GLOBAL ACADEMIC RESEARCH INSTITUTE

COLOMBO, SRI LANKA

GARI International Journal of Multidisciplinary Research

ISSN 2659-2193

Volume: 05 | Issue: 05

On 31st December 2019

http://www.research.lk

Author: Anuruddika Upul Bandara Ratnamalala

ICBT, Sri Lanka

GARI Publisher | Big Data | Volume: 05 | Issue: 05

Article ID: IN/GARI/ICET/2019/110B | Pages: 11-25 (15)

ISSN 2424-6492 | Edit: GARI Editorial Team

Received: 27.11.2019 | Publish: 31.12.2019

IDENTIFYING THE CHALLENGES OF BIG DATA USAGE IN THE EDUCATIONAL SECTOR OF SRI LANKA

Anuruddika Upul Bandara Ratnamalala

Faculty of Business Management

International College of Business and Technology (ICBT), Sri Lanka

aubupul@gamil.com, aubr84@mun.ca

ABSTRACT

The volume and velocity of data generated have increased substantially in the education sector of Sri Lanka. Although there is an increase, the majority of educational employees do not realize the importance or value of data and often discard without taking into consideration the numerous benefits big data can offer to students and educational institutions. The root causes that contribute to the lack of utilization of big data are mainly due to lack of infrastructure, trained workforce, standards of data usage, security issues, policies, and education of employees. The main objective of this study is to identify the challenges faced by the education sector of Sri Lanka when dealing with big data implementation and further recognize the incentives that encourage the use of big data to support the development, profitability, and decision-making capabilities of Sri Lanka. Qualitative data will be collected from interviews taken from managers, marketing professionals, accountants, and lecturers in various educational institutions across Sri Lanka using snowball and convenience sampling techniques. Furthermore, to establish how educational employees use big data, a survey will be distributed to various educational employees working in different regional educational institutions across Sri Lanka. SPSS is used to determine the correlation between independent variables (infrastructure, trained workforce, policies, education, security, and standards) and dependent variable (data usage). The findings of this research would help educational employees understand the value of big data. Moreover, this research would contribute to improving the profitability and standards of decision-making processes in the education sectors of Sri Lanka.

Keywords: Big data, data usage, education sector, Sri Lanka, data velocity

INTRODUCTION

In today's society, big data is a commonly discussed topic due to the availability and use of data for various purposes in different industries. Big data can be defined in simple terms as large volume of data which are organized and unorganized. Having ample amount of data does not matter for any industry, if there is lack of usage from it. Big data can be used to identify the market and create better strategic decisions that lead to competitive advantage of the business. An industry analyst called Laney (2001) identified three dimensions and properties of big data, namely Volume, Velocity and Variety. Volume is referred to as the amount of data that can be collected from variety of sources which include from daily business transactions, social media activities, data collected from sensor or from data transferred from one machine to another (Laney, 2001). In the past storing

of these huge volume of data would be a great difficulty, but nowadays open source software like Apache hadoop have provided a software framework for storing ample amount of data conveniently. It is expected to increase the data storage by 300 times from 2005 to 2020 which is 43 trillion gigabytes. Most of the companies in the world have a storage capacity of more than 90 Tb of data nowadays.

Nevertheless, velocity can be referred to as data processing speed. It is unbelievable that the way sensors, radio frequency identification and smart metering devices capture information and process them at rapid pace. It is identified that New York stock exchange captures 1TB of trade information daily. Even a vehicle that runs on the road has more than 100 sensors to capture data and drive away from traffic efficiently (Laney, 2001). Finally, variety can be defined the amount of different data available. Data can come in different formats, for instance one format is unstructured data such as financial data, stock ticker, audio, video data and text documents whereas another format is structured data such as numeric data in traditional databases.

In addition to the above mentioned Dong Lanely data classification, two other data dimensions are identified recently. They are Variability and Complexity. Variability means that with the increase of amount of variety of data the data flow can differ by maybe having peaks or by falls (McNulty, 2016). For instance, seasonal data received in online shopping sites tend to differ significantly. Finally, Complexity refers to data that comes from different sources tend to be difficult to managed and linked (Khan et al, 2014). Most difficult task today is to identify the connection that is having certain type of data. Data analyst spend plenty of time sorting these data and making it meaningful to the relevant parties. Data types can be structured, semi structured and unstructured data. This has to be analyzed and sorted in order to make them meaningful.

