Data Science and Engineering (DSE) Record, Volume 5, issue 1.

Student Academic Performance Prediction Model in Computer Related Courses Field.

Sumana Ganne [1,2] and Chartchai Doungsa-ard[3]

[1] Data Science Program, Chiang Mai University, Chiang Mai, Thailand
[2] Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand

sumana_ganne@cmu.ac.th

Abstract The phenomenon of student attrition is a pressing issue for higher education institutions globally. Universities aim to maximize their graduation rates, but maintaining a balance between enrollment and graduation has been challenging for decades. It's critical for universities to understand the rates and reasons behind student attrition, as well as when students are most at risk of dropping out, to implement effective strategies to address this issue. Most dropouts occur early in university life, often due to poor academic performance. This independent study aims to use data to identify factors affecting student performance and create a predictive model for their performance in advanced courses. The results will inform institutional policies and strategies to improve faculty-student interactions and increase retention rates. Identifying at-risk students early and creating support pathways are crucial steps toward reducing student attrition.

Keywords: student attrition; propensity model.

1 Introduction

Student attrition is a growing concern for higher educational institutions across the world and the issue of student attrition is becoming a central focus of study in higher education [1]. This is a phenomenon where students discontinue their studies before completing their degree programs. High attrition rates can be indicative of underlying issues within educational institutions, such as inadequate support systems or unengaging curricula.

One of the main goals of any university is to produce as many qualified graduates as possible [2], however, maintaining the right balance of students enrolled and graduated has become a challenge for educational institutions for the past few decades. A study in [3], analysed the determinants of student propensity to dropout at Prince of Songkhla University, Pattani campus. The overall dropout rate over the five-year period was 23.9%, and a decreasing trend in dropout rate was found from second semester and onwards.

Lower grades in initial semesters are a significant predictor of student attrition, highlighting the need for targeted academic assistance. Attrition not only hampers students' personal and professional development but also results in financial losses for both the students and the institutions. Universities are increasingly implementing data-driven strategies to identify at-risk students and provide timely interventions.

Addressing the problem of student attrition due to lower grades in the early university years requires a multi-pronged approach. Institutions must develop comprehensive support systems to identify at-risk students and provide the necessary academic support, mentorship, and resources to guide them through these challenging times. By focusing on this crucial aspect of student attrition, universities can foster a more supportive learning environment that encourages perseverance and academic success, leading to higher retention rates and more graduates poised to make meaningful contributions to society.

2 Literature Review

One foundational work in the field is Vincent Tinto's "Student Departure: Explaining the Causes and Cures," which was first published in 1975. Tinto's model of student departure has been highly influential; it argues that student retention is related to the individual's integration into both the academic and social systems of an institution. Other seminal authors and studies include John Bean's work on student engagement and persistence, as well as research by Spady (1970) and Pascarella (1980), who also contributed to developing theoretical frameworks for understanding student attrition.

Earlier published literature examines factors such as:

- Academic preparedness and performance
- Financial pressures and the cost of education
- Social integration and the sense of belonging within the campus community.
- Personal circumstances, including family support and personal health.
- Institutional factors, such as the quality of teaching, support services, and campus climate

The earlier literature on student attrition has laid the groundwork for ongoing research in the area, which continues to evolve with changing educational landscapes, student demographics, and societal factors. Researchers use these foundational theories and studies to build upon and refine strategies for improving student retention in universities.

