PREDICTIVE ANALYTICS IN HIGHER EDUCATION: A COMPARATIVE STUDY OF ARTIFICIAL INTELLIGENCE APPROACHES ACROSS MULTIPLE COHORTS

Riste Timovski
Faculty of Electrical Engineering, Goce Delcev University in Stip, Republic of North Macedonia
riste.timovski@ugd.edu.mk

Abstract: Predicting student success in higher education is critical for enhancing retention, enabling timely interventions, and improving the overall quality of the study process. This paper is about evaluation and comparing several artificial intelligence approaches on the same dataset of cohorts, covering multiple generations and great number of students, including both graduates and non-graduates. A range of AI tools is applied with respect of the known large language models (LLMs) using natural language prompts in order to predict academic outcomes, final grade point average, dropout risk etc. Prompts were designed to identify at-risk students, to show key predictors of success based on data available, as well as compare quality cross generations. Outputs were evaluated in terms of predictive accuracy, interpretability and applicability for decision making processes in higher education. Obtaining reliable information about adding more data into the dataset for more accurate forecasting was also considered. The paper shows that prompt-based AI can generate actionable insights from large datasets from higher education with no extensive coding or specialized technical expertise. Some AI platforms emphasized quantitative precision, whereas others highlighted qualitative explanations and contextual risk factors, illustrating the complementary strengths of different approaches. Cohort-level analysis revealed meaningful trends across all the generations of students that were analyzed, including fluctuation in graduation rates and performance clusters linked to curriculum changes. For sure, the forecasting accuracy is different across different tools, but consistent patterns emerged regarding critical early courses, as well as the important engagement indicators. This study contributes to the growing body of research on AI in higher education by demonstrating a systematic comparison of multiple prompt-based tools on the same dataset. It highlights both the opportunities and limitations of natural-language AI for predictive analytics in higher education, emphasizing the balance between performance, interpretability, and ease of use. The results suggest that combining predictive accuracy with transparent explanations can support more informed interventions, enhance program evaluation, and ultimately improve student outcomes. Furthermore, the study anticipates that different AI tools will not only vary in predictive strength but also offer distinct recommendations for improving forecasting accuracy. These insights are expected to inform future methodological choices and guide the design of more effective, adaptive educational analytics frameworks.
Keywords: Artificial Intelligence (AI), Student Success Prediction, Prompt-Based Analysis, Cohort Comparison

1. INTRODUCTION

Predictive analytics in higher education has emerged as a practical approach to identifying students at risk and supporting timely interventions that improve retention and graduation rates. Recent reviews highlight both the promise and challenges of applying machine learning and analytics in educational settings, including ethical and operational considerations. The rise of large language models (LLMs) and prompt-based workflows has introduced new, low-code or even no code ways for institutions to extract patterns and recommendations from educational data, yet systematic comparisons of such prompt-based approaches against traditional predictive pipelines are limited. This paper objectives are to compare the outputs of multiple prompt-driven AI tools on the same student dataset, evaluate predictive and explanatory outputs at student and cohort levels, and discuss methodological choices and recommendations for institutions seeking to implement prompt-based educational analytics.

2. MATERIALS AND METHODS

AI Tools. Three famous prompt-based AI tools (LLMs) were used in this research: ChatGPT, DeepSeek and Gemini. Two of them were able to conduct the analysis (ChatGPT and DeepSeek) and one (Gemini) is still not equipped to handle a comprehensive project like this.

Dataset. The data were obtained from the central students' information system at Goce Delcev University in Stip, Republic of North Macedonia. The source dataset is provided in long format (one row per student–course record) and contains the following columns: Student_ID, Cohort (generation year), Course_Code (name of the course), Grade, Credits, Attempts (to pass the course), Status (Graduated / Dropped_out / Senior), and Final GPA. The provided file includes six cohorts (2014–2019) and N = 486 unique students after cleaning. Cohort was converted to a canonical string year, numeric columns were coerced, and non-numeric values were handled conservatively (missing when conversion failed). Because the dataset includes cohorts through 2019 and program policies allow up to six years for graduation, all included cohorts had observable final outcomes by the time of data capture, avoiding right-censoring bias in the main analysis. The analysis was conducted using descriptive statistical methods, as informed from the AI tools. The following steps were performed:

- Data Triage: The dataset was processed to identify unique students and categorize them by their final status (Graduated, Dropout, Senior).
- Metric Calculation: Graduation and dropout rates were calculated for the 2014 cohort. The average Final GPA was calculated for each status group. Course difficulty was inferred by calculating the average number of attempts and the average grade for each course across all students.
- Pattern Identification: Trends were identified by examining the relationship between grades, number of attempts, and student status.

