A feasibility study on importance of using Big data Analysis in implementing E-Government services

Samane Ebrahimi^a, Yasser Elmi^a

^a. *Department of Information Technology, Sabzevar Branch, Islamic Azad University, Sabzevar, Iran (E-mail: samane.ebrahimi1994@gmail.com)*

^b. *Department of Computer Engineering, Sabzevar Branch, Islamic Azad University, Sabzevar, Iran (Corresponding Author, E-mail: elmi@iaus.ac.ir)*

Abstract

The growth and development of e-government communities is shown as a sovereign asset in developing societies, as a way for governments to employ new technologies, make people more accessible to government information and services, improve the quality of these services and provide the possibility of people's participation in democratic processes and activities. The purpose of this study is to evaluate the feasibility and importance of using metadata in the implementation of e-government services. In this research, which was conducted over a period of 6 months, 146 senior, operational and middle managers of Islamic Azad University in Khorasan Razavi province were selected through stratified random sampling. The research tool was a questionnaire whose validity was confirmed by experts and its reliability was confirmed by Cronbach's alpha coefficient. Based on partial least squares analysis (PLS) which is a model fitting based on validity and reliability analysis, the results showed that macro data volume, macro data speed, and macro data variability are based on three criteria of e-government implementation ie e-government core. , The extended layer of e-services and the outer layer of e-services based on the t-statistic, which is higher than 1.96, has a positive and significant effect.

Key words: Big Data Volume, Big Data Velocity, Big Data Variety, E-Government Services

1- Introduction

With big data, a wide range of information can be stored, managed and processed. Big data is a term used for a very large, diverse, complex (and difficult to store) data set. The process of researching large amounts of data to uncover hidden patterns and uncover correlations is called "big data analysis". This useful information helps organizations and companies gain deeper insights and succeed in competition. Big data is usually a collection of data that is large enough to be retrieved, managed, and processed in a reasonable amount of time by standard software.

Big data can play an important role, especially in developing countries. This article focuses on the potential of big data to be used in the developing country of Iran. It is worth noting that the security, protection and trust of users are among the challenges that exist in enjoying the benefits of big data.

Due to the importance of big data, many countries have launched their programs and initiatives on big data and its applications. Therefore, by using big data in the development of e-government and its application, long steps can be taken. On the other hand, big data in governments and especially in governments of developing countries can play an important role and reduce the knowledge and technology gap of these countries with developed countries. But its realization requires field-based studies. Governments can use big data to better serve citizens.

In this paper, in order to measure big data, three subscales of volume, speed and variability based on Patel et al. [1] questionnaire have been used. The validity and reliability of this questionnaire has been confirmed due to its standard. So that the total reliability is equal to 0.76. The method of analysis and fitting of modern utility is based on SMART PLS software. The inferential statistics of this research are measured using the analyzes related to modern validity and reliability, and considering the desirability of the modern structural framework, and based on that, the research hypotheses are tested. It is worth mentioning that the beta path

coefficient and T-statistic have been used to test the hypotheses, which according to the obtained results, the confirmed hypotheses should be higher than 1.96. The statistical population of this study is the managers of different levels of Islamic Azad University in Khorasan Razavi province who have been selected by simple stratified random sampling method. Considering that the statistical population of this research is 235 people, Cochran's formula has been used to determine the sample size.

2- Related works

In Keoduangsine et al.'s research [2], a new model was reviewed in order to identify the most influential factors based on Ruff theory analysis and it was found that infrastructure investments are considered as the most effective factor in the development of e-government readiness. In [3] a study is presented based on qualitative analysis approaches using content analysis that created a coherent structure in decision-making level of education in school systems. In this study, three structural changes including training, evaluation and accreditation are considered as the most important educational platforms for data production in educational institutions that can change the decision-making methods in educational institutions.