In [1], the author Vincent Tinto reviews research on student retention and suggests educational institutions should create supportive communities for all students, not just identify at-risk individuals. The paper calls for more complex, longitudinal research to understand different student experiences. In [2] Vincent Tinto emphasizes the importance of viewing college experiences through students' eyes to improve retention, highlighting the need for a sense of belonging and strong student-faculty relationships, especially in the first year. The study in [3], examines numerous factors that contribute to university dropout rates, aiming to provide data that can help develop strategies and support systems to improve student retention. In the 2012 AIR study [4] quantifies the budgetary impact of student attrition on US institutions, highlighting the lost revenue from dropouts and urging the implementation of effective retention strategies to improve outcomes and stabilize finances. In the [5], the author discusses the longitudinal nature of student departure from higher education, suggesting that leaving college is a process involving multiple stages over time rather than a singular event. A model is presented to illustrate the stages of departure, emphasizing that intervention strategies should be time-sensitive and tailored to the challenges students face at various points during their college journey. An attempt to explores methods to predict academic success in higher education is proposed in [7], examining predictors such as prior academic performance, socio-economic background, and psychological factors. It also reviews how institutions use predictive modelling to improve student outcomes.

R. Sittichai investigates the reasons for university student dropouts in Thailand [9], examining academic, financial, and personal challenges, as well as the broader socio-economic and cultural context. Recommendations are provided for reducing dropout rates by addressing the specific needs of Thai students. Finally, the study that examines the interactions between underrepresented students and faculty in the sciences across different college contexts is proposed in [10]. It investigates how these interactions vary by institution type, departmental culture, and other factors, with the aim of understanding how to better support these students in their academic pursuits.

3 Methodology

This research applied the CRISP-DM method of data mining. CRISP-DM is iterative, meaning that these phases are not strictly linear, and it is common to circle back to previous steps as new insights are gained or challenges are encountered. The framework is also adaptable, allowing it to be used in a variety of industries and with different types of data and business problems.

The methodology consisted of five major phases:

1. Defining the Objectives: This initial phase will focus on understanding the project objectives. The primary objective of this study is to predict if a student has a propensity to dropping out based on his/her performance in earlier semesters of a course.

2. Data Understanding: This phase involves collecting initial data, describing it, exploring it, and verifying its quality to uncover potential issues that could affect the analysis. The student grades in computer science courses from Chiangmai University for courses highlighted in red below between academic years 2017 to 2019 are the basis of this study.

3. Data Preparation: The data collected from CMU needed to be of one batch of students that traversed through all the levels of Computer Science courses above. In the dataset received, after intense analysis and exploration, 74 student IDs who started level 1 course Programming Logic Thinking" in 2017, 2018 and Level2 course "OOP" in 2018,2019 were normalized to get data. Any grade D or below was considered as a candidate who can potentially drop out. All Attendance and Marks scored were converted to percentage and grade were converted to Dropout Flag (Y, N). Grade A, B, C are considered Dropout (N) and D, F are considered (Y).

4. Modeling: Various modeling techniques were selected and applied. Multiple models were created to identify the most effective approach for the problem at hand. The following modeling were used – Decision Tree Classifier, Logistic Regression and Random Forest classification.

5. Evaluation: The models are evaluated with respect to the objectives established in the first phase. This involves assessing the results of the modeling process to determine if there is a satisfactory model that meets the initial objectives. If not, the previous phases are revisited.

4 Result

We explored multiple grading datasets to draw and analyze the correlation between variables. A lot of results have been obtained. Attendance in Lecture/Lab, Midterm score, Attendance score, Project score, Final Grade were importance factors considered. Students with lower attendance seemed to have lower overall total. Students with low mid-term scored tend to get lower grades in general.
Student ID was used a unique identifier to follow the path of grades a student took between Level1 and Level 2.

The following models were created to predict the

4.1 Decision Tree Classifier

Accuracy: 0.578

4.2 Logistic Regression

Accuracy – 0.947

4.3 Random Forest Classifier

Accuracy – 0.947

The research attempted at predicting if a student will dropout at the end of Level2 course itself so the university can take corrective action and provide the necessary help and guidance needed.

