

Predictive Analytics in Higher Education

 newamerica.org/education-policy/policy-papers/predictive-analytics-higher-education/

Introduction

Colleges are under increasing pressure to retain their students. Federal and state officials are demanding that those who enter their public institutions— especially students from underrepresented groups— earn a degree. Over two dozen states disburse some state funding on how many students an institution graduates, rather than how many it enrolls. Students and families are more anxious than ever before about crossing the degree finish line, as the financial burden of paying for college has increased significantly in recent years. And retaining students is becoming more crucial to the university bottom line. As recruiting and educating students becomes increasingly expensive, colleges hope to balance the resources they use to recruit students with revenue generated when those students are retained.

Because of these pressures, institutions have begun analyzing demographic and performance data to predict whether a student will enroll at an institution, stay on track in her courses, or require support so that she does not fall behind. Using data in this way is known as predictive analytics. Analyzing past student data to predict what current and prospective students might do has helped institutions meet their annual enrollment and revenue goals with more targeted recruiting and more strategic use of institutional aid. Predictive analytics has also allowed colleges to better tailor their advising services and personalize learning in order to improve student outcomes.

But while these are worthwhile efforts, it is crucial for institutions to use predictive analytics ethically. Without ethical practices, student data could be used to curtail academic success rather than help ensure it. For example, without a clear plan in place, an institution could use predictive analytics to justify using fewer resources to recruit low-income students because their chances of enrolling are less sure than for more affluent prospective students.

Our framework here aims to lay out some important questions to consider as administrators formulate how to use predictive analytics ethically. Examining the ethical use of data is an iterative process; colleges will continue to use student and institutional data in new and innovative ways and will therefore have to occasionally reassess whether their ethical standards address current data practices.

Using data ethically is complex, and no magic formula exists. This ethical framework is meant to start conversations on campus. It cannot address all possible issues surrounding the use—and potential abuse—of institutional data.

Predictive Analytics In Higher Education: Five Guiding Practices For Ethical Use



- Convene key staff to make important decisions.
- Consider the purposes, unintended consequences, and outcomes to measure when developing the plan.

- Communicate the benefits of using predictive analytics and create a climate where it can be embraced.
- Develop robust change management processes.
- Assess institutional capacity.

- Ensure data are complete and of high enough quality to answer targeted questions.
- Ensure data are accurately interpreted.
- Guarantee data privacy.
- Monitor data security.

- Design predictive models and algorithms so that they produce desirable outcomes.
- Test and be transparent about predictive models.
- Choose vendors wisely.

- Embed predictive-driven interventions into other student success efforts.
- Communicate to staff and students about the change in intervention practices.
- Recognize that predictive-driven interventions can do harm if not used with care.
- Carefully communicate when deploying interventions.
- Train staff on implicit bias and the limits of data.
- Train students to use their own data.
- Evaluate and test interventions.

#DataEthics

For more information, visit newamerica.org/dataethics.



Guiding Practice 1: Have a Vision and Plan

Developing a vision and plan for data use will help steer the direction of a predictive analytics effort. Without such planning, predictive analytics may be used in a way that does more harm than good for students, leaves out key staff who should be included in the planning process, and/or fails to identify how success of this effort will be measured.

To develop a vision and plan, take the following steps:

Convene key staff to make important decisions.

In developing a plan, include key staff and stakeholders in decision making, and get their support. Including these individuals in the planning process can help ensure that you are using predictive analytics in a way that does not intentionally harm those whose data are being used and analyzed.

Consider the following three factors when developing the plan.

1. The purposes of predictive analytics

The plan should include the questions you hope to answer and the goals you aim to achieve. It should also explore the potential pitfalls of using student and institutional data for the purposes intended. The team should make sure that data will not be used for discriminatory purposes.

2. The unintended consequences of predictive analytics

The plan should also include a discussion about any possible unintended consequences and steps your institution and its partners (such as third-party vendors) can take to mitigate them.

3. The outcomes to measure

The plan should also lay out the measurable outcomes you hope to achieve as a result of using predictive analytics.

Guiding Practice 2: Build a Supportive Infrastructure

A supportive infrastructure ensures the benefits of predictive analytics are understood and welcomed by campus stakeholders, and that processes and other supports are put in place to assist the data effort.

Communicate the benefits of using predictive analytics and create a climate where it can be embraced.

Predictive analytics uses student and institutional data to create change almost immediately. Many institutions may not be experienced with using data in this way, at this pace, and perhaps with such high stakes like ensuring students complete their degree in a timely manner. You should take the lead in communicating with campus leaders, staff, and students about why using predictive analytics is critical to institutional and student success. The head of communications and marketing could help in these efforts. Without a clear articulation of how using predictive analytics can benefit the campus, well-devised plans may fail to receive the support they need to be successful.