In [4] an approach which was conducted in the period of one year 2017-2018, 400 managers of active companies in the field of private service sector in China participated. In this research, partial least squares analysis (PLS) method was used to fit the model. The results showed that increasing the level of dynamism in recognizing technological management styles and big data analysis in line with the dynamic capabilities of companies can help to develop and effectiveness of decision-making level and cause the level of quality services to be developed. Shonhe et al. [5] is focused on information technology participation. The analysis method in their study was basic data analysis based on three stages of open coding, selective coding and axial coding. The results showed that the use of social media, use of update networks and websites with useful information content for consumers can increase the level of interaction of the audience.

Torrecilla et al. [6] was attended by 104 IT managers and R&D managers in the automotive industry. In this study, which was conducted over a period of one year, the method of experimental analysis using the experimental and control groups was used. One group was examined from the perspective of big data analysis theory only and one group was examined after passing a 6-month course. The results clearly showed the difference in the effectiveness of learning between the experimental and control groups, so that the group that had received practical training used more knowledge-based approaches in analyzing the problems of the organization. Patel et al.'s method [1] was based on simulation analysis, examined more than 30 scenarios at specific time intervals based on characteristics such as the type of governance, how to manage resources, the effectiveness of optimal allocation, and so on. The results showed that optimizing the data storage process in the form of a comprehensive database while increasing the speed of information flow increases the level of transparency. It was also found that big data analysis is very important to gain insight and predict the future of government governance, but extensive infrastructure investments need to be made in this regard.

In [7] a standard questionnaire was used to collect research data and participants Research 117 small company managers had an activity life of less than 5 years. The results showed that there is a direct relationship between technical training, knowledge training and human resource skills training with the effectiveness of IT implementation, but it turned out that investing in IT infrastructure in small companies, at least for the short term, does not affect the provision of effective technological services in these companies.

3- Research Method

The validity of the questionnaires of this research was standardized using the opinions of experts and specialists. To evaluate the reliability of the research, Cronbach's alpha method was used according to the following equation:

$$r_{\alpha} = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_j^2}{\sigma^2} \right) \quad (1)$$

where r_{α} represents the reliability coefficient of the whole test, k represents the number of test questions, σ_j^2 is the variance of the scores of question j , and σ^2 is the variance of the scores of the total questions. If this coefficient is greater than 0.7, the test has acceptable reliability. The Cronbach's alpha value is as shown

in Table. 1. Therefore, considering that the alpha value for all variables of the questionnaire is greater than 0.7, the reliability of the questionnaire is confirmed. Table.1 shows the Cronbach's alpha coefficient of research variables.

Table.1 Cronbach's alpha coefficient of research variables

variable	Number of questions	Cronbach's alpha coefficient
volume	4	0.76
velocity	4	0.73
variety	4	0.80
The core of e-government services	5	0.92
Extended layer of e-government services	5	0.93
External layer of e-government services	4	0.89
Total number	26	0.83

In this study, stratified random sampling method was used. In this method, based on the number of each class and the percentage of the specified sample volume, the sample is determined. According to the statistics and information announced by the Ministry of Science of Iran, there are 12 branches of the Islamic Azad University in the province of Khorasan Razavi, whose total number of managers is equal to 235 people at different levels. Based on Cochran's formula, the sample size required for this study was 146 samples.

Table. 2 Descriptive statistics related to education variable

Respondents education	Frequency	Valid relative frequency percentage	Valid relative CPF
Bachelor	31	21.23	21.23
Master	51	34.93	56.17
Ph.D	64	43.83	
Not answered	0	100%	100%
Total sample size	146		

According to the hypotheses presented and the statistical method used, the research findings are presented in the form of descriptive statistics tables and inferential statistics. For this purpose, first the descriptive findings are expressed, then in order to test the hypotheses, the partial least squares analysis (PLS) method is used, which is based on model fitting based on the validity and reliability. Demographic statistics are used to explain the data collected in the research. The frequency distribution of the respondents based on their education, age and managerial level is shown in Table, 2 - 4, respectively.

