

Comparative Analysis of Big Data Analytics and BI Projects

Gloria J. Miller
SKEMA Business School
Lille, France

Email: gloriajean.gloria@skema.edu

□

Abstract—Decision support systems such as big data, business intelligence (BI), and analytics offer firms capabilities to generate new revenue sources, increase productivity and outputs, and improve competitiveness. However, the field is crowded with terminology that makes it difficult to establish reasonable project scopes and to staff and manage projects. This study clarifies the terminology around the data science, computation social science, big data, business intelligence, and analytics and describes their meaning relative to decision support projects. For BI and big data projects, it identifies the critical success factors, empirically classifies the project scopes, and investigates the similarities and differences between the project types. This comparative analysis provides unique insights into the factors and criteria that influence BI and big data project success. These results should inform project sponsors and project managers of the contingency factors to consider when preparing project charters and plans.

Index terms: Big data analytics, data science, business intelligence (BI), project success factors

I. INTRODUCTION

DATA Science and computational social science are emerging interdisciplinary fields that overlap in content with big data, business intelligence (BI), and analytics. As more data have become available on the internet, social media, and other source organizations have begun to collect it in growing volumes, new business models and algorithms are emerging, and data sales have become potential revenue sources [3]. Despite the increased attention to big data, the critical success factors for decision support projects have received little attention in the project management literature. Decision support projects are implementation projects that deliver data, analytical models, analytical competence, or all three, for unstructured decision-making and problem-solving. They include subspecialties such as big data, advanced analytics, business intelligence, or artificial intelligence. Without insights into the project's critical success factors, it can be challenging for project sponsors and managers to establish a reasonable project strategy and to achieve the desired benefits efficiently.

The emerging status of these fields means that terminology is not standardized. Scant research exists about the application and scientific and commercial implications of these domains to project management. There is little understanding of the impacts of those differences in transitioning organizations in order to benefit from those scientific areas and data platforms. This paper empirically investigates the critical success factors

for decision support projects and provides a comparative analysis of big data and BI projects. This comparative analysis provides unique and previously unpublished results on the structural factors that contribute to decision support project success. The findings of this study add to technology and project management practices. In particular, it provides in-depth insights into what factors influence big data and BI project success. It provides information on the similarities and differences with regard to the criteria for measuring success. These results should inform project sponsors and project managers of the contingency factors to consider when preparing project charters and plans.

II. LITERATURE REVIEW

A. Business Intelligence

Business intelligence is used to refer to technology, processes, and software used to transform raw data into intelligence for computer-aided decision-making. The BI process includes the collection, evaluation, analysis, and storage of data and the production and dissemination of intelligence [6]. Davenport and Harris [7, p. 7] define analytics as a sub-category of BI that includes “*the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions.*” Otherwise, the term analytics is not standardized in literature.

B. Big Data

As data volumes have grown and computing requirements have increased, technologies and tools used to manage, manipulate, and understand data have evolved [8, 9]. On the one hand, there is a consensus that “Big Data” refers to the attributes of data—velocity, variety, validity, and volume—and on the other hand, it refers to innovative technologies and processes that allow for the use of data in novel ways [3].

C. Data Science

Data science is an interdisciplinary field that includes data analysis, statistics, data mining, and models; it has the goal to transform data into knowledge by finding patterns and trends in the data [8, 10]. The terminology originated as a role description for a single person who could act as a Business

¹ This work forms part of a research project on “Decision Support Project: Project Success and Organizational Performance” conducted by the author.

Analyst, Statistician, Engineer, or Research Scientist [8]. Davenport and Patil [10] popularized the data scientist's role in their "Harvard Business Review" article; they described it as "a high-ranking professional with the training and curiosity to make discoveries in the world of big data" [10, p. 72].

D. Computational Social Science

Computational social science refers to revealing patterns related to group and individual behavior. It is based on the emergence of scientific research methods that can leverage big datasets [11]. Computational intelligence refers to approaches where algorithms intimate human information processes and reasoning [12].