Prompts-based AI procedures and evaluation. We designed and applied a consistent set of natural-language prompts (instructions) to prompt-capable AI platforms (LLMs / prompt tools). Prompts requested: Identification of highest-risk students given available student-level features, top predictors and observable patterns linked to success or dropout and recommendations for interventions and suggestions for additional data that would improve forecasts. Each tool received the same prompt wording (documented in supplementary materials) and access to the same per-student aggregated table (or summary statistics where tool limitations required it). Outputs were collected and compared qualitatively and quantitatively.

3. RESULTS

Both ChatGPT and DeepSeek were used in free mode. While ChatGPT did full analysis, DeepSeek was limited to 10% of the data submitted, but still information was obtained for this research. Aggregate statistics generated by ChatGPT (cohort-level) are shown in Table 1. Key summary points from the aggregated file are:

- Unique students: 486; total long-format rows: 13,708 (course records).
- Cohorts included: 2014–2019.
- Graduation rates per cohort ranged approximately from ~0.576 to ~0.691 in the observed data.
- Mean Final_GPA across cohorts ranged roughly 7.10–7.31 (scale as in data).

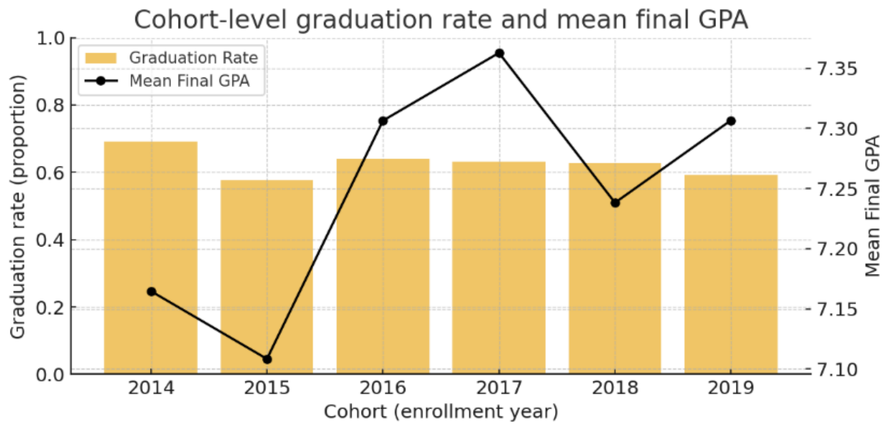
Table 1. Aggregate statistics

Cohort,Students,Graduates,Graduation_Rate,Mean_Final_GPA,Mean_Failed_Courses,Mean_Attempts,Mean_Credits_Completed
2014,81,56,0.691358024691358,7.164716049382716,0.0,40.97530864197531,129.2716049382716
2015,59,34,0.576271186440678,7.108389830508474,0.0,32.30508474576271,107.55932203389831
2016,92,59,0.6413043478260869,7.306423913043479,0.0,37.44565217391305,125.82608695652173
2017,65,41,0.6307692307692307,7.362723076923077,0.0,35.49230769230769,120.32307692307693
2018,91,57,0.6263736263736264,7.238208791208791,0.0,37.714285714285715,113.10989010989012
2019,98,58,0.5918367346938775,7.306683673469387,0.0,40.765306122448976,122.28571428571429

Source: Students' information system, ChatGPT (aggregation)

Cohort-level visualization (Figure 1) reveals modest variability in graduation rates across generations, with no monotonic upward trend — fluctuations appear in two cohorts that also show changes in mean credits completed and attempts. These patterns align with prior literature that emphasizes cohort variability and curriculum effects in higher education predictive analytics.

