Forecast

The Purdue Office of Institutional Research, Assessment, and Effectiveness (OIRAE) and Teaching and Learning Technologies (TLT) have partnered to positively impact student success through a robust data platform and web-based application called Forecast. The Forecast application provides nudges or suggestions towards positive behaviors which correlate with student success. The nudges are presented in a graphical view which supports the relationship of the positive behavior to success metrics of on-time graduation and higher GPA.

Overview of the Application

Purdue's Forecast was designed as a student-focused platform providing transparent information and data for West Lafayette undergraduate students on their individual progress and success. The data and information provided compare individual progress during the active semester to historical data from prior successful students as well as the current student body. Forecast has a very specific and singular focus of communicating academic progress in an actionable language with data visualizations of personalized feedback to undergraduate students at Purdue.

The Forecast application is currently comprised of two components: Campus Forecast and Course Forecast. Campus Forecast focuses on correlations between student behaviors and student success, independent of the specific course the student is taking. Course Forecast, on the other hand, looks at course-specific behaviors that correlate with success in that specific course, and provides an overall risk score for each student for each course in which they are enrolled. The data model development and the transformation of model outputs into the nudges were approached slightly differently between Campus Forecast and Course Forecast.

Application Design

A student-centered approach was taken to the design of Forecast. Early feedback from student focus groups helped define the features and functionality included in the app. An advisory group consisting of faculty, advisors, staff, and student leaders was created to provide input and feedback on the application as it was built. The work of the advisory committee enabled Forecast to be inclusive of the perspectives, excitement, and concerns of the Purdue community.

An effort was made to keep the most relevant information prominent at the time it is needed by the student. To this end, the behavior students see first when visiting the site is determined by what behavior is most important at that time in the semester. This contributed to the decision to focus on one key behavior per page. The site was also designed to provide value to students who choose not to log in to see their personalized information. General data and information about each behavior is displayed regardless of the authentication status. Logging in allows students to see their personal data and how they compare to their peers.

Forecast's navigation follows the common user interface design pattern of a sliding drawer menu, which provides some level of consistency and familiarity to most users. This menu style was chosen because it remains hidden when not needed, thereby lessening distraction to the user and allowing them to better focus on the site content. Icons were included with each navigation item and remain visible when the menu is closed to enable quick navigation between behaviors and courses. The site has a responsive design which allows it to be usable from a variety of device sizes, from phones all the way up to large desktops.

All pages in Forecast share a consistent layout. The top portion of each page is devoted to the information provided by the model, which illustrates why that item was identified as a key behavior to student success. The remainder of each view is devoted to the clear presentation of actionable steps that a student can take to help them achieve success in that behavior.

Forecast's accent colors were limited to a subdued palette of blues and greens for site navigation. Blue and greens were selected based on the common association of those colors with security, responsibility, and calmness. A bright contrasting purple was used to help call-to-action buttons stand out throughout the site. The site background remains largely white and off-white to allow other colors to be used to convey information in graphs and to minimize distraction.

In an effort to make information easily understood by a wide audience and at a glance, simple bar, bubble, and line graphs were used to display data.

The Data Modeling Effort

In early 2015, OIRAE envisioned a desired state: 1) Expand institutional knowledge of the existing relationship between student success and student characteristics and behaviors, and 2) Examine potential new data sources that might also have a relationship with student success. It was hoped that the modeling relationships would enable a student-facing application that would nudge students toward behaviors that might enhance their academic success.

In June 2015 Purdue consulted with Pivotal Software Inc., a consulting group associated with Dell-EMC big data analytics, to build an analytics environment and an array of initial predictive models for Purdue to use in studies of undergraduate student academic success. The team worked to identify features which had significant relationships with one of three target dependent variables: course grades, semester GPA, and on-time graduation. Three initial models were built with the above-mentioned dependent variables using three approaches: 1) Logistic regression, 2) Random forest, and 3) XGBoost (Extreme Gradient Boosted Decision Trees). XGBoost outperformed the other approaches in terms of prediction power. In addition, it was found to be easy to use both for development and production because it is robust to multicollinearity and it handles null values without a need to transform them (Chen & Guestrin, 2016).

Campus Forecast

In the Campus Forecast component, four variables were identified to have a large relative effect size with a related behavior that is malleable by a student or which could bring benefit to the student simply by raising awareness. These variables were:

1. On-time Registration - average days the student waits to register after the date when they are eligible to register for courses.