E. Project Success Factors

Projects are a form of temporary organization for introducing changes and transitions into organizations. The transition is the change or transformation expected as a consequence of the temporary organization's tasks. Project success refers to the project delivering its expected output and achieving its intended objective [13, 14]. Success criteria and success factors are the main elements of success. Success criteria are used to judge the outcome of the project, and success factors influence the likelihood of achieving a successful outcome. That is, the success criteria determine measures or indicators for success, while the success factors refer to the circumstances, conditions, and events that support the project in achieving its objectives [13, 14].

F. Summary

There is significant overlap in the implementation processes and technologies used between big data, analytics, BI, decision science, and computational social science. We can summarize the terminology as follows for further use in this paper. BI, big data, and computational social science have similar technical processes and techniques but differ in use cases. BI focuses on the platforms, architectures, and tools for the provision of data and intelligence and has an enterprise orientation. Big data seeks use cases to monetize data directly through its sale or indirectly through data-driven business models or algorithms. Computational social science has a scientific research focus. Its use case focuses on leveraging big datasets for learning. Data science refers to the responsibilities of the people who use BI, big data, or analytical techniques; it describes a role or job.

III. RESEARCH METHODOLOGY

The research used a web-based survey with quantitative methods to collect data on decision support projects and to explore the difference between BI and big data analytics' project characteristics. The measurement items and composite variables are based on a review of the literature and quantitative analysis conducted by [15]. The survey collected the data over a ten-week period (September 2017 to December 2017) from a single informant. Project managers, team members, and sponsors from completed decision support projects were asked to take the survey. The responses were checked for scope, completeness, consistency, ambiguity, missing data, extreme responses, outliers, and

leverage. Validity checks for common method bias, response bias, and reliability were conducted. No bias was found, and the data were considered to be reliable and valid.

The survey sample was comprised of 82 usable responses as follows: 76% of the respondents have a master's degree or higher; 38% perform IT roles; 18% are from a project management office; 48% are project or program managers; 5% agile coaches; 22% project team members; and 5% are project sponsors. The organizations sponsoring the projects were mostly publicly traded (51%) and were large, with more than 249 employees (83%) and US\$50 million in revenue (78%). They are spread throughout 22 different industries and 24 countries. The majority of the participants were from Europe (74%).

Latent Class Analysis (LCA) was used to classify the types of projects. Descriptive statistics, mean ranking, Wilcoxon score, and correlation analysis were used to explore the characteristics, establish the validity and reliability, and explain the relationship between the variables.

IV. CONSTRUCT OPERATIONALIZE

Project critical success areas based on [1] and decision support success systems factors were used to formulate the measurement instrument. Table I includes an abbreviated description of the measurement items. Most of the items are based on a five-point Likert scale (1 = Not at all; 5 = To a great extent). To ensure completeness of data, "Don't know" or "Not Applicable" (N/A) was added to the Likert scale. Items that used different scales are described in the relevant sections.

A. Project Mission

The project mission represents the clarity of goals and directions [1]. The following deliverables proposed by [7] as valuable components of analytics projects were used in the analysis to define the project strategy and classify the project types: *Proprietary algorithms* or business models, *new data* that was not previously available in the company, deliverables embedded into *distinctive business processes*, and data science or *analytic competence*. The measurement instrument did not directly ask a question on the project's objectives. Thus, the organizational impacts of the project were used as proxies for defining the business strategy and vision. The impacts include the *cost* and effort-saving or increases in productivity, increases in *revenue*, and *strategic benefits* such as providing new and reusable learning or improving forecast and prediction accuracy.

B. Top Management Support

Top management support is needed to authorize the project and to provide the resources and authority to execute the project [1]. User contribution is divided into user participation and user involvement [17]. *Participation* represents an active role in the development process and *involvement* represents the importance and personal relevance an individual places on the system or project. Top management support was measured based on top management and senior management involvement.

