https://doi.org/10.48009/2_iis_2007_383-388

ORGANIZATIONAL DATA MINING IN KOREA

Eun-Jeong Cho, rasberry2302@hotmail.com, Oregon Health and Science University, OR
Jae Ho Han, jhhan@pknu.ac.kr, Pukyong National University, Korea
C. Christopher Lee, chris.lee@plu.edu, Pacific Lutheran University, Tacoma, WA
Hyun Kyu Lee, hyunqlee@pknu.ac.kr, Pukyong National University, Korea
Joseph Bradley, josephb@umr.edu, University of Missouri, Rolla, MO

ABSTRACT

This study investigates organizational data mining in Korea. A survey instrument was developed based on a review of the organizational data mining literature and e-mailed to 600 professionals in the field of data mining or data analysis. 90 usable responses were analyzed using stepwise regression. Data quality, process documentation, understanding of data, data integration, data mining outsourcing strategy, understanding of organizational data mining and end-users requests were found to be related to organizational data mining project outcomes.

INTRODUCTION

Data mining is the search for valuable information in large volumes of data. Many companies have recognized data mining as an important technique that will significantly impact their performance, however, few firms are actually using this decision making tool [4]. In addition, managerial intuition and instinct are more prevalent than hard facts in driving organizational decisions. Organizational Data Mining (ODM) is defined as leveraging data mining tools and technologies to enhance the decision-making process. This tool can transform data into valuable and actionable knowledge to gain a competitive advantage [13]. A growing trend in worldwide organizations is for firms to leverage their data resources by developing and deploying data mining technologies to enhance their decision-making capabilities. To address this need, organizations are implementing ODM. Currently, data mining techniques, tools, and research are expanding into the various fields. Many researchers across the country throughout the years have studied data warehousing projects.

However, there are areas where no research has been done, especially in Korea, even though ODM can add values to the information assets of organizations in different sectors through effective induction of large corporate data warehouses into a client-server. In the US, Nemati & Barko [14]) studied key success factors of ODM in international organizations via an Internet survey. This study is one of the very few existing studies of the factors effecting ODM. However, each organization has its own unique culture that can affect business processes and organizational behavior. Different results could be obtained with the same economic and cultural conditions.

The factors influencing international organizations cannot explain the Korean organization behavior and factors. Most of the existing data mining implementation studies were undertaken in Western countries, and only a few have examined data mining practice in Asian countries. Furthermore, existing empirical research on data mining implications on organizations focuses on factors that involve resources (finance, human resources, skills, and others) and information (business strategy, visions and objectives, and others) perspectives. These studies ignore the effects of cooperative competence, organizational context and structure, as the data mining implication process becomes more complex and more difficult to handle as technology continues to change. Researchers have shown that IT management style may differ between Eastern and Western countries, due to cultural and political context [6[. Cultural differences exist between Korea and other countries. This research explores the factors affecting organizational data mining in Korea.

LITERATURE REVIEW

The success factors of data mining have not been widely studied in Korea, the US or Europe, yet some reports and articles have surfaced that mention the success factors of data mining. Nemati & Barko [14] used the TSR framework, which includes Iron Triangle, information systems, and organizational and stakeholder communities to identify key project factors and their relationships. The Iron Triangle (cost, time, and quality) are necessary factors in measuring project success. In

their studies, they used data quality, data integration, technology integration and the level of technological expertise as variables of data and technological issues. They used: 1.) the presence of a clearly defined ODM strategy aligned with corporate strategies, 2.) reengineering of business processes to support ODM systems, 3.) the presence of new incentive plans to support ODM systems, 4.) the presence of an outsourcing strategy for ODM as organizational issues, 5.) the presence of an influential executive project sponsor, 6.) the level of end-user ODM expertise and 7.) the presence of a non-IT (end-user, business analyst, etc.) ODM advocate as people issues. The iron triangle variables were used as success measures. Among those variables, the presence of a clearly defined ODM strategy aligned with corporate strategies, the reengineering of business processes to support ODM systems, the presence of new incentive plans to support ODM systems, the presence of an influential executive project sponsor and the presence of an non-IT ODM advocate, were denied their significance.

In open-ended questions Nemati & Barko [14] asked for additional factors affecting ODM outcomes. Respondents answered that 1.) the proper selection of ODM tools and algorithms; 2.) an in-depth knowledge of suggested as determinant factors; and 3.) understanding the data for the ODM system were required to be successful and offer valuable benefits. Thuraisingham [16] described the steps in the data mining process. He stated that improving data quality and data integration from various sources are things that must be done in the process of data mining. Communication with a data mining consultant (or data mining tool vendor) is needed before the actual mining job begins to achieve a desirable outcome. He also said that choosing a proper outcome and interpreting the data are important.