Table 3 Descriptive statistics related to the age variable

Age of respondents	Frequency	Valid relative frequency percentage	Valid relative CPF
Less than 30 years	0	0	0
Between 30 and 40 years	78	53	53
Between 40 and 50 years	66	45.20	98.2
More than 50 years	2	1.8	100%
Not answered	0	0	
Total sample size	146	100%	

Table. 4 Descriptive statistics related to the management level variable

Respondents' management level	Frequency	Valid relative frequency percentage	Valid relative CPF
Senior Managers	24	16.43	16.43
Middle managers	79	54.10	70.53
Operational managers	43	29.47	100%
Not answered	0	0	
Total sample size	146	100%	

Examining the average of the calculated variables, it was found that the highest average is related to the big data diversity variable (4.39), which indicates that the existence of different levels of data in the form of

technological software can help improve the level of analytical efficiency and cause Provide more reliable and reliable information for users. It was also found that the core of government services (4/29) is considered as one of the consequences of implementing e-government services, which can reflect the nature and philosophy of services that can be provided and the need for it in the community. The highest rate of standard deviation is related to the speed of big data, which with a deviation of about (0.95) indicates that the participants in this study had different views on the speed of big data in the analysis and service delivery, and this issue It has increased the level of standard deviation.

Table. 5 Descriptive Statistics

	External service layer	Extended service layer	Government services core	Bigdata diversity	Bigdata Velocity	Bigdata volume
Average	4.22	4.19	4.29	4.39	4.15	4.28
Standard deviation	0.91	0.91	0.75	0.74	0.95	0.78
Minimum	2.25	2.40	2.80	2.50	2.25	2.25
Maximum	5.00	5.00	5.00	5.00	5.00	5.00

One of the tests used to determine the normality of data distribution is the Kolmogorov-Smirnov test, which is a non-parametric test. In this test, if the significant level of Z statistic value is greater than 5% (Sig <0.05), the statistical null hypothesis that the distribution of the variable under study is normal is accepted with 95% confidence, but otherwise the hypothesis is confirmed that States that the data is not normally distributed. Examination of the results of Kolmogorov-Smirnov test showed that the significance level of Z statistic of all research variables is significant at the 5% error level (Sig <0.05), hence the percentage assumption that the data is normal is rejected and the acceptance hypothesis is rejected.

Table. 6 Experimental results for normality of research variables

Variable	Absolute value	Positive	Negative	Z	Significant level
External service layer	0.10	0.10	-0.14	1.18	0.001
Extended service layer	0.09	0.09	-0.15	1.07	0.000
Core of government services	0.10	0.10	-0.16	1.14	0.005
Big data variety	0.07	0.07	-0.11	0.84	0.012
Big data Velocity	0.07	0.07	-0.09	0.78	0.000
Big data volume	0.17	0.17	-0.19	2.03	0.0001

According to the results of Table. 6 and considering that the structural equation modeling method with partial least squares (Smart-PLs) approach has been used to analyze the research data, and in this method, the absence of normal data distribution does not create a limit. The model is modeled in the form of validity and reliability of dimensions. It is noteworthy that the software that uses structural equation modeling based on this statistical method is compatible with conditions such as the alignment of independent variables, the normality of the data and the small size of the sample.

Table. 6 shows the correlation matrix of the main components of the research model. Based on this matrix, since the correlation between the model variables is less than 0.9, so the non-overlap between the research model variables in the form of differential validity is also confirmed.

Table. 6 Correlation matrix for main dimension of research model

	Big data volume	Government services core	External service layer	Extended service layer	Big data Velocity	Big data variety
Big data volume	0.84					
Government services core	0.52	0.91				
External service layer	0.63	0.47	0.88			
Extended service layer	0.55	0.47	0.52	0.86		
Big data Velocity	0.54	0.57	0.53	0.53	0.88	
Big data variety	0.46	0.48	0.48	0.50	0.53	0.85

4- Analysis of research hypotheses

Path coefficients as well as t-statistic were used to test the hypotheses. The path coefficients were used to determine the strength of the relationship between the two variables. To evaluate the significance of relationships, t-value was used, which is obtained in Smart PLS software using the automatic start-up algorithm.

