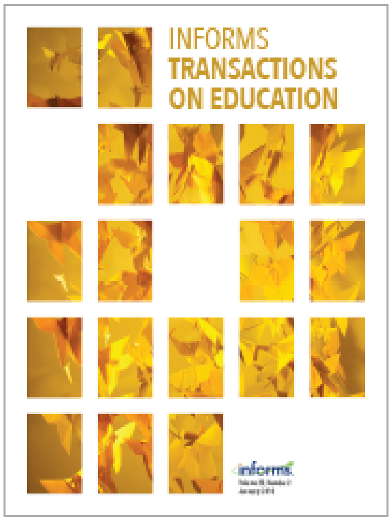


INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Case Article—Business Value in Integrating Predictive and Prescriptive Analytics Models

David Kopcso, Dessislava Pachamanova

informs.
http://pubsonline.informs.org/journal/ited/

INFORMS TRANSACTIONS ON EDUCATION
Vol. 19, No. 1, September 2018, pp. 36–42
ISSN 1532-0545 (online)

Case Article

Business Value in Integrating Predictive and Prescriptive Analytics Models

David Kopcso,[a] **Dessislava Pachamanova**[a]

[a] Division of Mathematics and Science, Babson College, Wellesley, Massachusetts 02457

Contact: kopcso@babson.edu (DK); dpachamanova@babson.edu (DP)

Received: August 2, 2016
Revised: February 25, 2017; July 20, 2017; July 25, 2017; July 26, 2017
Accepted: July 26, 2017
Published Online in Articles in Advance: October 27, 2017

https://doi.org/10.1287/ited.2017.0186ca

Abstract. This article suggests ways to frame classroom discussion around the business value of models in data science, predictive analytics, and management science classes. We consider an example in which predictive analytics is used to determine the inputs to prescriptive models for customer service, and illustrate how calculations of business value enter the process of creating recommendations for business stakeholders. A review of predictive and prescriptive techniques and how they map to business problems is provided to explain the context for the exercise, and the level of analytics maturity of organizations is discussed in connection with the use of predictive and prescriptive analytics. This example presents a unified view of concepts from traditionally disparate areas of analytics, making it suitable as a capstone or an ongoing project in a data science or business analytics course.

Supplemental Material: The teaching note is available at https://www.informs.org/Publications/Subscribe/Access-Restricted-Materials.

Keywords: business value • models • predictive analytics • prescriptive analytics • classification • ranking • accuracy • confusion matrix • lift chart • ROC curve • profit curve • yield management • simulation

1. Introduction

Most of us with backgrounds in quantitative fields do not need to be convinced that using models when addressing real-world problems introduces discipline and rigor in the search for a solution. Undergraduate and graduate curricula in the quantitative disciplines trains students to build complex models, and emphasize model accuracy as a virtue. Even in business school environments where students are supposed to be trained to be "wise judges" able to "assimilate . . . synthesize . . . assess . . . formulate and test . . . and decide well" (Bruner 1998), management science or traditional statistics core courses focus on methods and tools. (See also Grossman 2001.)

Once technically trained students leave university, however, they join organizations where they work side-by-side with people who are not necessarily like-minded. The communications gap between "techies" and "business users" is well known (see, for example, Davenport and Kim 2013), and it is to an organization's advantage to narrow it. Possible ways to address this gap in a quantitative curriculum have included:

- Teaching better managerial communication of analytical work by discouraging narrative presentations that chronicle a quantitative approach to a problem, and instead reporting a set of insights and recommendations derived from analysis (Grossman et al. 2008);
- Teaching how to focus on summarizing business insights over reporting of facts, better storytelling, and evoking emotions in the audience (Glen 2011).

Additionally, a piece missing from many quantitative curricula is an explanation of how to identify the business value resulting from using models. It is not only students but also quantitative methods instructors who need training in understanding the business point of view and presenting the value of quantitative models in dollar terms. As aptly summarized by Clauss (1997, p. 35), "We found it easy to teach mathematics. A generation later, we're still finding it difficult to teach management."

