Artificial Intelligence Based Model for Prediction of Students' Performance: A Case Study of Synchronous Online Courses During the COVID-19 Pandemic

Margaret Stadlman
Rowan University

Seyyed M. Salili
Rowan University
University of Pennsylvania

Ashish D. Borgaonkar *
New Jersey Institute of Technology

Amir K. Miri *
Rowan University
New Jersey Institute of Technology

* *Co-Corresponding Authors*

Abstract

Lack of student persistence and retention is significantly hurting the US in producing the required number of qualified graduates, especially in STEM fields. Although many factors contribute to students falling off track, one of the controllable factors is the identification of at-risk students followed by early intervention. Predicting the performance of students enables educators to single out struggling and highly talented students. Struggling students are often identified very late into an academic year, thus leaving little to no time for seeking consultation and determining the best course of action to improve performance. Some of such struggling students resort to dishonest means to catch up or make up at the last minute resulting in a higher number of academic integrity violations being observed and reported. Recently, the COVID-19 pandemic further corroborated the presence of such challenges. This research explores the possibility of using artificial intelligence to identify key elements in small datasets which could contribute to the development of a predictive student performance solution. A small set of data obtained through systematic data collection was used to train a predictive algorithm and aid in the analysis of in-class learning, which would lead to a viable student performance predictive solution. The data was collected for 133 students from a total of four sections of three different courses. With a limited amount of data, we were still able to construct a predictive solution able to produce valuable insights into the behaviors of students. The model's resulting accuracy on the test set is 0.85 and the model indicates that the earliest time to begin predictions is right after the midterm exam. The model performs well in its task to predict student performance and identify correlations between different variables. However, it is at this time subject to limited data which although treatable, can affect the accuracy and its ability to predict a final score numerically. This work paves the ground for future studies on the use of machine learning using in-class learning data, analyzing student learning as a function of time within each session rather than by grades alone.

Keywords: Artificial intelligence; Machine learning; Learning analytics; In-class learning, At-risk students, Synchronous online teaching.

Background and Motivation

The U.S. Bureau of Labor Statistics (USBLS) states that in the last 10 years, employment in the Science, Engineering, Technology, and Mathematics (STEM) fields has grown twice as fast as that in non-STEM fields (US Dept. of Labor, 2019). In particular, jobs in engineering and related fields are experiencing the most growth (US Dept. of Labor, 2019). In addition, there is also a rise in non-STEM jobs that require STEM skillsets (NSF Business-Higher Education Forum, 2019). This may read as highly encouraging for students pursuing STEM degrees, the matter of concern is that there are not enough students in engineering or related fields to fill all the current and potential job openings (National Science Board, 2018). This is made worse by the fact that student retention, persistence, and graduation rates are lower in STEM fields as compared to many other fields (National Science Board, 2018). It is also alarming that all the factors that contribute to such low rates are even more critical for women and minority students and negatively impact their chances to succeed in STEM fields (NCES, 2019; National Science Foundation, 2017). One of the contributing factors to such low rates is the student performance in core STEM courses. To address this, student performance measurement has taken center stage in a multitude of educational research communities.

One of the major side effects of students finding it difficult to adjust to challenging coursework is them resorting to dishonest means to improve performance and sometimes even just to stay afloat. Some of the common reasons cited by students as motivators for engaging in academic integrity violations are the intense and overwhelming pressure (internal as well as external) to succeed, lack of time to complete assignments, finding the need to get ahead, self-control, and low self-esteem issues (Borgaonkar et al. 2020; Simkin and McLeod 2010; Van Zyl and Thomas 2015; Walker and Townley 2012; Yu et al. 2018). Then there are organizational and systemic components such as professors not taking the time to emphasize the importance of academic integrity or observing their peers get away with cheating (Borgaonkar et al. 2020; Peters et al. 2019). Faculty perception of the common motivators of academic integrity violations by students is also as diverse as the student perception. Some of the reasons cited by faculty members on why students cheat include students willing to take extreme majors to maintain grades as well as lack of formal training, education, and hence understanding of academic dishonesty (Borgaonkar et al. 2020; Walsh et al. 2021). It is therefore educators' obligation to embrace a teaching philosophy that will reduce the burden and pressure that the students feel as well as to adopt assessment methods that focus on student learning. Efforts such as these clearly need to be informed by students' abilities and student performance measurement early in the semester.

Over the past few years, academic environments have experienced a volatile period, transiently adopting various platforms for learning management (Barrett et al. 2019; Hao 2019; Lakkaraju et al. 2015). This has resulted in the extensive use of various instructional delivery methods including, face-to-face, hybrid, hi-flex, and remote instruction. More recently, due to the COVID-19 pandemic, instructional delivery of courses in the US universities has been a roller-coaster ride with many rapid and major changes being forced on instructional staff with very little or no preparation time (Borgaonkar et al. 2021; Crawford et al. 2020; Walsh et al. 2021). In particular, the forced move to online or remote instruction for a vast majority of courses in Spring 2020 has had a profound effect on student perception and performance (Borgaonkar et al. 2021; Crawford et al. 2020). Some of this can be attributed to the fact that many instructors who were forced to move their courses online had little or no online teaching experience, whereas, even those with experience found it difficult to move courses online that were not optimized for remote delivery (Walsh et al. 2021). Even before the pandemic, studies had indicated that students found cheating to be easier in traditional online classes than in face-to-face courses, and studies conducted during the pandemic confirmed these earlier findings (Crawford et al. 2020; Walsh et al. 2021). Most instructors have given online lectures to classes of students who are non-responsive muted black screens throughout the entire session. In this study, we explore the viability of student performance prediction, in addition to the correlation of in-class learning (assessed through end-of-class questionnaires) with the performance results. Therefore, the urge for the development of a comprehensive yet pragmatic student performance assessment system has been intensified due to the failure of in-class learning in remote learning environments (Aucejo et al. 2020; Daniel 2020). Not to mention that before the pandemic-imposed collective online transition, in-class learning within a physical setting was not performing at the highest level (Dhawan 2020). Many academic instructors holding synchronized online classes often complain about students' disengagement in the classroom (Islam et al. 2015).

