Predicting and Understanding Success in an Innovation-Based Learning Course

Lauren Singelmann, Enrique Alvarez, Ellen Swartz, Ryan Striker, Mary Pearson, Dan Ewert
North Dakota State University
Lauren.N.Singelmann@ndsu.edu

ABSTRACT
In order to keep up with the rising demand for new and innovative solutions in an evolving world, an even greater importance is being placed on training engineers that can tackle big problems. However, the process of teaching engineering students to be innovative is not straightforward. There are multiple ways to demonstrate innovation and problem-solving abilities, meaning traditional educational data mining methods aren't always appropriate. To better understand the process of problem-solving and innovation, this work collected data from students working on innovation projects within a course and determined appropriate ways to gain information and insight from the data. Students wrote and categorized learning objectives in an online portal, which generated log data when they created, updated, and completed personal learning objectives and corresponding deliverables. Classification models that were both robust (ROC AUC $>$ .95) and interpretable were applied to both the language used in the objectives and the quantifiable features such as number of objectives, time of completing certain milestones, and number of deletions and edits. By extracting the most significant features, we are able to see which variables are most likely to lead to student success in innovation-based learning. This would aid instructors in offering impactful support to students or eventually lead to an online tutoring system. The conducted analysis will help students develop and grow throughout the innovation process in this course or in other open-ended problem-solving environments.

Keywords
Classification, open-ended learning, innovation, problem-solving

1. INTRODUCTION
Thomas Friedman describes the current era as the *Age of Accelerations*, the time at which technology, the climate, and globalization are all evolving at a rate like we've never seen before [5]. As these areas progress, engineers need to be able to identify and solve problems more quickly and effectively than ever before. ABET [1], the National Academies of Engineering [10], and experts in both engineering and education [12] all stress the growing importance for training engineers that can use their problem-solving skills to create new and innovative solutions. This work explores how students work on these skills and solve real-world problems in an Innovation-Based Learning (IBL) course. IBL students apply their content knowledge and skills to work on a real-world project with the goal of creating value external to the class. For example, successful students have presented their work at conferences, published papers, participated in invited outreach activities, or submitted invention disclosures. Students are required to write their own learning objectives and show evidence of work. However, because there are so many possible approaches to the course, predicting and understanding student success can be challenging.

In order to better understand what makes students successful in this type of course, data were collected from the online learning portal from the class. Because of the open-ended nature of the course, classification and knowledge discovery can be challenging. We wanted to build classifier models that were robust, but also interpretable in order to better predict and support future students in IBL-style courses. Therefore, three main research questions were explored:

RQ1: What feature sets and models work best for IBL data?
RQ2: How early can student success be predicted?
RQ3: What features are most likely to differentiate between top-performers and lower-performers?

Finding answers to all of these questions can help guide instruction in the course and potential development of an online tutoring system for innovation-based learning courses and other open-ended problem-solving environments. This paper will present literature about open-ended learning environments, give details about the course and data collected, elaborate on how the research questions will be tested, and share results and takeaways.

2. OPEN-ENDED LEARNING
Open-ended learning is a pedagogical approach that allows students to use their own motivations and approach to learn about the world [8]. Students are taking part in authentic problem-solving, practicing metacognition, and creating unique pathways through their learning. Examples of open-ended learning environments include computer pro-

Lauren Singelmann, Enrique Alvarez, Ellen Swartz, Ryan Striker, Mary Pearson and Dan Ewert "Predicting and Understanding Success in an Innovation-Based Learning Course" In: *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, Anna N. Rafferty, Jacob Whitehill, Violetta Cavalli-Sforza, and Cristobal Romero (eds.) 2020, pp. 662 - 666

gramming exercises, project-based learning, and inquiry-based learning. Educational research exists about the potential benefits of these experiences [11], but there are still gaps in understanding about how students progress through open-ended learning environments. Educational data mining (EDM) has shown great potential in being able to help shed light on student trajectories and habits within open-ended learning environments [20].

EDM has shown preliminary successes in open-ended learning environments such as learning computer programming [9], project-based learning courses [17], online tutoring platforms [3,4], learning-by-teaching platforms [6], and language tutoring systems [13]. Continuing to make strides in understanding learning in open-ended environments is imperative because it will be an important step in implementing evidence-based practices and assessment in these environments.