At one end of the process of preparing technically minded students to be effective within companies is teaching them how to map common business problems to analytical techniques (Pachamanova 2015a). At the other end is teaching them how to prove the value of their analysis. Business value is not necessarily well defined. It is typically measured in economic terms but can also be thought of as value to various stakeholders, such as customers, employees, suppliers, and society as a whole. (See, for example, Kiron and Shockley 2011, Acito and Khatri 2014, or Ali et al. 2018.) A holistic understanding of both the monetary and nonmonetary aspects of business value is an important part of student training.

The case accompanying this article, "Managing Staffing Inefficiencies Using Analytics (B): Business Value in Integrating Predictive and Prescriptive Models," brings together concepts from predictive analytics, operations, and yield/revenue management in a single continuous application. It allows instructors to explain approaches for calculating the business value of predictive models and expand the discussion to the case of operational models in the context of applying prescriptive analytics. The link between predictive and prescriptive models is particularly important to establish, as was recently argued by Dybvig and Cokins (2017). This link is not often addressed in the literature, as predictive and prescriptive models are the focus of different analytics fields. A detailed Teaching Note with accompanying spreadsheet models is provided with the Case.

The rest of this case article is structured as follows. Section 2 explains how the topic of mapping business problems to analytic methods is covered in the case. It also reviews the concepts of analytics maturity and different types of analytics. Section 3 presents the set-up for the case to be used to illustrate the integration of predictive and prescriptive models and the estimation of business value, lists metrics and graphs for assessing the performance of predictive models, and gives pointers for discussing those in the context of operational business value. In Section 4, we share our experience teaching this case to undergraduate students, MBA students, and executives. Section 5 summarizes our observations about integrating predictive and prescriptive models and incorporating business value calculation in courses with analytics content.

2. Business Problems and Analytic Methods

Analytically trained students should be able to understand not only the methodology and implementation behind an analytical technique but also the purpose of the analytical technique and how it can be mapped to a specific business problem context. In fact, although there is an infinite number of business problems, many of them can be framed within a small number of categories of analytic methods. The process of framing itself is not necessarily straightforward, and students need practice. Appendix A in the Case sets up a potential discussion on applications of different analytical methods, and introduces important terminology to students, such as descriptive, diagnostic, predictive, and prescriptive analytics.

Table 1. Correspondence Between Common Business Questions and Analytics Techniques

Business problem	Analytical technique
Observe trends and relationships; formulate hypotheses	Data visualization
Decide if a new situation is "better" than the current situation	Hypothesis testing
Determine the impact of various factors on an output variable of interest	Linear regression, logistic regression, classification and regression trees
Identify groups of observations that are similar in terms of specified characteristics	Cluster analysis
Identify transactions that tend to occur together	Association rules (Market basket analysis)
Assign a score (likelihood, ranking) to an observation (customer, loan, transaction, etc.)	Logistic regression, classification and regression trees
Classify an observation (customer, loan, transaction, etc.) into a category	Logistic regression, classification and regression trees, Naïve Bayes, *k* Nearest Neighbors, Support Vector Machines (SVM), Neural networks
Analyze data over time; forecast	Time series analysis
Analyze text data (customer call transcripts, social media posts, etc.); perform sentiment analysis	Text analytics, Natural Language Processing (NLP)
Find an optimal mix of products given limited resources; find an optimal route/schedule/price	Optimization

Table 1, also provided as part of the background reading for students in Appendix A of the Case, summarizes some common questions and analytical methods that could be used to address them. (See also EMC Education Services 2015 as well as Kokina et al. 2017 for applications to managerial accounting and strategy maps.)

The sophistication of the analytic methods applied at an organization is one of the factors that determines the level of analytics maturity of the organization. Figure 1 in Appendix A in the Case summarizes the stages of analytics maturity based on analytics maturity models suggested in the literature. (See Davenport and Harris 2007, who suggested the first Analytics Maturity Model, or Reitter and List 2016). The first two stages of analytics maturity involve data collection as well as using data for describing and observing events. The third stage uses data and basic models for setting alerts and making observations about factors driving events. The fourth stage is predictive: Predictive models are used to forecast and identify opportunities. The fifth stage is prescriptive: Models of increasing sophistication are used to determine the best course of action for a particular situation. (For a helpful introduction, see also Ingram Micro Advisor 2017.)

