Communications of the Association for Information Systems

Volume 37 Article 28

9-2015

Teaching Predictive Model Management in MIS Classrooms: A Tutorial

Sule Balkan
National Chiao Tung University, sule@nctu.edu.tw

Haluk Demirkan
The Milgard School of Business, University of Washington

Communications of the Association for Information Systems

Research Paper ISSN: 1529-3181

Teaching Predictive Model Management in MIS Classrooms: A Tutorial

Sule Balkan

Associate Professor Institute of Business and Management, National Chiao Tung University, Taiwan

sule@nctu.edu.tw

Haluk Demirkan

Associate Professor of Service Innovation and Business Analytics

The Milgard School of Business, University of Washington – Tacoma, WA

Abstract:

Analytics has become a key element of the business decision process over the last decade. In today's competitive business world, organizations have found out that their data and how they use it can make them much more competitive. According to many research institutions (e.g., Gartner and McKinsey), the worldwide market for business analytics solutions in practice, research, and education is growing exponentially. As the use of analytics become widespread, business school graduates need to gain the necessary knowledge and skill sets to use these assets effectively. In the spirit of analytical thinking, we developed a practice-oriented business case that uses a sample scenario, managerial dashboards, betting templates, model repository and model performance management metrics that teaches predictive analytics concepts and decision making with incomplete information intended for MIS courses. Through exercises and interactions, students gain the skills, knowledge and experience necessary to be become effective decision makers through applying analytical thinking. Digital copies of workshop lesson plans with dashboard and data entry templates can be downloaded free of charge from the Teradata University Network.

Keywords: Business Intelligence, Business Analytics, Predictive Modeling.

This manuscript was solicited by the Editor-in-Chief. It was received 11/04/2015 and was with the authors 1 month for 1 revision.

1 Introduction

According to a recent Gartner Report, the business intelligence and analytics market is growing nine percent per year and will exceed $80 billion by end of 2014, with about fifty percent from predictive analytics by that time. The same report indicates that, even though 73 percent of companies intend to increase spending on analytics and making data discovery a more significant part of their business intelligence and analytic platforms, 60 percent feel they don't possess the skills to make the best use of their data (Moore, 2011; Murphy, 2013). There is a significant amount of skill gap in organizations regarding to analytics.

On the other side, a growing number of discussions in academic communities (e.g., the Association of Information Systems (AIS), the Institute for Operations Research and the Management Sciences (INFORMS), and the Decision Sciences Institute (DSI)) exist regarding how to develop and integrate new concepts, methods, and tools on analytics in their existing curriculum. Institutions offer various conferences and workshops, and various new special interest groups have been established (e.g., AIS Special Interest Group on Decision Support and Analytics (SIGDSA), Informs Analytics) to define road maps for research and teaching pedagogies on this new emerging topic of analytics—especially in predictive analytics. Furthermore, various software vendor companies offer their solutions for education purposes (e.g., IBM SPSS, Cognos, MicroStrategy, Tableau, SAS, SAP Business Objects, R, Rapid Miner) free of charge.

While business analytics and business intelligence are ranked number one on the list of the top 10 technology priorities for chief information officers in 2014—ahead of such areas as mobile and cloud, and growth of academic institutions' interest, and the software vendor companies' support—little educational content teaches non-software vendor-specific concepts and fundamentals of predictive analytics (Gerneglia, 2013). While education programs on predictive analytics exist under the umbrella of data mining, almost all of these programs are geared toward teaching the mechanics of using the software tools (e.g. SAS, SPSS), not fundamentals of predictive analytics and modeling. Also, most of these education programs do not teach how to integrate predictive modeling in an organization's business processes and how the market (economic vs. economic + social market) affects the decision making processes. Especially in large corporations, the development and deployment functions are mostly fragmented and distributed to silo departments; that's why it is almost impossible for students to practice this complex environment just by learning predictive modeling vendor software.

However, as the use predictive analytics becomes widespread in disciplines such as information systems, decision sciences, operations management, supply chain management, marketing, finance, accounting, and healthcare, graduates need to learn the fundamentals of predictive analytics and gain the necessary knowledge and skill sets to use as predictive analytics effectively. Predictive analytics drive business transformation only when embedded into existing business processes and dynamic decision making processes (Demirkan & Delen, 2013). Today, there are more than seventy business schools with a type of course, program, and/or degree in business intelligence and business analytics that includes curriculum about predictive analytics. Most of these classes are offered by the information systems discipline (Institute for Advanced Analytics, 2013; KDnuggets, 2014; Power, 2011).

In this paper, we offer a tutorial that teaches concepts and fundamentals of predictive analytics and decision making under uncertain market conditions. We base our tutorial on a scenario-based simulation teaching methodology designed to engage students as stakeholders in using new and existing predictive models in a corporate setting. This tutorial complements the predictive analytics and business intelligence training by showing students the "so what?" factor after developing and deploying predictive models. Students gain skills, knowledge, and experience necessary to become effective decision makers by learning key concepts, doing hands-on exercises, sharing knowledge with each other, and through applying design thinking. Some of the learning objectives are to:

- Learn fundamentals of predictive analytics model development and deployment
- Evaluate campaigns with performance management
- Analyze differences between business process-enabled predictive model development versus IT's role
- Comprehend benefits of using a centralized model repository rather than having many fragmented models
- Interpret real time performance dashboard for predictive models, and

- Practice data-driven decision making and decision making under uncertainty.

Instructors who desire to use hands-on decision making dashboards with a sample business case in IS undergraduate and graduate classrooms and professional education workshops aimed to foster a collaborative and dynamic learning environment may find this tutorial especially useful.

This tutorial also proposes several IT artifacts, some of which professionals do not currently use: a conceptual centralized model repository, standardized performance management metrics (pie chart, stock price, revenue, cost, model builder, number of championships), and a real-time dashboard that shows the model interaction and performance. The dashboard also shows other visual and numeric performance metrics that incorporates all prediction market stakeholders' risk and return.

Readers can download the case and associated materials as an Integrated material set from the Teradata University Network (TUN) at http://www.teradatauniversitynetwork.com (Watson, 2009; Watson & Hoffer, 2003). TUN is a free, online software-as-a-service education platform that includes sample cases, PowerPoint presentations, data, and other content for data and information management courses (Demirkan, Goul, & Gros, 2010; Gros, Goul, & Demirkan, 2011). The Sanders integrated material set comprises a case and other documents that explain how several departments are doing business in the gaming industry. At The Sanders, while customers use the company's trademarked "Play-to-Win" card , customers' preferences, gaming patterns, and expenditures are constantly analyzed to build predictive models to score customer profitability given certain types of special offers by the Sanders.

