Early Forecasting of Student Performance in STEM Using Language Models

Helen Martinez, Dr. Mohammad R. Hasan, Ahatsham Hayat

Department of Electrical & Computer Engineering

University of Nebraska-Lincoln

Abstract

Retention rates and academic performance in undergraduate STEM (science, technology, engineering, and mathematics) courses are critical to building a robust and competitive workforce. Despite various intervention efforts, university support systems have proven inadequate in addressing the diverse needs of STEM students, often resulting in low retention rates and suboptimal academic performance in STEM degrees. To better identify students who might be at risk, pre-trained Language Models (LMs) were used to predict students' performance. This study assesses how well LMs process and learn from a natural language dataset that includes student grades & student background information to predict academic performance early on in the semester. The data was then run through our LM to identify students as "vulnerable" or "proficient". To improve the adaptability of LMs for forecasting tasks within this domain, we have developed a data enrichment technique. This technique includes strategies for replacing missing values, augmenting text sequence data, and incorporating specific task instructions and contextual cues. We explore the model's capability to make accurate early predictions with data collected as early as two weeks into the semester. Our findings suggest that integrating background information significantly improves prediction accuracy, highlighting the potential for early intervention strategies to support at-risk students.

1. Introduction

Improving student outcomes in STEM (Science, Technology, Engineering, & Mathematics) disciplines has emerged as a critical objective in higher education. More than one-half of freshman STEM majors left those fields before graduation causing the graduation rate to be 20% below their counterparts in non-STEM majors (Leary, et al.). Traditional university advising systems and support centers are beneficial to students. However, for retention efforts to be successful at large, they would be more effective at a smaller scale; for instance, within individual programs (Leary, et al.). Often, these advising systems fail to address the individual needs of students. This study explores **how well can Language Models (LMs) learn from a natural language dataset that includes student grades & student background information to predict academic performance early on in the semester.** This dataset includes experiential & non-experiential time-series data focused on student learning, capturing various aspects of their academic journey, such as cognitive, non-cognitive, and distal factors. Traditional forecasting methods rely heavily on cognitive data, such as grades and test scores, which do not fully capture the complexities of a student's academic experience. Early forecasting of academic performance is particularly challenging, especially when the data is limited to the first two, three, and even four weeks of the semester. Our approach is to add background information such as socioeconomic factors to explore the accuracy of LMs in early forecasting. We experimented with three types of features from the students' academic trajectories: (i) distal factors (academic meta-information and socioeconomic status), (ii) proximal cognitive factors (formative and summative test scores), and (iii) proximal non-cognitive factors (repeated measures of non-cognitive attributes like engagement). Because the data is numerical, it was converted into natural language for the LM to understand. Our experiments include one LM and three Non-LMs to understand how well they can predict accuracy with various combinations of the features mentioned above.

2. Background

2.1 Importance of STEM Education

STEM education is essential for cultivating a robust workforce capable of driving innovation. The journey through STEM disciplines is often met with challenges such as low student enrollment and high attrition rates (Sithole, et al.). Studies indicate that improving retention and performance in STEM courses is not only necessary for students' academic success but also for meeting the broader societal need of skilled STEM professionals (Sithole, et al.). STEM jobs in the U.S. are projected to increase by 10.8 percent between 2022 and 2032 (Bureau of Labor Statistics), and the demand for manufacturing jobs requiring STEM skills remains high. However, studies have consistently shown that enhanced support systems can positively impact student success in STEM, highlighting the need for improved intervention strategies (Leary, et al.)

2.2 Limitations of Traditional Interventions

Traditional university support systems, such as advising centers, often fail to address the specific and diverse needs of students. These systems typically provide support after issues have been identified. Specifically, in STEM, the rigor and the complexity of the coursework require more nuanced and targeted support. Furthermore, many students may feel reluctant to seek help from traditional advising centers due to perceived stigmatization or accessibility issues. Additionally, office hours and individual appointments can further reduce the likelihood of a student reaching out for help. As a result, students who are considered as 'vulnerable' often do not receive the help that they need, exacerbating retention and performance issues in STEM programs (Rodriguez, et. al). Even when traditional interventions are implemented, their effectiveness can vary across the student population. This variability is often caused by the lack of consideration for differences in a student's background (Rodriguez, et. al). However, when early interventions target students' psychosocial factors, it could positively impact their persistence in STEM fields (Nostrand & Pollenz).

