Predicting Factors Affecting the Readiness of Big Data Adoptions: An Application of Data Mining Algorithms*

Giang Thi Nguyen [1][0000-0002-8951-5937], Shu-Yi Liaw [2,**][0000-0002-5542-9096], and Thi Mai Le[3,**] [0000-0001-9720-308X]

[1] Department of Tropical Agriculture and International Cooperation, National Pingtung University of Science and Technology, Taiwan
nguyenthigiang@tuaf.edu.vn

[2] Department of Business Administration, Director of Computer Centre, National Pingtung University of Science and Technology, Taiwan
syliaw@mail.npust.edu.tw

[3] Department of Social Science, Economics and Management, International School, Vietnam National University, Ha Noi, Vietnam
mailt@isvnu.vn

Abstract. The purpose of this study is to predict the factors affecting the readiness to adopt big data in small and medium enterprises (SMEs) in Vietnam. Data collected from 240 managers at SMEs encompasses 13 input variables that impact the readiness (high/low) to adopt big data. Partitioned, training, and testing data were analyzed by three Data Mining algorithms techniques (CHAID, Bayesian networks, and Neural Network). The accuracy results of evaluation statistics on the training and testing data of the three models are 70% higher. The area under the Receiver Operating Characteristics (ROC) curve (AUC) value on the training data ranged from 0.827 to 0.908, while it ranged from 0.777 to 0.898 on the testing data. The results of this study highlighted that top management support, data quality, data security, partner pressure, and budget resources are the five most important factors to predict readiness. The findings of this study contribute important implications for managers, service vendors, and policymakers to understand the factors that influence readiness to adopt big data. Hence, managers can establish a clear strategy to enhance the readiness to adopt big data in SMEs in the future.

Keywords: Big data adoption, Data Mining, Manufacturing sector, Readiness, Service sector,Vietnamese SMEs

** Corresponding Authors.

1 Introduction

The rapid development of internet technology has created a huge amount of data from many different sources such as Media, Cloud, Web, Internet of Things, and Databases that are called big data. Big data is a "5V" characterized data source, exhibiting volume, velocity, variety, verification, and value characteristics [1]. Big data adoption is useful in helping companies seek new business opportunities, decrease costs, and minimize risks [2,3]. The use of big data in e-commerce context provides many benefits such as prediction trends for future product development and improvement of company-customer relationships [4]. Further, big data plays an important role in helping businesses develop a sustainable economy [5]. However, firms are facing many challenges when implementing big data including, information technology infrastructure, financial resources, data security, organizational culture, lack of skills, etc. [2,6-8]. In Vietnam, SMEs account for 98% of total enterprises, 30% of total export value, and create 500,000 new jobs annually [9,10]. Moreover, Vietnam is considered to be the place where large data sources are available with 66% and 60% of the population using the internet and using social networks, respectively [11]. However, SMEs in Vietnam have not yet taken full advantage of new information technology such as cloud computing, internet of things, and big data, thus faced with low business efficiency [12]. The big question is, what are the factors that affect the adoption of big data by SMEs in Vietnam? Leveraging this appeal, this research sought to predict the factors that impact the adoption of big data by SMEs in Vietnam. Understanding the factors that affect the readiness to adopt big data helps businesses to be well prepared before adopting big data. As a result, the likelihood of successful big data adoption will be high. Therefore, the adoption of big data will contribute to improving enterprises' business performance.

2 Literature Review

Previous studies have mentioned several factors influencing big data adoption (BDA). Yadegaridehkordi, *et al.* [13] applied the approach of decision-making trial and evaluation laboratory (DEMATEL) and adaptive neuro-fuzzy inference systems (ANFIS) to show that big data adoption in Malaysian manufacturing firms was mostly affected by factors such as perceived advantage, complexity, technology resources, big data quality, and integration. Verma and Bhattacharyya [14] collected data from 22 different businesses and service providers in India to explore the factors influencing the use and adoption of big data analytics among Indian businesses. Through the use of a qualitative approach, it was emphasized that complexity, compatibility, IT aspects, top management support, data environment, cost, external and industry pressures are factors that influence the implementation of big data adoption. Sun, *et al.* [15] made use of the Diffusion of Innovation theory (DOI), the institutional theory, and the Technology–Organization–Environment (TOE) framework to review 62 articles. The findings exposed 26 factors that impact big data adoption. Recently, Baig, *et al.* [16] also used the TOE framework and DOI theory to explore and found 42 important factors influencing BDA. Motau and Kalema [17] used a quantitative approach to assess the readiness of