We organize the paper as follows: In Section 2, we explain the demand for predictive analytics and modeling related education. In Section 3, we explain the tutorial with a case study and sample predictive models. In Section 4, we discuss evidence of student learning, and, in Section 5, we conclude the paper.

2 Overview

To survive in an increasingly competitive information-centered economy, today's organizations must constantly assess and update their strategies, techniques, and tools for effective information management. Organizations in every industry are now realizing the benefits of using data to align their current actions with their future objectives. Most companies, while interested in implementing predictive analytics, are challenged with hiring new employees who understand what predictive analytics actually is and with knowing how to develop, apply, and maintain the right predictive models in organizations (Sahoo, 2010). McKinsey Global Institute forecasts that, by 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills and 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions.

As a result, efficient and effective decision making processes with data-driven discoveries (e.g.. the practice of using quantitative and statistical models to identify trends and relationships that might otherwise not be apparent) and predictive analytics (e.g.. the practice of forecasting future trends) are becoming mainstream applications for companies that need to run smarter, more agile, and more efficient businesses (Davenport, 2006).

As the growth of data and analytics continue, Gartner predicts that, by 2015, 4.4 million IT jobs (business intelligence and analytics) will be created to support big data, with 1.9 million of them landing in the United States (Hall, 2012). It is important for the next generation of managers to gain an understanding of analytics. Increasingly, data analytics is becoming an essential skill for managers[1]. According to Manish Parashar, director of Rutgers Discovery Informatics Institute, economic advantage depends on the available data plus one's ability to transform that data into meaningful insights. The leaders are industries nimble enough to interpret and use the data in new ways to add value. Traditional decision making structures must adapt to incorporate data scientists in business and research.

Universities must get involved and educate the future technology-driven business leaders in how to best harness new technology and develop more efficient processes (Bell, Mills, & Fadel, 2013). Almost every business school implements a type of data and information management, business intelligence, and decision-analysis course that demonstrates the power of using data and information for effective decision making. More universities now offer a master's degree or certificate programs in analytics or even just predictive analytics (e.g., North Carolina State University, Arizona State University, University of Texas

[1] See http://www.itbusinessedge.com/cm/blogs/hall/data-analytics-the-new-must-have-skill-for-managers/?cs=50647

Austin, University of Chicago). One of the biggest challenges is having the right tools and techniques to practice the concepts and theories covered in course texts.

Some faculty members use case studies from Harvard, California Management, or similar resources, but these exercises are typically higher level and managerial.

These case-study-oriented course curriculums are great tools to simulate a work place environment. However, these exercises target more strategic approaches; they usually do not include hands-on exercises that cover fundamentals. As companies intend to increase spending on analytics, hands-on skills to build, maintain, and deploy predictive models become more necessary. Business school students are hard-pressed to find cases or tutorials on decision making using predictive models. Developing and maintaining predictive models are costly and time consuming investments for organizations. To use any solution, these predictive models need to be embedded into business processes, and managers need to understand how to drive value from using these solutions.

Predictive modeling[2] is the process of analyzing data to create a statistical model of future behavior. Predictive modeling solutions are a form of data-mining technologies that work by analyzing historical and current data and by generating a model to help predict future outcomes. One can use these technologies to generate a score (for example, a credit score), to assess behavior (for example, fraud detection or customer acquisition), or to analyze needed reserves. Insurers can apply this to key activities, such as customer service, pricing, actuarial, underwriting, and claims, to improve outcomes. Vendors such as SAS, IBM, TIbco, and R offer many software tools and methodologies. Table 1 depicts a sample list of software available to higher education institutes to teach predictive analytics.

Table 1. A Sample List of Software Available for Teaching Predictive Analytics

Software	Method	Capabilities
SAS-OnDemand	SEMMA (sample, explore, modify, model, assess)	Provides a no-cost, online delivery model to professors for teaching and to students for learning data management and analytics. By connecting to an SAS server in the cloud, users access the analytical power of SAS software through a user-friendly, point-and-click interface.
SPSS Academic Initiative	CRISP-DM (business understanding, data understanding, data preparation, modeling, evaluation, deployment)	By providing many flexible software options, SPSS facilitates integrating analytical content into undergraduate and graduate level courses.
R	KDD (selection, preprocessing, transformation, data mining, interpretation-evaluation)	R is an open source, free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows, and MacOS. It provides a wide variety of statistical tools (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, etc.)

Unfortunately, most education programs focus on the mechanics of learning the software and developing models with these tools. For the most part, they have already cleaned and transformed data sets with known results to teach predictive models. Whereas, in practice, many predictive models are built across different business units with several different software solutions.

These software-driven education programs are not designed to teach basic conceptual understanding, deployment, return on investment, business process integration, and management of predictive models. Most of them do not provide a real-life decision making scenario where the manager decides which model to deploy and which model needs to remodel. Only KDD from R includes the deployment phase as a definition: it notes that creating the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it (Chapman et al., 2000). However, the deployment phase does not include how to organize and to present the model in a way the customer can use. Their definition assumes that knowledge discovery through predictive modeling is totally outside of the system and the “customer” is actually the decision maker. These exercises become standalone technical skills and are not

[2] http://www.gartner.com/it-glossary/predictive-modeling-solutions

perceived as strategic weapons. In addition, they don't support "work-based pedagogy". Work-based pedagogy notes that "management education should focus on the imperfect realities of the workplace rather than on the idealized models of best practices" (Schlenker & Mendelson, 2008).

Through an extensive literature review, we assessed alternate methods for teaching data-driven discoveries and predictive modeling and analytics. For example, some studies demonstrate how to teach advanced decision making with analytics using tools such as MS Excel and simulation (Li, 2007; Cronan, Douglas, Alnuaimi, & Schmidt, 2011; Ragsdale & Zobel, 2010). These studies are excellent in terms of traditional perspectives of decision making. We believe teaching pedagogy for predictive analytics needs design thinking in addition to decision making with case analysis (Barnes, Christensen, & Hansen, 1994) and team learning (Rassuli & Manzer, 2005).

This tutorial explains the use of a hypothetical business scenario where the vice president of business intelligence solutions of Sanders company, Yolanda Wales, deployed predictive analytic models to analyze and make decisions about the company's marketing campaigns. Sanders developed and deployed several different predictive models (including an outsourcing, in-house logistic regressions, neural networks techniques, and some others) for many years. The company already has set of pre-developed models in their model repository. Table 2 briefly describes the Sanders' predictive analytics model environment.