Many qualitative and quantitative methods have been used to varying degrees of success to study student performance and to identify at-risk students (Hebert 2014; Norrish et al. 2014; Redmond et al. 2011; Sarra et al. 2019). Several of such techniques rely on pre-college assessment and academic performance data. Some have also used pre-tests and pre-surveys to identify students, who are likely not to do well in certain courses. Factors such as high school GPA or early college GPA have shown a strong correlation to student performance (Bala and Ojha 2012; Sarra et al. 2019). Studies have also established the need to perform such analysis early in the semester or early in students' degree progress, especially in the first year (Meedech et al. 2016). However, data and resources needed to duplicate many of these studies are not always readily available to faculty members. Therefore, there is a critical need to explore novel but reliable methods and techniques that rely on data that can be easily collected in the classroom.

Researchers have been curious about the relationship between students' classroom behavior and academic achievement for many decades (McKinney 1975; Morris et al. 2005). Furthermore, students' classroom behavior has been studied quite extensively for its relation to student performance and course grades in both in-person and online courses. In particular, cellphone use during class has been found to have a strong correlation to student performance to the point that the predictor value of in-class texting behavior on final course grade has also been determined. (McDonald 2013; Bjornsen and Archer 2015). Multiple studies also document the impact of various initiatives such as caring behavior on part of the instructor and classroom management strategies and programs to influence student behavior and in turn their performance in a positive way (Miller 2008; Korpershoek et, al. 2016). Since student persistence and performance have been linked to their behavior, studies have also been focusing on the variation of this impact on low-income and minority students as well as the effectiveness of out-of-the-box methods for influencing student behavior (Sheldon and Epstein 2002; Black and Fernando 2014). In the past decade, many new tools and techniques, including Artificial Intelligence (AI) have emerged and it is worth exploring how they can be leveraged to build on the existing work to further study this relationship and build predictive models to enable instructors to increases their students chances of success in their classes.

Introduction

Artificial Intelligence (AI) has affected many aspects of society, for example, technology taking over simple factory jobs and replacing some cashiers, and in the advertising industry, firms use users' information to decide what ads should be presented to you. Despite it repeatedly proving to be more than useful, there are still some places for the full adaptation of AI in the educational field. Recently, there has been considerable interest in leveraging various forms of AI such as machine learning, deep learning, Natural Language Processing (NLP), anomaly detection, conversational AI, and computer vision to augment education to better student performance and measure class engagement (Barrett et al. 2019; Dewan et al. 2019; Ginda et al. 2019; Lakkaraju et al. 2015; Rastrollo-Guerrero et al. 2020; Thomas and Jayagopi 2017; Whitehill et al. 2014). However, these forms of AI are rarely utilized within the popular learning management solutions commonly adopted by undergraduate programs across the United States. Therefore, colleges and universities are currently using traditional methods to carry out necessary pedagogical interventions to ensure students' success.

AI has immense potential to better the way educational facilities function as we know them. Places that have begun implementing these methods have seen great results. Squirrel AI founded by Derek Li claims the ability to identify a student's needs more rapidly than any teacher could alone (Hao 2019). Squirrel AI as well as many other educational technology companies utilize a concept known as Knowledge Space Theory (KST), to develop their AI. Essentially KST is a process that allows them to describe and model specified areas of knowledge (Kulkarni 2019). With this, the AI can determine a student's current placement in the material and develop a custom curriculum for them that will update itself as they learn. The main separation between Squirrel AI and the current work is data application and purpose. Squirrel AI utilizes Knowledge Space Theory to map the student's current knowledge, while the current method utilizes the data for the prediction of grades. As for purpose, Squirrel AI strives to improve student success rate by identifying and filling in knowledge gaps through assessments it provides. The current solution aims to help by identifying struggling students by not only their knowledge level, but their behavior such as attendance, and will alert them before it is too late to act.

Student in-class learning is a key element in achieving satisfactory results at the end of an academic semester. Accurate and precise measurement of in-class learning not only contributes to a better education but also better teaching outcomes. Also, human behavior measurement is a key factor in better assessing student learning. Over the years, some researchers have made endeavors to use computer vision and other sensor-based techniques in classrooms to measure the attention and engagement levels of students from their facial expressions, head pose, eye gaze, heart rate, and so forth (Duraisamy et al. 2019; Monkaresi et al. 2017; Thomas and Jayagopi 2017). However, these methods of measurement have raised serious ethical and privacy concerns (Pardo and Siemens 2014; Zhang et al. 2018). Additionally, almost none of the techniques can be applied in online learning environments.

This work aims mesoscale approach, a midway solution surpassing the macro-scale approach by accounting for time-dependent engagement and not drilling so far as to become micro-scale. From existing studies on gauging student performance, we understand the differences in approaches and the role of this mesoscale approach. The current methods are generally based on utilizing data that students generate throughout the semester, such as students' grades or attendance records. They might distribute additional gauging assessments; however, they are not re-occurring. The mesoscale methods are easier for students in data collection and have low commitments. The broad data leads to wide-ranging results, such as the prediction of whether students can finish school (Lakkaraju et al. 2015). On the other side, micro-scale studies are those that take invasive measures in gauging student engagement, such as individuals' movement tracking (Bui Ngoc Anh 2019; Duraisamy et al. 2019) or other similar motion tracing methods. Such studies still require broader data, such as yearly student surveys, to reach proper predictions. While the information gained from computer vision tracking is required, they need student behavior in the classroom. In this case, the students and instructors may feel uncomfortable being constantly tracked by an object. This short communication aims to introduce a novel approach to in-class learning based on collecting chronological questionnaires and utilizing them for predicting students' performance. Unlike many other studies and tools, this approach surveys in-class learning at neither macro-scale, collecting data through un-ordered session quizzes and surveys (Lakkaraju et al. 2015; Zainuddin et al. 2020), nor micro-scale, tracking student movement in hopes of measuring engagement and in-class learning (Duraisamy et al. 2019; Whitehill et al. 2014).