3. METHODS

3.1 Cardiovascular Engineering Course

Data were collected from an upper level cardiovascular engineering course at a medium-sized research university. Students learned about the main concepts of cardiovascular engineering including functional block diagrams of the heart, arterial systems, and ECG. The students were assessed on their ability to apply these concepts to a project they worked on during the semester. After identifying a project and a team, they wrote learning objectives and corresponding deliverables that would share what they needed to learn, when they would learn it, and how they would demonstrate it. Students could adjust their learning goals as needed, but they were expected to share their progress on their project multiple times during the semester [14]. To get an *A* in the course, students needed to work on an innovation project and achieve high external value. High external value is demonstrated by sharing your work outside the course and getting review from a subject-matter expert, e.g. publishing a paper, presenting at a conference, or submitting an invention disclosure [2, 19].

3.2 Online Learning Portal

A custom online learning management system (LMS) was created for students to keep track of their learning objectives and deliverables [18]. When students add a learning objective, they give it a title, description, assign it to a level of Bloom's Revised 3D Taxonomy [7], and categorize it using the list of objective categories in Appendix A. When adding a deliverable, students give it a title, description, level of external value, estimated completion time, and status (not started, in progress, or completed). An example of a learning objective and corresponding deliverables is shown in Figure 1. Every time a student adds, edits, or deletes an objective or deliverable, the action is recorded as a log entry, allowing us to see not only the completed products, but also early iterations of the students' objectives [16].

3.3 Data Set

28 students agreed to share their data during the semester. The average student logged approximately 8 objectives, 32 deliverables, and visited the platform more than 65 times during the semester. 17 students achieved high external value, 10 students worked on an innovation project but did not achieve high external value, and 1 student made some learning goals but did not complete any.

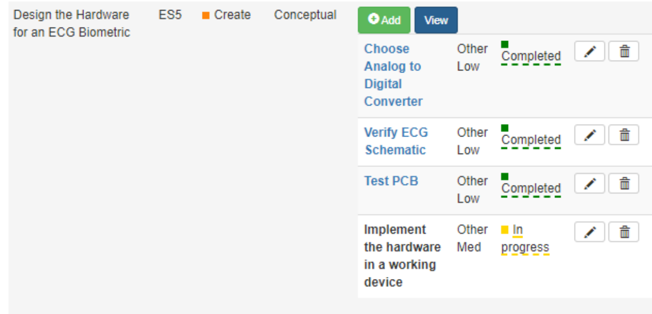

Figure 1: Example of collected learning objective and corresponding deliverables.

3.4 Feature Collection

Two main types of features were used and compared: quantitative data and text data. The quantitative features that were extracted from the data include countable features (e.g. number of planned learning objectives, number of logins, etc.), quarter-based progress (e.g. number of deliverables completed during quarter 2, number of learning objectives deleted during quarter 4, etc.), presence of the specific learning objectives as seed in Appendix A (e.g. presence of *Invention Disclosure* objective, number of *Fundamentals of Research* objectives, etc.), and the level of learning as defined by Bloom's Revised Taxonomy and the level of external value.

For the text data, all learning objective and deliverable titles and descriptions were extracted for each student. Using the scikit-learn library in Python, all the words that students wrote in their objectives and deliverables were tokenized, counted, and scaled.

4. EXPERIMENTS

4.1 Models and Feature Sets

In order to predict which students would achieve high external value during the course of the semester, three classifier models were tested: Support Vector Machine (SVM), Logistic Regression (LR), and K-Nearest Neighbors (KNN). These three models were chosen because they have some level of interpretability, an important feature in EDM [15]. In order for instructors to use the discovered information, they need to be able to understand where it was derived from. The baseline model was a Majority Class (MC) classifier.

In addition to comparing the models, both the text and quantitative features were compared. For each set, we also compared using all features to using the top K features. K was optimized and set at 24 for text and 15 for quantitative.

4.2 Evaluation Metrics

Each model was evaluated by calculating accuracy, recall, F1 score, and Area Under Receiver Operating Characteristic Curve (AUC). Accuracy is the proportion of correctly classified students to all students. Recall is the proportion of students that the model identified as not being on track to success to the number of total students that did not achieve high external value during the course. F1 score is a performance metric that takes the harmonic mean of precision and recall. AUC is the area under the Receiver Operating Characteristic (ROC) curve which shows how well the model can differentiate between the two classes. All models were evaluated using ten-fold cross validation.

4.3 Trajectory

In addition to exploring models that were developed by using each student's final learning objectives and deliverables, we were also able to explore how prediction power of the models changed during the course of the semester. Models were created using daily snapshots of all students to see when the model can begin predicting student success.