Develop robust change management processes.

With new tools often come new processes, reporting structures, people, and partners who bring new skills. This can, at best, create confusion for those charged with rolling out predictive analytics on a campus, and, at worst, chaos. Leaders convened to make important decisions about data use could also help ensure that processes are put into place to support the change taking place on campus.

Assess institutional capacity.

Assess your school's capacity to use predictive analytics. Having the appropriate technology, data infrastructure, talent, services, financial resources, and data analysis skills are essential. Maintaining a sound infrastructure can help ensure that cleaning, sharing, and using large amounts of data for making decisions institution-wide can be carried out smoothly and that different data systems can "speak" to one another. Experts in information technology, student data laws, and staff with experience drafting contracts with vendors would help ensure the success of the project.

Guiding Practice 3: Work to Ensure Proper Use of Data

Predictive models (showing how different data points are related) and algorithms need data to build predictive tools that will support enrollment efforts or help students make academic progress. To build and use these tools ethically, consider the quality of your data and data interpretation, as well as issues around privacy and security.

Ensure data are complete and of high enough quality to answer targeted questions.

Data about students and institutional processes should not only be accurate but also comprehensive. Comprehensiveness also means considering all relevant data about the students who are being examined.

Beyond being accurate and comprehensive, quality data is also timely, derived using consistent tools and processes, and is well defined.

Ensure data are accurately interpreted.

Include staff members who are knowledgeable about your institution's data and can accurately interpret predictive models derived from this information. It is essential that those analyzing the data take context into consideration. It is also important to train faculty so that they can easily interpret dashboards that explain how students using adaptive tools are faring in their courses. Lastly, look for ways to ensure that data used solely for reporting purposes is sound even though they may also be included in data sets that are used for predictive analytics. If institutional researchers are responsible for both compiling data sets for reporting purposes as well as for conducting analysis for predictive analytics projects, the integrity of data for reporting should not come into question because information is being used on campus in innovative ways. Put simply, predictive analytics should not diminish the quality of data your institution is required to report to remain in compliance for federal funding.

Guarantee data privacy.

Communicate with students, staff, and others whose data are collected about their rights, including the methods used to obtain consent to use the data for predictive analytics and how long the information will be stored. Make students and staff aware that their data are going to be used for predictive analytics and get consent to use highly sensitive information like health records.

Be vigilant that data are well protected so that the information does not get into the hands of those who intend to misuse it. It is especially important to protect the data privacy of vulnerable student groups, such as high school students who are minors and enrolled in dual-enrollment programs, undocumented students, and students with disabilities.

In addition, make school policies on ownership of and access to student and institutional data clear.

Monitor data security.

Security threats occur without notice. As colleges collect and store more data on students and staff, and more devices that store data on the teaching and learning process are used in classrooms, security becomes an ever more pressing issue. For this reason, schools need to be vigilant about assessing data privacy and security. Monitoring threats and risks should be a regular undertaking. Data security requires you and your vendors to have security protocols that adhere to student privacy laws and meet industry best practices.

To keep institutional data secure, involve your information technology (IT) department. Information security and privacy officers help keep institutional data safe. Providing regular training to IT and other staff about keeping these data secure should be a top priority.

Guiding Practice 4: Design Predictive Analytics Models and Algorithms that Avoid Bias

Predictive models and algorithms can help determine the interventions an institution uses to support students or meet recruiting goals. Therefore, it is crucial that predictive models and algorithms are, at the very least, created to reduce rather than amplify bias and are tested for their accuracy. You should also ensure that models and algorithms are created in consort with vendors who can commit to designing them in a way that does not

intentionally codify bias and so that they are able to be tested for veracity.

Design predictive models and algorithms so that they produce desirable outcomes.

It is crucial to address bias in predictive models, ensure the statistical significance of predictions beyond race, ethnicity, and socioeconomic status, and forbid the use of algorithms that produce discriminatory results. An algorithm should never be designed to pigeonhole any one group.

Therefore, design or know how predictive models and algorithms are created in order to ensure desirable outcomes as determined by their vision and plan. Failing to take this approach may lead to inadvertent discrimination.

Test and be transparent about predictive models.

Before predictive models can be used to develop algorithms, test them for accuracy, perhaps by an external evaluator. Predictive models should also be updated or refreshed to reflect new campus realities and goals. You may also want to limit the variables used in predictive models to those that can be easily explained and you should work to ensure algorithms can be understood by those who will be impacted by them. Such practices foster transparency and makes it easier to hold individuals accountable for creating poorly designed models or algorithms that produce discriminatory outcomes.