With the volume, verity, velocity, variability and complexity of data it is difficult to understand what we are going to do with it (McNulty, 2016). Most of the companies in various industries have used these data to take advantages such as cost reduction, new product development, time reduction and smart decision-making, but ample amount of data go to waste without any usage. Although most industries use big data to receive comparative advantage, education sector of Sri Lanka tend to show a negative attitude towards big data usage. Education sector is where there is so many skilled professionals such as doctors, accountants, engineers, mechanics and businessman are made, but attention of big data use seems to be greatly less. In education industry, there is currently a regular interaction of students with the technology. As a result of regular interaction, most of the big data created in educational industry are student assignments records uploaded to the Moodle, teachers notes uploaded to the system, metric systems to monitor students' attendance and performance, digital libraries to issue and return books as well as entrance and graduation records of students (Shacklock, 2016). Each of these mentioned data records is responsible for bringing in many students and developing the education sector in the economy. Big data usage can enable education institutes to retain students, improve the graduation rates, improving quality of the teaching and curriculum (Economist Report, 2008). Furthermore, big data would help to receive necessary funding to archive organizational efficiency (Petkovics et al, 2014). With all the above mentioned benefits, value of big data is not understood by the educational sector employees in Sri Lanka. They tend to discard and ignore the data available without getting proper usage of the data. By realizing that there is a problem to

clarify and verify the availability of problem, pilot study was conducted with 20 employees in education sector of Sri Lanka. The study was done through telephone interview and discussion using convenient and snowball sampling where mostly open-ended questions were asked from participants.

From the Pilot study, it was identified one of the major root cause for ignoring big data is lack of infrastructure. Most of the places in Sri Lanka lacks telecommunication network to link and connect to the required networks to share information. Moreover, special servers and multiple processors are required to handle and store the educational data, but due to the cost factor most of the educational institutes in Sri Lanka cannot afford the servers and processors. Secondly, it was identified that in order to use big data, there should be trained and skilled workforce. These workforce should be aware of the current technological development and should process developer skill set. Not only IT knowledge they also should possess the ability to analyze data using statistics and mathematics. To find the appropriate skills required in the education industry is especially a huge challenge. Only few possess all round skill set and others even lacks ICT knowledge. Therefore, to adopt and use big data is a major challenge in the education industry.

Thirdly, ample amounts of data are captured regularly by the student databases but there is no proper standards of maintaining these data. All the data are stored in one place without any categorization of data. When a particular need arises educational administrators find it difficult to retrieve the data that is relevant to the issue. As a result most of educational administrators tend to discourage the use of big data. Additionally, security issues are a major challenge to the professionals in the educational industry. Mainly, because antivirus software that is need to protect data need regular updating and need up to date protection. If the data protection malfunction the whole system can bring down to a standstill position and hackers could use those information to their own advantage. To purchase and update these virus guards to protect the whole databases in education sector also cost so much money. Therefore, due to limited budget and complexity educational administrators consider maintain security as a challenge to the big data usage. Most of the policies implemented in Sri Lankan educational institutes does not support the big data usage. For instance, there is a policy that student exam records are kept on manual books for maximum 5 years, after that they are considered no longer useful. If a particular student require those details after 5 years destroyed data cannot be retrieved from the system.

Finally, ICT education literacy level of Sri Lanka is declining from 28.5% in 2017 to 27.5% in 2018 from overall population according to department of census of Sri Lanka (Census, 2018). Lack of computer literacy suggests that lesser education level to cope up with huge databases in the society. Therefore, due to lack of education in ICT is a challenge to the use and implementation of big data in education sector in Sri Lanka. It could be seen that there are many root causes that lead to major problem of lack of big data usage in education industry. Neglecting without taking any necessary precautions for improvement of big data usage would impact our standards of education with the world and also would make poor decisions in education industry. As a result, of poor decision making it will weaken the profitability of education services in Sri Lanka. Therefore, necessary corrective actions need to be implemented for improvement of big data usage.