BDA in South Africa's public service sector and discovered that technological infrastructure, security, reliability, finances, competitors, customers, and vendor support factors are prerequisites for assessing the readiness to analyze big data. Klievink, *et al.* [18] conducted a study to evaluate the readiness of businesses in the Dutch service sector as affected by organizational alignment, maturity, and capabilities. Mneney and Van Belle. [19] evaluated four categories (technology, organization, environment, and task technology fit) that influence retail organizations in South African to apply big data.

Based on the literature review, it can be seen that there are many related studies on the factors affecting the adoption of big data. However, the factors influencing the big data adoption readiness in SME manufacturing and service sectors in South East Asia have not been evaluated. Moreover, the application of Data Mining technique to predict factors affecting readiness to adopt big data is still rare.

3 Methodology

3.1 Data Collection and Sample

Questionnaires were divided into two parts. In the first part, 35 items were included to evaluate factors that impact the readiness to apply big data in SMEs, and 9 items to assess the readiness to apply big data. In the second part, the questionnaires inquired about the firms' information and respondents' socio-economic characteristics. The seven-point Likert scale used ranged from 1 for "strongly disagree" to 7 for "strongly agree". The questionnaire's reliability and validity was assessed through a pilot test. The 30 respondents were composed of big data experts, Chief Executive Officers of companies, and professors. Cronbach's alpha was used to evaluate the internal consistency within the data. The pre-test analysis revealed Cronbach alpha values greater than 0.7 for all questionnaire items. Subsequently, the questionnaire was slightly modified to fit the reality of the company. Responses from this pilot study were not included in the final sample.

The subjects of this study are SMEs which are distributed over six areas of Vietnam's main sub-industries: food and beverages, construction, garment, wholesale, retail, accommodation services survey. The questionnaires were sent by emails to qualified individuals in SMEs in Vietnam. Finally, 240 managers were chosen from manufacturing and service companies, and data were collected at the end of 2020. According to the descriptive statistics, the majority (46.7%) held Bachelor's degrees and 39.2% had Master's degrees or higher. The respondents were 72.5% male and 27.5% female, with more than half (57.9%) between 31 to 45 years of age, and belonging to either small enterprises (82.5%), medium enterprises (17.5%).

3.2 Measurement of the Readiness to Adopt Big Data

Cronbach's α value was calculated to evaluate the reliability of the scale. The analysis results showed that the Cronbach's α value of all variables ranged from 0.626 to 0.867 (Table 1). Hair, *et al.* [20] suggested that Cronbach's α value should be higher than 0.7,

but 0.6 is acceptable. Therefore, the results prove that all variables in the study are consistent and reliable enough for further analysis.

Table 1. Variables and measurement

Variables	References	Item numbers	Cronbach α
Budget resources	[21]	3	0.798
Technical competence	[22]	4	0.867
Data security	[23]	3	0.716
Data quality	New measurement	3	0.691
IT infrastructure	[24]	3	0.798
Relative advantage	[21,25]	4	0.807
Top management support	[26,27]	3	0.707
Organizational culture	[6]	3	0.746
Competitive pressure	[13,15]	3	0.856
Partner pressure	[13,15]	3	0.626
Government support	[1]	3	0.719
Readiness to adopt big data in SMEs	[15,16] and seft-developed	9	0.773

To predict the factors' influence on readiness, the output variable (readiness to adopt big data) was categorized into two levels: high and low. Based on an average of 9 items used to identify the readiness to adopt big data in SMEs, the output variable was code conducted. "High" with the mean value of 9 items higher than or equal to 6.0 and "low" with the mean value of 9 items lower than 6.0. The thirteen input variables include: relative advantage, top management support, organization culture, technical competence, budget resources, data quality, data security, competitive pressure, IT infrastructure, partner pressure, government support, firm size, and type of industry.