The concept of analytics maturity as well as a mapping of the different stages of analytics maturity to the types of analytical methods are reviewed as part of the student reading in Appendix A in the Case. An important takeaway from this background reading is that a sign of analytics maturity for an organization is the ability to use the output of predictive models as input to prescriptive analytics models. To realize the business value from analytics models, analytics professionals at the company also need to estimate the value of using these models for various stakeholders, such as the company's customers and employees, and to adjust those models to achieve the desired value. There are many challenges with integrating predictive and prescriptive analytics effectively. (See, for example, Basu 2013 or Dybvig and Cokins 2017, who present a practitioner's perspective.) The goal of the remainder of the Case accompanying this article is to help students think through the methodological issues associated with this integration and identify the business value added by the methods used.

3. Context: Managing Staffing Inefficiencies at AdviseInvest

To illustrate the process of calculating business value, we use the case described in Pachamanova (2015b). The set-up for the case is a staffing problem that a small company, AdviseInvest, is facing because of the customer call scheduling system in place. (AdviseInvest is based on a real company, a venture-backed startup whose mission is to provide ordinary people with access to affordable financial advice.) Potential AdviseInvest customers arrive at the AdviseInvest web site and fill out a customer profile with information about their goals and background. They then schedule a call with a sales representative for a particular one-hour window. AdviseInvest has four sales reps on staff who are assigned two customers each during each particular one-hour window. The problem is that about 50% of the customers who schedule a call do not answer. Not only does this leave sales reps underutilized, it does so unevenly. For example, it is possible that one sales rep will not get to do any sales calls during that hour, whereas another will be busy the entire hour with two calls. The Director of Sales has approached a small team of analysts asking for a solution to the problem. Approaches for breaking down the problem into smaller problems that can be addressed with analytics are discussed in the teaching note for Pachamanova (2015b).

One approach to solving the problem posed by the Director of Sales is to use predictive models to identify customers who would be more likely to answer the call scheduled with the sales reps. Because of the binary nature of the variable to be predicted (Answer (1)/Not Answer (0)), classification and regression trees or logistic regression are possible choices (see Table 1). The classification method that is used is not important to the discussion in the Case. The important part is to record how the method classifies new observations. This information can then be used to assess the business value of the predictive model. As an example, a decision tree created with the `rplot` package in *R* for the data provided in Pachamanova (2015b) is shown in Figure TN.1 in the Teaching Note accompanying this Case. The tree is built using 90% of the data as a training set; 10% of all observations are set aside as a test data set. The output from classification models such as decision trees is a score, interpreted as a probability, which is then used to classify the observations based on a cutoff (threshold) value. AdviseInvest customers in the test data set are scored between 0 and 1 based on the decision tree output. A base case cutoff level of 0.50 can be used for deciding whether to assign an observation to the "Answer" (if the score is ≥ 0.50) or "Not Answer" (if the score is < 0.50) category. The data set, the *R* code for creating the tree, and the interpretation of the rules are provided in the teaching note for Pachamanova (2015b). Again, understanding those is not necessary for the Case accompanying this article.

3.1. Calculating Predictive Model Value

There are various ways to calculate the business value of the predictive model mentioned above. Business value is related to the accuracy of the predictive model, but one of the lessons that can be learned from the Case accompanying this article is that applying business value as a criterion for making decisions can lead to very different conclusions from applying model accuracy as a criterion. The case asks students to review two different approaches to using the outcomes of predictive models: classification of observations and ranking of observations. Students are also introduced to approaches for calculating the monetary benefits from applying the models. Brief tutorials on these concepts are provided as Appendices to the Case accompanying this article that can be made available to students. Specific numeric examples and tips for leading the class discussion are provided in the Teaching Note accompanying the case. (See Sections TN.3.1 and TN.3.2 of the Teaching Note.) When using classification of observations, instructors may introduce students to the concepts of classification accuracy (total accuracy, true positives, true negatives, false positives, false negatives), confusion matrices, ROC curves, and profit curves. When using ranking of observations, instructors may introduce students to concepts such as cumulative gains charts, decile-wise lift charts, and cumulative profit curves. See also Evgeniou (2016), Shmueli et al. (2016), Provost and Fawcett (2013), or Lo and Pachamanova (2015).