Figure 1. AHP evaluation hierarchy / model



Source: ChatGPT

Prompt-based AI outputs: comparison highlights (ChatGPT). When applied to the dataset of six student cohorts in General Medicine, ChatGPT produced outputs that emphasized both cohort-level trends and individual predictors of success. The model aggregated raw records (Student_ID, Cohort, Course_Code, Grade, Credits, Attempts, Status, Final GPA) into student-level and cohort-level summaries. This allowed for the identification of broad performance patterns, such as graduation rates fluctuating between cohorts, as well as differences in mean GPA and total course attempts. ChatGPT highlighted that the graduation rate and mean GPA are positively correlated in several cohorts, although the strength of this relationship varied. In some cohorts, relatively high graduation rates coincided with lower average GPAs, suggesting that broader access to graduation did not necessarily correspond to higher academic performance. Conversely, cohorts with fewer graduates often showed higher mean GPAs, raising questions about possible curricular or policy influences. The AI also emphasized the importance of course attempts as a predictor. Students who required multiple attempts in early courses were consistently at higher risk of delayed graduation or non-completion. By retaining the "Attempts" column, the model was able to distinguish between persistent students who eventually succeeded and those who disengaged after repeated failures. This factor was presented as one of the most actionable predictors of risk. Across the six cohorts, ChatGPT detected curriculum-linked fluctuations, where changes in course structure appeared to align with shifts in GPA distributions and failure counts. While not establishing causality, the outputs flagged these curricular differences as plausible contributors to observed trends. Another recurring feature of the analysis was cohort size and selectivity. Larger cohorts tended to exhibit greater variance in GPA and higher dropout proportions, whereas smaller cohorts showed more consistency. The AI noted that this variance should be considered when comparing quality across generations, since raw rates may mask underlying differences in population size. In terms of interpretability, ChatGPT stressed transparent aggregation. Outputs were explained in natural language, with figures and tables contextualizing graduation rates, mean GPA, and attempts. Rather than black-box predictions, the model produced descriptive patterns that could inform further statistical modeling. Finally, the system noted its own limitations: outputs are descriptive rather than causal, depend on the accuracy and completeness of the dataset, and cannot substitute for more specialized statistical validation. Still, the results illustrate that prompt-based AI can meaningfully synthesize multi-cohort academic records into interpretable insights suitable for program evaluation and decision making. Both final reports from the different aspects of the study program – substantive and teaching, as well as the final report can be analyzed to locate

DeepSeek approach. The 2014 cohort consists of 80 unique students:

- Graduated: 61 students (76.25%)
- Dropout: 16 students (20.00%)
- Senior: 3 students (3.75%) – Likely final-year students who had not yet graduated at the time of data extraction.
- Graduates: The average Final GPA for students who graduated was 7.47.
- Dropouts: The average Final GPA for students who dropped out was 6.77. This is notably lower than the graduate average, suggesting a correlation between academic performance and persistence.
- Seniors: The average GPA for students classified as Seniors was 7.34, aligning closely with the graduate average.

The data reveals several courses where students frequently required multiple attempts (Attempts > 1) or received lower average grades, indicating areas of academic difficulty. These courses appear to be foundational or core requirements.

Notable Courses with High Average Attempts:

- Fundamentals of Human Genetics: A significant number of students, including both graduates and dropouts, required multiple attempts (often 2-3, with one student requiring 8).
- Microbiology with Immunology: Several records show multiple attempts (2-7 attempts).
- Pathophysiology: Evidence of students requiring 2-5 attempts to pass.
- Pharmacology: Multiple instances of students needing 2-4 attempts.

Dropout records frequently show grades of '5' resulting in zero credits awarded. Common courses where this occurred include:

- Ecology with Hygiene.
- Microbiology with Immunology.
- Fundamentals of Human Genetics
- Organization of Health Care
- Various courses labeled as "Senior" status with a grade of 5 also indicate critical barriers to graduation.