5. RESULTS

5.1 Comparing Models and Feature Sets

Table 2 shows the accuracy, recall, F1 score, and AUC for each of the models and feature sets explored. These classifiers used all available data during the semester. Almost all models performed better than the MC baseline test. The text features consistently performed better than the quantitative features, and using feature selection usually improved the model as well. The top models are SVM and LR, both using the top 24 text features. In addition to having low performance, the quantitative models are also difficult to assess in real time. The most relevant features of the quantitative models can give us some information, but they are not as helpful when making predictions. Therefore, we'll focus on using the text models moving forward.

Feature Type	Model	Accuracy	Recall	F1	AUC
Baseline	MC	.6	-	-	.5
All Text Features	SVM	.783	.85	.758	.831
	LR	**.883**	.85	**.866**	**.972**
	KNN	.583	**.95**	.533	.700
Top 24 Text Features	SVM	**.917**	**.85**	**.9**	.937
	LR	**.917**	**.85**	**.9**	**.952**
	KNN	.783	**.85**	.767	.832
All Quantitative Features	SVM	.7	.6	**.648**	**.704**
	LR	**.717**	.5	.612	.697
	KNN	.567	**.7**	.482	.523
Top 14 Quantitative Features	SVM	.667	.5	.563	**.851**
	LR	**.7**	.5	.597	.798
	KNN	.667	**.9**	**.615**	.65

Table 1: Performance metrics for each of the models using end of semester data

5.2 Exploring Model Trajectory

Because the model performs well at the midpoint in the semester, the next experiment explored at what point in the semester top-performers can be differentiated from lower-performers. All models used the 24 top text features. Figures 2 and 3 show the accuracy and AUC of the models over time, respectively. The SVM and LR models improve as the semester goes on, with the AUC for the models leveling out at about day 55. Therefore, the midpoint of the semester seems to be an appropriate time to use the model, but using it earlier might give mixed results.

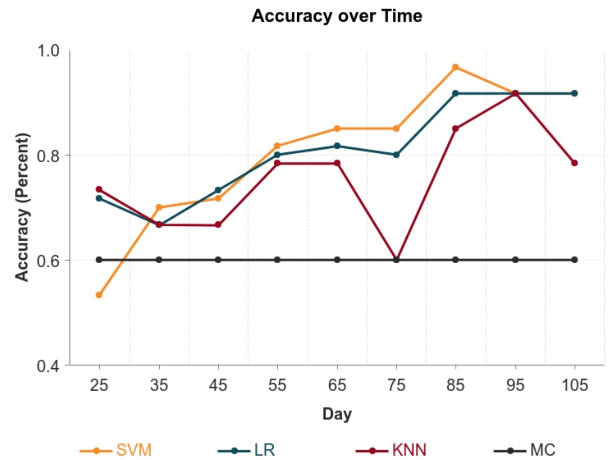

Figure 2: Accuracy of the text-based models over time compared with the baseline MC classifier

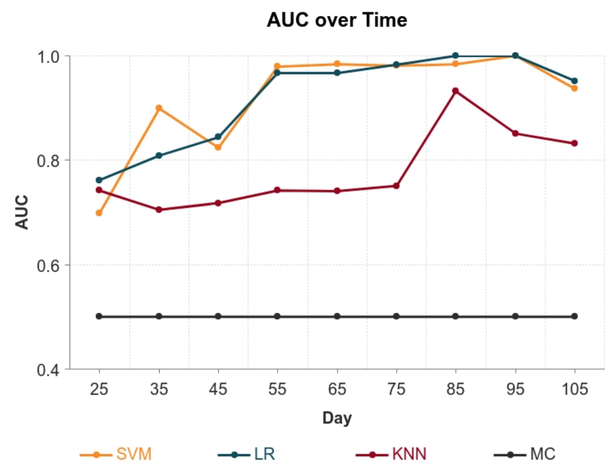

Figure 3: AUC of the text-based models over time compared with the baseline MC classifier

5.3 Knowledge Discovery

In order to better understand what features are most significant in predicting success, the most important features were extracted. By using linear classifier models instead of black-box models like neural networks and other deep-learning models, Chi-Square and the weights of each feature could be calculated. Chi-Square tells us which features are not independent of their classification, meaning they are more likely to differentiate between classes. The greater the Chi-Square value, the greater dependence on classification, meaning that feature is a strong differentiator. Weight can tell us which class a feature is more likely to be found in.

Figure 4 shows the 24 features with the largest Chi-Square value. If the word was more likely to be found in a low-performing student, the Chi-Square value was multiplied by -1 to allow for easier interpretation. The top words that differentiated low-performing students were *information*, *presentation*, *engineering*, *website*, *loops*, *review*, and *feedback*. The top words that differentiated high-performing students were *sensor*, *signal*, *model*, *device*, *idea*, and *symposium*.

