

Editorial

Dear Colleague:

Welcome to volume 19 supplementary issue of the Intelligent Data Analysis (IDA) Journal.

This special issue consists of ten articles, some of which have been selected from presentations at the conference BAFI 2014, which was held in Santiago, Chile in January 2014.

The following five papers present different applications of intelligent data analysis tools to the area of finance.

The paper by Stecking and Schebesch, “Classification of Credit Scoring Data with Privacy Constraints” deals with the pressing matter of finding a compromise between ensuring the privacy of financial records and sustaining a significant discrimination capacity of models which utilize the information contained in these records. The authors propose a micro-aggregation procedure to find a set of representative cases that are fabricated from the data, incorporating clustering methodologies, and then training the models over these prototype objects.

Kamaruddin *et al.* apply text mining tools to detect sentences that deviate from the typical jargon found in the financial statements of Malaysian banks. The tool is aimed at regulators, government agencies, and investors that seek to quickly analyze a large and complex corpus, and detect potentially useful information quickly and efficiently.

Avanzini and Jara consider the complex and, given the events of 2008, timely issue of identifying systemic risk, using the Chilean financial market as a case study. They employ data reduction techniques and Principal Component Analysis. The paper shows how data analytics tools can assist – or should assist – in financial modeling and in the practice of financial risk management. These two very active research areas have been coming together rapidly, and it is to be expected that applications such as this one will become increasingly useful in the future.

Staines and Barber show in their paper “Topic Factor Models: Uncovering Thematic Structure in Equity Market Data” how Topic Factor Modeling (TFM) can be employed to find thematic structure in a data corpus comprising text and time series. Several examples using equity market data underline the potential of their work to uncover shared topics in both text and returns data.

The work “How Much Effort Should Be Spent to Detect Fraudulent Applications When Engaged in Classifier-Based Lending?” by Chong *et al.* aims at detecting false information in loan applications, a relevant issue in credit scoring. The authors demonstrate that it can be profitable for banks and other institutions to spend resources to identify possible liars in loan applications, and under certain assumptions there is an optimal effort that can be assessed mathematically. The following two papers show methodological developments relevant for Business Analytics.

Lingras and Haider propose a clustering approach employing an ensemble method combined with Rough Sets for uncertainty modeling. Daily price patterns of commodities are used to show the potential of this new approach. A subsection deals with the robustness of the obtained cluster solution while the comparison with alternative cluster techniques hints at important advantages of the newly proposed method.

The work by López and Maldonado studies two important methodological aspects in business analytics and statistical learning, namely multi-class learning and attribute selection. An adequate identification

of the relevant variables may improve predictive performance while enhancing the understanding of the underlying process that generated the data. The authors propose novel multi-class SVM classifiers based on second-order cone programming. These new formulations provide a robust and efficient framework for classification, leading to superior performance in high-dimensional data sets.

The following three papers deal with different kinds of applications to diverse areas of Business Analytics.

The work by Maldonado provides an empirical comparison of various Support Vector Machine (SVM) approaches for the important business analytics application of churn prediction. The author explores the four main strategies identified by (He and García, 2009) to deal with the class-imbalance problem that arise: data resampling, cost-sensitive learning via Cost-Sensitive SVM, feature selection, and one-class Learning via Support Vector Data Description (SVDD). Experiments on four customer churn prediction datasets show the advantages of SVDD in terms of performance compared to the traditional SVM formulation.

Pinto *et al.* develop models to optimize space allocation in the retail industry. The authors combine regression models for sales forecasting with metaheuristics that search different hypotheses of space allocations for multiple product categories. For each such hypothesis the regression models provide sales forecasts based on the assigned space.

Seret *et al.* show how advanced segmentation allows the discovery of new knowledge about online banking customers. They combine an unsupervised quantitative technique with domain knowledge to obtain novel information from a customer base. An extensive description and discussion of the new insights illustrates the complementarity of the two approaches.

We feel that this special issue contributes to move forward the frontier of existing knowledge regarding business analytics both in terms of developing new techniques and via underlining the applicability of existing techniques to new areas of this important subject.

Our special thanks go to Fazel Famili, Editor-in-chief of the IDA journal who accepted this special issue and motivated us during the review process. Support from the Chilean “Instituto Sistemas Complejos de Ingeniería” (ICM: P-05-004-F, CONICYT: FB016; www.isci.cl) and the Center for Mathematical Modeling (<http://www.cmm.uchile.cl/>), is gratefully acknowledged. Last but not least, we would like to thank all our reviewers who helped us to ensure the acceptance of high-quality papers.

Dr. Cristián Bravo, Universidad de Talca, Chile

Dr. Matt Davison, Western University, Canada

Dr. Alejandro Jofré, Center for Mathematical Modeling, Chile

Dr. Sebastián Maldonado, Universidad de los Andes, Chile

Dr. Richard Weber, Universidad de Chile, Chile

Guest Editors