TABLE I.
MEAN RANKING AND KRUSKAL-WALLIS TEST OF CRITICAL SUCCESS FACTORS (N = 82)

Critical Success Factor			Big Data Analytics		Bus Intelligence		Kruskal Wallis	
Project [1]	BI & big data	Measurement Item	Mean	Rank	Mean	Rank	H	p-
Project Mission	BI/BD Strategy [2], [4], [5]	New Data	3.38	29	3.30	21	0.19	0.67
		Distinctive Business Processes	4.1	8.5	3.25	22	15.2	0
		Proprietary Algorithms	3.52	26.5	2.90	25	5.1	0.02
		Analytic Competence	3.67	22.5	2.67	29	11.45	0
	Business Strategy / Vision [2], [4], [5], [9]	Cost Performance	3.46	28	3.56	11	0.02	0.90
		Revenue Performance	3.67	22.5	3.62	9.5	0.07	0.80
Strategic Benefits		3.89	17	3.44	14	2.48	0.12	
Top Mgt Support	Top Mgt Support [5], [9]	Sr Mgr Involvement	4.05	11	3.35	16.5	2.86	0.09
		Top Mgt Involvement	3.99	14	3.31	19.5	3.32	0.07
Project Schedule/Plans	Approach [4], [9], [16],[5],	Technological Uncertainty	2.14	39	2.28	32	0.36	0.55
		Pace	2.48	34	2.10	35	1.75	0.19
		Complexity	2.33	35	1.92	38	4.45	0.03
		Product Novelty	2.52	32	1.69	39	12.15	0
Client Consultation	Mgt Participation [7]	Sr Mgt Prj Directing	3.60	24	2.78	28	5.27	0.02
		Top Mgt Prj Steering	2.57	31	2.06	36	2.57	0.11
	User Participation [2]	Business User Participation	3.94	16	3.08	24	7	0.01
		Business User Prj Acceptance	2.50	33	2.23	33	0.67	0.41
Personnel	Specialized Skills [7], [10], [2], [16],[5], [19]	Analytical Competence	3.71	20.5	2.84	26	4.77	0.03
		Business Competence	4.05	11	3.62	9.5	1.69	0.19
		Data Competence	3.95	15	3.31	19.5	4.85	0.03
		Data Scientist in Team	4.10	8.5	2.79	27	10.65	0
		Technical Competence	4.19	3.5	3.77	2	6.13	0.01
Client Acceptance	Mgt Participation [7], [5]	Sr Mgt Fact-Based Decision	2.24	36.5	2.17	34	0.01	0.93
		Top Mgt Data Driven Action	2.14	38	1.99	37	0.06	0.81
	User [17], [21]	System Use	3.81	18	3.86	1	0.18	0.67
Technical Tasks	Analytic Tools [7], [20], [5]	Analytical Sophistication	3.33	30	3.33	18	0.07	0.79
	Data Architecture [2], [24], [10]	Data Availability	4.14	5.5	3.63	8	4.83	0.03
		Data Privacy	3.81	19	3.20	23	4.7	0.03
		Data Quality	4.21	2	3.66	5	5.72	0.02
		System Security	4.00	13	3.64	6.5	2	0.16
	Data [2], [24]	Data Velocity	2.24	36.5	2.44	31	0.12	0.72
		Data Volume	4.38	1	3.64	6.5	7	0.01
		Data Variety	3.52	26.5	2.64	30	6.12	0.01
	IT Infrastructure [7], [24]	Ease of Operations	3.54	25	3.37	15	0.45	0.50
		Performance Quality	4.19	3.5	3.73	3	4.92	0.03
	Service Quality [5], [22]	Business Support Quality	4.05	11	3.50	13	4.67	0.03
		Personal Qualities	3.71	20.5	3.35	16.5	2.17	0.14
		Technical Service Quality	4.13	7	3.72	4	3.91	0.05
Monitoring & Feedback	Prj Mgt [2], [4], [19]	Prj Mgt Competence	4.14	5.5	3.54	12	2.65	0.10