Choose vendors wisely.

Most colleges rely on an outside vendor to help them build models and predictive tools. To ensure models and algorithms are sound, transparent, and free from bias, you must be intimately involved with or knowledgeable about how predictive models and algorithms are built. Partnering with third-party vendors may make this harder.

Some vendors are transparent about their models and algorithms, and allow colleges to have a hands-on approach in the design process, or even let institutions take the lead. Not all vendors, however, take this approach. Many consider their models and algorithms proprietary, meaning institutions are not involved in the design process or are deliberately kept out. You should make transparency a key criterion when choosing to work with any vendor.

Guiding Practice 5: Meet Institutional Goals and Improve Student Outcomes by Intervening with Care

How your institution acts as a result of what it learns from predictive analytics is where the rubber meets the road. Students will experience these actions or interventions firsthand, even if they do not see or understand how the algorithmic-based decisions are made. Despite the use of technology, humans primarily still have to deliver interventions. Therefore, it is important that interventions are thought about in the context of other supports offered at your institution and are disseminated with carefully communicated messages. Staff deploying interventions should be trained on how to intervene appropriately, and you should test the effectiveness of interventions once deployed.

Communicate to staff and students about the change in intervention practices.

Adding predictive analytics to the student success toolbox may spark a culture change as interventions informed by data become central to your institution. To get the campus to embrace this change, it is important to communicate how faculty, staff, and students will benefit from using interventions that are informed by predictive analytics, and allow them to guide the change as well.

Embed predictive-driven interventions into other student success efforts.

Despite being a powerful tool, predictive analytics is still only one part of a suite of tools—like first-year orientation

programs—that can ensure student and institutional success. Look for opportunities to leverage predictive analytics in ways that further advance other activities so that all student success efforts are connected and build upon one another.

Recognize that predictive-driven interventions can do harm if not used with care.

Even when institutional data, predictive models, algorithms, institutional practices, and training are as good as they can be, mistakes can be made when acting on information. This is why interventions used in response to predictive analytics should be carefully calibrated to avoid harming students. Interventions used in response to data generated by students and predictive models can range from targeting a particular student for increased outreach based on his predicted chances of enrolling, requiring a meeting with an adviser based on the recommendation of an early-alert system, to changing the type of practice problem a student is assigned based on an adaptive technology system.

However, these tools should not be used without examining their potentially negative effects. Algorithms used for strategic enrollment management, early-alerts, recommender systems, and adaptive technologies require that colleges understand where they can do more harm than good. Viewing students from a wellness or asset mindset rather than an illness or deficit mindset may help ensure students are not harmed. This approach values all students as full of potential. In addition, it leaves room to consider institution-specific characteristics or barriers that have an impact on a student's risk of dropping out. Finally, it will be wise to determine how individuals will be sanctioned for misusing or mishandling student and institutional data, as well as how to rebuild trust after a harmful incident has occurred.

Predictive tools can be used with care in the following ways:

- Early-alert systems
- Recommender systems
- Adaptive technologies
- Enrollment management

Carefully communicate when deploying interventions.

The messages you send should not demoralize students, and dissemination strategies should ensure that students are able to access interventions with relative ease. Craft messages in the right way, and ensure interventions are accessible to target populations.

Train staff on implicit bias and the limits of data.

Staff should be trained on how implicit bias and the limitations of data can impact how they intervene with targeted students. Personal biases and an over-reliance on institutional data can negatively affect the students they hope to serve. With the proper training, staff should eagerly embrace their obligation to use student and institutional data to produce positive results for students. Combat implicit bias, and understand data's limits.

Train students to use their own data.

Staff may also wish to train students to use their own data to guide their experiences on campus. For example, students can use data they generate in adaptive learning tools to understand the conditions under which they learn best.

Evaluate and test interventions.

Do not declare an intervention successful until it has been tested and evaluated for its effectiveness.

What interventions work when, for whom, and why?

Test the efficacy of interventions you are using. Such testing can uncover whether these interventions have differential impacts across different groups and allow recalibration as necessary. Efficacy research could also help reveal whether the interventions are having any unintended consequences.

Test tools based on vendor claims before committing to them long-term.

Insist that vendors partner with independent researchers to validate the effectiveness of their tools and services. Whether tools are effective is a particular concern for adaptive technologies. Many third-party vendors claim their products are adaptive and will accelerate student achievement despite having little external validations for these claims. As the field of technology-enabled student learning continues to develop, new tools and validation of claims should go hand in hand.