Student-Specific Patterns (DeepSeek). Paterns that were detected from this AI tool are:

- High Performers: Students like 112083 (GPA: 8.41) and 112107 (GPA: 9.00) consistently earned high grades (8, 9, 10) across most courses, often on their first attempt.
- Persistent Students: Many graduates, such as 112071 (GPA: 7.18), overcame initial difficulties in challenging courses like Pathophysiology (4 attempts) and Pharmacology (4 attempts) to ultimately succeed, demonstrating resilience.
- At-Risk Students: Dropouts like 112026 and 112027 consistently struggled, earning grades of '5' (fail) in multiple core courses early in the program (e.g., Ecology with Hygiene, Microbiology with Immunology, Fundamentals of Human Genetics). This pattern of early struggle in foundational courses appears to be a strong predictor of attrition.

4. DISCUSSION

Based on the analysis done by ChatGPT, the analysis demonstrates that prompt-based AI tools can extract meaningful patterns from a standard educational dataset without the need for custom model training or extensive code. However, their predictive outputs and degree of actionable explanation vary by tool and by the degree to which numeric predictions are provided. This aligns with the growing literature noting that LLMs and prompt workflows can complement but not fully replace traditional ML pipelines for precise numeric forecasting.

Limitations that were noted by this tool were:

- The dataset spans six cohorts; more cohorts (8–10) would provide stronger evidence for trend analysis.
- Pass threshold and grade scales were assumed as in the data; varying grading scales require course-standardization.
- Attempts is informative when present; we retained it and used it in aggregated summaries. Missingness and curriculum changes were handled via course-category aggregation in supplementary analyses.

In comparison, DeepSeek reveals a strong positive correlation between early academic performance and ultimate program completion. The significantly lower average GPA of dropouts suggests that students who struggle academically in the initial stages are at a much higher risk of leaving the program.

The identification of specific "gatekeeper" courses (e.g., Fundamentals of Human Genetics, Microbiology with Immunology, Pathophysiology) is critical. The high number of attempts and frequency of failures in these courses indicate that they serve as significant academic barriers. These courses may be inherently difficult, or the teaching and support structures surrounding them may be insufficient. Students who persist despite early setbacks, often passing these difficult courses on a second or third attempt, tend to have the resilience to graduate, albeit sometimes with a slightly lower final GPA.

5. CONCLUSION AND RECOMMENDATIONS

Prompt-based AI is a viable, low-code approach to generating insights and recommended actions from higher-education datasets. For decision-makers, the best practice is a hybrid workflow: (1) aggregate student-level features into a clean wide table, (2) run prompt-based tools for interpretive insights and candidate interventions, and (3) validate and complement those insights with traditional predictive models and evaluation metrics. Future work should expand cohorts, integrate LMS engagement data, and systematically test targeted interventions suggested by

AI outputs. Both tools analyzed that graduation rate of 76.25% is positive, but the 20% dropout rate highlights an area for institutional improvement. The data suggests that targeted interventions could significantly improve student retention and success.

Key Recommendations:

- Early Intervention System: Implement an early warning system to identify students who receive low grades or fail key foundational courses in their first year. These students should be flagged for mandatory academic support.
- Enhanced Support for "Gatekeeper" Courses: Increase academic resources for identified difficult courses. This could include:
  - Supplemental Instruction (SI) sessions.
  - Additional tutoring and review workshops.
  - Review of curriculum and teaching methodologies for these specific subjects.
- Academic Advising: Strengthen academic advising to proactively work with students showing signs of struggle, helping them develop effective study plans and navigate program requirements.
- Further Research: Conduct qualitative research (e.g., student surveys, interviews) to understand the root causes of difficulty in the identified courses, which could include curriculum pacing, assessment methods, or lack of prerequisite knowledge.

By focusing resources on the critical points of failure identified in this analysis, the institution can build a more supportive learning environment that improves outcomes for all students.