   Average course GPA

   Students that register early are more likely to succeed

   Average days waiting

   What should I do right now?
In the Course Forecast component, four variables were identified that focused on student behaviors in an individual course that differentiate among course grade outcomes.

1. Attendance in class - This module displays the relationship between a student’s course attendance and their course grades. Historical attendance figures were calculated for each student-class pair over two academic semesters. To calculate attendance, wireless network logs were inspected to determine if a student’s wireless device(s) had network activity in the building where their class was scheduled during the time of the class.

2. Network usage while in class - This module displays the relationship between the relative network usage for students in the class (compared to the rest of the class) and their course grades. For each course for each week in the semester, the mean bytes transferred by all students who attended a class session was calculated. A z-score for each student for each week was then calculated. A student’s semester usage was the median z-score for all weeks in a term.

3. Blackboard first view of assignments - This module shows the relationship between the course grade and elapsed days between posting of course material and the first view. It also displays the relative time to view the course materials, which is how quickly a student views the materials in comparison to other students.

4. Blackboard discussion forum participation - This module displays the relationship between the number of discussion board posts a student makes and their course grade. As with the previous Blackboard module, it displays both the historical relationship to the student, the amount of posting activity they perform, and how that compares to their classmates’ activity.

Communication to Students
The primary communication tool was direct email, but other vehicles were explored (poster sessions, information booth, etc.). At predetermined times in each semester, emails featuring different Forecast modules were sent to all Purdue West Lafayette undergraduates. A total of 6-8 emails were sent in both the fall 2016 and spring 2017 semesters. The timing and theme of each message was determined by relevance to when a successful behavior was believed to be most important. For example, the “Register on Time” message was sent at the beginning of the open course registration window.

During the spring 2017 semester, the email communications team implemented an A/B style test, splitting the undergraduate recipients into two (2) separate groups receiving different emails with the hope of measuring and identifying the more successful campaign. Students in the A group would receive the traditional, less visually appealing, and more formal email. The B group would in turn receive more contemporarily-styled emails that would use persuasive language, shorter messages, and flashy imagery such as GIFs or funny photos, to grab attention. A statistical test of the click-to-open rates of 4.55% and 4.39% showed no difference in student response to the differently-styled emails. The “compare your GPA” email had the biggest (by far) open/click ratio. The experiments also expanded to test time of day, subject line, sender, formal/informal voice, and graphics/no graphics. The goal was to incorporate the module message with a style of communication that best fit the demographic (18-22-year-old college students).

A process was created to assist those students who have concerns about the email campaign, the personalized recommendations, or data privacy. For those who wish to remove themselves from future email campaigns about Forecast, the tool has a built-in function to allow individuals to “unsubscribe” from the mailing list. Students wanting more information about their personalized recommendations and visualizations may email tlt@purdue.edu. Those wishing to opt-out of their data being collected were engaged in in-depth conversation so that the student understood the specifics on data collection. While the University collects all types of student data to operate Purdue, it is bound by federal law not to release an individual student’s data to anyone but that student. The data used in and generated by Forecast is likewise restricted.

External Review of Forecast
Four external reviewers visited Purdue in June 2017 to review Forecast and its methods. These comprised of two academicians with strong backgrounds in data analytics and student behavior, one who works in industry but also serves as an Associate Editor of Psychometrika, and one who worked for the Educational Advisory Board in marketing tools to students. Prior to arriving on campus, the reviewers received an extensive summary of Forecast and its methods, which was prepared jointly by OIRAE and TLT. They then spent 1.5 days on campus meeting stakeholders and asking questions about all aspects of Forecast.

Immediately following the review, the evaluators provided informal feedback to the Forecast leadership team. They also submitted a formal written report roughly one month after the end of the review. Their comments were similar across both reviews. They felt that the modeling work was appropriate, and that the scope of the overall effort was extremely impressive. They felt that the effectiveness of the nudges would improve if the model were segmented to look at sub-populations within the Purdue student body, rather than looking at the overall student body. Such groups might include first generation students, at-risk students, in the murky middle, or groups defined by race or gender. They also recommended that the Purdue team pursue explanatory modeling to help make the nudges more robust. Finally, they recommended more engagement with students to gauge their response to Forecast and to make it more responsive to their communication preferences.