Methodology

Data Collection

In this study, the records of 133 undergraduate students in Mechanical Engineering were collected at Rowan University, New Jersey from Fall 2020 to Spring 2021 (during the COVID-19 pandemic). The data were collected from four different classes including three major/core mechanical engineering courses: *Engineering Mechanics:*

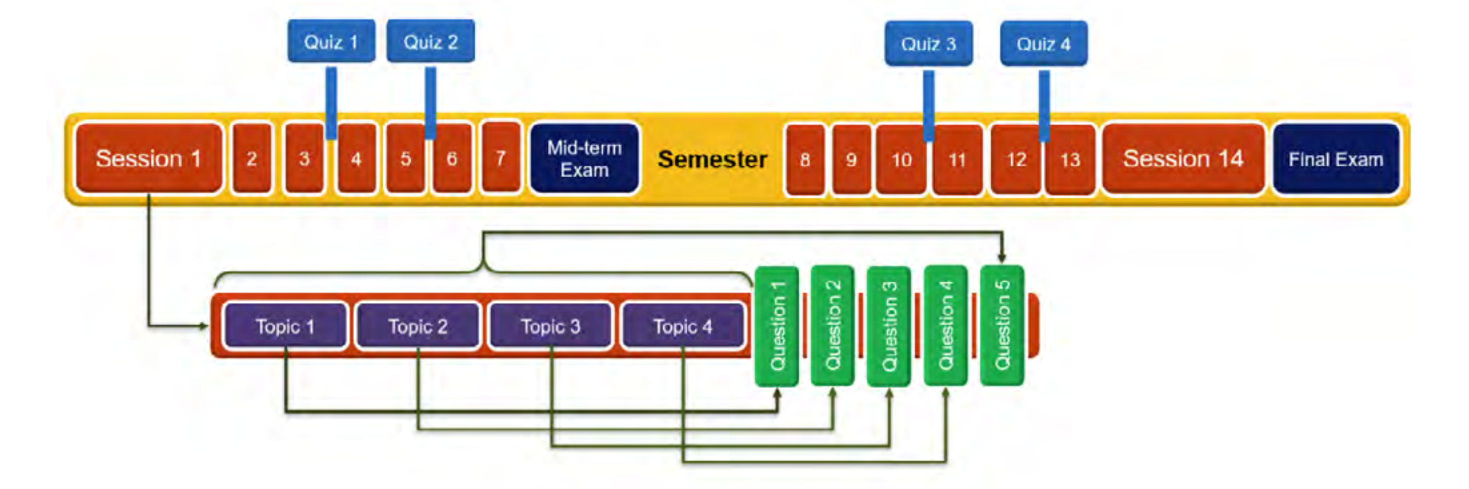

Figure 1. Schematic representation of in-class learning and thru-the-semester performance data collection approach. At the end of each session during the semester, the students are required to answer five questions, four of which are multiple-choice and associated with topics covered during the session in chronological order. The last question is open-ended and is about students summarizing what they have learned. (The session represents a single week in a semester)

Statics (Sophomore), *Engineering Mechanics: Dynamics* (Junior; two sections), and *Machine Design* (Junior). The class size was around 33 students per class, and the ratio between men and women was slightly higher than 2. 30% of the samples in the data were randomly selected to be set aside as a test dataset (excluded during the model development), and the remaining 70% were used as the training set for the model development. All students involved in the data collection were a part of four classes lectured by the same instructor during two consecutive semesters. During COVID-19, all lessons and instructions were distributed virtually. The collected data for each session was comprised of students' attendance, response time, response to an open-ended question, and attention span based on a true-false and multiple-choice assessment of class materials in chronological order. For new concepts presented in the lecture, the instructor selected true-false questions; for example, is "bodyweight" a result of Newton's third law (*Engineering Mechanics: Statics*). For new definitions, the instructor selected multiple-choice questions; for example, which item will be the definition of linear acceleration (*Engineering Mechanics: Dynamics*). For analytical topics, the instructor used an open-ended question as the last part of the questionnaire; for example, explain the new design criterion in a short statement (*Machine Design*). Data was collected through a mobile application discussed later. In addition to the session's data, the collected gradebook data was comprised of quizzes, homework, the midterm exam, and the final grade in the course. While all aspects of the collected data were utilized, the accuracy of the data was of main importance as it reflected how well the student retained information chronologically over the course of the lecture.

