
Predicting Risk of Failure in Online Learning Platforms Using Machine Learning Algorithms for Modeling Students' Academic Performance

Fazel Keshtkar

Southeast Missouri State University, MO 63701 USA

FKESHTKAR@SEMO.EDU

Jordan Cowart

Southeast Missouri State University, MO 63701 USA

JRCOWART1S@SEMO.EDU

Andrew Crutcher

Southeast Missouri State University, MO 63701 USA

ALCRUTCHER1S@SEMO.EDU

Abstract

Online learning platforms such as Moodle and MOOC have become popular in higher education. These platforms provide information that are potentially useful in developing new student learning models and predicting outcomes, such as pass/fail and final grade prediction. Rather than grade book, another source of information provided by these platforms are in the form of metadata, namely data describing how students interact with the learning platform. In this paper, our goal is to develop a model by which we can predict and detect the students that are at risk of attaining a negative learning outcome (or in risk of failure) early in the semester. We evaluated classification machine learning algorithms on this problem. We conducted two problems: prediction of pass/fail(risk of failure) by predicting the final grade. In this research we used a mix of online-only courses and face-to-face(offline-only) courses dataset from computer science major in Moodle. Our results show that in the classification methods, our algorithms are finding useful patterns that we can use to predict risk of failure in students' outcome.

1. Introduction

Note: We should attempt to find more resources related to our work for both here and for the related works section.

Detecting students performance is one the most crucial task

Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 2016. JMLR: W&CP volume 48. Copyright 2016 by the author(s).

in online learning and educational data mining. Saying this, recently, some universities put priority to recognise students failures or success in courses (Strecht et al., 2015). Similarly our goal is to find out a model to not only predict the individual success and performance or failure, but also to conduct a model in online platform such as Moodle to help instructors, department head, etc and officials such as Registrare office, university admin, HR, etc. To have better understating of students failure or success we conduct to apply our model in both online only courses and offline only (face-to-face) courses. This will help us to determine the features and factors that associated with good outcomes and failure or low performance.

In term of application level, our model can be plugged into Moodle to help instructors and directors to predict students whom are at risk of failure. Therefore, they are able to find a solution or strategy to induce the risk of failure. On the other hand, curriculum committees or director may apply this model to design useful course handouts such as a syllabus. Moreover, university officials in higher level may find this model useful for better understanding of general trends and students' learning outcomes that can be turn out to implement useful pedagogical models.

Needless to say, each individual course implies different subjects and topics. This variance makes it very hard to obtain a single standalone model to determine academic failure or success for all subjects and majors. Therefore, for for our model we considered the aggregation of different features that is mostly common in all courses and can easily be applied to all courses in different subjects and majors. However, we apply our model for each course separately to find out the differences. We assume this will lead us to understand our model in various angles in general.

Our experiments and results that conducted on specific domain report satisfactory for offline course using Logistics

Model Trees (LMT). And for online courses, Decision trees and naive Bayes perform satisfactory. However, this should be noted that through these results we aim to predict the negative learning and risk of failure comparing with success. In addition regression algorithm shows promising outcome in classification of predicting risk of failure and success.

In addition, various learning methods have been applied to detect course results and academic performance with each learning algorithm performing differently with different datasets. (Romero & Ventura, 2013) The No Free Lunch Theorem states that it is difficult to choose a specific model or classification algorithm for this difficult task. (Ho & Basu, 2002) Therefore, discovering and applying appropriate methods for a specific dataset should yield a significant improvement in the effectiveness of a given learning algorithm. Our approach will be to apply learning algorithms based on metadata, as they have proven to be sufficient in addressing this problem. (Hmlinen & Vinni., 2010) These meta-learning algorithms have been studied by exploring metadata to adopt suitable algorithm based on data mining and machine learning algorithms. In (Song et al., 2012) research they propose to apply various classifications/clustering models, evaluation measurements, and statistical analysis test to predict the performance of students learning outcomes based.

The contributions of our research in this paper are as following: i) comparison of various machine learning algorithm to find out which fit best for our model and problem to determine students' learning outcomes in both failure and pass. ii) to evaluation of features that used in classification, to find out weather positive or negative results. This can be useful for detecting the final grades as well.

The rest of paper organized as following: Section 2 explores related work. Section 3 explains the methodology, dataset, and features. Section 4 presents the results followed by section 5 with the conclusions and future work.