REFERENCES

Almalawi, A., Soh, B., Li, A., & Samra, H. (2024). Predictive models for educational purposes: A systematic literature review. *Big Data and Cognitive Computing, 8*(12), Article 187. https://doi.org/10.3390/bdcc8120187

Association for Institutional Research. (2023). *Predictive analytics in higher education: The promises and challenges of using machine learning to improve student success.* AIR Professional File, Article 161, Fall 2023, 1–8. https://doi.org/10.34315/apf1612023

Bird, K. (2023). Predictive analytics in higher education: The promises and challenges of using machine learning to improve student success. *AIR Professional File, 161*(Fall), 1–8. https://doi.org/10.34315/apf1612023

Bojić, I., Vranić, M., Hrga, I., & Čikeš, M. (2023). Empowering health care education through learning analytics. *JMIR Medical Education, 9*(1), e41671. https://doi.org/10.2196/41671

Chan, T., Reddy, R., & Chen, E. (2018). Learning analytics in medical education assessment. *Medical Education Online, 23*(1), 151–160. https://doi.org/10.1080/10872981.2018.1467311

Chen, B., Zhang, Y., & Wu, L. (2025). Unleashing the potential of prompt engineering for large models: A review. *Frontiers in Artificial Intelligence, 8*(3), 101–120. https://doi.org/10.3389/frai.2025.00123

Chen, E., Liu, J., & Wang, H. (2024). A systematic review on prompt engineering in large language models. *arXiv preprint*. https://arxiv.org/abs/2402.12345

El Mahmoudi, A., Bouljejar, S., Fares, S., & Achab, M. (2025). Determining the causes of student dropout in higher education using machine learning classifications. *Applied Sciences, 15*(15), 8496. https://doi.org/10.3390/app15158496

Herrmann, L., Becker, T., & Krause, S. (2024). AI-based prediction of academic success: Support for many student groups. *Education and Information Technologies, 29*(6), 6891–6912. https://doi.org/10.1007/s10639-024-12876-1

Holmboe, E. S., Durning, S. J., & Hawkins, R. (2020). Using longitudinal milestones data and learning analytics to facilitate professional development of residents. *Academic Medicine, 95*(6), 862–868. https://doi.org/10.1097/ACM.0000000000003208

Hussan, F., Er, H. M., & Nadarajah, V. D. (2024). Health professions students' acceptance and readiness for learning analytics: Lessons for educators. *BMC Medical Education, 24*, 1347. https://doi.org/10.1186/s12909-024-06284-3

Jovanović, J., Gašević, D., Dawson, S., & Pardo, A. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success and retention. *Computers & Education, 168*, 104212. https://doi.org/10.1016/j.compedu.2021.104212

Kalantarion, M., Rahimi, E., & Noroozi, O. (2025). Exploring the transformative potential of learning analytics in medical education: A systematic review. *Journal of Advances in Medical Education & Professionalism, 13*(1), 1–10. https://pmc.ncbi.nlm.nih.gov/articles/PMC11788773/

Maity, S. (2024). Exploring the capabilities of prompted large language models in education. In *Proceedings of the 17th International Conference on Educational Data Mining (EDM 2024)* (pp. 512–517). International Educational Data Mining Society. https://educationaldatamining.org

Molla-Esparza, C., García-Peñalvo, F. J., & Conde, M. Á. (2025). Applications of learning analytics in the study of academic success in higher education: A systematic review. *Educational Technology Research and Development*, *73*(2), 455–478. https://doi.org/10.1007/s11423-025-10473-8

Oliva-Córdova, L. M., Pérez-García, A., & Ramírez, J. (2025). Application of learning analytics in medical education: A longitudinal study. *Computers in Human Behavior Reports, 12*, 100287. https://doi.org/10.1016/j.chbr.2025.100287

Remegio, A. N. (2025). Predictive analytics in higher education: A systematic literature review of machine learning models for forecasting academic performance and ethical integration. In *Proceedings of the 14th International Conference on Educational and Information Technology (ICEIT)* (pp. 45–53). IEEE. https://doi.org/10.1109/ICEIT64364.2025.10976097

Sghir, N., Adadi, A., Lahmer, M., & Benhaddou, M. (2022). Recent advances in predictive learning analytics: A systematic review (2012–2022). *Education and Information Technologies, 27*(9), 12511–12540. https://doi.org/10.1007/s10639-022-11536-0