Editorial to the Special Issue on Profit-Driven Analytics

Bart Baesens, Wouter Verbeke, Cristián Bravo

Businesses are gathering an unprecedented amount of data to gain deeper insights into customer behavior and markets with the bottom line in mind. Popular analytical applications are: churn prediction, response modeling, credit risk modeling, sales forecasting and anomaly detection. A number of analytical techniques have been developed to address such problems, where the focus has typically been on algorithmic complexity, statistical significance or detection power. However, to be successful from a business standpoint, analytical models should do much more, namely, add business value or profit, provide interpretability, enhance operational efficiency, and keep business compliant in following correct practices. Far too often these criteria are forgotten with research papers focusing too much on the development of new analytical techniques which are typically only evaluated in terms of classification accuracy, lift, area under the receiver operating characteristic curve, and/or R-squared on a limited set of small-scale data sets (most often less than five data sets with on average only a few thousands observations). Although these studies and measures definitely have their merits, we believe it is time to move on and start focusing on what really matters for analytics to yield a true competitive advantage: profit and/or business value!

The objective of this special issue was to publish high-quality papers that address the added value of an analytical model from a profit perspective. The issue focusses on methods, measurement, and practices that demonstrate profit or business value. The editors had a challenging job in selecting the most valuable papers. The submissions and the variety of application fields has been striking, but we believe that the very variety of submissions helped lead to a final set of papers that readers will enjoy and find valuable. In what follows, we provide a short overview.

In their paper, FraudBuster: Reducing Fraud in an Auto Insurance Market, Nagrecha, Reid and Chawla illustrate that state of the art fraud detection performs poorly when deployed at underwriting. A new framework FraudBuster is introduced which aims at predicting fraud at underwriting focusing both on interpretability and compliance. Their method allows to successfully identify drivers who are unprofitable and likely to be fraudulent risks. By merging the conventional likelihood of fraud with the profitability driven Loss Ratio, they succeeding at identifying segments that are demonstrably worst-affected by fraud. Their predictive models can successfully identify fraud with an area under the Precision-Recall curve of 0.63 and F1 score of 0.769.

Devriendt et al. provide in their article “A literature survey and experimental evaluation of the state-of-the-art in uplift modeling: a stepping stone towards the development of prescriptive analytics” an in-depth review and classification framework of the various uplift modeling techniques that have been introduced in the literature over the past two decades. Uplift modeling, which is situated at the core of prescriptive analytics, presents a significant potential to further increase revenues and profits generated by analytical models, by facilitating the next step in the evolution from descriptive over predictive to prescriptive analytical models. Additionally, the authors present the results of a broad experimental study in which they evaluate various uplift modeling approaches on four real-life data sets. Together, the presented classification framework and the findings from both the literature survey and the experimental study, provide a deeper insight in the value and practice of uplift modeling and a broad overview of the field, as well as an agenda for future research to further improve and extend the capabilities of the discussed approaches. The experimental framework will be published by the authors to facilitate reproduction and validation of the empirical results that are presented in this study. As such, this article is considered as a convergence point and catalyst, spurring further development of the field and practical applications in the industry.

Ting et al. study the relationship between geospatial variables and retail sales. Using the information from over 400,000 places of interest, combined with census and other demographic data, they construct a detailed profile of the attractiveness of existing 96 stores with respect to other businesses located around them. This dataset is then used to construct a predictive model for the sales of new stores, studying which geospatial factors are significant to construct this prediction. The results of the paper suggest that, for their market, the population in the area is not as relevant as the purchasing power of the potential consumers living within the area of influence of the store. To clearly determine this area, the authors suggest that a mix of different distance measures provide a much better profile than using just Euclidian distance. The results from this paper, contributing to the emerging interest in business applications of geospatial analytics, enrich our toolbox: When dealing with problems where our customers are spread across a geographical area, we should ask ourselves if this information should be included explicitly, using the powerful methodologies of geospatial analytics.

Óskarsdóttir et al. extend the earlier developed expected maximum profit (EMP) measure to take into account the variability in the lifetime values of customers, thereby basing it on individual characteristics. They demonstrate how to incorporate the heterogeneity of customer lifetime values when customer lifetime values are known, when their prior distribution is known, and when neither is known. By considering individual customer lifetime values, their method of measuring model performance gives novel insights when deciding on a customer retention campaign.

In terms of future research, we expect two things. First, we anticipate business relevant criteria such as profit to be more frequently used as ex-post evaluation criteria for the performance of analytical models. Interesting insights will emerge on any possible discrepancies between traditional statistical evaluation measures and profit. Building upon these insights, we also anticipate the development of new analytical techniques that directly embed profit in the analytical model building or estimation step rather than optimizing a traditional likelihood function or statistical error. These studies may further pinpoint interesting differences between statistical and profit optimization. We believe that by more directly embedding relevant business perspectives such as profit into the model evaluation and building, we will be able to more clearly demonstrate the added value of analytics from a business perspective.

Editing a special issue is a combination of joy and grief for authors, reviewers, and associate editors. We believe that the special issue represents a nice selection of the current state-of-the-art thinking in the realm of profit driven analytics. The issue's contents cover a wide range of topical areas and methods, suggesting that the issue might be useful in Ph.D. seminars seeking to provide a broad and well-rounded perspective on the topic.

To conclude, we would like to end by thanking Vasant Dhar, Editor-in-Chief of the Big Data journal, Craig Ryan and Sophie Mohin for giving us this opportunity, as well as the authors, reviewers, and associate editors, without whose valuable contributions, the special issue would not have seen the light of day!