2.3 Potential of Artificial Intelligence (AI) in Educational Interventions

AI can offer promising solutions to the limitations of traditional university support systems. LMs can analyze vast amounts of data to deliver personalized interventions tailored to the diverse needs of students. Prior research has demonstrated the potential of AI tools, such as virtual assistants, to aid students in their academic journey (Page & Gelbach). Therefore, it is possible to create a support system that is both scalable and highly effective. Specifically, AI can effectively identify at-risk students early on, and provide interventions that significantly improve retention rates & academic performance (Chen et. al). For example, the use of learning analytics at Purdue University demonstrated how AI could predict which students were at risk of failing their enrolled course and provided the students with targeted interventions (Arnold & Pistilli, 2012).

3. Methods

We have compiled a 48-dimensional dataset that provides a comprehensive view of college students' academic journeys in an introductory programming course over the course of the Fall 2023 and Spring 2024 semesters at the University of Nebraska-Lincoln. This dataset is divided into two main parts:

1. **Experiential Component:** It includes repeated qualitative measurements of non-cognitive attributes like student motivation and engagement. These factors offer insights into students' perceptions of their academic experience throughout the semesters.

2. **Non-Experiential Segment:** This includes (i) a quantitative analysis of cognitive attributes, which covers students' formative and summative assessment scores, and (ii) a qualitative assessment of static data concerning students' background factors, such as academic meta-information and socioeconomic status.

Both the experiential (non-cognitive) and non-experiential (cognitive) data components are organized as time-series. Through this, we aim to evaluate whether pre-trained LMs can effectively predict a student's academic performance at the end of the semester. Specifically, we will analyze the model's ability to learn and integrate the complex correlations between experiential and non-experiential data elements.

3.1 Data Collection

At the start of the Fall 2023 & Spring 2024 semesters, course-related metadata (major, class standing) and socioeconomic factors (gender, race, family yearly income) were gathered; this is the distal data. The experiential data (i.e non-cognitive data) consists of repeated survey questions and answers that measure the students' motivation and engagement throughout the semester. This data was sourced from a cellphone application that was downloaded on the participating student's phone. This application preserved each student's privacy and tailored daily questions for the students to answer. Additionally, the non-experiential component contains cognitive data from students' assessment scores throughout the semester, taken directly from Canvas, the University of Nebraska-Lincoln's learning management system and distal data collected at the beginning of the semester.

3.2 Experiments

To address the research question, all of the data was converted into text form to be processed by the LM. The Flan-T5 Language Model was fine-tuned using this dataset. This approach allows the LM to learn not only from cognitive data, but also from personal and psychological factors that influence student success. The study then employed a time-series analysis to evaluate the data that was collected two, three, and four weeks into the Fall 2023 & Spring 2024 semesters. Next, we conducted a series of six experiments, i.e., Experiment 1 (C), Experiment 2 (C + D), Experiment 3 (C + NC), Experiment 4 (NC), Experiment 5 (NC + C + D), and Experiment 6 (NC + D). These experiments utilized different combinations of the three feature types: Cognitive (C), Non-Cognitive (NC), and Distal (D). In each experiment we fine-tuned the Flan-T5 LM using the three language datasets created from proximal data of varying lengths: 2 weeks, 3 weeks, and 4 weeks. For comparison, non-language models (Transformers, 1D CNN, and LSTM) were used to process cognitive data only; non-LMs can only process numerical data, hence why we used the cognitive dataset. Finally, the Flan-T5 LM predicted two classes: “At the end of the semester, the student will be vulnerable” and “At the end of the semester, the student will exhibit proficient performance.”

4. Results

The results demonstrated the effectiveness of the Flan-T5 language model in early forecasting of academic performance. Two key findings include:

1. **Accuracy Improvement:** The Flan T5 model achieved higher accuracy when using all three types of data (C+NC+D), compared to cognitive data alone
2. **Outperformance:** The Flan-T5 model outperformed the non-LMs

Dataset	Week	Stand. Dev	Mean
C	2	0.120291	0.642
C	3	0	0.73
C	4	0.026833	0.718
C+D	2	0.038341	0.842
C+D	3	0	0.87
C+D	4	0.031305	0.744
C+NC	2	0.058566	0.614
C+NC	3	0.032863	0.694
C+NC	4	0.032863	0.706
NC	2	0.038341	0.558
NC	3	0.026833	0.718
NC	4	0	0.6
NC+C+D	2	0.031305	0.814
NC+C+D	3	0.076811	0.64
NC+C+D	4	0.031305	0.786
NC+D	2	0.026833	0.718
NC+D	3	0.071204	0.652
NC+D	4	0	0.73

Figure 1: Results

Figure 1 represents all of the experiments, the standard deviations of the accuracy, and numerical mean of the accuracy. The Flan-T5 language model predicted whether the student is vulnerable or proficient based on the data that was fed to the LM.

