How do leaders support data mining and analytics projects?

Richard A. Huebner, Ph.D.
Houghton Mifflin Harcourt

ABSTRACT

There are challenges associated with implementing data science, data mining, and analytics projects. Previous literature has explored the variety of technical and individual factors that relate to data mining success. However, a gap exists in the examination of the organizational factors that relate to data mining. These organizational factors have not been widely explored to date, which has left top leaders with little guidance on how to lead and manage these types of projects. This paper provides an examination of the role of the executive leader and their relationship and involvement in data science and analytics projects. The main research question focuses on whether top management support is a critical factor in the successful implementation of data science and related projects. This research effort is designed to provide leaders with specific strategies when implementing data science and analytics projects. The paper examines leadership from the perspective of the Theory of Planned Behavior (TPB), initially developed by Ajzen (1985), and provides specific recommendations that leaders can follow for ensuring that data science, data mining, and analytics projects are successful within their organizations.

Keywords: top management support, data mining, data science, analytics projects, IS success model, theory of planned behavior
INTRODUCTION

During the last five years, there has been significant growth in the amount of data that organizations are collecting and storing. Data mining is a series of techniques and algorithms that can be used to assist organizations in making sense out of this data. Since data science and analytics projects are so relevant today, it makes sense to find means of ensuring their success. For leaders, it can be a challenging endeavor to ensure these types of projects are a success. A gap exists in the literature that investigates what leaders can do to improve the chances for a successful analytics or data science project. This gap exists because a significant proportion of data science research and practice are focused on the development of algorithms and specific techniques to gain insight from data, leaving managers to figure out what they should do to manage data science projects. There is much less research that focuses on how leaders influence and guide analytics projects.

There is limited empirical research on organizational factors that relate to data science, data mining, and analytics projects (Bole, Popović, Žabkar, Papa, & Jaklič, 2015; Huang, Liu, & Chang, 2012). Even though some empirical work exists on technical, organizational, and individual factors that relate to analytics and data science project success, there are uncertainties regarding which specific factors are related to analytics project success (Ang, 2009; Wook, Yusof, & Nazri, 2014). While top management support (TMS) has not been explored as an organizational factor that relates to data science success, researchers have demonstrated that TMS is an essential factor during information systems implementations (Davis, 2014; Dong, 2008; Tarhini, Ammar, Tarhini, & Masa’deh, 2015). There is a wealth of research focused on top management support for projects, but the research has been unclear as to which specific actions leaders can or should do to support those projects. Some example of the ways leaders have supported other types of IT projects include providing financial and human resources, monitoring a project, providing guidance during decision-making processes, and staying engaged throughout the project to ensure that it remains on track.

Top managers may work closely with project managers during steering committee meetings to ensure projects remain on track as well. Senior managers are also responsible for removing roadblocks to a successful project. There are numerous suggestions, but there are still countless IT projects that fail. The challenge for leaders is that every single project is unique – each with its own set of benefits and challenges. Therefore, it is also a challenge to devise any single piece of wisdom or recommendations that can apply to every single project. Even so, this paper will establish some general guidelines specifically for leaders who are in the process of implementing analytics, data science, or business intelligence projects and do so from a Theory of Planned Behavior (TPB) perspective.

As organizations continue to implement data mining, data science, and business intelligence, there is a continuing need to determine which factors, if any, can help lead organizations to successful implementations and realize significant benefits and outcomes. To that end, the research questions explored in this study include: To what extent is top management support and organizational size related to data mining and data science success? What is the relationship, if any, between top management support and data mining success? What is the relationship, if any, between organizational size and data mining success? The goal of this research was to determine if there is a relationship between top management support and the success of a data mining project. In other words, is top management support a critical factor when implementing data science and data mining projects?
In many ways, the terms data science and data mining are often interchanged due to the very subtle differences between these terms. Data science and data mining, in this context, refer to any data mining, analytics, or business intelligence project whose outcome is the extraction and presentation of new knowledge and insight mined from data sets, using various statistical and machine learning algorithms. The algorithms could include a wide range of regression, association rules, clustering, classification, or neural network approaches, all of which are designed to gain new insight into the data.