This paper would recognize the importance of big data usage to education sector of Sri Lanka and also would

highlight the need for actions to be taken to encourage big data usage. With the finding of this paper most of the professionals in education sector will realize the value of big data usage which in return improve the decision-making, profitability and the competiveness in the education sector of Sri Lanka. Primary objective of this paper is to determine the main challenges of big data usage in education sector Sri Lanka. Secondly, to evaluate and see weather identified Challenges have any correlation with the big data usage. Furthermore, to understand whether the realized challenges would continue to be challenges in future as well. Finally, to recognize the recommend actions to be implemented in order to encourage data usage.

With lack of big data usage becoming a major problem the research would identify what are the main challenges faced by education sector of Sri Lanka of big data implementation? Further, need to determine whether there is a relationship between the identified challenges and big data usage? Additionally, to predict whether identified challenges would contribute to the future big data usage? Finally, this research would try to recognize what are the appropriate recommend solutions in order to encourage big data usage in education industry Sri Lanka?

This paper mainly focuses on answering the above mentioned questions. The boundaries are set forth in order to achieve a more valid and accurate outcome. Research mainly uses managers, marketing professionals, accountants and lecturers in selected areas such as Kurunegala, Kandy and Colombo in Sri Lanka within 8-year time period (2010-2018).

LITERATURE REVIEW

There are various advantages of utilizing big data, however it can be seen that the implementation of big data is a major challenge for the different sectors in one's economy. Hence, empirical evidences from various studies have highlighted the actual and perceived challenges of big data usage in diverse industries. According to Stephen Kaisler et al. (2014), there are numerous issues and challenges of moving forward with big data implementation for education sector as well as some other sectors. Some of the most common challenges that were highlighted are issues of storage, attitudes of management, affordability, issues of processing, not having tools to analyze and also not having powerful algorithms to sort the big data from their systems or processes.

Other different studies conducted by Daniel (2014); Shitut (2017); and Kernochan (2013) also emphasize certain challenges that act as barriers for the implementation of big data systems. Mainly, they have identified such challenges as difficulty of acceptance (most of the management are not willing to change from traditional systems of data analytics), difficulties in accessing relevant data (sorting from ample amount of data and obtaining required data is a problem), obtaining expertise knowledge to operate (lack of relevant skills is a problem) and barriers in the organization environment (linking all the departments for big data implementation is a problem).

Another study by Long and Siemens (2011) demonstrate that not accurate data, mismanagement of data, culture, privacy issues, lack of skills, not enough return on investment, lack of training and resources as well as difficulty in data standardization are major challenges in application of big data in the industry. Further, studies conducted by authors such as Dan and Roger (2010); Jayasree (2013); and Rachana and Guruprasad (2014) debate that security, reliability, data quality, cost,

performance and data storage facilities are key issues in big data implementation in major economies. Some of the major challenges that are identified from different literatures have been discussed below for better understanding.

Privacy and security

Most of empirical studies have identified that when it comes to raw data and information, securing and the privacy of data are as a biggest threat for any industry and any organization. When ample amount of data are used for various purposes they should be securely stored in order to prevent it from usage of various third parties. Ferguson (2017) identifies that some data are valuable for making vital decisions in the organization, but there are possibility of targeted cybercrimes and hacking of such valuable data due to competitive edge and various other reasons. Eynon (2013) also suggests that fear of misuse of valuable data lead to lack of implementation of big data in most of the companies. Additionally, a comprehensive study conducted by Broeders et al. (2017) also state that misuse of sensitive data and information fraud is a major concern in big data utilization. Hence, it is evident that majority of organizations have fear of using big data due to the lack of privacy and security of valuable data and information.

Infrastructure

Infrastructure is a vital component in big data usage, especially in today's society. According to Barroso, Clidaras and Hölzle (2013) telecommunication is considered as a vital infrastructure or a factor for implementing big data projects. In order to connect with big data storage systems such as cloud systems, telecommunication systems are considered highly necessary for any organization. Furthermore, specially designed architecture is required to process millions of nodes with multiple of disks and processors at high internet connection speed (Shapiro and Varian, 2010). Moreover, maintaining all the infrastructure leads to number of side effects such as rise of huge costs and additional resources and support systems. Hence, only large companies have the necessary capacity to bear those costs and other required resources (Barroso, Clidaras, and Hölzle, 2013). Therefore, obtaining necessary infrastructure to implementing big data usage is a massive challenge in any industry.