Abbreviations: BI-Business Intelligence, Bus-Business, BD-Big data analytics, Prj-Project, Mgr-Manager, Mgt-Management, Sr-Senior

C. Project Schedule/Plan

The project plan represents the steps needed to reach the project goal [1]. The Shenhar and Dvir [18] project classification model was used to define the attributes of the project. The three levels of *Technological uncertainty* refer to the degree to which the company is using a technology it has never used before: no, some, almost all, or all new technology. *Complexity* is described on three levels: assembly, system, and array. *Pace* describes the sense of urgency and the four levels are: regular, facts/competitive, time-critical, and blitz. Product *novelty* represents the uncertainty of the project goals and in the market: derivative, platform, and breakthrough.

D. Client Consultation

Client consultation involves engaging the internal and external stakeholders to give them the opportunity to air their views, influence the project plans, and know what has been decided [13]. Top management, senior manager, and business user involvement were evaluated using their participation in *steering* (establishing criteria), *directing* (steering and solving conflicts), *acceptance* (evaluating outcomes), and *participation* in projects tasks, specifically setting requirements and building models.

E. Personnel

Competency attributes were used to evaluate the specialized skills required for the staff [2, 8, 10]. The items measure the effort required to deliver quality services during the project for competencies highlighted by Debortoli, Müller and Vom Brocke [19] as being relevant to BI and big data, including *technical*, *business*, *data*, and *analytical* competence. The *data scientist* item considers the involvement of an analytically competent person or persons on the team.

F. Client Acceptance

Lucas Jr [20] suggests that the decision-making style of the users is a factor in their ability to comprehend and accept the results of decision support systems. Thus, *system use* is a measure of client acceptance. System use covers the mandatory and voluntary use as a measure of system success [21]. Top management and senior management participation include their usages of system results in *decision-making* or decision-making and *acting* on the results.

G. Technical Tasks

Success factors for technical tasks are addressed from multiple perspectives such as data sources and IT infrastructure, which includes the availability and quality of data, as well as the heterogeneity and sophistication of IT infrastructure [2]. New and existing items previously used in the DeLone and McLean (2003) [22] information systems success model were used to create factors by [15] and used to evaluate the technical tasks.

H. Monitoring and Feedback, Communication, Troubleshooting

The project manager and team members, the organization, and the external environment are four interrelated groups of

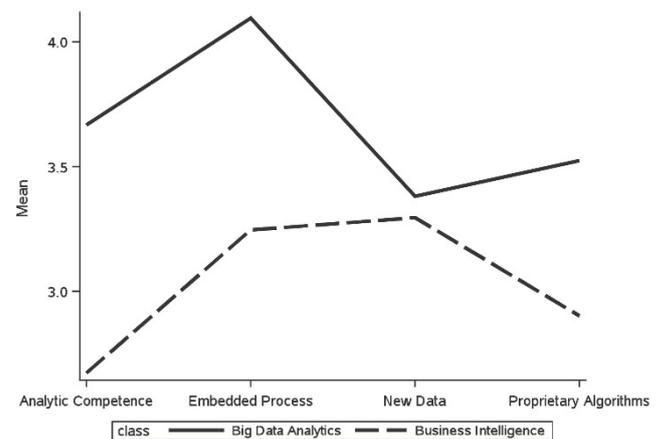


Fig. 1. Two-Cluster Model Structure

project success factors [1, 23]. *Project management competency* is a quality measure for change management, planning, or agile competency [19]. Other environmental factors evaluated include: project demographics, budget, duration, number of departments involved, number of organizations involved, and team size. Organizational demographics include revenue, number of employees, market share position, and revenue position.

V. ANALYSIS

The latent class clustering technique was used to identify the homogeneous clusters. The dimensions were modeled in RStudio 1.0.153 using poLCA for the analysis. The four items for the BI/BD strategy in Table I were specified as indicators in the models. The two-cluster model was selected, given that it had the lowest Bayesian Information Criterion (BIC) measures and highest maximum likelihood. Fig. 1 includes the mean distribution for the two-cluster solution. The delivery of algorithms and analytic competency provide the most differentiation between clusters. Thus, each cluster has a strong differentiating feature. Cluster 1 was named Big Data Analytics and Cluster 2 was named Business Intelligence.