3.2. Beyond Predictive Models: Prescriptive Analytics and Value

The output from predictive models ultimately needs to be operationalized; how it is done should be part of the recommendations to the stakeholders in an analytics project. The next stage of the case can be used to extend the discussion to operational issues and to touch on techniques in the realm of prescriptive analytics. For example, the instructor could walk the students through various alternatives to the status quo to think about the business value of different systems for scheduling calls. Because these systems involve staff schedules and workload, instructors can talk not only about the monetary value but also about fairness and the human aspect of different options. The instructor could then discuss methods for framing the problem using results from the yield management literature as well as simulation models.

Appendix C in the Case sets up the operational discussion. Students are given information on revenues and costs and asked to think about how the current system performs according to several different metrics such as contribution to earnings, capacity utilization, and average revenue per customer. Students are then prompted to think about alternative scheduling schemes and their effect on these metrics. Section TN.4.1 in the Teaching Note suggests some ways to lead the classroom discussion.

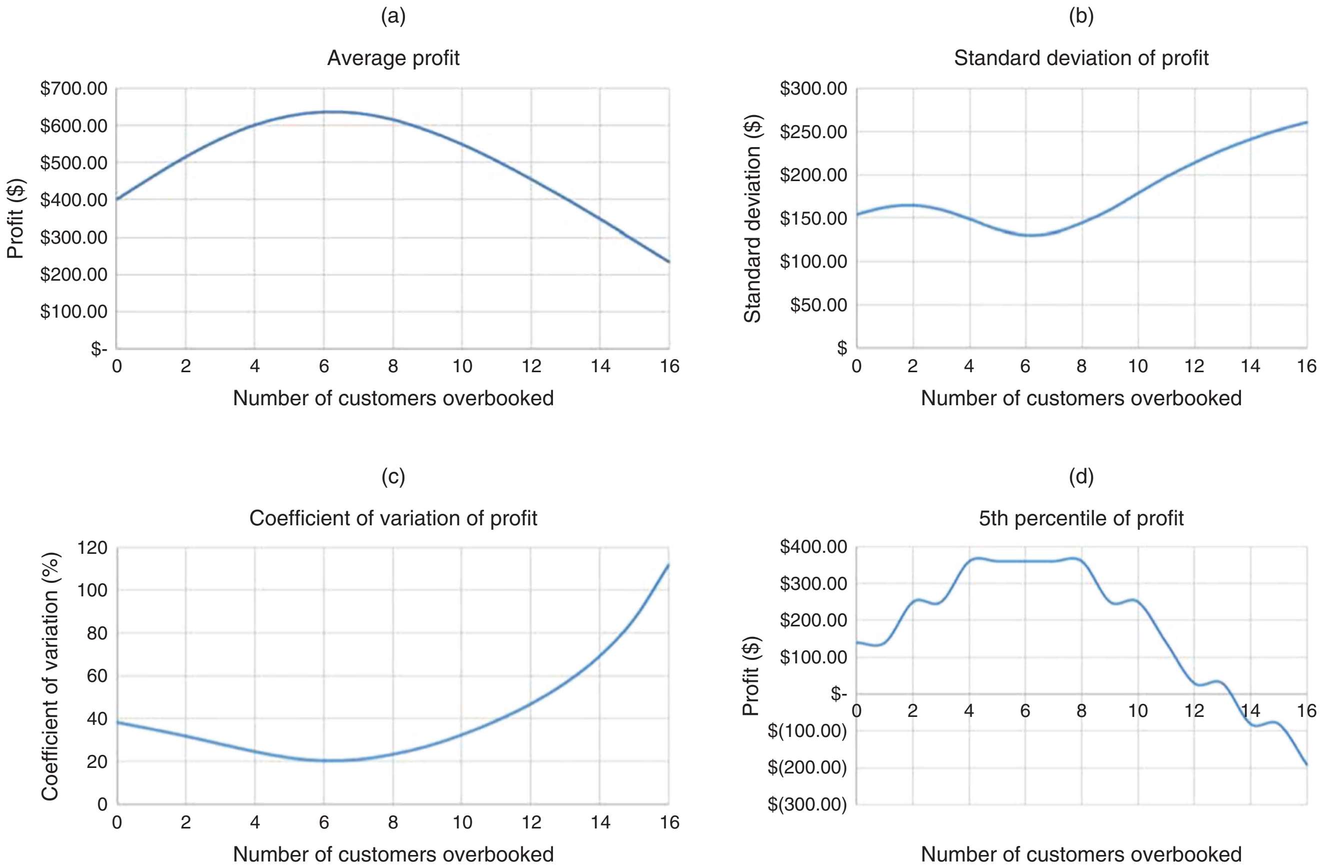
The context of the Case allows for mathematically rigorous framing of the operational problem. Specifically, the AdviseInvest call scheduling problem has similarities with the famous newsvendor problem, first mentioned in Morse and Kimball's (1951) book; see also Chen et al. (2016) for an up-to-date overview of research on this problem. While in the newsvendor problem the goal is to select the optimal number of newspapers to have on hand to maximize revenue subject to uncertain demand, in the AdviseInvest case the goal is to select the optimal number of demand-for-service requests to accept to maximize revenue for each time slot. As explained accessibly in Netessine and Shumsky (2002), there are well-known parallels between the newsvendor problem and yield management models.

