Racing for Success

What every business student needs to know about analytics

- Case-studies course for real-world corporate applications
- Continuing education: essential skills for analytics professionals
- Engineering, business schools team up to benefit students
- Contest builds industry experience without leaving classroom
Industry-academia partnerships

By Galit Shmueli, Rajib Saha and Reema Gupta

Business analytics courses are an excellent ground for industry-academia partnerships, and when executed right can offer a combined experience of classroom and industry without requiring students to leave the school premise. A few teaching components such as in-class guest lectures by industry professionals, student internships or site visits in industry do offer some exposure to the real challenges faced by real companies, but they are not always commensurately blended with the classroom experiences.

With the right technology and meaningful industry partnership, business analytics courses can bring the industry experience into the classroom. Furthermore, a business analytics project conducted by student teams (with faculty guidance) can create two-way value for both the industry partners who share the data, and students who embark on such learning experience. Student projects that are based on real data and real problems can offer not only solutions to existing problems, but also identify new uses and opportunities that companies have not considered.
The learning experience attains another dimension when such projects are conducted on a global competition platform, where students not only get an opportunity to evaluate themselves against data analytics professionals around the globe, but also draw motivation and learn more in the process of competing with more experienced experts.

The Indian School of Business initiated data mining contests as part of their Business Analytics for Data Mining (BADM) course in winter 2012. The contests were designed as components within the larger framework of a real business analytics project in industry. In the spirit of mutual learning and sharing, and to give students a taste of real world applications in the regional context, ISB Professors Galit Shmueli and Rajib Saha partnered with several companies that provided data and a business goal.

Teams of students worked with the data provided towards the stated business goals. The first step, of identifying an analytics opportunity or challenge, was set up through an “ideation contest,” where students were presented with data and the company goals. The second contest component was designed as a data-mining contest, where teams were presented a particular dataset and analytic goal. The contest submissions and learning were used by the student teams to develop the end-to-end business analytics project, which offered a business problem, converted into an analytic problem, and then used data mining applied to real data to come up with solutions and recommendations.

Another defining feature of this course was to host the contest on an open platform such as Kaggle.com or CrowdANALYTIX.com where not only ISB students but also other professionals and data scientists could participate in the competition. This aspect of opening the competition to professionals took the learning to a new level. The contests offered not only a high-energy competitive aspect that the students enjoyed, but also offered peer learning so that teams could see other approaches and solutions. Students engaged through online forums with various stakeholders to discuss open issues and share industry examples of use cases.

To give a flavor of the contests, one contest was based on a dataset of electronic gadget information and pricing on Indian online shopping sites, provided by the startup www.theBargain.in. The dataset was collected over a period of two months, recording price changes and product information (category, brand, color, avg. rating, number of reviews, list price, etc.). Ideation submissions proposed different uses for such data and types of business opportunities that they can create. For example, one team proposed a data-based gambling platform to bet on price fluctuations. Another team proposed a price insurance scheme against price drops. The goal of the data-mining contest was to accurately predict the probability of a price increase of an electronics product on an online shopping site.

The students obtained real-world experience taking data from the field and applying their classroom learning one lecture at a time – from defining a data analytic problem from a business problem, to cleaning and exploring the data, to conducting appropriate analysis (in this case, using data mining), to evaluating performance and assessing the actual impact on the business.

In the online electronic gadget pricing data, students developed an understanding of the raw data by creating visualizations using state of the art interactive visualization software. This step helped them identify a subset of potentially informative predictors of price change as well as recognize and address the real world issue of missing values. Once students understood the basic data, they moved to the second step where they spent considerable time in: pre-processing the data. In this step they used their domain knowledge and analytic methods, hand-in-hand. The process included creating derived variables and dimension reduction (e.g., by choosing the top 12 of 30 brands and top six of 20 color choices). The last step in the data pre-processing exercise was partitioning the given data into a training set that was used to build the model and validation set(s) kept
aside to test the predictive performance of the model.

The final step in this project, which was achieved through the data-mining contest, was to build predictive models and analyze the results. Students used a variety of data mining techniques (naïve Bayes, K-nearest neighbors, classification trees, logistic regression, ensembles, etc.). They compared the results of the different models and also compared against a simple naïve approach, which was presented on the contest website. The relative advantage of each of these methodologies came to the fore.

For example, one team found that KNN yielded better result than the naïve approach (71 percent vs. 41 percent accuracy in correctly predicting price increase), while multiple teams found logistic regression to achieve the highest predictive power of more than 90 percent. Teams learned about over-fitting and the ability to generalize to new data, based on the discrepancies that they saw between the performance on their own validation set compared to a secret validation set in the data mining contest that was not available to them, but where their performance was measured and benchmarked. Most importantly, they realized how domain knowledge can complement different statistical and data mining techniques, and how focusing on one of the aspects alone limits the performance and applicability of a model.

The best part of this contest was that students not only had a chance to practice the concepts and techniques of data mining using real world data, but they could also put on their creative thinking hats to think of different business use cases for predictive modeling. The industry partners were impressed with the breadth of ideas and the possible solutions. They also learned more about the relationship required in business analytics projects, and of course, about the abilities of our students.

For more information on the contests that were conducted, see liberation.isb.edu/mobile-usage-in-india and liberation.isb.edu/online-shopping. The project reports and presentations are available at galitshmueli.com/student-projects.