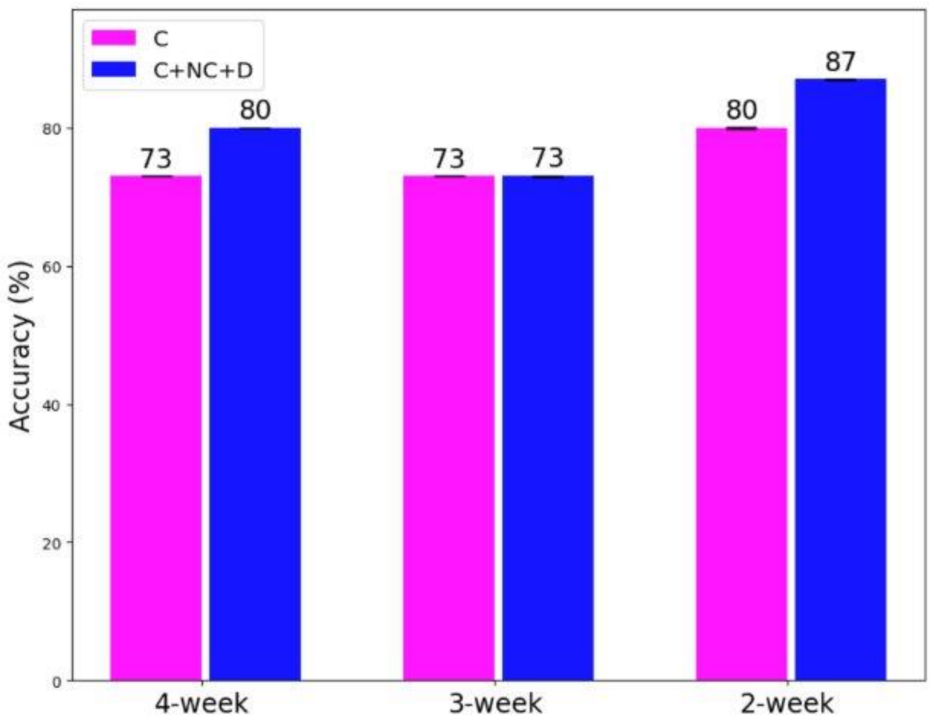

Figure 2: C vs C+NC+D

Figure 2 shows the best accuracy comparison among cognitive data only vs cognitive, noncognitive, and distal. When all three features are included, the Flan-T5 LM had better accuracy when predicting academic performance than when only cognitive data was included.

	precision	recall	f1-score	support
At the end of the semester, the student will be vulnerable.	0.00	0.00	0.00	4
At the end of the semester, the student will exhibit proficient performance.	0.73	1.00	0.85	11
accuracy			0.73	15
macro avg	0.37	0.50	0.42	15
weighted avg	0.54	0.73	0.62	15

Figure 3: Results with only non-LMs

All of the of the cognitive datasets were then run through the Non-LMs (Transformers, 1D CNN, and LSTM). As expected, all 3 of the non-LMs outputted the same result: 73% accuracy when predicting the proficient class; i.e, not performing well and not learning.

5. Discussion

Early forecasting of academic performance is crucial for timely support, particularly in STEM disciplines where the attrition rate is very high. Traditional forecasting methods rely only on cognitive data collected later on in the semester, making it difficult to identify vulnerable students early on. Our study demonstrates that integrating all three features (Cognitive, Non-Cognitive, and Distal), significantly improves the accuracy of early forecasts than when only using Cognitive data.

5.1 Implications

Additionally, the findings of this study have several implications for educational use.

1. **Comprehensive Data Integration**: Including cognitive, non-cognitive, and socioeconomic factors can enhance the predictive powers of LMs, allowing for a more accurate prediction of vulnerable students in STEM courses.
2. **Early Intervention:** Identifying vulnerable students early in the semester enables timely interventions that can significantly impact student outcomes, improving retention and graduation rates in STEM programs.