Framework

This framework as shown in Figure 1 is comprised of a few indicators throughout an academic semester, such as quizzes, midterm exams, final exams, and chronological in-class learning measurement. One important indicator which is not present in this framework is homework. Homework is a flexible component of this framework and depending on class needs, can be added anywhere between sessions. However, the most notable part of this framework is the chronological in-class learning measurements. For data collection during classes, a cell phone application was developed using an intelligent no-code platform (AppSheet.com) for ease of access and streamlined usability. The application was used by both students and instructors. An instructor had to create one questionnaire related to the topics covered during each lecture. For example, a question might ask for the definition of a term discussed during class. These questions were organized chronologically and uploaded into the app before the class started. Passwords to see the questionnaire and answer the questions were only given to the students at the end of each class to ensure students were only able to see the questions and were able to respond after the lecture is over but before they leave the classroom. The students were prompted to answer the questions quickly and notified that the correctness of their answers would not affect their final grades. Once finished, they were able to submit their answers to the questions using the cell phone application. The data collected from students' submissions were then accessible to the teacher in a live fashion in a Google Sheet whose link is provided to the instructor within the app. Similarly, an instructor can implement the same methodology manually (instead of using an app) using Google Forms as a replacement for app-based questionnaires in combination with Google Sheets. In that case, the instructor would share a Google form link with his/her students when the lecture is over. The Google Form automatically records timestamps along with responses to questions in an associated Google Sheet when the form is submitted by a student; therefore, the instructor can calculate students' response time, attendance (by answering the first question of the questionnaire: student ID number), and responses correctness (via a predefined answer sheet designed by the instructor). There was no statistical analysis before inserting the data into our code. At the end of the class, in-class learning measurements for each session are summarized in one unique Google Sheet for all students who participated in the class.

	Q_1	Q_2	Q_3	Q_4	Q_5	Time	Attend	HW	Quiz	Midterm	Grade
mean	0.43	0.27	0.34	0.28	17.60	69.45	88.44	84.25	84.78	81.98	3.06
std	0.39	0.41	0.36	0.41	7.17	32.04	15.89	14.13	16.37	17.02	0.85
min	-0.60	-1.00	-0.60	-0.75	4.40	8.87	25.00	26.03	22.50	36.36	1.70
25%	0.14	0.00	0.14	0.00	13.57	46.86	87.50	79.11	77.50	70.00	2.70
50%	0.50	0.25	0.43	0.25	16.37	64.87	100.00	88.33	89.50	85.00	3.00
75%	0.71	0.50	0.60	0.50	20.75	88.40	100.00	94.33	98.75	95.00	4.00
max	1.00	1.00	1.00	1.00	52.00	159.75	100.00	100.00	100.00	100.00	4.00

Q_1, Q_2, Q_3 and Q_4 are calculated as below:

$$Q_i = \frac{\sum_j^n p_j}{n}$$

$$p_j = \begin{cases} +1, & \text{if answer is correct} \\ -1, & \text{if answer is incorrect} \end{cases}$$

Table 1. The summary statistics of the full dataset.

Unlike many other systems which are under active research or already commercialized, in-class learning is assessed at the end of each session while considering the chronological order by which the topics were lectured to the students. Chronological in-class assessment enables the instructors to quickly assess the attention and involvement of the students at each session and throughout the semester. It should also be noted that in many class settings, including the setting where this study's data was collected, there is a short break right after the instructor is done teaching topic 2 (Figure 1). Moreover, the open-ended fifth question allows an opportunity for performing NLP in the presence of abundant data. The main objective of the present study is to predict students' end-of-semester performance immediately after the midterm. Table 1 shows the summary statistics of the full datasets (both training and test sets). It is comprised of mean, std (standard deviation), min, max as well as lower (25), median (50), and upper (75) percentiles.

where *i* is indicative of the class quarter to which the subject of the question pertains, j is the session number, *n* is the total number of sessions, and P*j* represents the answer accuracy. The 1-point penalty serves to create a difference between those who answer correctly, answer incorrectly, or skip the question. For example, considering that a semester has 10 sessions, a student that answers 9 Q_1 questions (P1 to P9) correct, and only gets one Q_1 question wrong (P10) will get a Q_1 score of 0.8. Q_5 is simply the average number of words a student uses in writing his/her summary of a session. *Time* is the average number of seconds it takes a student to submit his/her responses to the five questions. *Attend* is the percentage of a student's attendance. *HW* and *Quiz* are the averages recorded for a student's homework and quizzes, respectively. The *midterm* is the score a student receives in the midterm exam. The *grade* is the final grade a student receives after showing his/her end-of-semester performance. To comply with the university's data privacy requirements, the students' IDs and names were replaced with anonymous identifiers before performing any data analysis. To analyze the dataset, both Microsoft Excel and Python were used. Python has also been used for training the dataset with different classification algorithms (Lemaître et al. 2017). Evaluation of each algorithm is done scrupulously to identify and use the most predictive supervised machine learning classification algorithm. Initially, data pre-processing was performed on the raw data acquired by the instructors. Subsequently, the grades were mapped to the "Letter" and "Description" variants shown in Table 2.

The dataset was checked for missing values, and since no values were missing, no imputation was required. Normalization was done for the variables *Attend., HW, Quiz,* and *Midterm* such that the maximum possible score received for each becomes one hundred. In addition, standardization on the entire dataset except the Grade was done using the scikit-learn StandardScaler (Pedregosa et al. 2011). Due to an imbalanced dataset concerning end-of-semester grades, different sample balancing techniques were tested and SMOTE was selected because it showed better performance.

Letter	Point Value	Description
A	4.0 points per credit hour	Excellent
A–	3.7 points per credit hour	
B+	3.3 points per credit hour	
B	3.0 points per credit hour	Good
B–	2.7 points per credit hour	
C+	2.3 points per credit hour	
C	2.0 points per credit hour	Fair
C–	1.7 points per credit hour	
D+	1.3 points per credit hour	
D	1.0 points per credit hour	
D–	0.7 points per credit hour	
F	0.0 points per credit hour	Failure

Table 2. The grading system and its variants used in this study.

Training Methods

Multiple machine learning (multi-class) classification algorithms were used to train the dataset, including logistic regression, random forest, support vector machine (SVM), k-nearest neighbors (KNN), decision tree, and ensemble learning (soft majority voting). The idea was to evaluate algorithms and therefore identify the one which produces the most sensible and accurate prediction of students' performance for small datasets. Each algorithm was optimized using GridSearch, which tuned hyperparameters relevant to individual algorithms and Principal Component Analysis (PCA) (as a dimensionality-reduction technique). Additionally, both 10-fold cross-validation and Leave-One-Out Cross-Validation (LOOCV) were used during the machine learning algorithm selection phase.