3.3 Data Mining Techniques

To classify the readiness of SMEs to adopt big data, this study used three techniques of data mining including CHAID, Bayesian networks, and Neural Network with SPSS modeler 18 software. These algorithms are common algorithms in Data mining techniques.

4 Results

4.1 Prediction Accuracy of Models

In Table 2, the detection sensitivities of CHAID, Bayesian networks, and Neural network classifiers for training data were 83.54%, 79.88%, and 74.39%, respectively. The results of the prediction on the testing data were 71.05%, 81.58%, and 72.37%, respectively. The accuracy of all 3 classification models is higher than 70.00%. According to Tavakoli [28] these models have high prediction accuracy. This suggests that 13 input variables including relative advantage, top management support, organization culture, budget resources, technical competence, data quality, data security, competitive pressure, IT infrastructure, partner pressure, government support, type of industry, and firm size were well identified factors that impact Vietnamese SMEs' readiness to adopt big data.

Table 2. Comparative results of three algorithms

Techniques	'Partition'	Training		Testing		AUC	
						Training	Testing
CHAID	Correct	137	83.54%	54	71.05%	0.907	0.777
	Wrong	27	16.46%	22	28.95%		
	Total	164		76			
Bayesian networks	Correct	131	79.88%	62	81.58%	0.908	0.898
	Wrong	33	20.12%	14	18.42%		
	Total	164		76			
Neural Network	Correct	122	74.39%	55	72.37%	0.827	0.780
	Wrong	42	25.61%	21	27.63%		
	Total	164		76			

Area under the ROC curve (AUC) are used as appropriate measures to evaluate the performance of classification algorithms and have values ranging from 0.5 to 1. The model will be considered to have acceptable discrimination if the AUC value is higher than 0.7 [28]. In this study, the AUC value on the training data ranges from 0.827 to 0.908 and on the testing data range from 0.777 to 0.898 meaning that the models would be considered to have good discrimination [28]. The AUC for training data of CHAID, Bayesian networks, and Neural Network algorithms were 0.907, 0.908, and 0.827, respectively. While the respective figures for testing data were 0.777, 0.898, and 0.780. The results show that the AUC value was highest with Bayesian networks on both training data and testing data models.

4.2 Variables Importance

Table 3 presents the predictive results from the three algorithms for variables that have an important influence on the readiness to adopt big data. The results of predicting the importance of each algorithm are shown in order of importance from the most important to the least important. For example, the results of the CHAID model illustrate that top management support (rank 1) is the most important variable to predict readiness, and IT infrastructure (rank 6) is the least important variable. Similarly, in Bayesian networks and Neural Network models, top management support (rank 1) is the most important variable to predict readiness to adopt big data.

On table 3, based on the output frequency from the three models of the variables evaluated as the most important for predicting readiness, we can see that five variables including top management support, data quality, data security, partner pressure, and budget resources were found as the most important variables to predict the readiness of SMEs in Vietnam to adopt big data.

Table 3. The most important factors in predicting the readiness to adopt big data

CHAID		Bayesian networks		Neural network	
Variables	Rank	Variables	Rank	Variables	Rank
Top management support	1	Top management support	1	Top management support	1
Data security	2	Data quality	2	Data quality	2
Budget resources	3	Organizational culture	3	Partner pressure	3
Data quality	4	Partner pressure	4	Data security	4
Government support	5	Budget resources	5	Budget resources	5
IT infrastructure	6	Government support	6	Competitive pressure	6
		Competitive pressure	7	Technical competence	7
		Data security	8	Government support	8
		Type of industry	9	Type of industry	9
		Firm size	10	Organizational culture	10

5 Discussion

Today's adoption of big data offers many benefits to companies including increased performance, improved strategic direction, development of more reliable customer service, product identification and development, and reducing risks [2,3,29]. The aim of