BACKGROUND

The top management team (TMT) can have a significant impact on the overall success of IT-related projects (Young & Jordan, 2008; Young & Poon, 2013). According to Mahoney (2011), top management support (TMS) is defined as, “the extent to which a top manager personally engaged in behaviors that attempt to promote the success of an information technology project” (p. 10). There is a significant body of research that focuses on how top leaders support information systems and IT-related project implementations. The relationship between TMS and their effects on specific technology implementations has been explored, including a focus on service-oriented architecture (SOA) (MacLennan & Belle, 2014), accounting information systems (Anggadini, 2015), healthcare IT (Hung, Chen, & Kuan-Hsiuang, 2014), and ERP systems (Dong, Neufeld, & Higgins, 2009; Palanisamy, Verville, Bernadas, & Taskin, 2010). Each of these studies cited top management support as a critical factor for IT projects but neglected to provide any specific recommendations for leaders who need to support those projects. The absence of TMS can also have significant adverse effects on an IT project’s success. This lack of support can lead to resources being diverted to other projects, to the point that it can lead to a failed IT implementation altogether (Chua, 2009; Kappelman, McKeeman, & Zhang, 2006). Elbanna (2013) argued that TMS must be constant and consistent during an IT project implementation, otherwise the project can fail. The lack of top management support is often cited as one of the major reasons for project failure.

Even though TMS can significantly impact the success of a project (Ofori, 2013), it has been suggested that critical support processes do not receive the attention they deserve (Zwikael, 2008). Some of the essential support processes include developing project procedures, establishing a project management office (PMO), defining clear project success measures, involving the project manager early in the process, and forming a supportive organizational structure. Other ways TMS can be provided is by financing a project and ensuring appropriate human resources are allocated. While these processes may be beneficial during a project, it is still unclear what role top leaders are expected to play during the IT project implementation (Madanayake & Gibson, 2015). Since top managers play different roles during a project, it is possible that there are critical roles they should play, such as resource allocator, spokesperson, negotiator, and leader. The research conducted by Madanayake & Gibson (2015) also coincides with previous research that suggests top managers are expected to play the role of resource allocator (Haque & Anwar, 2012).

Top leaders sometimes withhold their support for IT projects. Providing such support for a strategic information systems initiative requires the commitment of top management, but this commitment is based on a variety of other factors. Factors contributing to top leaders providing support for IT projects may depend on the stage of a project, project characteristics, team configuration, industry factors, senior leader characteristics, and organizational factors (Mooney,
Researchers have also suggested that top managers should not be held fully responsible for the relative success or failure of these projects. Barclay (2015) suggested that the project team and project members are just as responsible for the outcome. Interestingly, there are authors that disagree with this indicating that project planning, user involvement, and project methodology are not critical for the success of an IT project (Young & Jordan, 2008).

**THEORY OF PLANNED BEHAVIOR**

According to Ajzen (1991), individual behavior steps from an individual’s intention to engage in a specific behavior. Aspects of the theory of planned behavior (TPB) include social norms, perceived behavioral control, normative beliefs, and subjective norm, which all contribute to the intention to engage in a given behavior. The theory has been used to predict and explain behavior within the information systems literature (Ifinedo, 2012; Oliveira & Martins, 2011), and has been widely used to study human behavior. It can provide a theoretical lens through which one can predict and explain an individual’s behavior. Furthermore, the TPB can help leaders with understanding how they can change people’s behavior.

The TPB provides a perspective that behavior is deliberate and planned. The theory does not consider any behaviors that are habitual, unintentional, impulsive, or automatic. It is also an extension of the Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1975) and addresses the limitations of the TRA by adding perceived behavioral control as an additional construct and predictor of behavioral intention. This addition of perceived behavioral control addresses the fact that behavior is not always voluntary or under a person’s control. Behavioral intention is essentially the most significant predictor of outward behavior, so if a person intends on doing something, they most likely will.