Table 2. Sanders' Predictive Analytics Models Development Environment

- In the past, one employee was developing predictive models; now, they have many people who develop models in a fragmented manner.
- In the past, the company deployed mostly one model for their campaigns; now, they need to develop and deploy multiple models simultaneously.
- The company needs a repository to store and manage multiple models.
- The company needs to be able to review the outcome of each deployed model.

Sanders realized that, while the model development team was quickly growing and more tools were available to develop advanced models, outcomes from the predictive models were not improving significantly. As such, they considered developing a centralized model management system to help them obtain: 1) a central repository that stores all predictive models that are currently developed and stored in employees' desktops in a fragmented way, 2) incorporated cost, proposed revenue, and actual results from the campaigns using these models, and 3) standardized metrics and measures to evaluate the performance of models.

3 How to Teach Predictive Analytics with the Wales Market Case

To teach predictive analytics with the Wales Market case, faculty members, once on the TUB website (www.teradatauniversitynetwork.com), should:

- Go to http://www.teradatauniversitynetwork.com/Library/Items/Solving-Business-Problems-with-Data-Driven-Discoveries-and-Predictive-Analytics/
- Review and download the below list of items (Table 3).

Table 3. Materials Used to Teach Predictive Model Management

Content	File name	Description
Overview of predictive modeling	1. What is Predictive Modeling.ppt	Brief explanation of predictive analytics and modeling, including basic concepts, such as decile, model lift, and model overlap.
Brief review of the case, dashboard and instructions for students	2. IntroToSandersCasino.ppt	After students read the case and reviewed the dashboard on their own. The instructor reviews and summarizes the business problem that needs to be solved using predictive models and answers students' questions.

Table 3. Materials Used to Teach Predictive Model Management

Case study that explains the business problem (Wales Market case)	3. WalesMarketCase.pdf	A mini-case about Yolanda Wales, VP of Business Intelligence at the Sanders' Casino. She needs a good way to use and manage predictive models that are built by different methods
Dashboard for predictive model performance reports	4. WalesMarketDashboard.xls	Informs the manager about the revenue and overlap from multiple predictive models with their performances.
Data entry forms for students (economic market)	5. BettorSheetEconomic.pdf	Betting sheets are provided to the students so that they mark their selection to see how the selections are made. Students are awarded $1000 play money to disburse among model deciles to show their decisions.
Data entry forms for students (economic & social market)	6. BettorSheetEconomicSocial.pdf	Betting sheets are provided to the students so that they mark their selection to see how the selections are made. Students are awarded $1000 play money to disburse among model deciles to show their decisions.
Evaluation of results	7. TemplateForBettingResults.xls	Combining the overall results to see which model deciles will be used in the next cycle and selecting the winners.

- Divide students into two groups: members of group 1 will be tasked to make their decisions without any discussion and collaborations with others (economic market). Members of group 2 will be tasked to make their decisions with discussions and collaborations with others (economic & social market). This will help students to see differences about making decisions with these two markets.
- Begin a class with the PowerPoint deck 1.What is Predictive Modeling.ppt. Make sure students understand predictive analytics and definitions of lift, deciles, scoring, and dashboards. The last slide in that deck is compelling—it is a toddler with two ice cream cones who is deciding which to eat. This is germane to the deployment decision: once a model is deployed, you can't tell with certainty what the model choice might have done for you: you've already invested in your best bet. The same is true of that ice cream cone on a hot day. Once you decide on a flavor, the other ice cream will melt away.
- Hand out the case study to students (3.WalesMarketCase.pdf—Appendix A), and go over the second PowerPoint deck titled 2.IntroToSandersCasino.ppt, which explains a business case titled "Wales Market". The dashboard demonstrates the pre-developed models with their performances. After reviewing the decks, instructors can ask students the following questions to start guided discussions:
    1. What is the marketing manager's dilemma?
    2. How do predictive models work? What is a decile? What is the importance of "lift" in decision making?
    3. How do companies make money from using predictive models?
    4. What does it mean to deploy a predictive model? Why is deploying a predictive model in the right manner often important to the bottom line?
    5. If the cost of mailing an offer is $1 and you have a budget of, for example, $200,000, do you have to mail the offer to the top 200,000 records that are most likely to respond? How do you know where to stop?
    6. How can an expected return on investment (ROI) for using a predictive model be calculated, and why is it important to determine this ROI?
- Display 4.WalesMarketDashboard.xls (Appendix B) that shows the results from earlier champion and challenger models. When a model is first used in a campaign—such as a mailing to encourage customers to come to a casino during a week that is known to be slow given the normal business cycle—that model is known as the initial champion model. It is difficult to

compare the predicted performance of a non-champion model ("challenger model") with the performance of the champion model.

- Ask students in each group to bet on each campaign's model performance. Students in the economic-only market are asked to fill out the betting sheet for economic market (5.BettorSheetEconomic.pdf—Appendix C) by allocating their fictional $1,000 across different models and deciles in those models by using the information on the dashboard. This group is not allowed to talk or collaborate with each other during this exercise.
- Ask students in the economic and social market to bet on each campaign's model performance (fill out the betting sheet for economic & social market (6.BettorSheetEconomicSocial.pdf—Appendix C) by allocating their fictional $1,000 across different models and deciles in those models by using the information on the dashboard. This group is allowed to talk or collaborate with each other during this exercise and give each other candies for exchanging information. Students are also asked to enter the candy exchange information on the second page of the bettor sheet. The count of candy is not used in any way for deciding the winners. It is used to measure the extent of interaction.
- At the end of the exercise, collect all the bettor sheets and enter the results in the Excel sheet template provided (7.TemplateForBettingResults.xls—Appendix D). Calculate the number and dollar amount bet for each stock (model decile) as described in detail in the Excel worksheet and present it to the class.

Students will be surprised that most of the bets moved away from the champion model deciles and actually no statistically significant difference between the way two groups' betting exists (economic vs. economic and social). Students realize that they were participating in the decision making process for using predictive models. When the enterprise offers multiple models, students, therefore, learn how models are used for decision making. Comparison of multiple models is made available by real-time dashboards. The student(s) who bet the top three most populist stocks wins the prize.

The gist of the case is to create a "predictive model market" where students can invest in models for deployment in a campaign. Campaign decision making is then informed by the bets that the students make in the class. This gives rise to two additional discussion questions:

1. If predictive models are becoming company assets, how can organizations manage them as a portfolio?
2. What are the similarities and differences between tracking the performance of model assets and tracking the performance of stocks in the stock market?