Results

Because of the small size of the dataset, a correlation map is another analytical result that can be used to better understand the conspicuous data structure and types of attributes used in a machine learning algorithm. As can be seen in Figure 2 which shows the correlation heat map

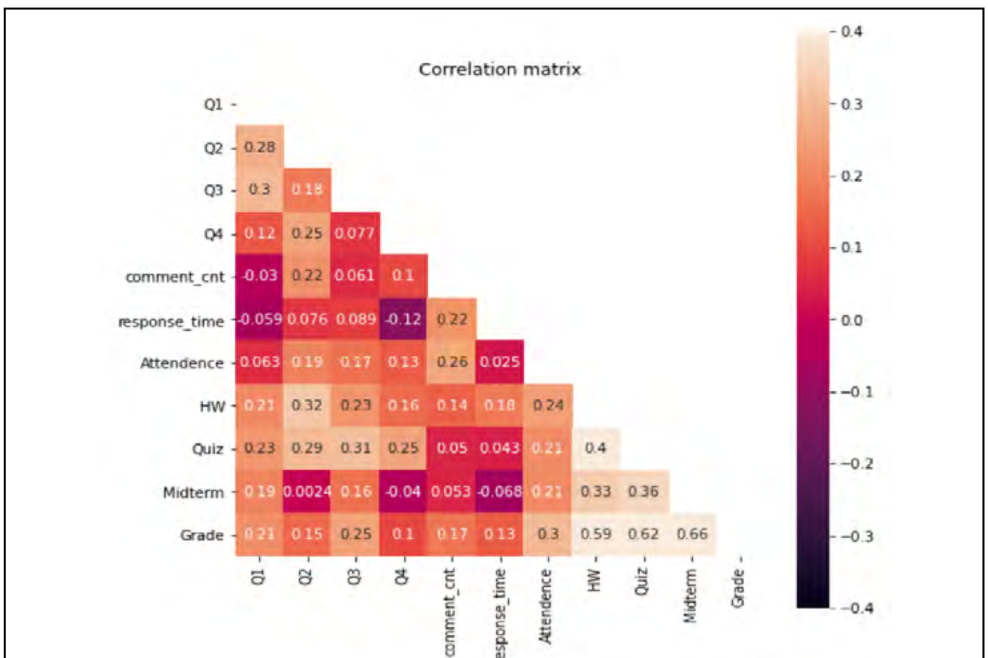

Figure 2. Correlation heat map of the class's dataset.

Method	Precision	Recall	f1-score	Accuracy
Ensemble Learning	0.84	0.81	0.82	0.85
Logistic Regression	0.65	0.68	0.64	0.78
SVM	0.67	0.68	0.66	0.78
KNN	0.70	0.67	0.68	0.75
Random Forest	0.67	0.64	0.65	0.75
Decision Tree	0.49	0.47	0.45	0.50

Table 3. Performance comparison amongst machine learning algorithms on the test set.

of the class's dataset; the lightest color of the heat map showed a strong correlation between the two indicators and the darkest color showed a minor connection between the factors. The correlation heat map showed unexpected information such as the non-negligible correlation between in-class learning in both the first and third quarters of a session and the end-of-semester grades of students. It also exhibited a high correlation between homework sets, quizzes, and midterm scores with the end-of-semester grades. Expectedly, students' attendance had a considerable correlation with the final grade. Interestingly, the response time and response word length also hold a non-negligible correlation with the final grade.

Because the dataset is small, neither the letter grade nor the grade point value was set as the target variable to be predicted (see Table 2), instead the "Description" of the grade is planned to be predicted via a machine learning model. In other words, prior attempts to successfully predict the "Letter" grade have not been successful due to the small size of the dataset. Also, in the current dataset, there has not been a student with a "Description" grade of "failure"; thus, the classification model is ternary. The resulting performance comparison (Table 3) amongst different methods on the test set has shown that Ensemble Learning outperforms other machine learning algorithms. Ensemble learning is a data training method unique in that it utilizes multiple other existing techniques to create the best possible model. In our study, Ensemble Learning creates an ultimate model from a soft majority voting of the developed methods: Logistic Regression, SVM, KNN, Decision Tree, and Random Forest models. A soft voting ensemble learning involves summing the predicted probabilities for class labels and predicting the class label with the largest sum probability.

Because there is a class imbalance in the data, all four precision, recall, f1-score and accuracy are taken into account to select Ensemble Learning as the best modeling method. The precision, recall and, f1-score shown in Table 3 are macro averages, accounting for the present class imbalance in the dataset.

The selected model's resulting accuracy on the test set is 0.85. Figure 3 shows the confusion matrix after assessing the selected machine learning model's (Ensemble Learning) performance with the test set. The confusion matrix presents an overall comparison of what the model predicted versus what the actual outcome came to be. The model predicts the final grade description based on the data available to it at that point in time during the semester. The earliest time to begin making predictions is right after the midterm exam. Even though the model has a few errors, the misprediction is only off by one category. For instance, a grade that is meant to be "Good" is mispredicted as "Excellent" and not "Fair.". Moreover, all the students' grades that are "Fair" are predicted correctly, therefore no struggling student is mispredicted within the test set.

Our current product comes with both advantages and setbacks. The current model performs well in its task to predict student performance and identify correlations between different variables. However, it is at this time subject to limited data which although treatable, can affect the accuracy and its ability to predict a final score numerically. With the continuation of this study, we plan to accumulate more data and improve the model's function further.