Trained workforce

According to Hilbert (2013), proper adaptation is required to big data project implementation. For a proper adaptation of big data software, it is necessary to acquire trained workforce within the organization. He also suggests that if data is distributed in several clusters, expertise knowledge can be obtained from several communities as well. For instance, companies like Google and Facebook use open source software like Apache hadoop to distribute their data among several clusters and obtain necessary expertise knowledge (Hilbert, 2013). Hence, it can be seen that it is essential to have a trained workforce with some expertise knowledge in order to implement the big data systems and processes in organizations.

Education level of employees

According to Villars et al. (2011), it is crucial to acquire adequate knowledge of information technology and application developing skillset in order to effectively apply big data systems and processes. People who have the relevant knowledge are scares and difficult to acquire today in terms of information technology, especially when it comes to developing countries like Sri Lanka. Moreover, studies piloted by Manyika et al (2011) suggest that key subject areas such as statistics, mathematics and computer science are needed to be enhanced regularly by employees in order effectively integrate with big data implementation. His study prove that

having lack of knowledge in the mentioned subject areas would make employees difficult to analyze and interrupt the big data and data would be meaningless to conduct the further business operations. Thus, in order to implement the big data systems successfully, organizations should have right skillset and knowledgeable employees who can be identified as difficult to acquire in today's job market.

Policies and standards of data usage

Majority of empirical evidences have identified another challenge of implementing big data systems which is lack of policies and standards of data utilization. Protecting data privacy in health services report (2000) stresses out the necessity of collecting data in a standard procedure, mainly because of big data projects are vulnerable for possible security threats. Report also elaborates that strategies should be implemented by companies as a part of their policy guidelines and code of conduct in order of methods data should be collected, how should be protected and the procedure they should be used. Moreover, the action plan needs to be constructed in order of how data should be managed in order to protect from possible security breaches and data losses. In addition, Campbell (2007) identifies regulatory framework is necessary to protect against possible data breaches and security losses and also would create trust among the users of big data, if there is proper regulatory procedures.

THEORETICAL FRAMEWORK OF BIG DATA USAGE

Technology Acceptance Model (TAM) which was developed by Davis in 1989. This theory determines that new technology is mainly accepted by users due to perceived usefulness and the ease of use. When people perceive that new technology is not beneficial for them and complicated to use they tend to neglect the acceptance of usage of new technology (Davis, 1989). Therefore, from this theory we can deduce that perceived usefulness and ease of use contribute to acceptance of big data usage and implementation in organizations.

Technology Task Fit Model (TTF) assumes that user require necessary technology that fit their working environment to increase employee performance as well as organizational performance. When user lacks the necessary equipment, they find it difficult to perform to the expected standards by the organization which creates a visible gap of expected performance and actual performance. So in that sense, the performance is heavily correlated with the appropriateness of technology (Goodhue, and Thompson, 1995). Therefore, it can be recognized from technology task fit model that employees require necessary equipment in order to boost their actual performance levels to be successfully implement the big data systems within their working environment. Theory of Planned Behavior (TPB) which was developed by Icek Ajzen in 1988 which predicts the behavioral intention of humans to make decisions. Main factors that lead to behavior intention are attitudes, subjective norms and perceived behavioral control. People tend to make positive decisions based on agreement of all the factors mentioned (Ajzen, 1988). For instance, if people have positive attitudes, cultural views tend to support and also if they believe it is easy to use, then people tend to accept new technologies, particularly like big data processes. Hence, it is vital to understand the perception and intention of management as well as employees when implementing the big data processes and systems in the organization.

Unified Theory of Acceptance and Use of Technology (UTAUT) is a developed

theory that discusses the intention to use technology in any industry (Venkatesh et al, 2003). This theory identifies behavioral intention of people to use the different systems is mainly driven by expectancy of the performance of the system, amount of effort, influences by the society and conditions that support the system. Thus, a study by Venkatesh et al. (2003) state that people's behavioral intention to use the system diminishes, if any of the conditions mentioned do not provide the necessary facilitation. Therefore, it can be identified that there has to be a clear and proper intention of implementing big data systems in every industry as well as the people who relates to it should be clearly understood these intentions to productively utilize the big data in their daily operational activities.

