Hawai'i International Conference on System Sciences 55
Decision Analytics and Service Science
Big Data and Analytics: Pathways to Maturity Minitrack

Introduction to Minitrack

Stephen H. Kaisler, D.Sc. SHK & Associates Laurel, MD 20723 Skaisler1@comcast.net (301) 498-4244	Frank J. Armour, Ph.D. Kogod School of Business American University Washington, DC 20016 fjarmour@gmail.com (202) 251-3554	J. Alberto Espinosa, Ph.D. Kogod School of Business American University Washington, DC 20016 alberto@american.edu (202) 885-1958

Abstract

The Big Data and Analytics minitrack ("the Original Big Data and Analytics Minitrack") enters its eight year with five papers that cover a variety of interesting topics.

Introduction

The first paper, *The Risk Management Process for Data Science: Gaps in Current Practices*, by Sucheta Liri and Jeffrey Saltz, focuses on risk management in data science, a topic that has been neglected until recently. The authors explore potential risks and risk management processes through a focused review of existing literature. They review risk management frameworks that have been used in other disciplines such as construction, and financial and regulatory environments where managing risk is critical to success. From this analysis, they draw insights that suggest how risk should be managed in a data science project. Using these insights, the authors formulated a set of questions about risk management processes which they used to conduct interviews with representatives from 16 private and public organizations. Analysis of the interviews generated 16 findings that represented the dominant themes mentioned by the interviewees, which yielded six common actions across the organizations that affected identifying risks and minimizing risks. Further analysis of these actions showed that in many cases – outside financial organizations – that risk management was an ad hoc process, in part due to a lack of a framework to apply within the data science project. Further, in many data science projects, risk was focused on technical or project cost factors. Several limiting factors were discussed, such as only male participants. Their conclusions included that risk needs to be considered from the beginning of a data science project and that a standard framework and associated methodology needs to be developed to assist data scientists in managing risk.

The second paper, *A Novel Population Analysis Approach for Analyzing Financial Markets under Crises – 2008 Economic crash and Covid-19 pandemic*, by Zahra Hatami, Prasad Chetti, Hesham Ali, and David Volkman, explored the impact of the COVIUD-19 pandemic on the financial markets. Their goal was to determine how crises affect stock portfolios and the behavior of stocks over time. They developed the population analysis approach to compare the behavior of groups of companies within the stock market and applied it to the historical data from the 2008 economic crash and the pandemic-influenced crash of the early months of 2020. Population analysis examines the behavior of individual elements as they compare to a group of peers. They developed

URI: https://hdl.handle.net/10125/79474
978-0-9981331-5-7

correlation networks to compare the behavior of different economic sectors to identify similarities and differences between the reactions of these sectors to decisions based on the characteristics of the crises. The authors used 48 months of data for the 2008 economic crash and eight months of data for the 2020 crash. The financial sector suffered significantly in both crises, but government intervention helped to mitigate the effects over time and to initiate a recovery. The utility and energy sections also suffered, but for different reasons related to problems in the financial sector – the utility sector in the 2008 crisis and the energy sector in the 2020 crisis. These differences were realized from the use of the correlation network which indicates that is a potentially powerful model for examining future crises as these unfold.

The third paper, *Deep Learning Strategies for Industrial Surface Defect Detection Systems*, by Dominik Martin, Simon Heinzel, Johannes Kunze von Bischhoffshausen, and Niklas Kühl, applies deep learning techniques to image processing of structures to locate and identify defects. The authors noted three factors affect the use of deep learning in industrial surface defect analysis: cost of data generation, the small size of the data sets, and the rarity of surface defects. This paper identifies challenges to applying deep learning to detecting surface defects and strategies to overcome them. A case study is presented that discusses an approach to resolving some of these challenges. The authors interviewed several experts to identify the challenges and used the results to generate design requirements for an artifact as a solution to the problem. From these requirements, they developed a set of design principles features for an artifact, which - as presented in the case study – is a deep learning model. The authors generated four design requirements which led to five design principles and yielded 12 design features. Eight experiments were carried out using different techniques and different knowledge transfer strategies. The authors conclude that industrial knowledge transfer offered only slight improvement over general knowledge transfer for the experiments that yielded the best results. The authors point out several opportunities for improving on their work, but the need for large, labeled data sets is essential to this effort.

The fourth paper, *Application of the Technology Acceptance Model to an Intelligent Cost Estimation System: An Empirical Study in the Automotive Industry*, by Frank Boedendorf and Joerg Franke, note that cost estimation models based on machine learning and deep learning have not been widely accepted within industrial organizations. They studied this problem at a large automotive manufacturer through 50 questionnaires and participant observation through surveys of experts who regularly used an implemented deep learning-based cost estimation system. They identified opportunities for further improvement using the Technology Acceptance Model. The authors identified 15 hypotheses based on five factors to explain the use of the implemented system. They developed a structural equation model to analyze the data extracted from their survey. Of the 15 hypotheses, nine were confirmed by the data they collected. Their conclusion was that the intelligent cost estimation system was not fully accepted due to insufficiently perceived usefulness and low usability. Usefulness was based on the lack of understanding of how to use the system and how its results would help them in their work. This led to low usability of the system because of perceptions about it. The authors suggested that mitigation of these issues could be achieved by having the system better explain what it did and how it achieved its results, and making it more transparent, e.g., inspectable by potential users.

The Last paper, Applying Forensic Analysis Factors to Construct a Systems Dynamics Model for Failed Software Projects, by Stephen Kaisler, William Money, and Stephen Cohen, revised and extended a forensic analysis framework that had been developed in earlier research. As is known, IT projects fail for many reasons, but how they fail had not been examined in depth in many cases. The authors believed that using forensic analysis methods on a continuing basis from project inception could shed light on how IT projects fail. A literature survey of over 60 papers identified 14 causal categories that could lead to an IT project failing or failed. A brief case study based on the experience of one of the authors (Cohen) showed that in many cases potential failure could be mitigated and project recovery could be accomplished – but not necessarily to the original project goals and objectives. The authors developed an initial framework and proposed building a Systems Dynamics model to explore the relationships among factors leading to potential failure based on capturing fata during project execution. This model will provide an understanding of how different factors interact and provide an assessment of potential failure as a figure of merit based on three primary factors: cost, schedule, and functionality. The model will also be able to support "what-if" analysis to determine under what conditions failure is forced by changes in critical project elements. The authors suggest that the model can be expanded to include elements for assessing recovery mechanisms that can mitigate the potential for failure.

Each of these papers has proposed different modeling techniques for assessing data that could lead to understanding of new approaches to solving problems via data science. The key takeaway from the papers presented here is that there is a rich suite of analytics that can be applied in many domains and problem situations.