2. Related Works

Wang has indicated a need for the examination of log analysis within online learning platforms, namely the examination of indicators of participation such as use of discussion forums, quiz completion rate, and video usage. Wen et al (Wang, 2014) have shown that a combination of sentiment analysis and survival analysis can somewhat reliably predict if a student will or will not drop from a MOOC class. Notably, they suggest that "...sentiment analysis should be used with caution in practice, especially when the text are noisy and in limited quantity. The research of Yudelson et al. indicates that finding and analyzing certain sub-populations within a student body can produce a better pre-

dictive model than that of examining the entire of the population (M. Yudelson, 2014).

The research of Coffrin et al (C. Coffrin, 2014) indicates that student interactivity and success during the first two weeks of a course strongly related to their outcomes at the end of the course. They also suggested that identifying students based on their patterns of engagement presents the opportunity of tailored feedback to these sub-populations.

Predicting students' performance has been an issue studied previously in educational data mining research in the context of student attrition (Zafra & Ventura, 2009; Zimmermann et al., 2011). Minaei-Bidgoli (Minaei-Bidgoli, 2003) used a combination of multiple classifiers to predict their final grade based on features extracted from logged data in an education web-based system.

Pittman (Pittman, 2008) performed a study to explore the effectiveness of data mining methods to identify students who are at risk of leaving a particular institution. Romero et al. (Romero, 2008; Romero & Ventura, 2013) focused on comparing different data mining methods and techniques for classifying students based on their Moodle (e-learning system) usage data and the final marks obtained in their respective programmes. The conclusion was that the most appropriate algorithm was decision trees for being accurate and comprehensible for instructors. Kabakchieva (Kabakchieva, 2013) also developed models for predicting student performance, based on their personal, pre-university and university performance characteristics. The highest accuracy is achieved with the neural network model, followed by the decision tree model and the kNN model.

Strech et al and Mendes-Moreira et al (Strech et al., 2014; 2015) work predicted the failure of students in university courses using an approach to group and merge interpretable models in order to replace them with more general ones. The results show that merging models grouped by scientific areas yields an improvement in prediction quality.

Keshtkar et al (Keshtkar et al., 2015) had analyzed student interaction based on their response to determine learning outcomes. Their goal was explore trends in how students interact with their course over the duration of a semester and, more specifically, how quickly they react to activities performed by their professor. They noticed that when a professor interacts with Moodle, students typically perform a lot more than one action. professor made in a continuous block of interactions.

3. Methodology

Our method contains the following steps process that were developed: The first step prepares the dataset from Moodle

platform that contains course grade book, and other meta-data information. The dataset are extracted into online and offline courses. These datasets were used to create classification models for each course using different algorithm. Then, we evaluated our models using different classification algorithms and are able to analyze and compare by the final process.

3.1. Dataset

Our dataset contains student and professor interactivity metadata and course gradebook on Moodle, an online learning platform. We extracted 11 programming courses over two semesters at Southeast Missouri State University. The following sections will contain an overview of our data and actions that we took to work it into a more manageable form. The target variables are pass for final in classification. The final grade in these dataset is stored as a alphabet (A, B, C, D, and F). We considered A to C as passing grade. We conducted a case study to analyse the pass and failure of students. The overview of dataset presented in Table 1. As it shown in Table 1 the number of instances that we analysed are 195 with 157 pass and 37 fail. These instances are extracted from different courses.

Table 1. Overview of dataset

Category (online/offline)	# Instances	Pass	Fail
Offline	88	64	23
Online	107	93	14
Total	195	157	37

3.2. Feature Selection

Our initial data was in the form of raw Moodle logs and gradebook for each class, with each log entry containing the following: course id, a unique student identifier, timestamp of activity, and the IP address from which the action originated.

We then used the timestamp to divide the classes into two week sessions to allow a narrower scope for our analysis. We then used this timestamp information paired with the unique student identifier to determine how many times a student would login during a given session and how many times a student interacted with the platform per login, as it was set to log a student off after fifteen minutes of inactivity. Using this information we were able to calculate the average and total number of interactions per login.

Using the IP address, we could determine an approximate location of where a student was interacting with the learning platform from. To protect student privacy, we scrubbed the IP address using the following method: If the IP originated from the school's campus, we marked the IP as ON,

and likewise OFF for origination from elsewhere.

Each course was either online or offline, where online denotes a course in which there is no physical meeting location and all of the course activity is entirely online and offline meaning that it is not online. Because the rates of interaction were drastically different between the two groups, we decided to analyze online and offline courses separately. Furthermore, due to a somewhat low sample size for each of the classes (with an average of 23 students per class), we combined the class data into two larger datasets corresponding to which group they were in.

Our processed dataset contains the following features: average and total number of interactions per login session, whether an interaction was performed from on-campus or off, first exam score, and final course grade.

4. Experiments and Results

This section presents the results obtained by running experiments to train models for both classification and regression.