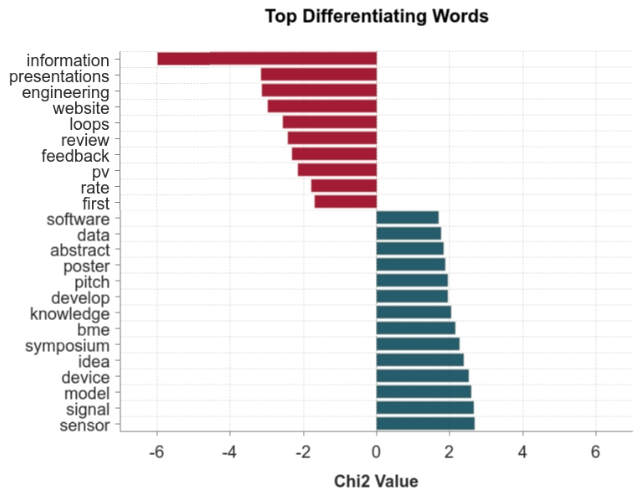

Figure 4: The Top 24 text features that differentiated the most between successful and unsuccessful students. Words with positive Chi-Square values were more associated with successful students. Words with negative Chi-Square values were more associated with unsuccessful students.

Table 4 shows the quantitative features that had the highest Chi-Square values. The weights were used to know which group the variable was more likely to be present in. Top students were more likely to have data analysis, data collection, journal manuscripts, and general Mechanisms of Research learning objectives. Unsuccessful students were more likely to have providing critique and outreach communication learning objectives.

Variable	Chi-Square	Group
Presence of MR4: Data analysis	3.882	Successful
Presence of RM3: Providing critique	3.091	Unsuccessful
Total number of Mechanisms of Research Learning Objectives	2.146	Successful
Presence of MR3: Data collection	1.941	Successful
Presence of PC5: Journal manuscript	1.941	Successful
Presence of PC7: Outreach communication	1.807	Unsuccessful

Table 2: Quantitative features with the highest Chi-Square values

6. DISCUSSION

6.1 Insights Gained

Unsurprisingly, top students were more likely to mention work on their abstracts, posters, pitches, and presence at the BME Symposium (an on-campus biomedical engineering conference). Low-performing students were more likely to have deliverables like websites and outreach activities. Although websites could be high impact deliverables, they can also just be a report of students' lower-level learning. For outreach activities, this can be interpreted broadly and could be outreach to a classmate or small group rather than a visit of high impact. In addition, successful students were more likely to have words related to the design process such as *idea*, *develop*, and *data*. Unsuccessful students were more likely to mention words like *information*, *presentations*, *review*, and *feedback*. We believe these words appeared in low-level students because they were activities required by the class. Therefore, top students did not see the need to write specific learning objectives about them, but lower performing students added them in an attempt to have more items logged.

6.2 Limitations

Just as the world around us is accelerating, so are our students. Therefore, these models will need to continue to evolve and improve as students change their approach to the class. Aiming for consistently high performing models is not a realistic goal for this work. Rather, we can use the knowledge discovery from these models to better understand how students move through these environments and aim to better support them.

6.3 Future Work

In addition to collecting data during more semesters and at more universities, we would also like to explore both clustering and sequential modeling moving forward. By clustering similar students and finding patterns that emerge in successful students in that cluster, we can give personalized feedback that allows students to find success while staying true to their own learning goals.

7. CONCLUSION

Modeling student learning in open-ended learning environments can be challenging, but SVM classifiers show potential in being able to predict which students will be successful in an IBL course. Models had accuracy of over 80% and AUC of over .95 by the midpoint in the semester. This accuracy increased to over 90% by the last few weeks of the semester. By using linear models, we could also gain insight as to what features differentiated between successful and unsuccessful students. Using these results can help instructors know which students could use extra support and lead to more understanding about how students progress through problem-solving environments in general. By understanding how to better support our students in the innovation process, we can foster the next generation of problem-solvers to take on the *Age of Accelerations*.

8. REFERENCES

[1] ABET. Criteria for accrediting engineering programs.

[2] E. Alvarez Vazquez, M. Pearson, L. Singelmann, R. Striker, and E. Swartz. Federal funding

opportunity announcements as a catalyst of students' projects in mooc environments. In *6th International Conference on Learning with MOOCs*. IEEE, 2019.