A person’s behavior can be predicted by three main constructs, which include attitude, subjective norm, and perceived behavioral control. As people develop their beliefs, people also develop attitudes around those beliefs. In other words, the person’s attitude is a function of his or her beliefs about whether the behavior is good or bad, and if the outcome of the behavior would be of value. Attitudes come from beliefs, which may be formed over a person’s lifetime. Beliefs typically remain stable during a person’s lifetime, but a person’s overall attitudes about those beliefs can shift. However, as beliefs change, people’s attitudes also change. Subjective norm refers to whether a specific behavior will be socially acceptable, or if the behavior is something that others expect a person to do.

There is a perceived social pressure to engage or not engage in a specific behavior, and these perceptions have an influence whether a person will engage in a behavior. When subjective norms influence a person to behave a certain way, a person could form an intention to comply with those expected norms. Perceived behavioral control refers to how the person perceived the overall difficult of the behavior. A person may not have the confidence that they could behave in a certain way. In other words, the lack of perceived behavioral control can limit a person’s intention and ability to exhibit an expected behavior. These constructs do not predict actual behavior, only an individual’s intentions. However, the TPB suggests that when a person intends to engage in a specific behavior, they probably will exhibit that behavior.

In relating the theory back to the problem at hand, there are internal and external pressures on the top manager to behave a certain way. That is, he or she provides the necessary support to a project. One important factor here is the subjective norms and social pressures to
support an important IT-related project. More specifically, analytics and data science-related projects are unique endeavors that have important implications for strategy, planning, and even competitive advantage. As analytics and data science projects are increasingly tied to organizational goals, it becomes increasingly important for leaders to support these types of projects in various ways. Simply stated, for important projects, there is pressure on the leader to provide the financial and human resources necessary for the project to have successful outcomes.

**INFORMATION SYSTEMS SUCCESS MODEL**

The literature on information systems success focuses on finding and measuring the factors associated with a successful IS project implementation. The literature on IS Success is primarily empirically based, to find relationships between a variety of predictor variables and the outcome variable of IS Success. The most widely used model of information systems success was developed by DeLone and McLean (1992), known as the IS success model. This model, which appears in Figure 3, provides a framework for measuring IS success, which includes five dimensions.

DeLone and McLean updated the model in 2003, in which they added another dimension, service quality, and modified organizational impact and individual impact into a single construct they called net benefits (DeLone & McLean, 2003). Each of these variables in the model are interrelated and are important determinants of IS success. The model has been empirically tested in many studies, which has revealed mostly positive results (Dwivedi, Kapoor, Williams, & Williams, 2013; Iivari, 2005; Xu, Benbasat, & Cenfetelli, 2013).

![DeLone & McLean IS Success Model](image-url)


The aspect of the IS success model most applicable to this study are the net benefits. According to DeLone and McLean (2003), net benefits are context-dependent and are related to the objectives of the information system. Additionally, net benefits are the key success metrics in that they measure both positive and negative impact of the system on employees, suppliers, customers, markets, industries, organizations, economics, and society (DeLone & McLean, 2003). The IS success model provides a way to measure the overall impact that an information system has on an organization. In the initial IS success model (1992), the construct was called
organizational impact. The more recent model refers to this as net benefits. One of the more significant challenges in this area is that there is little consensus on how to measure IS success and the impact that IS has on organizations. Success and impact measures have different dimensions and can be examined from an individual level or organizational level. The introduction of the net benefits construct allows researchers to examine the benefits of a system from multiple perspectives, including individual, organizational, and societal.