When predictive analytics are taught, most students believe that, once a predictive model is constructed, then the work is done. In this exercise, we set out to provide students with a better understanding of how models are deployed in companies. We wanted to show how multiple models might be in play in the context of a campaign (targeted marketing, fraud detection, etc.), why decisions about deploying models (and changing the models deployed mid-campaign) are complex, and why they are made without complete information. We also want to make this a fun, challenging, and a cooperative learning experience.

4 Evidence of Student Learning

The exercise starts with reviewing the concepts of predictive modeling, reviewing the Wales Market case, explaining the dashboard that shows comparative performance of the predictive models in the portfolio, and making a decision based on the interpretation of the model dashboards. The sequence of these events keeps the students engaged in the class. The results are compiled and shared with students in the next class by compiling betting results in a chart that shows the count of bets and amount of bets for each bet (a sample of results is provided to instructors at TUN website). As the graphs show, a collective intelligence forms for the bets. Students repeatedly tend to pick the same stocks as favorites as a group even though their individual bets vary. This phenomenon is explained to students as using prediction markets as decision support systems.

Students are excited to find out how well prediction markets worked, especially as they see that the results from the economic and economic and social markets yield the same top (most favorite) stocks. The business intelligence section of an introduction to information systems class consistently uses this exercise. We taught approximately 40 students in the 2012-13 academic year, approximately 160 in the

2011-12 academic year, and approximately 100 students in the 2010-11 academic year with these case materials. Students completed the assignment successfully. In the context of a recent class experiences, we asked students to provide feedback regarding the use of Wales Market. The following comments and discussions provide support for the effectiveness of this tutorial in promoting knowledge of how predictive models work and how dashboards can be used in reporting performance of multiple models and how to interpret model deciles:

- Eighty percent of the students acknowledged that the case helped them to become familiar with the concepts used in assessing the performance of predictive models such as "decile" or "lift".
- Seventy-six percent of the students indicated that this exercise helped with the knowledge of how predictive models are used and managed as a portfolio. They mentioned acquiring the ability to compare models by viewing performance data and analyzing/picking the best models.
- More than half of the students mentioned that working with dashboards helped them to interpret the information provided in dashboards.

We issued a quiz to students to measure their learning for concepts.

We also asked students to rank order the following information when they were making decisions (what they paid the most attention to in making their allocation decision):

____ Pie charts
____ Stock prices
____ Revenue
____ Cost
____ Model builder
____ Prior championships
____ Creation date
____ Profit
____ Model overlap tables

Table 4 shows the results for the economic vs. economic and social groups.

Table 4. Dashboard Information Ranking

Criteria	Economic	Economic and social
Pie charts	2.92	2.71
Revenue	2.92	2.90
Stock prices	3.18	4.08
Profit	4.31	3.90
Model overlap	4.54	4.87
Cost	4.97	4.87
Prior championships	6.06	6.28
Model builder	6.92	7.05
Creation date	7.72	7.89

These results also informed us about what we should include in the predictive analytics management dashboards.

5 Discussion and Conclusion

This tutorial should be of interest to educators who desire to teach students how to make decisions in the context of applying predictive analytics. Students learn the power of teamwork (knowledge sharing), and they apply design thinking using case studies. The important skills gained are effective decision making and problem solving with dashboards and prediction markets.

This tutorial also takes the students through the process of building models, managing models, seeing the results of the models, and incorporating these results into the next campaign cycle by creating a lively and active work environment and using the role-playing technique. As the manager realizes the need to bring all these models together to create transparency across stakeholders and keep the performance metrics up to date, the students also realize that they are part of the decisions and how the decisions they make as “employees” or “stakeholders” impact their company’s and their own bottom-line.

Ninety-eight percent of employers believe business school graduates need to know how to use data to drive decisions according to the 2013 “Year-End Poll of Employers”(GMAC, 2013) by the Graduate Management Admission Council (USNews.com). In the last two years, we have seen a big expansion of business analytics and predictive analytics programs in the United States higher education institutions. These programs use and create alignment with BI and big data institutions such as SAS and IBM and teach students analytics tools and concepts. This tutorial is an excellent integrated and self-contained material that can be used in all these programs regardless what modeling technique or software is used. This tutorial involves managing multiple predictive analytics solutions through an IT-enabled solution and students (as “employees for the Sanders”) as stakeholders for selecting the right model for the right campaign using the right metrics and dashboards. Because predictive model use proliferates in analytics-driven business campaigns, students need to understand how exactly these advanced analytics tools are used in real-life decision making. Students may be taught *what* predictive modeling is and *how* to build predictive models using different techniques in a lecture environment. However, with the Wales Market case and a classroom activity enforcing participation in decision making, students understand *why* it is important to deploy predictive models and *why* we must treat predictive models as company assets and displayed in a comparative dynamic dashboard.

References

Bell, C., Mills, R., & Fadel, K. (2013). An analysis of undergraduate information systems curricula: Adoption of the IS 2010 curriculum guidelines. *Communications of the Association for Information Systems, 32,* 73-94.

Barnes, L. B., Christensen, C. R., & Hansen, A. J. (1994). *Teaching and the case method: Text, cases, and readings.* Boston: Harvard Business School.

Chapman, P., Clinton, J., Kerner, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0—Step-by-step data mining guide*. Retrieved from https://the-modeling-agency.com/crisp-dm.pdf

Cronan, T. P., Douglas, D. E., Alnuaimi, O., & Schmidt, P. J. (2011). Decision making in an integrated business process context: Learning using an ERP simulation game. *Decision Sciences Journal of Innovative Education*, *9*(2), 227-234.

Davenport, T. J. (2006). Competing on analytics. *Harvard Business Review*, *84*(1), 99-107

Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems and Electronic Commerce*, *55*(1), 412-421.

Demirkan, H., Goul, M., & Gros, M. (2010). A reference model for sustainable e-learning service systems: Experiences with the joint university/teradata consortium. *Decision Sciences Journal of Innovative Education*, *8*(1), 151-189.

Gerneglia, B. (2013). Top 2014 IT priorities for state CIOs. *CIOZONE*. Retrieved from http://www.ciozone.com/index.php?option=com_myblog&show=Top-2014-IT-Priorities-for-State-CIOs.html&Itemid=713

GMAC. (2013). “2013 year-end poll of employers” summary report. *Graduate Management Admission Council*. Retrieved from http://www.gmac.com/~/media/Files/gmac/Research/Employment%20Outlook/2013-year-end-poll-summary-report.pdf

Gros, M. E., Goul, M., & Demirkan, H. (2011). Promoting effective decision making using analytics in a virtual technology lab. *Decision Sciences Journal of Innovative Education*, *9*, 119-127.