It should be noted that if the path coefficient between the independent latent variable and the dependent variable is positive, we conclude that by increasing the independent variable we will see an increase in the dependent variable. A summary of the hypothesis test is given in Table 7.

Table. 7 Summary of the hypothesis test

Causal relationships between research variables	Hypothesis result
Big data volume has a significant impact on the core of big data e-government services.	Confirmed
Big data velocity has a significant impact on the core of big data e-government services.	Confirmed
Big data variety has a significant impact on the core of big data e-government services.	Confirmed
Big data volume has a significant effect on the developed layer of electronic big data services.	Confirmed
Big data velocity has a significant effect on the developed layer of big data electronic services.	Confirmed
Big data variety has a significant impact on the developed layer of big data electronic services.	Confirmed
Big data volume has a significant effect on the external layer of big data electronic services.	Confirmed
Big data speed has a significant effect on the external layer of big data electronic services.	Confirmed
Big data variety has a significant impact on the external layer of big data electronic services.	Confirmed

5- Conclusions

Big data diversity, known as the diversity of data analysis processes and their conversion into information, has a positive effect on the external layers of electronic services. Layers ranging from user satisfaction to culture and acceptance of the use of information technology also reduce overall social costs. In fact, e-

government systems in Islamic Azad University have the necessary dynamism when the diversity of big data at the technological level of the university in a coherent manner and in line with the needs of analysis and information to create a level of satisfaction and satisfaction for users. Mere data analysis in the simplest possible way may not be very effective in increasing the level of social convergence of academic users and may not lead to the willingness of the community to use it effectively. Therefore, due to the sensitivities at the university level, it is necessary to provide a stronger infrastructure in this area to lead to more favorable social outcomes and increase the level of motivation.

Government use of information technology, especially the Internet, to increase the level of access of citizens, government agencies, public and private sector employees to online services and information has been referred to as the definition of e-government (Lane and Lee, 2001). But it is noteworthy that communities that, even with the development of information technology in the last decade, still do not have a stable infrastructure in creating databases based on the desired features such as speed, feedback and information flow, etc., seek to create E-government can be. E-government as a communication between governments, customers and suppliers using electronic means and ICT to support government operations and service preparation, are other definitions in this area. According to another definition, e-government is defined as a way for governments to use new technologies, provide people with good access to government information and services, improve the quality of these services, and enable people to participate in democratic processes and activities. In the definition of government and governance, the provision of services is often the responsibility of the government and the decision to provide or not to provide services is often the responsibility of the government.

References

1. Patel B, Roy S, Bhattacharyya D, Kim T-H (2017) Necessity of big data and analytics for good e-governance. *International Journal of Grid and Distributed Computing* 10 (8):11-20
2. Keoduangsine S, Robert G (2007) Measuring e-readiness assessment: The case of Laotian organisations. *Literacy* 68 (7)
3. Zeide E (2017) The structural consequences of big data-driven education. *Big Data* 5 (2):164-172
4. Shamim S, Zeng J, Shariq SM, Khan Z (2019) Role of big data management in enhancing big data decision-making capability and quality among Chinese firms: A dynamic capabilities view. *Information & Management* 56 (6):103135
5. Shonhe L (2019) An assessment of the technology readiness of public librarians in Botswana. *Global Knowledge, Memory and Communication*
6. Torrecilla JL, Romo J (2018) Data learning from big data. *Statistics & Probability Letters* 136:15-19
7. Dyerson R, Spinelli R, Harindranath G (2016) Revisiting IT readiness: an approach for small firms. *Industrial Management & Data Systems*