SAS Studio Release: 3.6 (Basic Edition) was used to perform the mean score ranking. The means were computed and ranked. The Wilcoxon test was used to compare the means of variables between the two classes of projects and to provide the significance of the comparison. The mean ranking and Wilcoxon score and significance for the variables for the two project classes are shown in Table I. The Wilcoxon scores with a p-value of less than 0.05 indicate that there are significant differences in the project types. The rank indicates the relative position—equated to importance—of the item for that project type.

VI. RESULTS

The 13 items for project attributes and demographics were not significant and are excluded from Table I for space reasons. Next, many of the results were expected. Specifically, *analytic competence* is ranked higher for big data analytics than for BI. Innovative technologies such as MapReduce- and Apache Hadoop-based systems exist specifically to process significant amounts of data and store

structured and unstructured data such as text, sound, images, video, etc. [9]. Thus, *data variety* was more highly associated with big data analytics, as expected. However, an unexpected result was that *data quality* was higher for big data analytics than BI. Given the *novelty* of big data analytics and its innovative uses, it is unusual that data quality would rank higher. Perhaps it can be explained by the big data solutions being used for the monetization of data. Monetizing big data means being able to process or derive intelligence from data that results in additional revenue from the sale, use, or reuse of the data or intelligence [3]. However, the organizational performance measures for revenue, costs, and strategy were not significantly different. The empirical results suggest BI and big data projects are differentiated based on analytics competence, providing models or algorithms. However, both project types have a similar level of technology uncertainty and pace. Next, while top management contribution is a critical success factor for all type of projects, senior managers and business users are significantly more involved in big data projects than they are in BI projects. The reason is elusive as the organizational performance measures do not differ significantly. However, the newness of the expected outcome, the technical, data, and analytical competence of the team, the data quality and variety, and the performance of the technical infrastructure are significant success factors for big data analytics.

VII. CONCLUSIONS

The practical implication is that sponsors and project managers should use this information to plan projects and to establish success criteria. First, based upon the literature review, it establishes the contingency factors for BI and big data analytic projects. That is, the measurement items offer a guide for defining the infrastructure, personnel, technical tasks, and governance for a project. For example, data privacy would be a critical factor in projects that produce proprietary algorithms and embed them in the business process. Thus, the project should include a specialist and activities for following data protection and privacy regulations. Consequently, the items can be used to facilitate discussions to assign accountable persons and human and financial resources to the project goals. The second area for using the results is to formulate success criteria that can be measured and monitored during the project. For example, measures could be defined around important aspects of service quality. This study contributes by adding clarity to BI and big data analytic project differences. The results of this study are not generalizable beyond decision support projects, and the findings are limited due to the small sample size.