The Case also allows for exploring the idea of overbooking, which often leads to a lively classroom discussion because of media coverage of overbooking in the context of airline seat reservations. Instructors could use the empirical probability distribution of number of people who did not answer the phone, shown in Figure TN.8 in the Teaching Note, to derive the optimal number of customers to overbook, Y^*, using a result from Belobaba (1989) that is also explained in Netessine and Shumsky (2002). Once the optimal number of customers to overbook has been determined, one can calculate the expected profit from overbooking. Section TN.4.2 in the Teaching Note walks the instructor through the actual calculations.

Belobaba's (1989) result is based on the assumption that customers who are signed up are interchangeable; it also assumes that all overbooked customers would answer the call. Given the individual customer information from the predictive model in the Case, these assumptions are no longer realistic. There is scant literature on integrating predictive models with prescriptive models in the revenue management literature, although interest in this area has been increasing. (See, for example, Wittman and Belobaba 2016.) We have found that spreadsheet simulation is a simple and effective tool for illustrating the effect of incorporating predictive model estimates in the scheduling process. The instructor could demonstrate the characteristics of the (variable) profit per hour, starting with a simulation in which each customer is assumed to have the same probability of answering the call, and following with a simulation in which each customer is assigned a probability of answering the call based on the customer score determined from the predictive analytics model. (There is a caveat: As mentioned earlier, the actual scores calculated from the predictive models might not be as useful when treated as probabilities. In practice, they are more useful as relative rankings of customers; hence, some calibration may be necessary.)

Section TN.4.3 in the Teaching Note provides instructions on how to implement the simulation. Spreadsheet model templates are also provided. A simple simulation template can be assigned to the students as preparation for the class if the instructor so desires; see Item 5. from the Assignments questions in Appendix TN.A in the Teaching Note accompanying this article. The file ProfitSimulation-Excel.xlsm provided with this Case provides a simulation model with no overbooking using Microsoft Excel functions only. The file ProfitSimulation-AtRisk.xlsx provides a simulation model with no overbooking using the Excel add-in @RISK, which makes it easier to track simulation output. Additional spreadsheet model files accompanying the Teaching Note allow instructors to generate output that demonstrates how a decision about the optimal number of customers to overbook can be made based on analysis of the expected hourly profit and the variability of profit. Figure 1 shows examples of graphs that can be analyzed. Note that the conclusions from the simulation model might not be the same as the conclusions from the Belobaba (1989) model, which

Figure 1. Simulation Output Helpful to Determine the Optimal Number of Customers to Overbook



Notes. (a) Expected hourly profit vs. number of overbooked customers; (b) Standard deviation of profit vs. number of overbooked customers; (c) Coefficient of variation of profit vs. number of overbooked customers; (d) 5th percentile of profit vs. number of overbooked customers.

could provide context for additional interesting class discussion on the business value of models. See Section TN.4.3 in the Teaching Note for interpretation of the simulation models output and more detail.