METHODOLOGY

A survey was sent to a random sample of 500 data scientists and technology leaders to gather their opinions on the importance of specific behaviors and actions that leaders can do to demonstrate their support for a data science project. The sample consisted of data mining professionals and data scientists in the United States who had at least one year of data mining or data science experience. Thong, Yap, and Raman (1996) provided the permission to use a survey instrument that contained measurement scales for top management support, organizational impact, and information systems (IS) effectiveness. The instrument contained a 7-point Likert-type scale for measuring the variables. Thong et al. (1996) reported high reliability scores for each of the measurement scales.

Each scale contained alpha scores over .80, suggesting that the constructs and measurement scales are reliable. Thong also reported the specific coefficients for each scale, which included top management support (.92), organizational impact (.85), and IS effectiveness (.88). In particular there were 5 variables in the top management scale, which include: a) top managers attended data mining project meetings, b) top managers are involved in information requirements analysis, c) top managers are involved in reviewing the staff or consultant recommendations, d) top managers are involved in decision-making, and e) top managers are involved in monitoring the data mining projects.

RESULTS

A total of 339 responses were received via the SurveyMonkey survey. The response data was analyzed, and several preprocessing steps were done on the data to find complete responses. A final count of 175 usable responses were retained for this analysis. Demographic questions on the survey included sex, age, level of education, and organization size. For sex, there were 46 females and 129 males, which represented 26.3 and 73.7% of the sample, respectively. For education, most participants were highly educated. 41% of the participants indicated that they had a master’s degree, and 7% had a doctorate. Next, about 46% of the participants indicated that they have a bachelor’s degree. The remaining participants either had some college, an associate degree, or some other professional degree such as a JD or MD. Not surprisingly, 48% of the respondents earned an advanced degree, which is common among data scientists and data mining professionals.

The survey included a question about organization size, because past research indicated that organizational size could potentially be a factor that contributes to the relative success of an information systems project. Although the researcher did not expect this to be a major factor, it was included within the analysis to either confirm or refute prior work. Today, it is unlikely that organizational size would be a significant contributing factor in these types of projects, since any size organization could reap the benefits of analytics or data science efforts. For organizational
size, the mean size was 3,340, but the median size organization was 155, and the mode 100. This indicated that the data was highly skewed, but that most participants worked within smaller organizations.

Prior to calculating a linear multiple regression model to address the research question and subsequent hypotheses, it is important to establish if the data are normally distributed and check if the predictor variables are correlated with each other. The K-S test for data normality showed that the outcome variable, data mining success, $D(175) = .10, p < .05$ was significantly non-normal. Field (2009) explained that statistically significant results for the K-S test do not necessarily mean that the deviation from data normality is serious. Collinearity diagnostics, such as the variance inflation factor (VIF), were assessed to check for the absence of multicollinearity. One established standard for VIF is to check for values greater than 10. If any VIF values are greater than 10 it would indicate a high degree of collinearity between predictor variables (Myers, Montgomery, Vining, & Robinson, 2012). VIF values were very close to 1, indicating little to no multicollinearity between predictor variables. When there are no VIF values greater than 10, this indicates low collinearity between predictors. Therefore, for this analysis, no multicollinearity was present in the data.

Results for Hypothesis 1

$H_{10}$: There is no statistically significant relationship between top management support and data mining success.

$H_{1a}$: There is a statistically significant relationship between top management support and data mining success.

To test hypothesis 1, linear regression was used to calculate top management support as the predictor variable and data mining success as the outcome variable. Top management support and data mining success variables provided interval-type data for measuring and testing hypothesis 1. The results of this model indicate that top management support is a statistically significant predictor of data mining success. Therefore, $H_{10}$ is rejected in support of the alternate hypothesis. The resulting model for top management support is significant at $p < .001$. The results show that the correlation coefficient for top management support as it relates to data mining success is $R = .434$, which indicates a moderate positive relationship between top management support and data mining success. In this model, all the top management support variables were entered together in a single step. A significant regression equation was found, $F(5, 169) = 7.823, p < .000$, with an $R^2 = .188$. Top management support accounts for approximately 18.8% of the overall variability of data mining success. Thus, top management support is a predictor of data mining success. The adjusted $R^2 = .164$. Results of the hypothesis test were used to answer to research sub-question #1, which is that there is a statistically significant relationship between top management support and data mining success.

