources

	Marks + PASTA + Piazza
Mid sem. (week 7)	87.00 (8 rules)
End sem. (week 15)	87.00 (9 rules)

Figure 1 shows the two trees. Although equally accurate, the two DTs are different: they have different rules, using common and different attributes from the three sources. Both use *prac_quiz_mark* from assessment marks and *early_task* from PASTA but the other attributes are different, as shown below.

```
Mid semester (week 7)
prac_quiz_mark <= 81.875
|  prac_quiz_mark <= 54.375: F (51/3)
|  prac_quiz_mark > 54.375
|  |  w7_piazza_active_viewer = no
|  |  |  w7_homework_mark <= 70
|  |  |  |  early_task = no
|  |  |  |  |  task2_mark <= 70: notF (3)
|  |  |  |  |  task2_mark > 70: F (4/1)
|  |  |  |  early_task = yes: notF (2)
|  |  |  w7_homework_mark > 70: F (13/1)
|  |  w7_piazza_active_viewer = yes
|  |  |  task_finish <= 0: F (4)
|  |  |  task_finish > 0: notF (32/6)
prac_quiz_mark > 81.875: notF (114/3)

End of semester (week 15)
prac_quiz_mark <= 81.875
|  assignment_total_submissions <= 15
|  |  prac_quiz_mark <= 45: F (32)
|  |  prac_quiz_mark > 45
|  |  |  early_assignment = no
|  |  |  |  piazza_active_questioner = no
|  |  |  |  |  early_task = no: F (28/7)
|  |  |  |  |  early_task = yes: notF (3)
|  |  |  |  piazza_active_questioner = yes
|  |  |  |  |  assignment_finish <= 0: F (14/4)
|  |  |  |  |  assignment_finish > 0: notF (9/1)
|  |  |  early_assignment = yes: F (8/1)
|  assignment_total_submissions > 15
|  |  prac_quiz_mark <= 50.9375: F (2)
|  |  prac_quiz_mark > 50.9375: notF (13.0)
prac_quiz_mark > 81.875: notF (114.0/3.0)
```

Figure 1. DTs produced using all three data sources

The most important attribute in both cases is *prac_quiz_mark,* which is selected as a root of both trees and classifies correctly a large number of examples (e.g. If *prac_quiz_mark* > 81.875, then notF (114/3) in both DTs). This is expected as the practical quiz tests both theoretical and practical skills, and, similarly to the final exam, is conducted in a supervised environment, within time limits (in this case directly at the computer).

We highlight some interesting rules using attributes from PASTA and Piazza. From the mid-semester tree, the following rule shows the importance of following the discussions on Piazza, in addition to having relatively good marks on the practical quiz and homeworks:

If prac_quiz_mark ∈ (54.375, 81.875] &
w7_piazza_ active_viewer = no &
w7_homework_mark > 70
then F (13/1)

The following rule, also from the mid-semester tree, shows the importance of viewing the posts on Piazza, and also finishing the tasks earlier than on the due day, in addition to having a relatively good mark on the practical quiz:

If prac_quiz_mark ∈ (54.375, 81.875] &
w7_piazza_active_viewer = yes &
task_finish > 0
then notF (32/6)

From the end-of-semester tree, the following rule shows the importance of submitting the assignment and tasks early and asking questions on Piazza:

If prac_quiz_mark ∈ (45, 81.875] &
assignment_ total_submissions <= 15 &
early_assignment = no &
piazza_active_questioner = no & early_task = no
then F (28/7)

The rules in the two DTs generally make sense. The counter-intuitive ones (e.g. the two rules in the mid-semester tree that include *task2_mark* as their last condition and predicting F if *task2_mark* is greater than 70 and vice-versa) cover a very small number of instances (5/223 in this case) and represent coincidences in data rather than patterns.

Finally, both the mid-semester and end-of-semester trees are small (8 and 9 rules respectively), therefore easy to use by teachers.

In summary, the produced rules are compact, useful and actionable. They show the importance of the practical quiz, good practice such as starting and finishing assessments early and regularly reading the posts on the discussion board.

6. IS THE INFORMATION FROM PASTA AND PIAZZA USEFUL FOR PREDICTION?

We investigate if the information from the automatic marking system PASTA and the discussion board Piazza helps to improve the predictive accuracy, in comparison to just using the assessment marks. Table 4 shows the results when using marks only, and marks and PASTA only. The results using all three sources - marks, PASTA and Piazza - are given in Table 3.

Table 4. Accuracy and number of rules using assessment marks alone, and assessment marks and PASTA

	Marks	Marks + PASTA
Mid sem. (week 7)	84.30 (8 rules)	84.70 (13 rules)
End sem. (week 15)	82.96 (9 rules)	83.41 (14 rules)

We can see that using the assessment marks only provides a very good accuracy of 83-84%. The addition of information from the automatic grading system PASTA improves the accuracy by about 1%. Adding the information from the discussion board Piazza (Table 3) further improves the accuracy by about 3%, raising it to 87%. Hence, using information from PASTA and Piazza improves the predictive accuracy, in comparison to just using the assessment marks. However, this improvement is small in this case as the marks alone already provide high accuracy and there is a ceiling effect.

7. CONCLUSIONS

In this paper we investigate whether students at risk of failing can be identified early enough in the semester for timely intervention, using machine learning prediction methods and information from three different sources: automatic marking system, discussion board and assessment marks. We define useful attributes from each data source, to characterise student performance and activity. Using these attributes, we built a decision tree that achieved 87% accuracy in predicting whether students will pass or fail their final exam, from information available in the middle of the semester.

The produced rules are useful and actionable, and indicate the importance of starting and finishing assessments early and reading the posts on the discussion board, in addition to performing well on key assessment components. We show that using information from the automatic marking system and discussion board improves accuracy, compared to only using the assessment marks.

Our results can be used to detect students at risk of failing early in the semester and provide them with simple preventive feedback about remedial actions. Having an early flagging of students at risk also allows teachers of large classes to approach these students and provide more personalised remedial actions. At the beginning of the semester all students can also be made aware of the characteristics of the failing and passing students, to encourage better learning, good practice and improved student engagement.

An important aspect of our work is that we exploited different data sources capturing various facets of student activity during the course. This allowed the DT results to provide some concrete suggestions of remedial actions. The methodology we have followed can be applied to other contexts combining similar types of data sources. We are currently applying it to another very large course.

8. ACKNOWLEDGMENTS

This work was partially supported by the Human Centered Technology cluster at the University of Sydney.

9. REFERENCES

[1] Kotsiantis, S., Patriarcheas, K., and Xenos, M., 2010. A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based Systems 23*, 6, 529-535.

[2] Romero, C., Ventura, S., Espejo, P.G., and Hervás, C., 2008. Data Mining Algorithms to Classify Students. In *Proc. Int. Conference on Educational Data Mining (EDM)*, 8-17.

[3] Romero, C., López, M.-I., Luna, J.-M., and Ventura, S., 2013. Predicting students' final performance from participation in on-line discussion forums. *Computers & Education 68*, 458-472.

[4] Agnihotri, L., Ott, A., 2014: Building a student at-risk model: an end-to-end perspective. In *Proc. Int. Conference on Educational Data Mining Conference (EDM)*, 209-212.

[5] Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., and Loumos, V., 2009. Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education 53*, 3, 950-965.

[6] Witten, I.H., Frank, E., Hall, M.A., 2011. *Data Mining - Practical Machine Learning Tools and Technique*s. Morgan Kaufmann.