Hall, S. (2012). Gartner: Big data will generate 6 million U.S. jobs by 2015. *ITBusinessEdge*. Retrieved from http://www.itbusinessedge.com/blogs/charting-your-it-career/gartner-big-data-will-generate-6-million-u.s.-jobs-by-2015.html

Institute for Advanced Analytics. (2013). Survey Of graduate degree programs in analytics. *NC State University*. Retrieved from http://analytics.ncsu.edu/?page_id=4184

Li, X. (2007). Intelligent agent-supported online education. *Decision Sciences Journal of Innovative Education*, *5*, 311–331.

KDnuggets (2014). Education in data mining, analytics and data science in USA/Canada. *KDnuggets.com*. Retrieved from http://www.kdnuggets.com/education/usa-canada.html

Moore, S. (2011). Gartner forecasts global business intelligence market to grow 9.7 percent in 2011. *Gartner*. Retrieved from http://www.gartner.com/newsroom/id/1553215

Murphy, I. B. (2013). Gartner researchers: Predictive analytics to gain traction in business. *DataInformed*. Retrieved from http://data-informed.com/gartner-researchers-predictive-analytics-to-gain-traction-in-business/

Power, D. (2011). What universities offer masters degrees in analytics and data science? Retrieved from http://dssresources.com/faq/index.php?action=artikel&id=250

Ragsdale, C. T., & Zobel, C. W. (2010). A simple approach to implementing and training neural networks in Excel. *Decision Sciences Journal of Innovative Education*, *8*(1), 143-149.

Rassuli, A., & Manzer, J. P. (2005). Teach us to learn: Multivariate analysis of perception of success in team learning. *Journal of Education for Business*, *80*, 21-27.

Sahoo, S. (2010). *Power of predictive analytics*. Hexaware Technologies.

Schlenker, L., & Mendelson, A. (2008). Technology at work. *BizEd*, *7*(1), 22-26.

Watson, H. J. (2009). Tutorial: Business intelligence—past, present, and future. *Communications of the Association for Information Systems*, *25*, 487-510.

Watson, H. J., & Hoffer, J. A. (2003). Teradata university network: A new resource for teaching large data bases and their applications. *Communications of the Association for Information Systems*, *12*, 130-144.

Appendix A: Excerpt from Wales Market

Yolanda Wales is Vice-President of the Business Intelligence Solutions Area of an international conglomerate named “The Sanders”. The Sanders comprises several divisions doing business in the gaming industry. With casinos on four continents, the company also manages numerous American Indian Nation casinos in the U.S.A. Business intelligence (BI) is the primary enabler of customer relationship management whereby customers enjoy seamless cross-casino experiences. Using the company’s trademark “Play-to-Win” card, customers’ preferences, gaming patterns, and expenditures are constantly being analyzed to build predictive models designed to help score customer profitability given certain types of special offers. Such special offers include complimentary room-nights, dining coupons, free slot machine play, and so on.

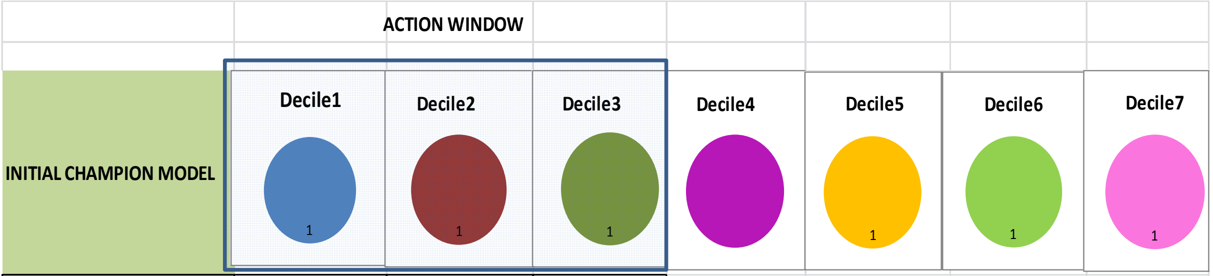
Because the company’s BI capabilities have advanced to the point where analysts are proliferating numerous high-quality predictive models, Yolanda needs to come up with a new “model dashboard” to help all of her analyst teams, the marketing group, and other managers to quickly ascertain which models are the best to use in particular marketing campaigns. These groups need a single source of the truth when it comes to the available model assets. However, achieving a single source of the truth is more complicated than one might first envision. When a model is first used in a campaign—say to encourage customers to come to a casino during a week that is known to be slow given the normal business cycle—that model is known as the initial champion model. It is difficult to compare the predicted performance of a non-champion model (called a challenger model) with the known performance of the champion model. For example, if the champion model for a campaign targeted 30 customers and 10 actually came and spent money using their Play-To-Win card, then, for that champion model, we know with certainty the profitability. In contrast, a challenger model might have targeted a different set of customers, and we could only speculate on their profitability if we tried to compare the champion and non-champion models. However, some customers were targeted in both models, known as the model overlap.

Yolanda invented a unique strategy for the dashboard she envisions. She explains:

> *Why not create a market of models where each model can be described, explained and its performance can be shown to all of the main stakeholders in a campaign. Those stakeholders could be the analysts who build models, the managers who run the campaigns - and even the marketing group composed of people who experience in running many other campaigns. We’ll stake them each $1,000 (in real money) and ask them to allocate that money to buy shares of the different models in a bet they are making on each campaign’s model performance. Once everyone makes their bets, we’ll pick the model combinations with the highest investment and make it the model for the next cycle of the campaign. After we know the results of the campaign, we’ll let everyone know the actual and predicted performance, and then we’ll start another round of betting. The market will react by changing the share prices of the models. New share prices will reflect model performance and the demand from the prior cycle of the campaign. Each “bettor” will have a portfolio of the shares it owns in the market continuing as long as we are running the campaign. This unique market will provide an informative and innovative way to guide our management of this complex and growing set of model assets at our disposal. Our collective bets will guide our actions, and those who make money from their bets will, over time, influence the models we deploy in our campaigns—because they will have more to bet with!*

After all, Yolanda thought, “We all work in the gaming industry, and betting is in our blood!