Lee & Siau [9] described the requirements and challenges of data mining. These include the ability to handle different types of data, graceful degeneration of data mining algorithms, valuable data mining results, representation of data mining requests and results, mining at different abstraction levels, mining information from different sources of data and protection of privacy and data security. In relations to the generalization of data mining algorithms, the searching, mining, or analyzing time of a data mining algorithm should be predictable and acceptable as the size of the database increases. The data mining tool should allow users to discover information from their own perspectives. The information should be presented to the users in forms that are comfortable and easy to understand. To facilitate the mining process, the systems should allow the users to mine at different abstraction levels.

Nemati & Barko [14] used the Iron triangle to measure the results of survey participants ODM project and actual outcome. The participants gave their answers of the outcome in relation to expectations. For project outcomes, choices ranged from 1 (miserable failure) to 5 (overwhelming success). The study of the factors effecting success of data warehousing were actively going on in and outside of Korea. Data warehousing is a valuable, proven approach to providing business users at all levels of an organization with the information they need to make high-impact decisions—only if the right technologies, architecture, and techniques are used.

Watson & Haley's study [17] showed that the success of data warehousing affected by the support of executive managers, participation of end-users, necessity in business operations, end-user support, use of methodologies, specific project plans, quality and quantity of data and end-user expectation management. Also pointed out, were the bad quality of data, lack of resources, unspecific project plans, limited technical skills, lack of understanding of existing data, lack of end user support, lack of project team members' professional knowledge and unsuccessful management of end-user expectations effect negatively on the successful data warehousing.

As there are few ODM studies done previously, and IT and organizational context relationship study is necessary for this paper, organizational factors related to IT in organizations were reviewed before executing this study. IS/IT strategic planning (ISSP) has been identified as essential in integrating IT into an organization to increase a firm's strategic competitive advantage [5, 7, 12]. Recent experience suggests that despite the many spectacular e-business successes, there have also been numerous failures [15]. Numerous companies fail not for technical reasons related to IS/IT implementation, but rather because they neglect the effects of organizational context on e-business and e-commerce strategic planning [2]. Introducing an e-strategy not only involves a change in technology, but also an impact on various organizational perspectives, such as culture, skills, styles, structure, and beliefs, as well as various

organizational considerations that may influence the ISSP process.

Through our literature review, we could see the success and failure factors of data mining, data warehousing, IT implementation and related studies. Inducing failure factors in success factors study might bring about a problem. However, since data mining related studies in Korean organizations have not yet been done, we included both the success and failure factors in this study. Moreover, all the factors that were resulted as not significant in each study have included as factors in this study for the means of cross-cultural study. Additionally, there were various factors named differently even though they represented the same thing. This causes confusion; thus we changed some of them to a different word and then reorganized. For example, the word 'understanding of existing data' used by Watson & Haley's [17], and 'thorough data preparation' used by McGee's [11] have the same or similar meaning with Watson & Haley's [17] 'quality of data source' [8]. And also, the word 'support' used by Adelman & Moss [1] has the same meaning as 'possession of proper resources' used by Watson & Haley [17], and 'prototyping' and 'specific data requirements' used in Little's [10] study was the same as 'prototyping' used by Watson & Haley [17]. Bai et al. [2] uses the phrase "CEO & CIO relationship" in which is referred to as "executive sponsor" in almost every other study. Through this empirical study, a base of building a new framework of organizational data mining in Korean organizations will be maddened. Rearranging each factor according to the relations of each is a mandatory part of this empirical research.

METHODOLOGY

Five issues of ODM variables – data, technical, organizational, project, environmental – are developed from previous studies of related fields. Each of the organizational variables and research hypotheses are described and rationalized below. The first four hypotheses examine the link between data issue and ODM project outcomes. These main hypotheses are as follows: role of data quality, data integration, data variety and understand data used in project.

*H1.*Data factors have a positive influence on ODM project outcomes.
H2. Technical factors have a positive influence on ODM project outcomes.
*H3.*Organizational factors have a positive influence on ODM project outcomes.
*H4.*Project factors have a positive influence on ODM Project outcomes.
H5. There is a positive relationship between the level of understanding the organization's environment and the actual outcomes of ODM project.

The objective of this paper is to find the relationship between the five key implementation areas (data, technology, organization, project and environment) with the actual outcome of ODM projects through surveying organizations in Korea. For ODM project success criteria, we asked the respondent to choose on a scale from 1 to 7 (1 being a strongly disagree and 7 being a strongly agree for a questionnaire given each time a related hypothesis was made) what the actual (not perceived) outcome if the ODM project in question was. The Iron Triangle factors (times, scope, and resources) were chosen to measure success of an ODM project. These factors have been researched extensively in determining their role on project outcomes. They were included in this study to measure the ODM project outcomes. Resources are defined as the length of time an ODM project took compared with the initial expectations. Moreover, scope is defined as the amount of system functionality implemented in comparison with initially defined scope.