We propose that applying data mining techniques, namely classification, to the metadata from an online learning platform will allow us to develop a model that can detect potential for a negative learning outcome for a student early within a course semester. Because of this goal, we only analyze the first four two-week sessions (the first eight weeks) of the semester.

It also needs to be discussed what we consider a good model in this context. Because our eventual goal is to develop a model by which to provide feedback to students that are at risk for a negative learning outcome, we consider the False Positive rate to be of more importance than the True Positive rate. Our reasoning is this: providing feedback to a student that isn't at risk of failing is a less serious matter than not providing feedback to a student that is at risk. However, it is possible that providing feedback to a student that is not actually at risk could lead them to taking actions which ultimately undermine their end result, indicating that a high True Positive rate is still of great concern.

We ran our analysis using Weka, a collection of machine learning algorithms and tools (Hall et al., 2009). We created a baseline for each group utilizing the well established Naive Bayes classifier; in order to establish this baseline, we applied the classifier to all of our metadata for each of the first four two week sessions. Because our results were radically different for online and offline classes, the discussion of the results is broken into two sections. We will first go over the results obtained for the offline group, followed by the online.

4.1. Offline and Online Results

Table 2. Offline Baseline

Sess	Cls	Acc	Prc	Rec	FP
1	NB	0.65	0.77	0.66	0.29
2	NB	0.76	0.77	0.76	0.34
3	NB	0.75	0.75	0.75	0.37
4	NB	0.72	0.70	0.72	0.55

The baseline results already formed a somewhat decent model on their own, particularly the Session 1 baseline (see Table 2). While we felt a 29.1% False Positive rate was particularly acceptable, we were curious to see if we could keep such a rate while simultaneously increasing the True Positive Rate. Using the Weka tool InfoGainAttributeEval paired with Ranker, we set out to determine which attributes were more likely to contribute to a better model. This tool consistently indicated that Exam 1, the students' on/off campus designation, and average number of interactions contributed more to the model than other attributes.

Table 3. Best Offline Models; ALL - All attributes used. ECA - Exam 1 Score, On/Off Campus, Average Interactivity

Sess	Cls	Acc	Prc	Rec	FP
1	LMT (ALL)	0.86	0.86	0.86	0.25
2	LMT (ECA)	0.82	0.81	0.82	0.32
3	LMT (ALL)	0.85	0.85	0.85	0.30
4	LMT (ECA)	0.78	0.77	0.78	0.41

With the knowledge of which attributes contributed more towards a better model, we began testing classifiers with various combinations of attributes; the best of these results for each session are listed in Table 3. We were surprised that logistic model trees (LMT) seemed to consistently outperformed the other classifiers we tried, amongst these where Naive Bayes (NB), SMO, and J48. It's also important to note that the best results were again in Session 1 with an 86.2% Accuracy and 24.5% False Positive rate, as this supports the finding by that the first two weeks of class have the most significant impact on a student's learning outcome. Perhaps of more significance, this finding appears to carry over to metadata. Logistic Model trees (LMT) introduced by Landwehr et al (Landwehr et al., 2005) is an efficient and flexible approach for building logistic models and uses the well-known CART algorithm for pruning. LMT show that it is often more accurate than C4.5 decision trees and standalone logistic regression on real-world datasets, and competitive with boosted C4.5 trees. Like other tree induction methods, it does not require any

tuning of parameters. LMT produces a single tree containing binary splits on numeric attributes, multiway splits on nominal ones, and logistic regression models at the leaves, and the algorithm ensures that only relevant attributes are included in the latter. Another advantage of LMT is that it can be used for predicting negative learning outcomes, that fits our goal in this paper.

Table 4. Online Baseline

Sess	Cls	Acc	Prc	Rec	FP
1	NB	0.81	0.77	0.80	0.82
2	NB	0.83	0.82	0.83	0.63
3	NB	0.83	0.81	0.83	0.70
4	NB	0.80	0.84	0.80	0.46

As with the offline courses, we performed our baseline analysis using Naive Bayes with the results shown in Table 4. Unfortunately, the results obtained in modeling the offline courses did not carry over to the online ones. In contrast to the offline results, which obtained the best results in the first session, online obtained its best in the last session that we analyzed. Table 5 shows us that there is a fairly sharp decrease in the false positive rate, about 17%, between Session 4 (our best model) and Session 2 (our next best). It's possible that an increase in activity around the midterm generated better metadata for analysis, but this will need to be studied further. Ultimately, these results are unsatisfactory though it is possible that a better set of metadata may yield better results.

Table 5. Online Best Models; ALL - All attributes used.