Concluding Remarks

There is significant demand for STEM graduates in the US and it is only increasing in recent years. A literature survey indicates a strong need to develop tools to predict student performance as early as possible. Such analysis followed by early intervention has great potential to help more students succeed in STEM courses and ultimately increase their retention and persistence. Prediction of students' performance halfway through the semester is important and can be accomplished using the method proposed in this study. Collecting engagement data for the machine learning model over the course of a lecture with respect to the order of topic allows for an accurate gauge of where engagement was or was not. The data collected not only aided in the development of the model but also gave an insightful view of student behavior over the span of a semester. Referencing back to the correlation heat map, interesting correlations emerged that could lead to further discussion. The model has only begun to show-

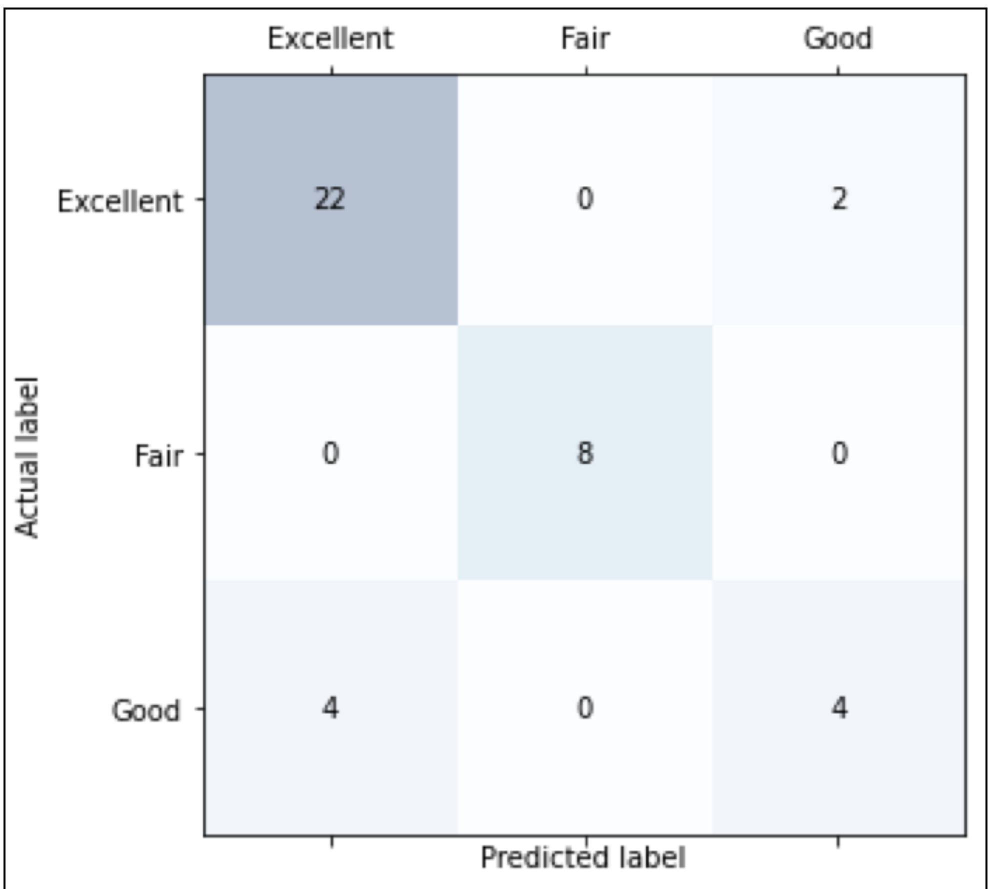

Figure 3. Confusion matrix over the test dataset.

case its capabilities. As the dataset continues to grow, the model will continue to increase in accuracy and provide us with more insights into the behavior of students. With this, both the instructor and student will have a higher likelihood of having sufficient time for improving the student's performance for the remainder of the semester. Despite an ongoing need for such a predictive solution, one has yet to be developed and trusted in academic environments. One salient factor in the nonexistence of an overarching solution is the lack of cooperation and standardization amongst instructors and schools etc., in acquiring and recording student data especially vis-à-vis in-class learning. The other important factor is the feasibility of attaining proper measurements of in-class learning with the growing use and methods of online learning and the accompanying controversies. We proposed a simple collection of short answers from students and an application of AI to provide a chronological approach. This has been tested in the study vis-à-vis in-class learning and has shown relevance to the end-of-semester grades of students. In the future, one can expect that when more data is acquired for the continuation of this study, numerical or letter grades can also be prognosticated with decent accuracy. The next step for the progression of this study is to have a pilot stage in which multiple instructors across different institutes use the solution we have described to further educate the model and reassure its predictive performance in a variety of educational settings. To apply the method discussed here in your own classroom, we stress the importance of administering the assessments consistently and encouraging students to participate in them. In doing this you are obtaining the most accurate data possible and should be able to gain more accurate insight into student behavior from the model's predictions. A predictive system similar to one proposed in this study can efficiently identify at-risk students at a time in the semester when something can still be done to help these students. However, it is important that institutions of higher education equip their instructors with such predictive systems and back such efforts with strong educational policies.

Machine learning analytics is an entirely different process from other methods of study. Machine learning automates the entire data analysis workflow to provide deeper, faster, and more comprehensive insights when there is no exact solution to a problem. On the contrary, an analytical solution involves framing the problem in a well-understood form and calculating the exact solution. Therefore, in this study, Machine learning analytics is solely providing the solution without the need for an analytical solution due to the non-existence of an exact solution. If an analytical solution was already available to predict the end-of-semester outcomes of students, then there would be no need for this study.

