# Big Data And Analytics:
# Concepts, Techniques, Methods, and Applications

Stephen H. Kaisler, D.Sc.
SHK & Associates
Laurel, MD
skaisler1@comcast.net

Frank J. Armour, Ph.D.
Kogod School of Business
American University
Washington, DC
farmour@american.edu

Alberto J. Espinosa, Ph.D.
Kogod School of Business
American University
Washington, DC
alberto@american.edu

## Abstract

The Big Data minitrack features a number of papers addressing methods and techniques, issues and challenges, and organizational approaches to processing and managing Big Data within an organizational environment. This year three papers focus on the financial domain, one on competitive sports, one on mobile advertising, and one paper on a framework for Big Data coordination and governance. These papers contribute to a growing body of practical technical techniques for using a variety of analytical methods to process Big Data in a business operations environment. And, they contribute to a framework for managing Big Data activities within a business environment.

## Introduction

Our first paper by T. Wang, H-L. Lin and J-C. Yun addresses the impact of investor-related attention to companies' disclosures on Facebook and its influence on financial analysts following the firm. They analyzed a sample of S&P 500 companies with a Facebook presence. 566 pieces of information including earnings, profit, and sales were disclosed. They correlated this information with the number of financial analysts following the firm. A positive correlation was found on the number of analysts, but a negative correlation was found regarding individual investor's holdings in the firm. This is one of the first studies to examine the influence of Facebook-posted data on how analysts regard firms and how investors use/do not use such data.

The second paper by J-O. Jandl uses analytics to examine the stock market movement of United Kingdom Real Estate Investment Trusts (REITs). He extracted news sentiment as a proxy for information supply from newspaper articles an online search behavior as an indicator of information demand. Dictionary-based and machine learning techniques were used to perform sentiment analysis on the news articles. However, these data are only a partial indicator. He also used Google Trends data, the Search volume Index (SVI) to analyze investor attention to stocks as an indicator of information demand. He found that significant changes in news sentiment about a REIT was reflected in a significant change in SVI data.

Our third paper by A. Ratku, S. Feuerriegel, and D. Neumann how changes in financial news regarding firms is reflected in changes in stock prices. Because public companies are required to disclose any information that can influence their stock prices, corporate press releases are a good source for predicting stock market returns. They analyzed forty different topics on ad hoc announcements from the German stock market using Latent Dirichlet Allocation. They focused on identifying key topics and assessing which if these topics affected abnormal stock returns. Of the forty topics, drug testing seemed to be the strongest positive influence whereas corporate development seemed to be the largest negative influence. Several topics had minimal or no influence.

These three papers represent different approaches to predicting stock prices and, thereby, financial market movement, using public knowledge about a firm's business activities. The mix of techniques shows that future predictive and prescriptive approaches may not (should not? ) rely on a single technique but on an amalgam of techniques that cross-correlate and reinforce their respective results.

The fourth paper by O. Caya and A. Bourdon applies analytic techniques to competitive sports data. Such data has been analyzed before, notably in the book *Moneyball*.

But, as the authors note, it has not been the subject of much academic research. The authors motivate the need for and describe a framework for using business intelligence and analytics to support analysis of competitive sports data. Numerous statistical techniques have been applied to several major sports such as baseball, soccer, and basketball. Their framework is focused on how different techniques contribute to providing value at the individual, organizational, and institutional levels. In effect, they seek to move from sports analytics to sports intelligence with the intent on improving both the organization and the competitive advantage.

The fifth paper by D. Kridel and D. Dolk uses an innovative approach – lattice and device histories - to target customers using mobile advertising campaigns. Because customers on the move, such campaigns have to operate in real-time with adaptive model management to dynamically target customers as they approach locations of interest (such as prior behavior). The authors use a lattice approach to tile a geographic area. They then assign households, points of interest, and demographics to those cells. Using device histories, they map customer behavior to cells. Using several techniques, they compute customer propensity models based on cell data to score a user as a potential target for one or more advertisements that can be dynamically sent to the mobile device. This approach uses customer profiling on an individual device rather than a customer segment, so is more specific in its targeting and less annoying to users who might have no interest in a particular advertisement.

The last paper, by F. Armour and A. Espinosa, develops a research framework for coordination and governance of Big Data activities. This aspect of Big Data has not been addressed heretofore as the focus has been on concepts and techniques and case studies. The authors, using coordination theory as the basis for their framework, because Big Data Analytics (BDA) is a multi-disciplinary, team-based activity. This framework provides the basis a future research using qualitative studies to elicit a deeper understanding of how BDA is treated within an organization as an essential function of business operations.