5 Discussion

This study identifies academic performance, particularly in the early stages of education, as a key predictor of student dropout rates. It highlights that low grade, especially in crucial foundational courses, may signal a higher risk of dropping out. The trend of a student's grades over time also provides insight into their likelihood of continuing their studies, with improving grades suggesting engagement and declining grades indicating potential problems. Establishing GPA thresholds for intervention can help identify students who need extra support. Additionally, disparities in performance across subjects and class participation are factors that can influence dropout rates. Early prediction of at-risk students allows for more timely interventions. A pattern of course withdrawals and the difficulty or quantity of courses taken are associated with higher attrition rates. Students' perceptions of their performance relative to peers can also affect their likelihood to stay.

However, the study's accuracy is limited by the lack of demographic data, which is crucial in predicting attrition since factors like age, gender, ethnicity, and socioeconomic status are significant indicators. Without this data, the model may miss important trends and produce biased predictions, potentially leading to ineffective intervention strategies and an overreliance on academic performance and engagement metrics. Demographic data is also essential for designing targeted support interventions for various student groups.

6 Conclusion

Our study was demonstrated that Attendance and midterm results were key factors in the dropout rate. Future enhancements of this model are possible if demographic data can be added to this along with income levels, age, and sex. The results of the model however even with 2 levels of scoring were impressive with good accuracy rates.

References

[1] Tinto, V. (2006). Research and practice of student retention: What next? Journal of College Student Retention, 8(1), 1-19. https://doi.org/10.2190/C0C4-EFT9-EG7W-PWP4
[2] Tinto, V. (2015). Through the eyes of students. Journal of College Student Retention: Research, Theory & Practice, 19(3), 254-269. https://doi.org/10.1177/1521025115621917
[3] Tentsho, Kelzang & McNeil, Rhysa & Tongkumchum, Phattrawan. (2019). Determinants of University Dropout: A Case of Thailand. Asian Social Science. 15. 49. 10.5539/ass. v15n7p49.
[4] The institutional costs of student attrition. (2012) Nate Johnson P Institution: American Institute of Research ISBN 2024035000 https://deltacostproject.org/sites/default/files/products/Delta-Cost-AttritionResearch-Paper.pdf
[5] Tinto, V. (1980). Stages of Student Departure: Reflections on the longitudinal character of student leaving. The Journal of Higher Education, 59(4), 438-455. https://doi.org/10.2307/1981920
[6] R. Vilanova, M. Dominguez, J. Vicario, M.A. Prada, M. Barbu, M.J. Varanda, P. Alves, M. Podpora, U. Spagnolini, A. Paganoni, Data-driven tool for monitoring of students performance ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2019.08.188.
[7] Alyahyan, E., Düştegör, D. Predicting academic success in higher education: literature review and best practices. Int J Educ Technol High Educ 17, 3 (2020). https://doi.org/10.1186/s41239-020-0177-7
[8] Synco, T. M. (2012). Background or experience? Using logistic regression to predict college retention (Doctoral dissertation, University of Alabama). Retrieved from http://www.mhsl.uab.edu/dt/2012p/synco.pdf
[9] Sittichai, R. (2012). Why are there dropouts among university students? Experiences in a Thai university. Inter. Journal of Educational Development, 32(2), 283-289. https://doi.org/10.1016/j.ijedudev.2011.04.010
[10] Hurtado, S., Eagan, M. K., Tran, M. C., Newman, C. B., Chang, M. J., & Velasco, P. (2011). "We do Science Here": Underrepresented students' interactions with faculty. in different college contexts. Journal of Social Issues, 67(3), 553-579. https://doi.org/10.1111/j.1540-4560.2011.01714.x

[11] Al-Radaideh, Qasem & Al-Shawakfa, Emad & Al-Najjar, Mustafa. (2006). Mining Student Data Using Decision Trees. The International Arab Journal of Information Technology – IAJIT

[12] Gilchrist M, Mooers D, Skrubbeltrang G, Vachon F. Knowledge Discovery in Databases for Competitive Advantage. Journal of Management and Strategy. 2012; 3:2.