Yolanda’s idea was assigned for further development to her junior executive, Mora Modeles. Mora had the difficult task of figuring out how to determine the performance of a challenger model from one cycle of a campaign to another, and, more importantly, how a challenger model would pay off when the results weren’t known with certainty. On one hand, the challenger model might perform better on the most profitable subjects that it had in common with the initial champion model. On the other hand, since the challenger model might target different subjects, the profitability of those subjects would never be known. Mora decided that what she could do was provide the market with information only about what was actually known. She could provide the market with a snapshot that addressed the scoring of subjects by the champion model; for example:



Revenue	$ 900.00	$ 900.00	$ 700.00				
Cost	$ 50.00	$ 50.00	$ 50.00				
Share Price	$ 65.00	$ 60.00	$ 55.00				
Model Builder	Bob Hamilton						
Prior Championships	5						
Creation Date	Jan-12						
Total Profit	$ 2,350.00						

Figure A1. Scoring of Subjects by the Champion Model

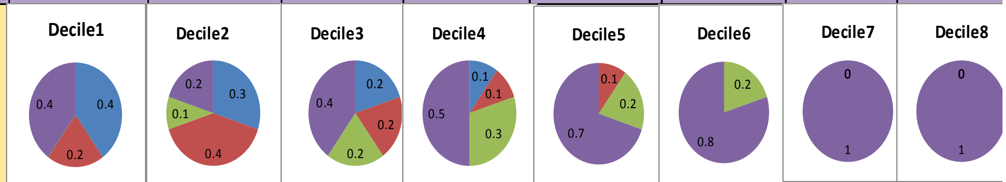
This table and the graphs show many things, and Mora would explain them to anyone who would listen. This set of graphs and numbers in the dashboard reflects the first stage of a campaign where the initial champion model was used. This model scored all of the possible people to be targeted in the campaign (i.e., the population). For example, if the model was a good model, it would score those people highest who The Sanders could make the most money from. The action window shows the graphs in the dashboard that are relevant to the first stage of the campaign.

After the first cycle of a campaign, we know the profitability of the initial champion model with certainty. So, if we color-code the percentage of respondents that conform to each of what are referred to as the scored "deciles", then we can tell when we compare this model to challenger models how dispersed the scored subjects might be among the different scoring capabilities of each model. A decile is a division of the population by their scores such that there are an equal number (one-tenth) of the subjects in each. The top decile (decile 1, for example) refers to the top ten percent of subjects who received the highest scores by the first model. In the initial champion model, the color blue signifies those who were targeted in the campaign (because they are inside the action window), and it indicates that this set of "blue" subjects were those in the top 10 percent of all subjects as scored by the initial champion model. Similarly, the color burgundy signifies those scored by the initial champion model to be in the second 10 percent of the scores given to all subjects. The cost of a model by decile refers to promotion cost, and that promotion cost is fixed by decile. The share price refers to the bet that Yolanda's stakeholders can make. A higher share price reflects the amount of profit that a model generated in its lifetime, and it takes into account the demand from the bettors for that model's performance for the particular decile. For example, a share price of $60 listed under decile 2 means that a bettor would pay $60 for each share of that decile for that model because they think the share price will go up after the next cycle of the campaign. Other important information about a model is also provided. The dashboard shows the model builder, it shows how many times that model was selected by the campaign manager to be the initial champion model, and it shows the date it was created. Some bettors rely on this information in making their decisions—almost as much as they rely on the pure profits a model produces. That profit is shown for the initial champion model at $2550. This figure is calculated by considering only the action window and the amount of revenue by decile, less the fixed costs for the deciles.

Now, if we want to look at how the initial champion model compared with challenger models (e.g., challenger model 1), we need to be able to see the dispersion of scored subjects as Figure A2 shows:

Model 1 overlap	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8
ovelap with Champ DECILE 1	0.4	0.3	0.2	0.1	0	0	0	0
ovelap with Champ DECILE 2	0.2	0.4	0.2	0.1	0.1	0	0	0
ovelap with Champ DECILE 3	0	0.1	0.2	0.3	0.2	0.2	0	0
no MATCH withCHAMP	0.4	0.2	0.4	0.5	0.7	0.8	1	1

CHALLENGER MODEL 1



Dispersed Revenue	$ 540.00	$ 980.00	$ 500.00	$ 210.00	$ 150.00	$ 140.00		
Cost	$ 50.00	$ 50.00	$ 50.00	$ 50.00	$ 50.00	$ 50.00		
Share Price	$ 67.00	$ 66.00	$ 59.00	$ 50.00	$ 40.00	$ 30.00		
Model Builder	NeuralNet Consulting							
Prior Championships	1							
Creation Date	Apr-12							
Total Dispersed Profit	$ 2,220.00							

Figure A2. Initial Champion Model vs. Challenger Models: Dispersion of Scored Subjects

Note that, in the first decile of the comparison, there are three different colors in the pie chart. The blue and burgundy correspond to those subjects scored in the initial champion model in their respective deciles, and it shows that there was a 40 percent overlap of subjects in decile 1 between the two models. This overlap means the initial champion model scored subjects in the first decile and so did challenger model 1. The purple area depicts those that were scored only by the second model, but without overlap with the initial champion model. In other words, the results are unknown for this 40 percent because they were not targeted in the campaign. However, "dispersed revenue" shows the actual amount of revenue made by the people targeted in challenger model 1 in the corresponding decile of that model. Mora always smiled when she got to this part of the discussion: she got to explain how those numbers might actually come out to be greater than the revenue of the initial champion model. She would note:

> *The customers we make the most money on are scored into multiple deciles by the initial champion model. However, a challenger model may make more profit within a decile because it happened to score those most profitable customers in the particular decile. For example, consider the revenue of the second decile of challenger model 1. It is $980.00, while the second decile for of the initial champion model is $900.00. The reason it made more revenue is because more of what they call "whales" in the gaming industry happened to be scored in the second decile of challenger model 1. In considering which model to use in the next cycle of a campaign, one must pay attention to the dispersion, because it helps determine which models scored the best customers in each decile.*

"However", Mora would carefully point out, "the total revenue is still the same total as in the use of the Initial Champion Model—it is the known revenue—but the dispersion of that revenue is important information.". After explaining it, most of the time Mora would wait a few minutes and then say:

> *So, your challenge in this market is to take your stake and allocate it to deciles for the next round of the campaign. For example, you may choose to invest in the second decile of champion model 1 because it made more profit than the initial champion model. But since you are making an investment, you need to take the share price into account as well. Remember, this is a market, and anything can happen in the next cycle. Investing everything in one bet on one outlier can get you in the long run—I know, I used to run a craps table. I've seen some big losers bet on a number just because it came up last time.*