5.2 Limitations & Future Work

Despite the promising results, we did have a major limitation.

1. **Small Sample Size:** The sample size of 48 students and the samples fed to the Flan-T5 model were relatively small, limiting the generalizability of the findings. Additionally, because of the small sample size, one high or low accuracy can skew the results. Future research should involve a larger sample size to validate these results.

Future work will involve scaling up the study to include a larger population of students at the University of Nebraska-Lincoln. Furthermore, we plan to investigate the long-term impact of early forecasting and interventions on academic performance and retention. This will involve longitudinal studies tracking student outcomes. The development of a robust and scalable forecasting system will ultimately contribute to the larger goal of developing a possible intervention system. Through this we aim to improve educational outcomes and support student success in STEM disciplines. We will continue collecting data in the Fall 2024 & Spring 2025 semesters.

6. Conclusion

To conclude, this study demonstrates the potential of pre-trained Language Models (LMs) in enhancing early forecasting of student performance in STEM education. Traditional forecasting methods often rely only on cognitive data (formative and summative scores) collected later on in the semester. As a result, these methods fail to identify vulnerable students early enough to provide timely interventions and support. Our approach addresses this gap by integrating 3 key features: Cognitive, Non-cognitive, and Distal data into the forecasting process. The insights gained from this study pave the way for future innovations in educational practice and policy, ultimately contributing to the advancement of STEM education and student success.

7. References

[1]

M. Leary, A. Morewood, and R. Bryner, “A controlled intervention to improve freshman retention in a STEM-based physiology major,” *Advances in Physiology Education*, vol. 44, no. 3, pp. 334–343, Sep. 2020, doi: 10.1152/advan.00038.2020.

[2]

S. Freeman *et al.*, “Active learning increases student performance in science, engineering, and mathematics,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 23, pp. 8410–8415, Jun. 2014, doi: 10.1073/pnas.1319030111.

[3]

D. Findley-Van Nostrand and R. S. Pollenz, “Evaluating Psychosocial Mechanisms Underlying STEM Persistence in Undergraduates: Evidence of Impact from a Six-Day Pre–College Engagement STEM Academy Program,” *LSE*, vol. 16, no. 2, p. ar36, Jun. 2017, doi: 10.1187/cbe.16-10-0294.

[4]

K. Arnold and M. Pistilli, “Course signals at Purdue: Using learning analytics to increase student success,” *ACM International Conference Proceeding Series*, Apr. 2012, doi: 10.1145/2330601.2330666.

[5]

D. Findley-Van Nostrand and R. S. Pollenz, “Evaluating Psychosocial Mechanisms Underlying STEM Persistence in Undergraduates: Evidence of Impact from a Six-Day Pre–College Engagement STEM Academy Program,” *LSE*, vol. 16, no. 2, p. ar36, Jun. 2017, doi: 10.1187/cbe.16-10-0294.

[6]

F. Rodriguez, M. J. Rivas, L. H. Matsumura, M. Warschauer, and B. K. Sato, “How do students study in STEM courses? Findings from a light-touch intervention and its relevance for underrepresented students,” *PLOS ONE*, vol. 13, no. 7, p. e0200767, Jul. 2018, doi: 10.1371/journal.pone.0200767.

[7]

L. C. Page and H. Gehlbach, “How an Artificially Intelligent Virtual Assistant Helps Students Navigate the Road to College,” *AERA Open*, vol. 3, no. 4, p. 2332858417749220, Oct. 2017, doi: 10.1177/2332858417749220.

[8]

Y. Chen, A. Johri, and H. Rangwala, “Running out of STEM: a comparative study across STEM majors of college students at-risk of dropping out early,” Mar. 2018, pp. 270–279. doi: 10.1145/3170358.3170410.

[9]

X. Chen, “STEM attrition among high-performing college students: Scope and potential causes,” *Journal of Technology and Science Education*, vol. 5, no. 1, Art. no. 1, Mar. 2015, doi: 10.3926/jotse.136.

[10]

A. Sithole, E. T. Chiyaka, P. McCarthy, D. M. Mupinga, B. K. Bucklein, and J. Kibirige, “Student Attraction, Persistence and Retention in STEM Programs: Successes and Continuing Challenges,” *Higher Education Studies*, vol. 7, no. 1, Art. no. 1, Jan. 2017, doi: 10.5539/hes.v7n1p46.