The survey was developed that contained 35 ODM project research and 5 demographic questions. The questions used to assess project performance and measure critical success factors were extracted from an extensive literature review. As previously noted, questions were developed to determine the impact with relevant ODM dimensions such as data, technological, organizational, project related, and environmental in the Iron Triangle elements [time, resources (cost), and scope (quality)]which are the successful implementation of ODM projects. The survey required participants to indicate their degree of agreement on a seven-point Likert scale (1=strongly disagree and 7=strongly agree) to 35 questions covering data, technological, organizational, project related, and environmental issues.

This study was conducted via an Internet survey. A request to visit the Web survey site was e-mailed to about 600 professionals in various fields who had prior experience of data-mining/data analysis projects. The mailing list database was constructed using a variety of sources, including personal

contacts, magazines, conference registration lists, journals, and member lists of Internet clubs for data mining. Replies were limited to only e-mailed professionals and introduced by them. Only authorized persons listed on the database could enter the survey site, and answer the survey questions. Respondents were asked to focus on a particular ODM project and measure the actual effect of specific factors in that project's outcome. A couple of subsequent reminder e-mails were sent to the individuals and organizations failing to respond to the initial round. 102 responses were received, but 90 usable were collected and analyzed. Twelve outliers were excluded. The resulting response rate was 15 percent.

STATISTICAL RESULTS

Descriptive statistics for the respondents show that the respondents were mainly consultants (28 percent), R&D (18 percent) and IT staff (12 percent), but all job categories surveyed were well represented. Industries most represented are consulting/system integration (23 percent), retail (8.9 percent) and entertainment (8.9 percent). 51.2 percent of respondents work for small organizations (under 100 employees). Most respondents are relatively new to ODM with 87.8 percent having less than five years of experience.

In order to test the 20 hypotheses of each factor's influence on project outcomes, a stepwise multiple regression analysis was performed. The results of the stepwise multiple regression analysis are displayed in Table 1.

Table 1. Results of Stepwise Multiple Regression Model

Model	Factors entered	Adjusted R-square	Standard Error	F	Sig.
A	data quality	.103	.9469	11.249	.001
B	a + process documentation	.222	.8820	13.701	.000
C	b + understanding of data	.253	.8464	11.044	.000
D	c + data integration	.303	.8349	10.666	.000
E	d + data mining outsourcing strategy	.342	.8114	11.235	.000
F	e + understanding of ODM	.371	.7930	9.756	.000
G	f + end-users' request	.411	.7676	9.864	.000
H	g – process documentation	.409	.7686	11.273	.000

Eight models were constructed from the stepwise multiple regression analysis. Regression results indicate that the overall model "G," consisting of seven factors, significantly predicts project outcomes, R^2adj = 0.411, standard error = 0.767, F = 9.864, $p < 0.000$. The seven factors included in the model "G" are data quality, process documentation, understanding of data, data integration, data mining outsourcing strategy, understanding of ODM and end-users' request. Data quality, data integration, and understanding of data were data issue factors that found to be significant in determining ODM project outcomes. Process documentation, understanding of ODM, and end users' requests were project issue factors that also found to be significant. Data mining outsourcing strategy was also found to be significant factor in determining ODM project outcomes, and it is an organizational factor issue. However, no factors of technology issues and environment issues were found to be significant.

Overall, some results were expected, but other hypotheses were found to be surprisingly insignificant. Technical issues and organizational issues, except external DM expertise help were not found to be significant in determining project outcomes. This may be because most Korean organizations were conducting IT solution and DM tools from major providers like Oracle, SPSS, etc. This brought few IT differences between organizations. Furthermore, it may be due to the

level of inexperience of respondents of Korean organizations, with a majority having less than five years experience. It is possible to say that technical issues were proved as insignificant because of multicollinearity problems due to multiple dependent variables being analyzed at one time. Department support, executive support, and DM expertise significance results differ compared to data warehousing significance in determining project outcomes. This result may be due to the difference of DM and data warehousing.

These results suggest that there are some differences between the ODM implementation in world organizations and Korean organizations, but not significant differences. Furthermore, the ODM project issues were more effective to the ODM implementation than organizational and environmental issues. However, technical issues need to select new variables to avoid multicollinearity problems, and create a new framework. The challenge for future research is to determine the impact of other factors in the framework in implementation success. It must be noted that a potential bias of these results could be presented due to the use of a subjective project performance scale. However, there might be some differences in important factors according to an organization's industry type. Studying the main factors of ODM success by each industry type should be incorporated in future studies.