[3] I. Arroyo and B. P. Woolf. Inferring learning and attitudes from a bayesian network of log file data. In *AIED*, pages 33–40, 2005.

[4] F. Bouchet, R. Azevedo, J. S. Kinnebrew, and G. Biswas. Identifying students' characteristic learning behaviors in an intelligent tutoring system fostering self-regulated learning. *International Educational Data Mining Society*, 2012.

[5] T. L. Friedman. *Thank you for being late: an optimists guide to thriving in the age of accelerations*. Penguin Books, 2017.

[6] J. S. Kinnebrew, K. M. Loretz, and G. Biswas. A contextualized, differential sequence mining method to derive students' learning behavior patterns. *JEDM| Journal of Educational Data Mining*, 5(1):190–219, 2013.

[7] D. R. Krathwohl and L. W. Anderson. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman, 2009.

[8] S. M. Land. Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development*, 48(3):61–78, 2000.

[9] Y. Mao, R. Zhi, F. Khoshnevisan, T. Price, T. Barnes, and M. Chi. One minute is enough: Early prediction of student success and event-level difficulty during a novice programming task. *The 12th International Conference on Educational Data Mining*, pages 119–128, 2019.

[10] N. A. of Engineering. *The engineer of 2020: Visions of engineering in the new century*. National Academies Press, 2004.

[11] E. Panadero. A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology*, 8:422, 2017.

[12] H. J. Passow and C. H. Passow. What competencies should undergraduate engineering programs emphasize? a systematic review. *Journal of Engineering Education*, 106(3):475–526, 2017.

[13] P. I. Pavlik Jr. Mining the dynamics of student utility and strategy use during vocabulary learning. *JEDM| Journal of Educational Data Mining*, 5(1):39–71, 2013.

[14] M. Pearson, E. Swartz, R. Striker, E. Alvarez Vazquez, and L. Singelmann. Driving change using moocs in a blended and online learning environment. In *6th International Conference on Learning with MOOCs*. IEEE, 2019.

[15] C. Romero and S. Ventura. Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1):12–27, 2013.

[16] L. Singelmann, E. Swartz, M. Pearson, R. Striker, and E. Alvarez Vazquez. Design and development of a machine learning tool for an innovation-based learning mooc. In *6th International Conference on Learning with MOOCs*. IEEE, 2019.

[17] D. Spikol, E. Ruffaldi, G. Dabisias, and M. Cukurova. Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning*, 34(4):366–377, 2018.

[18] R. Stiker, M. Pearson, E. Swartz, L. Singelmann, and E. Alvarez Vazquez. 21st century syllabus: Aggregating electronic resources for innovation based learning. In *6th International Conference on Learning with MOOCs*. IEEE, 2019.

[19] E. Swartz, M. Pearson, R. Striker, L. Singelmann, and E. Alvarez Vazquez. Innovation-based learning on a massive scale. In *6th International Conference on Learning with MOOCs*. IEEE, 2019.

[20] P. H. Winne and R. S. Baker. The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *JEDM| Journal of Educational Data Mining*, 5(1):1–8, 2013.

APPENDIX

A. LEARNING OBJECTIVE CATEGORIES

Category	Code	Objective
Discipline-Specific Knowledge	DSK0	Cardiovascular Concepts
	DSK1	Learning in student's program
	DSK2	Learning in student's College
	DSK3	Learning outside of College
	DSK4	Freeform learning
Fundamentals of Research	FR1	Research method
	FR2	Literature review
	FR3	Experimental design
	FR4	Experimental equipment
	FR5	Intellectual merit
	FR6	Broader impact
	FR7	IRB/IACUC
	FR8	Lab safety
Mechanisms of Research	MR1	Statistics
	MR2	Experimental controls
	MR3	Data collection
	MR4	Data analysis
	MR5	Drawing conclusions
	MR6	Knowing nature of results
Professional Communication	PC1	Conference abstract
	PC2	Conference poster
	PC3	Conference presentation
	PC4	Proposal presentation
	PC5	Journal manuscript
	PC6	Standard operating procedure
	PC7	Outreach communication
	PC8	Invention disclosure
Research Mindset	RM1	Personal statement
	RM2	Receiving critique
	RM3	Providing critique
	RM4	Metacognition
	RM5	Establishing requirements
	RM6	Team conduct
	RM7	Mindset
Entrepreneurial Skills	ES1	Business model
	ES2	Customer communication
	ES3	Customer segment
	ES4	Value proposition
	ES5	Product evaluation

Table 3: List of all learning objective categories