Motivation model (MM) theory was introduced in 1992 by Davis, Bagozzi and Warshaw in order to identify the adaptation and utilization of information and communication technology within an organizational environment. They stress out that usage is mainly depend on intrinsic and extrinsic motivators (Davis et al, 1992). Intrinsic motivators arises from person's inner motive to perform a task, for example satisfaction received from the computer usage is an intrinsic motivator. Extrinsic motivator arises from outside of a person, for instance factors from the society, perceived benefits and usefulness encourages people to use technology (Davis et al, 1992). Therefore, it can be determined that according to this theory intrinsic and extrinsic factors motivates people to use and implement big data for their daily operations.

Model of PC Utilization discusses about leading factors that lead to utilization of technology. Main factors such as social factors, complexity, job fit, long term consequences, affect towards PC usage and fascinating conditions greatly affect technology usage (Thompson et al, 1991). According to this study by Thompson et al (1991), they believe that people tend to absorb and adopt technology based on certain factors such as; if the social factors support, technology is less complex to use, day-to-day tasks fit their working environment, less consequences of personal computer usage and there should be supporting conditions to use personal computers. Hence, it could be understood all these six factors can contribute the big data usage and implementation in an organization.

METHODOLOGY

Based on the analysis of literature review and pilot study conducted, this research investigates whether the identified challenges can impact the big data usage in the education sector of Sri Lanka. Consequently, big data usage in education sector in Sri Lanka is considered as dependent variable and lack of infrastructure, trained workforce, policies, ICT education and lack of security is identified as independent variables. The research mainly tests weather there is a relationship between the independent and dependent variables and also research tries to predict whether the particular identified connection would continue to the future using regression analysis.

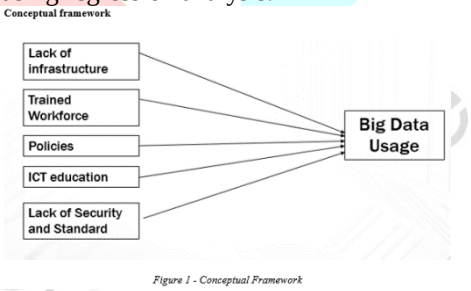

Figure 1 - Conceptual Framework

List of Hypothesis

H0: There is no relationship between Infrastructure and big data usage in Sri Lanka.

H1: There is relationship between Infrastructure and big data usage in Sri Lanka.

H0: There is no relationship between trained workforce and big data usage in Sri Lanka.

H2: There is a relationship between trained workforce and big data usage in Sri Lanka.

H0: There is no relationship between policies and big data usage in Sri Lanka.

H3: There is relationship between policies and big data usage in Sri Lanka.

H0: There is no relationship between ICT education and big data usage in Sri Lanka.

H4: There is a relationship between ICT education and big data usage in Sri Lanka.

H0: There is no relationship between security and standards and big data usage in SL.

H5: There is relationship between security and standards and big data usage in SL.

Population, Sample size and sample selection

It was identified that approximately 120,000 private sector employees working in education sector in Sri Lanka. Managers, marketing professionals, accountants, and lecturers in various fields are considered to be working in educational industry in Sri Lanka. From the whole population, 130 employees working in different sectors are selected using simple random sampling technique. Areas such as Colombo, Kandy and Kurunegala are selected for the research depending on the convenience.

Data collection, Questionnaire type, time and Data analyzing technique

Data collection was done using an online questionnaire mainly because it could be easier to reach a widespread population. Selected all 130 employees have access to internet and they could answer the questions in their own spare time. Questionnaire was designed in a simple manner with mostly close-ended questions. Questions are setup using Likert scale that ranges from 1=strongly agree, 2= agree, 3=neither agree nor disagree, 4= disagree and 5=strongly disagree. Questionnaire is designed in a manner to derive at specific answer in the questionnaire. This research is conducted within time frame of 01st October 2018 to 28th February 2019. Data analysis is done using IBM SPSS software to facilitate correlation and regression.

