A major issue that has persisted in higher education institutions is the student dropout risks. Not only do universities lose students but also face financial consequences related to the enrolment and retentions statistics. For students, dropping out can often lead to long lasting psychological and economic burdens. Some of those burdens may be an increase in debt without a degree, lower salaries for the majority of their life, and less opportunities to advance in their careers.

According to the Education Data Initiative, data shows that 39% of students seeking a bachelor’s degree for the first time do not finish their program within 8 years. Being aware of early identifiers for students at risk of dropping out allows for faster interventions. That can be done through academic support, mental health services, financial aid, or mentoring programs. These actions can potentially change the trajectory of a student’s academic career.

Predictive analytics provide a strong approach to addressing this issue. Through the analysis of previous student records, which could mean academic performance, demographic, economic background, and other factors, it’s possible to make a model that predicts which students are most likely to drop out. This will help universities to manage their resources more effectively and improve the graduation rate.

This project will explore a machine learning predictive model for identifying students that are at risk by using public educational datasets. The goal is to make an easy to understand system that educational institutions can use to decrease dropouts and improve academic success.

Multiple universities have explored predictive analytics in education. Georgia State University implemented an early alert system that helped decrease their dropout rate by using more than 800 predictive indicators. Civitas Learning and Blackboard also provide analytics platforms that use students’ data to flag academic risk.

Even though these solutions are quite effective, they are solely proprietary. This means they are inaccessible to smaller institutions and researchers. That’s not the only thing, but there are also a few that focus on transparency and interpretability. Those two attributes are crucial in education when lives are significantly impacted by decisions based on predictions. Unlike the similar solutions, this project will use public tools and interpretable models on datasets released to the public.

The model interpretability is important for educational settings. Decisions based on vague algorithms can wear down trust, even more so when it influences a student’s access to resources or forms of intervention. Using models like logistic regression or decision trees will allow faculty and advisors to understand why students were flagged to be at risk of dropping out. Having this clarity will ensure that interventions are justified, ethical and student oriented.

For the scope of the work, the project will focus on developing and applying classification models to determine wither a student is at risk of dropping out. It’ll start with exploring publicly available educational data sets to ensure the quality and relevance of the data. One key aspect of this phase has to do with identifying and choosing the most significant variables that affect dropout risk. The variables may have academic indicators like GPA, attendance records, number of courses, and standardized test scores. Other additional factors could be age, gender, financial status, and family income.

After the relevant factors have been selected, multiple machine learning classification models will be evaluated based on their performance and interpretability. Logistic regression will be set as the standard model because of its simplicity, easy interpretability, and common use in binary classification problems. Additional models like decision trees, random forests, and support vector machines could be tested and analyzed using values such as accuracy, recall, and F1 score.

As the process goes on, there’ll be an emphasis on transparency and explainable system that educational professionals can use confidently. The end goal of this is to have a tool that aids universities and schools in making informed decisions to successfully support students and reduce dropout rates.

References

Hanson, Melanie. “College Dropout Rates” EducationData.org, August 16, 2024,  
<https://educationdata.org/college-dropout-rates>

Georgia State University: Leading with Predictive Analytics

<https://success.gsu.edu/approach/>