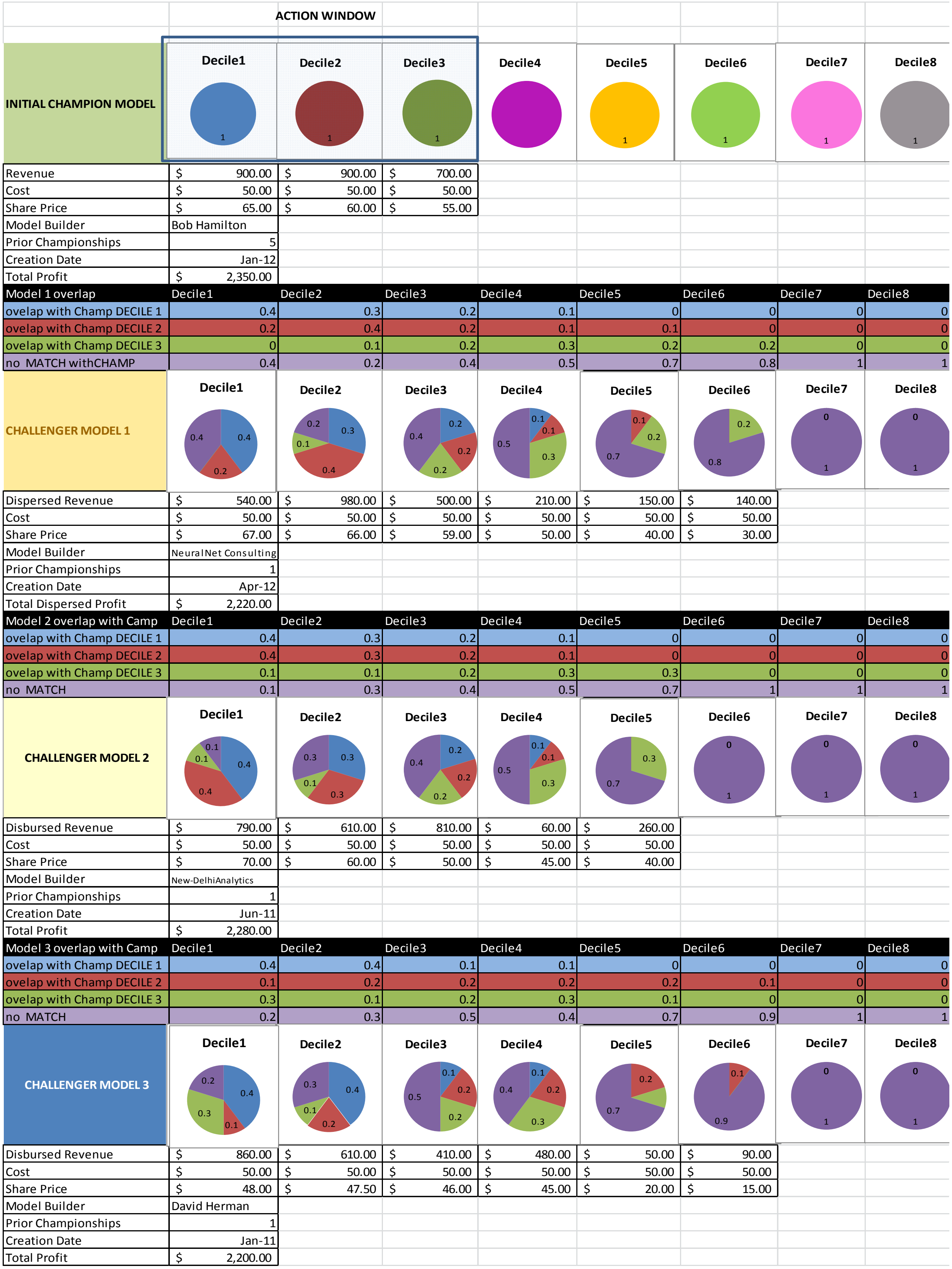
Mora prepared the attached sheet to represent the model market information she can provide. Code-named "The Wales Market" after her boss's innovation, the dashboard shows all of the comparisons subsequent to the first utilization of the initial champion model. This snapshot of the markets describes the actual performance of the champion model and the integrated performances of the non-champion models. There are three challenger models in the dashboard, and for each, the dispersion of revenues is shown as based on actual expenditures from the first cycle of the campaign when the initial champion model was used. The total profit for each of the challenger models is lower than that of the initial champion model because capturing all of the subjects that were targeted by the initial champion model would take more mailings, and so additional fixed costs are associated with each of these mailings. This is not an unusual dispersion: some highly scored subjects in one model are often scored much lower in another model.

In the second and subsequent cycles of a campaign, Yolanda uses what is referred to as a “fused model”. That is, she can use the subjects targeted by some model A as ranked in the first decile, some model B as ranked in the second decile, and so forth. In her experience, the bettors’ bets provide the best information on which models and their associated deciles to include in the next cycle. She noted: “It’s pretty interesting: the bettors do a pretty solid job of providing model selection guidance. We run with the model/decile mix associated with the highest bets.”.

Mora Modeles’ assistant Winnie Whiner was always assigned to give the new model builders a training class on investing in the Wales Market. When Winnie explained to the new staff how it all worked and how to make an investment decision, she made it plain and simple:

> *Look, you’ve got a thousand bucks and you going to buy shares in the particular deciles of models. Nothin’ more, nothin’ less. You’re hoping your share price will go up after the next cycle of the campaign. If the share prices go up, you’ve made some money in the market. That will happen if there’s lots of demand for those shares, and there’s usually lots of demand where there’s the highest profitability from what you know from the prior campaign cycle. You’ve got to figure out which model deciles are worthy of your investment. When you’re ready to make your bets, you fill in a betting sheet. Below is an example. Look forward to seein’ how well ya do! Doubt you’ll do too well. By the way, if you make stupid bets, the rest of the people around here will be all over you, so don’t screw it up.*

Appendix B: Dashboard



	ACTION WINDOW							
INITIAL CHAMPION MODEL	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8
Revenue	$ 900.00	$ 900.00	$ 700.00					
Cost	$ 50.00	$ 50.00	$ 50.00					
Share Price	$ 65.00	$ 60.00	$ 55.00					
Model Builder	Bob Hamilton							
Prior Championships	5							
Creation Date	Jan-12							
Total Profit	$ 2,350.00							

Model 1 overlap	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8
ovelap with Champ DECILE 1	0.4	0.3	0.2	0.1	0	0	0	0
ovelap with Champ DECILE 2	0.2	0.4	0.2	0.1	0.1	0	0	0
ovelap with Champ DECILE 3	0	0.1	0.2	0.3	0.2	0.2	0	0
no MATCH withCHAMP	0.4	0.2	0.4	0.5	0.7	0.8	1	1