Including simulation as an exercise contributes to the richness of the case, and makes it an attractive option for capstone projects in analytics programs.

More sophisticated models can be used to decide on the overbooking policy and the allocation of time slots. The whole system could be dynamically updated: As customers try to schedule calls, the optimal number of customers to book during a time slot could be changed (see, for example, van Ryzin and McGill 2000 for the solution of a similar problem in dynamic airline seat protection level assignments). Given the size of AdviseInvest and its current stage of analytics maturity, however, even simple heuristics could help the company realize significant business value.

3.3. Extensions

An additional line of discussion for the case is the type of product or plan a customer is likely to buy. This factor can have a great effect on the ultimate profit realized, and would make a difference for the performance of different customer prioritization systems. AdviseInvest does keep track of the outcome of the phone call with each customer. The data are shown in Table TN.10 in the Teaching Note. Section TN.5 in the Teaching Note suggests points of discussion on how to use these data and creating a predictive model to score a customer based on the product he or she would buy rather than on the likelihood of answering the call.

4. Student Background, Classroom Use, and Student Reactions

Versions of this Case have been used in sessions of 1.5–2 hours in an undergraduate data mining course (with a total of 60 students in two sections), an MBA business analytics course (with a total of 230 students in multiple sections), in a face-to-face session in an executive education program (16 participants), and in a synchronous online session of an executive education program (15 participants). The Teaching Note provides sample teaching plans.

Feedback from students and participants has been overwhelmingly positive. In a post-Case survey, 92% of respondents (93% of undergraduate and 92% of MBA students) agreed or strongly agreed that the exercise taught them valuable skills. The participants in executive programs filled out a standard questionnaire asking whether the subject matter was useful, and 100% of respondents identified the case subject matter as "Useful" or "Very Useful."

Interestingly, the case exercise was feasible regardless of the background of the participants, and was appreciated by the MBA students and the executives (who could understand the context and the business implications better). At the point of the course where the exercise was run, the undergraduate students had been exposed to basic statistics, basic optimization models, as well as tools for data visualization, data mining and predictive analytics. The MBA students had been exposed to basic descriptive statistics and regression models, as well as a brief overview of predictive analytics models. The executives' backgrounds varied widely but they had been asked to watch a few short introductory videos explaining how predictive models, such as regression and decision trees, work, and how the output would look.

The concepts of classification and scoring are the only information that is critical to understanding the takeaways of the Case; these concepts can be explained in assigned reading or a video introduction to the session. In our experience, the prescriptive analytics discussion at the level presented in the exercise (perhaps with the exception of the simulation model) is intuitive and does not require particularly strong modeling background or preparation.

5. Concluding Remarks

The Case described in this article illustrates the link between predictive and prescriptive analytics, with a focus on the estimation of business value for the case of a small company that is trying to solve a problem with its sales operations. The exercise can be used to demonstrate the dichotomy between recommendations that would be made by prioritizing predictive method solutions based on measures of accuracy and prioritizing solutions based on business value. Depending on pedagogical goals, the exercise can include a variety of predictive and prescriptive analytics tools, e.g., machine learning and statistical estimation models, yield management models, and simulation. However, the Case provides enough flexibility to omit some of these techniques and still facilitate a rich discussion about the business value of applying these models. The exercise can be expanded to include additional discussions on the "right" questions to ask, the "right" models to use, and the "right" way to summarize the information for various stakeholders in an analytics project (see, for example, the teaching note in Pachamanova 2015b).

Acknowledgments

We are grateful to the editor, the associate editor, and two anonymous referees for their helpful comments and suggestions.

References

Acito F, Khatri V (2014) Business analytics: Why now and what next? *Bus. Horizons* 57(5):565–570.

Ali A, Mancha R, Pachamanova D (2018) Correcting analytics maturity myopia. *Bus. Horizons* 61(2):211–219.

