CIVITAS® LEARNING

Key Terms

DATA SCIENCE GLOSSARY

Curated by

Civitas Learning

SEPTEMBER**2015**

PREDICTIVE ANALYTICS PRIMER

COLLEGES AND UNIVERSITIES ARE AWASH IN DATA.
TECHNOLOGY PRESENTS BIG DATA SOLUTIONS, PROMISING INSIGHTS AND AN OPPORTUNITY TO IMPROVE OUTCOMES.

What differentiates predictive analytics from descriptive analytics?

What do institutional leaders need to know about machine learning? What data sources matter and how are they used? Can time series feature extraction indicate when students are cramming and how they respond to adversities?

Predictive analytics uses data to forecast—at the individual student level—future success based on key indicators, allowing institutions to assess the likelihood of individual student success. Institutions can use predictive analytics to deliver the right type of support to the right students at the right time in the right modality.

Not all analytics are predictive. Predictive analytics is not about cohort level data: it involves developing specific features, preprocessing algorithms, and learning algorithms based on the unique data footprint for each institution. Our data science glossary is designed to help institutional leaders understand the lexicon of learning analytics and data science. Use the glossary to explore the concepts behind data science and learn more about how analytics can be used to improve institutional effectiveness.

A GLOSSARY OF KEY TERMS

DATA SCIENCE EXPLAINED

CANONICALIZATION

Process for translating raw data into a consistent and homogenous representation for analysis.

WHY IT MATTERS
Canonicalization, sometimes called standardization, allows data scientists to apply algorithms to massive quantities of learning data, developing insights into institutional patterns on a larger scale and recommending specific actions. This process creates a common data language for data scientists to compare and evaluate seemingly unrelated data features across institutions. For example, degree alignment scores measure whether students are taking the right courses to graduate on time, or wandering around taking courses unrelated to their majors, without regard to their degree requirements.

TIME-SERIES FEATURE EXTRACTION

Turns complex events into regularly sampled time-series data from which we can extract meaningful features.

WHY IT MATTERS
Time-series feature extraction allows institutions to identify complex trends and changes over time and collect analytic insights into how interventions and outreach by faculty and advisors are influencing these time-series features. For example, using time series feature extraction from the LMS, an institution could identify times when students are cramming or studying consistently, or identify changes in student sentiment over time.

DESCRIPTIVE ANALYTICS

Examines historical data and identifies trends or patterns over time from known facts to inform future decisions.

WHY IT MATTERS

Descriptive analytics allows institutions to understand trends such as enrollment, retention, and course selection, and use quantitative data analysis to understand the underlying factors that influence those outcomes.

PRESCRIPTIVE ANALYTICS

Examines the relationship between descriptive analytics and predictive analytics to determine the best way to achieve a desired outcome.

WHY IT MATTERS

Prescriptive analytics inform the decision-making process, allowing institutions to weigh the impacts and effects of certain decisions that will lead to desired outcomes.

REGRESSION

Predicts the outcomes of a continuous variable, such as the time it takes to master a topic over the semester in competency-based learning, or salary achieved after graduation.

WHY IT MATTERS

Regression allows for the institutions to have a greater depth and understanding of success factors that have continuous, not discrete, values. On a related note, classification refers to predicting a discrete number of student success outcomes, such as persisting or not persisting.

FEATURE RANKING AND OPTIMIZATION

Uses optimization algorithms to analyze features, and find the best combination of features to maximize prediction accuracy without over-fitting data.

WHY IT MATTERS

Feature ranking and optimization ensures that we have the most relevant and powerful features in predicting student success. **Through feature curation, we also ensure that these top features are insightful, and – where possible – actionable.**

LEARNING ALGORITHMS

Sophisticated mathematical equations that researchers use to predict future student success outcomes using historical data, and that continuously evolve according to accumulating data over time.

WHY IT MATTERS

Just as any student changes and adapts over time based on the environment, learning algorithms also change and adapt to make predictions about the future. They process ongoing, real-time inputs to provide accurate, timely predictions on the future success of students. Learning algorithms are retrained as new data becomes available. Furthermore, as part of continuous training, we surface each student's key success factors through hierarchical feature ranking and sensitivity analysis to ensure that advisors and faculty have the right talking points to drive student success. This ensures that predictions are up-to-date, specific to each institution, responsive to changing characteristics in student data, and personalized to each student.

PREDICTION-BASED PROPENSITY SCORE MATCHING

An algorithm that matches a pilot student with a non-pilot student using both prediction and propensity scores, thus creating an artificial control group for apples-to-apples comparison on intervention impact.

WHY IT MATTERS

Propensity score matching ensures that we can systematically identify control students matched to pilot students in a statistically robust manner. Furthermore, by using top predictors for each student segment, we maximize statistical power in intervention impact analysis, meaning that we can detect lower signal-to-noise ratio (SNR) impact size than what is possible with conventional propensity score matching.

COMPLEX EVENT PROCESSING

A process for tracking events, inferring patterns by linking them, and responding to them in an appropriate manner.

WHY IT MATTERS

Complex event processing is the underpinning of designing real-time student engagement triggers by leveraging multi-dimensional real-world events around students that influence student success outcomes significantly. For example, if a student exhibits withdrawal patterns after an adverse event, this could be an opportune moment for mindset coaching.

A GLOSSARY OF KEY TERMS

DATA SCIENCE APPLIED

STUDENT ENGAGEMENT PREDICTION

Uses influenceable activity and behavioral factors to predict student engagement, which is highly associated with student success.

WHY IT MATTERS

By knowing what influenceable factors are directly correlated to student success and when to reach out for receptivity and engagement, institutions can identify which students are at risk, know when to intervene proactively, and personalize interventions based on students' level of engagement and academic performance.

INTERVENTION AND INSPIRATION SCIENCE

Examines how to engage and motivate students to succeed, and measures the impact of current interventions.

WHY IT MATTERS

Intervention and inspiration science gives real-time feedback for institutions to determine the best way to re-engage each student, and shapes the likely impact of intervention efforts.

IMPACT PREDICTION

Creates a forecast of which interventions will be most impactful for what students in what situations.

WHY IT MATTERS

Impact prediction is the understanding of what micro-interventions and nudges are the most impactful for each student, based on their needs and institutional resources available. These predictions help institutions match the right student with the right support at the right time, enabling them to use resources effectively and provide support where and when it is needed most.

CIVITAS LEARNING

civitaslearning.com | partnersuccess@civitaslearning.com | civitaslearningspace.com