Sess	Cls	Acc	Prc	Rec	FP
1	J48 (ALL)	0.82	0.80	0.82	0.76
2	J48 (ALL)	0.86	0.84	0.86	0.63
3	NB (ALL)	0.83	0.81	0.83	0.70
4	NB (ALL)	0.80	0.84	0.80	0.46

5. Conclusion and Future Work

Positive learning outcomes results were conducted in our approach where the aim was to determine whether an individual student has success or fail in a course. Our results on a classification model, where the goal is to determine the grade of the student in an online course, were not satisfactory compared to offline courses but beat the baseline in the first two sessions. However, our results show that using Logistic Model Tree (LMT) can help us to predict the negative learning performance; doing so, we are able to detect which students are at risk of failure in offline courses.

For future work, improvements could be made in automating our methodology. In addition, feature selection and feature weighting can be applied to our data, as it has been shown to yield favorable results in educational data. We aim to utilize more metadata features in our model, such as scholarship, gender, financial status, part-time, full-time, age, nationality, etc.

Although the current feature set used in the experiments provided some interesting results for the offline dataset, particularly with regards to LMT, the same did not happen with the online. Therefore, investigating an online model that incorporates some of these other metadata features, discussed above, may prove to be worthwhile.

Acknowledgements

This research is funded by Graduate Research Funding Grant (GRFC) at Southeast Missouri State University.

References

- C. Coffrin, et al. Visualizing patterns of student engagement and performance in moocs. *ACM Proceedings of the fourth international conference on learning analytics and knowledge*, 2014.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. The weka data mining software: An update. *SIGKDD Explorations*, 2009.
- Ho, T.K. and Basu, M. Complexity measures of supervised classification problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:289–300, 2002.
- Hmlinen, W. and Vinni., M. Classifiers for educational data mining. *Handbook of Educational Data Mining, Chapman ’—& ’ Hall Data Mining and Knowledge Discovery Series*, pp. 57 – 71, 2010.
- Kabakchieva, D. *Predicting Student Performance by Using Data Mining Methods for Classification*. *Cybernetics and Information Technologies*, 13(1):61 – 72, 2013.
- Keshkar, F., Cowart, J., Kingen, B., and Crutcher, A. *Analyzing Students Interaction Based on their Responses to Determine Learning Outcomes*. In 8th International Educational Data Mining Conference (EDM2015), pp: 588 – 589, 2015.
- Landwehr, Niels, Hall, Mark, and Frank, Eibe. Logistic model trees. *Machine Learning*, 59(1):161–205, 2005. ISSN 1573-0565. doi: 10.1007/s10994-005-0466-3. URL <http://dx.doi.org/10.1007/s10994-005-0466-3>.
- M. Yudelson, V. Michael, et al. Better data beat big data. *Proceedings of the 7th International Conference on Educational Data Mining*, 2014.
- Minaei-Bidgoli, B. *Predicting student performance: an application of data mining methods with an educational web-based system*. In 33rd ASEE/IEEE Frontiers in Education Conference, pages 1 – 6, 2003.
- Pittman, K. *Comparison of data mining techniques used to predict student retention*. PhD thesis, Nova Southeastern University, 2008.
- Romero, C. *Data mining algorithms to classify students*. In 1st International Educational Data Mining Conference (EDM08), 2008.
- Romero, C. and Ventura, S. Data mining in education. *WIREs Data Mining Knowl Discov*, 3(1):12–27, 2013.
- Song, Q., Wang, G., and Wang., C. Automatic recommendation of classification algorithms based on data set characteristics. *Pattern recognition*, 47:2672 – 2689, 2012.
- Strecht, P., Mendes-Moreira, J., and Soares, C. *Decision Trees: A Case Study in Predicting Student Performance*. In X. Luo, J. Yu, and Z. Li, editors, *Advanced Data Mining and Applications, Lecture Notes in Computer Science*, pages 535 – 548. Springer International Publishing, 2014.
- Strecht, P., Cruz, C., Soares, C., Mendes-Moreira, J., and Abreu, R. *A Comparative Study of Classification and Regression Algorithms for Modelling Students Academic Performance*. In 8th International Educational Data Mining Conference (EDM2015), pp 392 – 395, 2015.
- Wang, Y. Mooc learner motivation and learning pattern discovery. *The Proceedings of the 7th International Conference on Educational Data Mining*, 2014.
- Zafra, A. and Ventura, S. *Predicting student grades in learning management systems with multiple instance genetic programming*. International Working Group on Educational Data Mining, 2009.
- Zimmermann, J., Brodersen, K. H., Pellet, J.-P., August, E., and Buhmann, J. M. *Predicting graduate-level performance from undergraduate achievements*. In 4th International Educational Data Mining Conference (EDM11), pages 357 – 358, 2011.