The emergence of the COVID-19 pandemic has shifted the attention to online/remote and collaborative learning and the personalization of education technologies. In the current education setting, the students often do not have a clear-cut judgment over their learning process, and they are not receiving enough feedback from their instructors. Any feedback on formative assessments can be used to improve their engagements. The proposed application of machine learning for in-class learning can lead to an effective education environment for students and allow them to have dynamic insight into their own learning process.

Acknowledgments

We acknowledge the comments and feedback from Dr. Derek Lehane. This research was funded by National Science Foundation (NSF; grant number 2024226).

References

Aucejo, E. M., French, J., Araya, M. P. U., and Zafar, B. (2020). "The impact of COVID-19 on student experiences and expectations: Evidence from a survey." *Journal of public economics*, 191, 104271.

Bala, M., and Ojha, D. B. (2012). "Study of applications of data mining techniques in education." *International Journal of Research in Science and Technology,* 1(4), 1-10.

Barrett, M., Branson, L., Carter, S., DeLeon, F., Ellis, J., Gundlach, C., and Lee, D. (2019). "Using artificial intelligence to enhance educational opportunities and student services in higher education." *Inquiry: The Journal of the Virginia Community Colleges,* 22(1), 11.

Bjornsen, C. A., & Archer, K. J. (2015). Relations between college students' cell phone use during class and grades. Schol*arship of Teaching and Learning in Psychology, 1*(4), 326–336.

Black, D. S., & Fernando, R. (2014). Mindfulness training and classroom behavior among lower-income and ethnic minority elementary school children. *Journal of child and family studies, 23*(7), 1242-1246.

Borgaonkar, A. D., Sodhi, J., Vijayabalan, R., and Nair, A. S. K. "A Comparative Study of the Impact of COVID-19 Pandemic on Student Participation and Performance in First-Year Engineering Courses." *Presented at 2021 ASEE Virtual Annual Conference Content Access*.

Borgaonkar, A. D., Zambrano-Varghese, C. M., Sodhi, J., and Moon, S. "Fantastic Cheats: Where and How to Find Them? How to Tackle Them?" *Presented at 2020 ASEE Virtual Annual Conference Content Access.*

Bui Ngoc Anh, N. T. S., Phan Truong Lam, Le Phuong Chi , Nguyen Huu Tuan, Nguyen Cong Dat, Nguyen Huu Trung, Muhammad Umar Aftab, Tran Van Dinh. (2019). "A Computer-Vision Based Application for Student Behavior Monitoring in Classroom." *Applied Sciences,* 9(22).

Crawford, J., Butler-Henderson, K., Rudolph, J., Malkawi, B., Glowatz, M., Burton, R., Magni, P., and Lam, S. (2020). "COVID-19: 20 countries' higher education intra-period digital pedagogy responses." *Journal of Applied Learning & Teaching*, 3(1), 1-20.

Daniel, J. (2020). "Education and the COVID-19 pandemic." *Prospects*, 49(1), 91-96.

Dewan, M. A. A., Murshed, M., and Lin, F. (2019). "Engagement detection in online learning: a review." *Smart Learning Environments*, 6(1), 1-20.

Dhawan, S. (2020). "Online learning: A panacea in the time of COVID-19 crisis." *Journal of Educational Technology Systems*, 49(1), 5-22.

Duraisamy, P., Van Haneghan, J., Blackwell, W., Jackson, S., Murugesan, G., and Tamilselvan, K. "Classroom engagement evaluation using computer vision techniques." *Presented at Pattern Recognition and Tracking XXX*.

Ginda, M., Richey, M. C., Cousino, M., and Börner, K. (2019). "Visualizing learner engagement, performance, and trajectories to evaluate and optimize online course design." *PloS one,* 14(5), e0215964.

Hao, K. (2019). "China has started a grand experiment in AI education. It could reshape how the world learns." *MIT Technology Review,* 123(1).

Hebert, A. P. (2014). "Identifying at-risk students: a study of freshmen and the beginning college survey of student engagement."

Islam, N., Beer, M., and Slack, F. (2015). "E-learning challenges faced by academics in higher education." *Journal of Education and Training Studies,* 3(5), 102-112.

Kim, B.-H., Vizitei, E., and Ganapathi, V. (2018). "GritNet: Student performance prediction with deep learning." *arXiv preprint arXiv:1804.07405.*

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. Review of Educational Research, 86(3), 643-680.

Kulkarni, A. (2019). "AI in Education: Where is It Now and What is the Future." *Lexalytics.*

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., and Addison, K. L. "A machine learning framework to identify students at risk of adverse academic outcomes." *Presented at Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining.*

Lemaître, G., Nogueira, F., and Aridas, C. K. (2017). "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning." *The Journal of Machine Learning Research*, 18(1), 1-5.

McDonald, S. (2013). The effects and predictor value of in-class texting behavior on final course grades. *College Student Journal, 47*(1), 34-40.

McKinney, J. D., Mason, J., Perkerson, K., & Clifford, M. (1975). Relationship between classroom behavior and academic achievement. *Journal of Educational Psychology, 67*(2), 198.

Meedech, P., Iam-On, N., and Boongoen, T. (2016). "Prediction of student dropout using personal profile and data mining approach", *Intelligent and Evolutionary Systems.* Springer, pp. 143-155.

Miller, R. M. (2008). The Influence of Teachers Caring Behavior on High School Students Behavior and Grades.

Monkaresi, H., Bosch, N., Calvo, R. A., and D'Mello, S. K. (2017). "Automated Detection of Engagement Using Video-Based Estimation of Facial Expressions and Heart Rate." *IEEE Transactions on Affective Computing,* 8(1), 15-28.

Morris, L. V., Finnegan, C., & Wu, S. S. (2005). Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education, 8*(3), 221-231.