CONCLUSION

This study presents a framework for applying ODM project implementations to better understand the factors influencing the success of ODM projects, and demonstrates that selected ODM project factors, when modeled under a framework built on this paper, have a significant impact on the outcome and success of ODM projects. These key project factors are data quality, data integration, understand of data, data mining outsourcing strategy, process documentation, understanding ODM, and end users' request. As expected, the findings from this study reinforced the ODM necessities of leveraging quality data and technology that are both well integrated and well understood. In addition, the examination of an external DM expertise's help suggests that organizations, especially not experienced with ODM systems or whose core competencies are not ODM, should look outside their organization when implementing these technologies. No matter how good the mining results are, if it is hard to apply in the real world, it is meaningless. The original purpose of data mining is providing a variety of information that is applicable to the practical world. Accordingly, it is very important to have a thorough grasp of what the end-user wants. Process documentation helps smooth operating of any project. For example, it helps to have easier communication between team members, managers and end-users. It also helps review and reexamination of projects at any time.

It is hoped that the empirical research and findings can contribute to the efforts of identifying more efficient and effective ways of improving ODM project implementations for Korean organizations. Having the right information at the right time is crucial for making the right decision. The problem of collecting data, which used to be a major concern for most organizations, is almost resolved. Industry surveys indicate that over 80 percent of Fortune 500 companies believe that data mining would be a critical factor for business success by the year 2000 [3] Obviously, DM will be one of the main competitive focuses of organizations. The gap is growing between more powerful storage and retrieval systems, and the users' ability to effectively analyze and act on the information they contain. Data mining can widen this gap.

Future research in the area of ODM implementations should investigate and identify other essential factors and sub dimensions that impact project success. Technology issues need to select new variables, considering multicollinearity problems. In addition, there might be some difference of important factors according to an organization's industry type. Studying the main factors of ODM success by each industry type should be incorporated in future studies. However, there are many issues that remain to be resolved and many research opportunities continue to present themselves.

REFERENCES

1. Adelman, S., and Moss, L., Indicators of Success. DM Direct, August 1999.
2. Bai, R.J. and Lee, G.G. (2003). Organizational factors influencing the quality of the IS/IT strategic planning process. *Management and Data Systems* 103(8), 622-632.
3. Baker, S. and Baker, K. (1998). Mine over matter. *Journal of Business Strategy*, 19(4), 22-7.

4. Davenport, T.H., Harris, J.G., DeLong, D.W., and Jacobson, A.L. (2001). California Management Review, 43(2), 117-138.
5. Huysman, M.H., Fischer S.J., and Heng, M.S.H. (1994). An Organizational Learning Perspective on Information Systems Planning. *Journal of Strategic Information Systems*, 3(3), 165-177.
6. Kunnathur, A.S. and Shi, Z. (2001). An investigation of the strategic information systems planning success in Chinese publicly traded firms, *International Journal of Information Management*, 6, 423-439.
7. Lederer, A. L., and Sethi, V. (1996). Key prescriptions for strategic information systems planning. *Journal of Management Information Systems*. 13(1), 35-62.
8. Lee, Y. S., Lee, D. M. and Suh, C. K. (2000). An empirical study of success factors in data warehousing. *Journal of Information Technology and Database*, 7(2) December, 61-85.
9. Lee, S. J. and Siau, K. (2001), "A review of data mining techniques," Industrial Management & Data Systems, Vol. 101, Iss. 1, p. 41.
10. Little, R.G., & Gibson, M. L. (1999). Identification of Factors Affecting the Implementation of Data Warehousing. Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences, Maui, Hawaii, Jan. 5-9, 1999.
11. McGee, K. R. (1997). The Bottlenecks of Implementing a Successful Data Warehouse. Unpublished Degree of Master Dissertation, University of Texas at Arlington, 1997.
12. Mentzas, G. (1997). Implementing an IS Strategy—A Team Approach. *Long Range Planning*, 30(1), 84-95.
13. Nemati, H. R. and Barko, C.D. (2001). Issues in organizational data mining: a survey of current practices. *Journal of Data Warehousing*, 6 (1), Winter, 25-36.
14. Nemati, H. R. and Barko, C. D. (2003). Key Factors for Achieving Organizational Data Mining Success. *Industrial Management and Data Systems*, 103(4), 282-292.
15. Phan, D. D. (2001). E-Business Management Strategies: A Business-To-Business Case Study. Information Systems Management, 18(4), 1-9.
16. Thuraisingham, B. (2000). A primer for understanding and applying data mining. IT Professional, 2(1). 28-31.
17. Watson. H. J., & B. J., Haley (1997). Data Warehousing: A Framework and Survey of Current Practices, *Journal of Data Warehousing*, 2(1), 10-17.