

Picking a Predictive Analytics Partner

By Iris Palmer

Campuses today are overflowing with data—data that, if used appropriately, can help colleges and universities operate more efficiently and effectively. With the sheer volume of data available, it is sometimes difficult for an institution to know where to start its analytics efforts. Many institutions seek a partner to help them with their analytics needs, and these partnerships can bring benefits to both parties.

Yet, choosing a vendor can itself be one of the more challenging aspects of implementing predictive analytics at your institution. Among other risks, a bad fit between the campus and its vendor may decrease the likelihood of the system boosting student success. Failure to ask the right questions can also result in ethical lapses in how data are used and particular groups are affected.

Business officers are in a powerful position to ask questions of potential vendors about model transparency, data ownership, and protections for student data. Campus leaders should consider asking the following questions of prospective predictive analytics vendors:

■ **What data go into your prediction?** Ask the vendor to provide a list of all data elements the company uses in its predictive models and algorithms. With this information, institutional leaders can determine the sensitivity of the data elements and document where those data are stored. Documenting the data elements provides an important record of what vendor predictions are based on. Vendors should commit to providing the institution with an updated list of variables whenever they make changes.

■ **What kind of data make up your training data?** One way bias creeps into algorithms is through nonrepresentative training data. Some vendors use data from multiple institutions to make their models more robust. Ask vendors to document the diversity of the institutions and student data that helped train their models. College and university leaders should also ask how vendors plan to customize the algorithm for their institutions. Many vendors create their model using only the institution's data. In that case, business officers should ask how far back the training data go, since it's important to balance having the most recent data possible to reflect the current population with ensuring that there is a valid sample size to train the algorithm.

■ **Can you test your algorithms on my institution's data?** The robustness of algorithms matters, but so does campus context.

Before signing a long-term contract with a vendor, campus leaders should ask if the vendor is willing to conduct a pilot test. Some vendors will agree to test their algorithms on your institution's data. This can help indicate if the system is discriminatory—overidentifying certain groups of students as “at risk” on campus or identifying all students as “at risk”—which would be of limited use to the institution. If a vendor doesn't agree to a pilot test, it should show how its models performed at institutions similar to yours and agree to a disparate impact analysis of its tool output, after getting the contract, to ensure that the algorithm is correctly identifying at-risk students.

■ **Who owns the cleaned source data, repurposed data, and byproduct data?** In some cases, the vendor will not agree to return the cleaned source data to the institution, instead destroying it at the end of the contract. The contract should clearly lay out who owns what data and under what terms. The contract should also prohibit the vendor, or any partner of the vendor, from selling student data.

■ **What are your physical, administrative, and technical data safeguards?** Vendors represent a serious potential for data breaches if they do not have the correct protections in place. The massive breaches of customer data at Target Corp. and Home Depot Inc. in recent years happened because of flaws in vendor systems. While most colleges and universities have IT staff who know what questions to ask vendors, given the liability at stake, it is a good idea for business officers to check this box as well. Ask vendors how the company controls access to the institution's data; if the data are encrypted using an industry-standard encryption both when stored at the vendor and when transferred between the institution and the vendor; and if the data will be stored in a physically secure location.

As demands for big data and artificial intelligence systems continue to increase on campus, business officers play an important leadership role in getting these collaborations right. For additional ethical and practical considerations for collaborating with a predictive analytics vendor, read *Choosing a Predictive Analytics Vendor: A Guide for Colleges* by searching the title at www.newamerica.org.

IRIS PALMER is senior policy analyst, education policy program, New America, Washington, D.C.
palmer@newamerica.org

Given the nature of analytics and the consequences of the black box technologies that deliver these data—not only to inform institutional priorities, but also to shape student outcomes—it is necessary for higher education administrators to approach the use of these data thoughtfully. Unfortunately, there are few existing, and no universally accepted, guidelines or policies for practice. Current policies or guidelines are often limited to explanations of federal privacy policies or articulation agreements centered on data security and access.

Data security and access are certainly central to using data ethically. While organizations are legally and ethically bound to secure data and establish rules for access, related policies and practices are often unclear. Similarly, and equally ill-defined, data ownership and consent are important and complex components of ethics and privacy. Ownership of data collected via analytics technologies often lies with data users (colleges and universities) and not with data producers (students). However, ideas about data ownership are evolving, with calls for students to become the owners or co-owners (alongside their universities or technology vendors) of the data they produce. If individual student users are potential data owners with the ability to consent to participate (or not) in producing and using data, then clearer data ownership and consent policies and terms of use agreements are needed to ensure ethical use.

In an effort to provide better guidance and address ethical issues, scholars have developed a number of codes of practice that focus on promoting data transparency, security, ownership, control, stewardship, and trust. Although these codes of practice provide useful guidance, they are still evolving and fall short of comprehensively addressing the contexts and needs of organizations and students from an ethical perspective.

To address the bias and potential for discrimination in data analytics, Linnet Taylor—a data analytics researcher at Tilburg University in the Netherlands—has

Data ownership and consent are important and complex components of ethics and privacy.

developed a data justice framework based on three pillars:

- 1. Visibility** includes access to information, representation, and privacy, and focuses on how individuals are represented, profiled, and monitored through analytics systems.
- 2. Engagement with technology** includes autonomy in making technology-related choices (including the choice *not* to use or be used by technologies) and sharing the benefits of data collection and use.
- 3. Nondiscrimination** includes the ability to challenge bias within, and to be free from discrimination in, big data algorithms.

While this framework for data justice provides a useful foundation for establishing ethical use of data analytics, it does not speak to the unique nature of higher education. As noted, higher education institutions have a legal duty of care. International professors and lecturers Paul Prinsloo and Sharon Slade have been pioneering research on the ethical use of analytics in higher education. They argue that care should be extended to the collection, storage, and use of student data. They also contend that care is required because current policies and data justice initiatives alone do not take into account the specific contexts within higher education or the complexities of individual students. Care-based use of analytics should incorporate codes of practice and pillars of justice from a student-centered perspective.

Guidelines for Ethical Use

Data justice, care, and associated codes of practice combine to form a touchstone for grounding policies, processes, and practices of ethical analytics-informed decision making. From enrollment planning, to benchmarking, to strategic initiatives, to student support, codes of practice that incorporate the principles of data justice and care can be used as a framework for centering students in data analytics processes, unearthing the contextual complexities that influence data-informed decision making, and providing a useful starting point for conversations about ways to improve data-informed processes.

We recommend the following guidelines for approaching data analytics from a more ethical and equitable perspective:

Consider context. Data are invaluable for facilitating efficient and effective solutions to many challenges campuses face. However, campuses are as varied as their students. Each campus has its own mission and priorities, and students bring their own myriad experiences and goals. These contexts matter—and data are only as valuable as the meanings campus leaders derive from them.

Campus audits that are focused on understanding an institution's readiness, capacity, and unique culture are helpful ways to better align analytics with an institution's needs and priorities. Analytics and institutional alignment are especially important when working with predictive or prescriptive data, as algorithms developed in one context will not easily translate to the unique context and needs of another environment. To help contextualize data, higher education should approach the purchase, use, and implementation of analytics tools with the mission and priorities of the institution and the specific needs, goals, and experiences of students top of mind.

Collaborate broadly. Improved collaboration across higher education divisions and departments can help focus attention on student needs and interests. At the