Norrish, M. I., Kumar, P. A., and Heming, T. A. (2014). "Interim identification of "at risk" students: A predictive model." *J Contemp Med Edu*, 2(4), 199-203.

Pardo, A., and Siemens, G. (2014). "Ethical and privacy principles for learning analytics." *British Journal of Educational Technology,* 45(3), 438-450.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). "Scikit-learn: Machine learning in Python." *the Journal of machine Learning research*, 12, 2825-2830.

Peters, M., Boies, T., and Morin, S. "Teaching academic integrity in Quebec universities: roles professors adopt." *Presented at Frontiers in education.*

Rastrollo-Guerrero, J. L., Gomez-Pulido, J. A., and Duran-Dominguez, A. (2020). "Analyzing and predicting students' performance by means of machine learning: a review." *Applied sciences*, 10(3), 1042.

Redmond, B., Quin, S., Devitt, C., and Archbold, J. (2011). "A qualitative investigation into the reasons why students exit from the first year of their programme and UCD."

Sarra, A., Fontanella, L., and Di Zio, S. (2019). "Identifying students at risk of academic failure within the educational data mining framework." *Social Indicators Research*, 146(1), 41-60.

Sheldon, S. B., & Epstein, J. L. (2002). Improving Student Behavior and School Discipline with Family and Community Involvement. Education and Urban Society, 35(1), 4–26.

Simkin, M. G., and McLeod, A. (2010). "Why do college students cheat?" *Journal of Business Ethics*, 94(3), 441-453.

Sweeney, M., Rangwala, H., Lester, J., and Johri, A. (2016). "Next-term student performance prediction: A recommender systems approach. arXiv 2016." *arXiv preprint arXiv:1604.01840.*

Thomas, C., and Jayagopi, D. B. "Predicting student engagement in classrooms using facial behavioral cues." *Presented at Proceedings of the 1st ACM SIGCHI international workshop on multimodal interaction for education.*

Van Zyl, A., and Thomas, A. (2015). "Academic honesty: Perceptions of millennial university students and the role of moderating variables." *koers*, 80(1), 1-15.

Walker, M., and Townley, C. (2012). "Contract cheating: A new challenge for academic honesty?" *Journal of Academic Ethics,* 10(1), 27-44.

Walsh, L. L., Lichti, D. A., Zambrano-Varghese, C. M., Borgaonkar, A. D., Sodhi, J. S., Moon, S., Wester, E. R., and Callis-Duehl, K. L. (2021). "Why and how science students in the United States think their peers cheat more frequently online: perspectives during the COVID-19 pandemic." *International Journal for Educational Integrity*, 17(1), 1-18.

Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., and Movellan, J. R. (2014). "The faces of engagement: Automatic recognition of student engagementfrom facial expressions." *IEEE Transactions on Affective Computing, 5*(1), 86-98.

Yu, H., Glanzer, P. L., Johnson, B. R., Sriram, R., and Moore, B. (2018). "Why college students cheat: A conceptual model of five factors." *The Review of Higher Education*, 41(4), 549-576.

Zainuddin, Z., Shujahat, M., Haruna, H., and Chu, S. K. W. (2020). "The role of gamified e-quizzes on student learning and engagement: An interactive gamification solution for a formative assessment system." *Computers & Education,* 145, 103729.

Zhang, C., Chang, C., Chen, L., and Liu, Y. "Online Privacy-Safe Engagement Tracking System." *Presented at Proceedings of the 20th ACM International Conference on Multimodal Interaction.*

Ms. Margaret Stadlman is currently a senior studying in Rowan University's Mechanical Engineering program (to be graduated in May 2022). She began her research in Machine Learning in the Fall of 2021 as a student clinic (Capstone) project. Since then, she has been working with the team to extract new insight from experimental data obtained by Dr. Miri's graduate project. For this project, she learned a lot about coding in Google Colab, a tool that uses Python.

Dr. Seyyed Salili is currently a Lead Data Scientist at Alteryx. He also serves as an Advisory Council Member of the Harvard Business Review. He received his Ph.D. from the Advanced Materials & Liquid Crystal Institute at Kent State University in 2016. After receiving his doctoral degree, he worked at the University of Pennsylvania as a Postdoctoral Research Fellow in the Physics Department and as a Mack Technology Fellow in The Wharton School. For his innovative entrepreneurial activities, he has been granted the National Science Foundation Innovation Award four times. Dr. Salili is also an alumnus of the MIT Sloan School of Management Executive Program on Management & Leadership.

Dr. Ashish D. Borgaonkar works as an Assistant Professor of Engineering Education at the New Jersey Institute of Technology (NJIT). He received his Ph.D. in Environmental Engineering from NJIT. Dr. Borgaonkar has won multiple awards for excellence in instruction; most recently the Saul K. Fenster Award for Innovation in Engineering Education. His research focuses on increasing diversity in STEM education and STEM workforce. He has received several grants to run workforce development training programs as well as undergraduate research experience programs to train underrepresented minority and first-generation students. He is also involved in various projects and initiatives in engineering education, integration of novel technologies into the engineering classroom, excellence in instruction, water, and wastewater treatment, and storm water management.

Dr. Amir K. Miri serves as the director of the Advanced Biofabrication Laboratory at the New Jersey Institute of Technology (NJIT). He received his Ph.D. from McGill University, where his doctoral research involved multiscale characterization of soft tissue biomechanics using analytical tools and a variety of mechanical testing methods. Dr. Miri's postdoctoral training at Harvard-MIT Health Sciences and Technology included advanced manufacturing of biomedical scaffolds and organoids using 3D bioprinting techniques. Dr. Miri's research group has worked on 3D bioprinting for creating microtissue models and has received multiple National Science Foundation and National Health Institutes grants. Recently, he has developed educational modules based on his research activities.