REFERENCES

- [1] J. K. Pinto and D. P. Slevin, "Critical Success Factors Across the Project Life Cycle," *Project Management Journal*, vol. 19, no. 3, p. 67, Jun 1988.
- [2] S. Olbrich, J. Pöppelbuß, and B. Niehaves, "Critical Contextual Success Factors for Business Intelligence: A Delphi Study on their relevance, variability, and controllability," in *45th Hawaii International Conf. on System Sciences*, Hawaii, 2012, pp. 4148-4157.
- [3] P. Géczy, "Big Data Management: Relational Framework," *Review of Business & Finance Studies*, vol. 6, no. 3, pp. 21-30, Aug 2015.
- [4] W. Yeoh and A. Koronios, "Critical Success Factors For Business Intelligence Systems," *The Journal of Computer Information Systems*, vol. 50, no. 3, pp. 23-32, Jul 2010.
- [5] S. Akter and S. F. Wamba, "Big data analytics in E-commerce: a systematic review and agenda for future research," *Electronic Markets*, vol. 26, no. 2, pp. 173-194, May 2016.
- [6] T. Gilad and B. Gilad, "SMR Forum: Business Intelligence - The Quiet Revolution," *Sloan Management Review (1986-1998)*, vol. 27, no. 4, pp. 53-61, Jun 1986.
- [7] T. H. Davenport and J. Harris, *Competing on Analytics: The New Science of Winning*. Boston, MA, USA: Harvard Business School Press, 2007.
- [8] J. Hammerbacher, "Information Platforms and the Rise of the Data Scientist," in *Beautiful data: the stories behind elegant data solutions*, T. Segaran and J. Hammerbacher, Eds., ed Sebastopol, CA, USA: O'Reilly Media, Inc., 2009, pp. 73-84.
- [9] S. Sun, C. G. Cegielski, and Z. Li, "Amassing and Analyzing Customer Data in the Age of Big Data: A Case Study of Haier's Online-to-Offline (O2O) Business Model," *Journal of Information Technology Case and Application Research*, vol. 17, no. 3/4, pp. 156-165, Dec 2015.
- [10] T. H. Davenport and D. J. Patil, "Data Scientist: The Sexiest Job of the 21st Century," *Harvard Business Review*, vol. 90, no. 5, pp. 70-76, Oct 2012.
- [11] R. M. Chang, R. J. Kauffman, and Y. Kwon, "Understanding the paradigm shift to computational social science in the presence of big data," *Decision Support Systems*, vol. 63, no. p. 67, Jul 2014.
- [12] R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, "Big Data analytics and Computational Intelligence for Cyber-Physical Systems: Recent trends and state of the art applications," *Future Generation Computer Systems*, Nov 2017.
- [13] R. J. Turner and R. Zolin, "Forecasting Success on Large Projects: Developing Reliable Scales to Predict Multiple Perspectives by Multiple Stakeholders Over Multiple Time Frames," *Project Management Journal*, vol. 43, no. 5, pp. 87-99, Oct 2012.
- [14] L. A. Ika, "Project success as a topic in project management journals," *Project Management Journal*, vol. 40, no. 4, pp. 6--19, Dec 2009.
- [15] G. J. Miller, "Decision Support Project: Project Success and Organizational Performance," DBA Thesis, Project and Program Management, SKEMA Business School, Lille, France, 2018.
- [16] J. Thomas and J. Kielman, "Challenges for visual analytics," *Information Visualization*, vol. 8, no. 4, pp. 309-314, Jan 2009.
- [17] H. Barki and J. Hartwick, "Measuring user participation, user involvement, and user attitude," *MIS Quarterly*, vol. 18, no. 1, pp. 59-82, March 1994.
- [18] A. Shenhar and D. Dvir, *Reinventing project management: The diamond approach to successful growth and innovation*. Boston, MA, USA: Harvard Business School Press, 2007.
- [19] S. Debortoli, O. Müller, and J. P. D. Vom Brocke, "Comparing Business Intelligence and Big Data Skills," *Business & Information Systems Engineering*, vol. 6, no. 5, pp. 289-300, Oct 2014.
- [20] H. C. Lucas Jr, "Empirical evidence for a descriptive model of implementation," *MIS Quarterly*, pp. 27-42, Jun 1978.
- [21] H. Barki and S. L. Huff, "Change, attitude to change, and decision support system success," *Information and Management*, vol. 9, no. 5, pp. 261-268, 1985.
- [22] W. H. DeLone and E. R. McLean, "Information Systems Success: The Quest for the Dependent Variable," *Information Systems Research*, vol. 3, no. 1, pp. 60-95, Dec 1992.
- [23] W. Belassi and O. I. Tukel, "A new framework for determining critical success/failure factors in projects," *International Journal of Project Management*, vol. 14, no. 3, pp. 141-151, Jun 1996.
- [24] M. Halaweh and A. El Massry, "Conceptual Model for Successful Implementation of Big Data in Organizations," *Journal of International Technology and Information Management*, vol. 24, no. 2, pp. 21-34, Dec 2015.