Basu A (2013) Five pillars of prescriptive analytics success. *Analytics Magazine*, INFORMS, March/April.

Belobaba P (1989) Application of a probabilistic decision model to airline seat inventory control. *Oper. Res.* 37(2):183–197.

Bruner RF (1998) Vision of the MBA program. Accessed October 3, 2017, https://faculty.darden.virginia.edu/brunerb/oldsite2004/Vision_MBA.htm.

Clauss FJ (1997) The trouble with optimal. *OR/MS Today* 24(1): 32–35.

Chen RR, Cheng TCE, Choi TM, Wang Y (2016) Novel advances in applications of the newsvendor model. *Decision Sci.* 47(1): 8–10.

Davenport T, Harris J (2007) *Competing on Analytics* (Harvard Business Review Press).

Davenport T, Kim J (2013) *Keeping Up with the Quants: Your Guide to Understanding and Using Analytics* (Harvard Business Review Press, Boston).

Dybvig A, Cokins G (2017) Enterprise optimization: A new application. *ORMS Today* 44(3):28–32.

EMC Education Services (2015) *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data* (John Wiley & Sons, Hoboken, NJ).

Evgeniou T (2016) Classification methods. Retrieved July 14, 2016 from http://inseaddataanalytics.github.io/INSEADAnalytics/Report_s67.html.

Glen P (2011) "Nothing but the facts" approach just won't work with business people. Accessed May 27, 2016, http://leadinggeeks.com/nothing-but-the-facts-approach-just-wont-work-with-business-people/.

Grossman TA (2001) Causes of the decline of the business school management science course. *INFORMS Trans. Ed.* 1(2):51–61. http://dx.doi.org/10.1287/ited.1.2.51.

Grossman TA, Norback JS, Hardin JR, Forehand GA (2008) Managerial communication of analytical work. *INFORMS Trans. Ed.* 8(3):125–138.

Ingram Micro Advisor (2017) Four types of big data analytics and examples of their use. Accessed June 14, 2017, http://www.ingrammicroadvisor.com/data-center/four-types-of-big-data-analytics-and-examples-of-their-use.

Kiron D, Shockley R (2011) Creating business value with analytics. *Sloan Management Rev.* (September 15), http://sloanreview.mit.edu/article/creating-business-value-with-analytics/.

Kokina J, Pachamanova D, Corbett A (2017) The role of data visualization and analytics in performance management: Guiding entrepreneurial growth decisions. *J. Accounting Ed.* 38:50–62.

Lo VSY, Pachamanova D (2015) From predictive uplift modeling to prescriptive uplift analytics: A practical approach to treatment optimization while accounting for estimation risk. *J. Marketing Analytics* 3(2):79–95.

Morse PM, Kimball GE (1951) *Methods of Operations Research* (The Technical Press/John Wiley & Sons, New York).

Netessine S, Shumsky R (2002) Introduction to the theory and practice of yield management. *INFORMS Trans. Ed.* 3(1):34–44.

Pachamanova D (2015a) Mapping business problems to analytics solutions: Surrogate experiential learning in an MBA introductory data science and business analytics course. *INFORMS Trans. Ed.* 16(1):15–22.

Pachamanova D (2015b) Managing staffing inefficiencies using analytics. *INFORMS Trans. Ed.* 16(1):23.

Provost F, Fawcett T (2013) *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking* (O'Reilly Media, Sebastopol, CA).

Reitter N, List B (2016) INFORMS analytics maturity model. Accessed July 18, 2016, http://analytics-magazine.org/informs-analytics-maturity-model/.

Shmueli G, Bruce P, Patel N (2016) *Data Mining for Business Analytics: Concepts, Techniques, and Applications with XLMiner*, 3rd ed. (John Wiley & Sons, Hoboken, NJ).

Van Ryzin G, McGill J (2000) Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Management Sci.* 46(6):760–775.

Wittman MD, Belobaba PP (2016) Personalization in airline revenue management—Heuristics for real-time adjustment of availability and fares. *J. Revenue and Pricing Management* 16(4):376–396.