the study is to predict the factors that significantly influence the readiness of SMEs in Vietnam to adopt big data. Research results show that the five factors that have the most important influence on the readiness to apply big data include top management support, data quality, data security, partner pressure, and budget resource. Top management support plays a key role in the adoption of big data as it guides resource allocation, integration of services, and redesign of processes [29]. Top management has clear goals for adopting big data, and encouraging building a big data-driven decision-making culture that will play an important role in improving the quality of the company's decision-making [27,30]. Therefore, if top management understands the benefits of big data adoption, they will invest resources and encourage employees to implement big data which is consistent with previous researches [15,29]. Data quality directly affects the results of the business analyses [31]. Big data come from different sources such as databases, text data, graph, documents, images, videos, audio files, emails, comments, tags, tweets, and clicks, etc. [6,32]. SMEs are abundant in data sources and high accuracy which should positively contribute to readiness in applying big data as proven by similar studies [29,31]. Data security issues are of particular concern to organizations when adopting big data [33]. Big data includes a lot of personal information and with rapidly increasing capacity, this is considered a valuable data source which may be exploited by unrelated third parties or cybercriminals [33]. Similarly, Motau and Kalema [17] highlighted that it is a precondition factor affecting the readiness to analyze big data. Adopting big data to keep up with partners and maintaining the firm's internal balance with them is a key readiness marker[15,16]. Some empirical research studies have suggested that trading partner pressure is an important determinant for IT adoption and use [34]. Financial resources refer to the enterprise's budget to invest in information technology infrastructure systems, investment in training high-quality human resources capable of analyzing big data, and financial resources to maintain the operating system when the company implements big data deployment. The financial resource is an important factor that influences the application of big data [15].

6 Conclusion and Implications

Predicting the factors that influence the readiness to adopt big data in SMEs is an important contribution to help managers better understand the factors affecting readiness to adopt big data in their enterprise. Hence, managers could build a big data adoption strategy that is right for their business. It will help SMEs increase revenue and consequently increase the contribution to the state budget. This research study is also a valuable reference for government agencies, for example, the Vietnam Association of Small and Medium Enterprises and the Ministry of Investment Planning to encourage the future implementation of big data by enterprises. In addition, service providers also could understand the factors that affect the readiness of enterprises to apply big data thereby also developing a strategy to provide suitable products for businesses. This study is useful for the implementation of BDA by SMEs in Vietnam as well as SMEs in some developing and underdeveloped countries.

References

1. Zhong, R.Y., Newman, S.T., Huang, G.Q., Lan, S.: Big data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. Computers & Industrial Engineering 101, 572-591 (2016).
2. Coleman, S., Göb, R., Manco, G., Pievatolo, A., Tort-Martorell, X., Reis, M.S.: How can smes benefit from big data? Challenges and a path forward. Quality and Reliability Engineering International 32(6), 2151-2164 (2016).
3. Raguseo, E.: Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. International Journal of Information Management 38(1), 187-195 (2018).
4. Le, T.M., Liaw, S.-Y.: Effects of pros and cons of applying big data analytics to consumers' responses in an e-commerce context. Sustainability 9(5), 1-19 (2017).
5. Wang, L., Yang, M., Pathan, Z.H., Salam, S., Shahzad, K., Zeng, J.: Analysis of influencing factors of big data adoption in chinese enterprises using danp technique. Sustainability 10(11), 1-16 (2018).
6. Alharthi, A., Krotov, V., Bowman, M.: Addressing barriers to big data. Business Horizons 60(3), 285-292 (2017).
7. Luna, D., Mayan, J.C., Garcia, M.J., Almerares, A.A., Househ, M.: Challenges and potential solutions for big data implementations in developing countries. Yearbook of medical informatics 9(1), 36-41 (2014).
8. Tabesh, P., Mousavidin, E., Hasani, S.: Implementing big data strategies: A managerial perspective. Business Horizons 62(3), 347-358 (2019).
9. Chau, N.T., Deng, H.: Critical determinants for mobile commerce adoption in vietnamese smes: A conceptual framework. Procedia Computer Science 138, 433-440 (2018).
10. Ministry of Planning and Investment: Vietnamese enterprises white book. Statistical publisher, (2021).
11. Google, Temasek.: E-conomy sea 2018 southeast asia's internet economy hits an inflection point. https://www.thinkwithgoogle.com/_qs/documents/6730/Report_e-Conomy_SEA_2018_by_Google_Temasek_v.pdf accessed 18 April 2021 (2018).
12. Le; Vu, X.-B.B., Nghiem, S.: Technical efficiency of small and medium manufacturing. Economic Analysis and Policy 59, 84-91 (2018).
13. Yadegaridehkordi, E., Hourmand, M., Nilashi, M., Shuib, L., Ahani, A., Ibrahim, O.: Influence of big data adoption on manufacturing companies' performance: An integrated dematel-anfis approach. Technological Forecasting and Social Change 137, 199-210 (2018).
14. Verma, S., Bhattacharyya, S.S.: Perceived strategic value-based adoption of big data analytics in emerging economy. Journal of Enterprise Information Management 30(3), 354-382 (2017).
15. Sun, S., Cegielski, C.G., Jia, L., Hall, D.J.: Understanding the factors affecting the organizational adoption of big data. Journal of Computer Information Systems 58(3), 193-203 (2016).
16. Baig, M.I., Shuib, L., Yadegaridehkordi, E.: Big data adoption: State of the art and research challenges. Information Processing & Management 56(6), 1-18 (2019).
17. Motau, M., Kalema, M.B.: Big data analytics readiness: A south african public sector perspective. 016 IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech), Balaclava, Mauritius, 265-271 (2016).
18. Klievink, B., Romijn, B.-J., Cunningham, S., de Bruijn, H.: Big data in the public sector: Uncertainties and readiness. Information Systems Frontiers 19(2), 267-283 (2016).