Results of Hypothesis 2

$H_{20}$: There is no statistically significant relationship between organizational size and data mining success.
H2a: There is a statistically significant relationship between organizational size and data mining success.

The resulting model for organizational size as a predictor of data mining success is not statistically significant at $p = .37$. Therefore, the statistical tests failed to reject the null hypothesis for hypothesis 2. This means that organizational size is not a contributing factor to data mining success, which means in theory, both small and large organizations can be successful with their data mining projects.

**Hierarchical Multiple Regression Model**

A hierarchical multiple linear regression model was used to analyze the extent to which top management support and organizational size are related to data mining success. Hierarchical multiple regression can be used to ascertain which scale variables contribute the most in a regression model. Since there were five variables within the TMS scale, it was prudent to determine which specific supportive behaviors contributed to data mining success. This allows us to gain additional insight into the ways top managers can support these types of projects. Results of the hierarchical multiple regression model are as follows.

A statistically significant regression model was found that included two out of the five variables from the TMS scale. The two variables: top managers are involved in decision-making, and top managers are involved in monitoring the data mining project, were found to contribute the most within the regression model, contributing a total of 17% to the model. Note that there was a slight reduction in overall $R^2$ when removing some of the variables from the TMS scale and only including the ones that contribute significantly to the model.

![Model Summary for Hierarchical Multiple Regression](image)

*Figure 2. Model summary for a hierarchical multiple regression test against the top management scale and organizational size.*

**CONCLUSION**

There are several practical contributions of this research in terms of how top managers view and support data mining and data science projects. The results suggest that leaders play a key role in data mining projects. Top management participation in monitoring the project and playing a role in decision-making are significantly related to data mining success. Leaders need to understand that these specific behaviors can lead to better outcomes and increase the chances for a successful project. Leaders can support projects by making sure that project monitoring is a part of the project from the start. Perhaps during each phase of a data mining project, top
managers could check in with the data scientist or analytics team. Status reports should be presented to managers to keep them informed. It is also clear that open communication take place during all phases of data mining projects. This is also true of any IT-related project. Leaders may also want to decide which metrics will be used to monitor a data mining or data science project. Developing success metrics is critical during early planning phases of a data science project. Monitoring these types of projects has some benefits including comparing actual vs. planned progress, team engagement, and learning from previous experiences. The study also found that top managers playing an active role in decision-making throughout the project would also lead to a successful data mining project. While the study did not determine the types of decisions or how frequently the leader should engage in decision-making on a project, this could be explored in a future study.

The findings of the study also suggest that the underlying theories could be expanded to include management action and behavior. One possibility is that subjective norm and perceived behavioral control could be added as mediating factors for a construct called management intentions or management support within Delone & McLean’s IS success model.

A limitation of the study is that survey participants could have experienced survey exhaustion because of the length of the survey. If the survey was shorter, it is possible that there would have been more responses. While the total number of responses was n = 339, only 52% of the responses were complete and usable. Nonetheless, the researcher was pleased with the number and quality of these usable responses (n = 175). Estimates of the number of actual data scientists varies, and it is not possible to reach every single data scientist or data mining professional via the SurveyMonkey Audience tool. While this study analyzed 175 responses, it is likely that this is a very small sample of the overall population of data scientists today. Future studies should try to gather higher numbers of responses.

One strength of the study is that it introduces top management support as an important factor that is related to data mining project success. Research that explores these organizational factors as they relate to data mining, data science, and analytics projects is lacking throughout the literature. Thus, a new stream of research exploring these factors on a deeper level, and exploring additional factors such as project management, skilled project team members, and other leadership traits. One possible avenue would be to explore how specific leadership styles contribute positively or negatively on data science projects.
REFERENCES