CHALLENGER MODEL 1	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8
Dispersed Revenue	$ 540.00	$ 980.00	$ 500.00	$ 210.00	$ 150.00	$ 140.00		
Cost	$ 50.00	$ 50.00	$ 50.00	$ 50.00	$ 50.00	$ 50.00		
Share Price	$ 67.00	$ 66.00	$ 59.00	$ 50.00	$ 40.00	$ 30.00		
Model Builder	NeuralNet Consulting							
Prior Championships	1							
Creation Date	Apr-12							
Total Dispersed Profit	$ 2,220.00							

Model 2 overlap with Camp	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8
ovelap with Champ DECILE 1	0.4	0.3	0.2	0.1	0	0	0	0
ovelap with Champ DECILE 2	0.4	0.3	0.2	0.1	0	0	0	0
ovelap with Champ DECILE 3	0.1	0.1	0.2	0.3	0.3	0	0	0
no MATCH	0.1	0.3	0.4	0.5	0.7	1	1	1

CHALLENGER MODEL 2	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8
Disbursed Revenue	$ 790.00	$ 610.00	$ 810.00	$ 60.00	$ 260.00			
Cost	$ 50.00	$ 50.00	$ 50.00	$ 50.00	$ 50.00			
Share Price	$ 70.00	$ 60.00	$ 50.00	$ 45.00	$ 40.00			
Model Builder	New-DelhiAnalytics							
Prior Championships	1							
Creation Date	Jun-11							
Total Profit	$ 2,280.00							

Model 3 overlap with Camp	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8
ovelap with Champ DECILE 1	0.4	0.4	0.1	0.1	0	0	0	0
ovelap with Champ DECILE 2	0.1	0.2	0.2	0.2	0.2	0.1	0	0
ovelap with Champ DECILE 3	0.3	0.1	0.2	0.3	0.1	0	0	0
no MATCH	0.2	0.3	0.5	0.4	0.7	0.9	1	1

CHALLENGER MODEL 3	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8
Disbursed Revenue	$ 860.00	$ 610.00	$ 410.00	$ 480.00	$ 50.00	$ 90.00		
Cost	$ 50.00	$ 50.00	$ 50.00	$ 50.00	$ 50.00	$ 50.00		
Share Price	$ 48.00	$ 47.50	$ 46.00	$ 45.00	$ 20.00	$ 15.00		
Model Builder	David Herman							
Prior Championships	1							
Creation Date	Jan-11							
Total Profit	$ 2,200.00							

Figure B3. Dashboard

Appendix C: Bettor Sheets

Directions: Fill in your bets on the following form as shown in the Wales Market case. Please remember that your total amount to bet is $1000. You can allocate your bet to any of the deciles shown in the form. Assume that the system to keep track of the bets will compute the number of shares you are purchasing in the Wales Market, so you only need to enter a dollar total. Yolanda, Mora, and Winnie thank you for testing the market. **Please don't forget to include your name in the box where it says "Bettor".**

Wales Market Betting Sheet

Campaign: Weekday Promotion Event

Offer: Free Night Stay on Tuesday or Wednesday

Campaign Manager: Mora Modeles

Bettor:

REMEMBER: YOU HAVE A TOTAL AMOUNT OF $1000 TO INVEST

Initial Champion Model

Share Price:	$65.00	$60.00	$55.00
	Decile 1	Decile 2	Decile 3
Investment			

Challenger Model 1

	$67.00	$66.00	$59.00	$50.00	$40.00	$30.00
	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6
Investment						

Challenger Model 2

	$70.00	$60.00	$50.00	$45.00	$40.00
	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5
Investment					

Challenger Model 3

	$48.00	$47.50	$46.00	$45.00	$20.00	$15.00
	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6
Investment						

Figure C4. Bettor Sheets

Appendix D: Betting Results

	This templete will be used for entering and analyzing the betting results.
In RESULTS Worksheet:	
COLUMN	
Student # (Column A)	Student idenfication number (name and student ID is withheld separately for privacy purposes)
Model Decils (Columns B to U)	Correspond to the bets (model deciles) that are on the dashboard as well as on the bettor sheets. Students on average bet 3 to 4 bets (according to the average of Column X). The amount of dollars bet are entered in the corresponding row for each stufdents for each bet column. The bets have to add up to $1000 for each student.
# Candies (Column V)	Shows the number of candy exchange for the Econ+Social Group and as can be seen this value is equal to zero or "No Candy" for the Econ group as interaction was not allowed.
Sum (Column W)	Shows the total dollar amount. As can be seen in this example few students did not follow the instructions and did not use all of their $1000 or exceede their budget and would disqualify
In ANALYSIS Worksheet:	
We used a separate work sheet called Analysis for evaluation of the results . Calculate the count of bets (stocks)and average amount bet for each stock separately for Econ and Econ+Social. Formulas are on .Rows 72-75 of the Analysis Sheet.	
As can be seen on average and coutwise students bet mostly on the stocks named 12, 21 and 31 (columns are higlighted in dark blue and results are circled in blue) The student or students who bet on the top 3 winning bets (most popular bets) win(s) the prize. Inthis case it is Student #42 because he/she bet only on the winner bets and nothing else. In order to share the results with students insert a 2D Column graph for each group (econ and econ+social) (highlighting B1-U33) that shows the amount on the Y axis and the bets on the X axis.	

Figure D5. Betting Results

About the Authors

Sule Balkan is an Associate Professor at the National Chiao Tung University, Institute of Business and Management in Taiwan. Her research and teaching interests include Predictive Modeling, Business Intelligence, and Application Development. She received her PhD in Economics from the University of Arizona in 1998 specializing in Applied Econometrics. She presented her research in a number of conferences such as AMCIS, ICIS and HICSS. She possesses more than ten years of professional experience in information management, predictive modeling and campaign execution fields. Prior to joining NCTU, she was a Clinical Associate Professor at Arizona State University for four years. She worked as a Director of Information Management at Ameriprise Financial. Her previous positions include being a Senior Manager/Econometrician at American Express International, and Research Associate for National Bureau of Economic Research.

Haluk Demirkan is a Professor of Service Innovation and Business Analytics, and the Founder of Center for Information Based Management at the Milgard School of Business, University of Washington. Haluk received his Ph.D. in Information Systems and Operations Management from the University of Florida. His main research interests are in service science and analytics. Some of his recent joint industry-academic research projects include industry experts from American Express, Intel and IBM. His research appears in many journals, including *JMIS*, *JAIS*, *EJOR, IEEE Transactions*, *ECRA*, *IS Frontiers*, *CACM*, *Journal of Service Research*. He has fifteen years consulting experience in business intelligence and analytics. In 2014, he is being ranked 5th in Top-100 Rankings of Worldwide Researchers according to the Association for Information Systems-sanctioned worldwide research rankings. He is affiliated with a number of research centers. He currently serves on a number of editorial boards