19. Mneney, J., Van Belle., J.-P.: Big data capabilities and readiness of south african retail organisations. 2016 6th International Conference - Cloud System and Big Data Engineering. 14-15 January 2016, Amity University, Uttar Pradesh, Noida, India, 279-286 (2016).
20. Hair, Black, W.C., Babin, B.J., Anderson, R.E.: Multivariate data analysis. Peason, (2014).
21. Lian, J.-W., Yen, D.C., Wang, Y.-T.: An exploratory study to understand the critical factors affecting the decision to adopt cloud computing in taiwan hospital. International Journal of Information Management 34(1), 28-36 (2014).
22. Kuan, K.K.Y., Chau, P.Y.K.: A perception-based model for edi adoption in small businesses using a technology-organization-environment framework. Information and Management 38, 507-521 (2001).
23. Ghasemaghaei, M.: The role of positive and negative valence factors on the impact of bigness of data on big data analytics usage. International Journal of Information Management 50, 395-404 (2020).
24. Noonpakdee, W., Phothichai, A., Khunkornsiri, T.: Big data implementation for small and medium enterprises. 2018 27th Wireless and Optical Communication Conference (WOCC), Hualien, Taiwan, 1-5 (2018).
25. Philip Chen, C.L., Zhang, C.-Y.: Data-intensive applications, challenges, techniques and technologies: A survey on big data. Information Sciences 275, 314-347 (2014).
26. Premkumar, G., Roberts, M.: Adoption of new information technologies in rural small business. OMEGA – International Journal of Management Science 27(4), 467-484 (1999).
27. Shamim, S., Zeng, J., Shariq, S.M., Khan, Z.: 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), 103-135 (2019).
28. Tavakoli, K.A., Rabieyan, R., Besharati, M.M.: A data mining approach to investigate the factors influencing the crash severity of motorcycle pillion passengers. Journal of Safety Research 51, 93-98 (2014).
29. Park, J.-H., Kim, M.-K., Paik, J.-H.: The factors of technology, organization and environment influencing the adoption and usage of big data in korean firms. 26th European Regional Conference of the Interational Telecommunications Society, Madrid, Spain, 24-27 June 3(1), 121-129 (2015).
30. Shamim, S., Zeng, J., Shariq, S.M., Khan, Z.: 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), (2019).
31. Kwon, O., Lee, N., Shin, B.: Data quality management, data usage experience and acquisition intention of big data analytics. International Journal of Information Management 34(3), 387-394 (2014).
32. Saggi, M.K., Jain, S.: A survey towards an integration of big data analytics to big insights for value-creation. Information Processing & Management 54(5), 758-790 (2018).
33. Kshetri, N.: Big data′s impact on privacy, security and consumer welfare. Telecommunications Policy 38(11), 1134-1145 (2014).
34. Zhu, K., Kraemer, K.L., Xu, S., Dedrick, J.: Information technology payoff in e-business environments: An international perspective on value creation of e-business in the financial services industry. Journal of Management Information Systems 21(1), 17-54 (2004).