Reliability test

Cronbach's Alpha reliability test was conducted in order to measure the complete stability of the outcomes of the statements in the questionnaire involving to the variables highlighted. It was found that more than 90% of data are reliable and satisfactory.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items
.974	.978

Table 1 - Reliability Statistics

DATA ANALYSIS

Correlation analysis

Pearson Correlation is used to identify weather there is any relationship between independent and dependent variables. In Pearson correlation, coefficient range determines the type of relationship between variables which further depicts in the following table.

Coefficient Range	Interpretation
0.90 to 1.0	Very strong positive correlation
-0.90 to -1.0	Very strong negative correlation
0.70 to 0.90	High positive correlation
-0.70 to -0.90	High negative correlation
0.50 to 0.70	Moderate positive correlation
-0.50 to -0.70	Moderate negative correlation
0.30 to 0.50	Low positive correlation
-0.30 to -0.50	Low negative correlation
0.0 to 0.30	Negligible correlation
-0.00 to -0.30	

Table 2 - Coefficient range for correlation

Regression analysis

Simple regression analysis examines when the value of dependent variable Y (big data usage) can be effectively predicted using the Independent variable X (Infrastructure, trained workforce, policies, ICT education and security and standard).

Relationship between Infrastructure on big data usage in Education sector Sri Lanka

Correlations

		Infrastructure at the educational sector has an effect on big data usage.	There are significant challenges on implementation on big data usage in educational industry
Infrastructure at the educational sector has an effect on big data usage.	Pearson Correlation	1	.810**
	Sig. (2-tailed)		.000
	N	130	130
There are significant challenges on implementation on big data usage in educational industry	Pearson Correlation	.810**	1
	Sig. (2-tailed)	.000	
	N	130	130

**. Correlation is significant at the 0.01 level (2-tailed).

Table 3 - Correlation between infrastructure and big data usage in Education sector

Based on the above analysis it indicates that Pearson correlation of r value is 0.810 which implies that there is a high positive correlation between variables. Moreover, Sig. (2-tailed) value is 0.000 which is less than 0.05 this means when one variable increase or decreases another variable also changes significantly. From the hypothesis, it can reject null hypothesis and accept H1: which shows there is a relationship between infrastructure and big data usage. Therefore, since there is a positive correlation, it suggests when there is lack of infrastructure there is lack of big data usage and also when people are provided with proper infrastructure, the big data usage tend to be high.

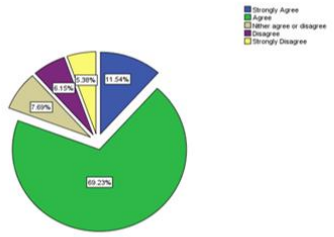

Figure 2 - Opinions about infrastructure and big data usage in Sri Lankan Education sector

According to the analysis of survey 69.23% agreed better infrastructure is required for higher big data usage in educational sector. Moreover, 11.53% of educational sector employees disagreed that infrastructure lead to improvement big data usage. Overall, majority agreed that infrastructure is a requirement in big data usage.

Regression analysis infrastructure on big data usage

Coefficients*

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.765	.085		9.008	.000
1	Infrastructure at educational sector has an effect on big data usage.	.546	.035	.810	15.626	.000

a. Dependent Variable: There are significant challenges on implementation on big data usage in educational industry

Table 4 – Regression on infrastructure and big data usage in Sri Lankan Education sector

Regression equation to determine Impact for big data usage from infrastructure is determined as follows: Y=0.546X+0.765. Where X is the people viewpoint of the need for infrastructure where Y is the change of big data usage. With the regression equation identified it could be determined that in future people believe proper infrastructure is necessary for big data adaptation.

Relationship between trained workforces on big data usage in education sector

Correlations

		ICT Trained workforce at the education sector has effect on big data usage	There are significant challenges on implementation on big data usage in educational industry
ICT Trained workforce at the education sector has effect on big data usage	Pearson Correlation	1	.834**
	Sig. (2-tailed)		.000
	N	130	130
There are significant challenges on implementation on big data usage in educational industry	Pearson Correlation	.834**	1
	Sig. (2-tailed)	.000	
	N	130	130

**. Correlation is significant at the 0.01 level (2-tailed).

Table 5 - Correlation between trained workforce and big data usage in Education sector

From the information Pearson R value is 0.834 which implies there is a high positive correlation. Additionally, Sig. (2-tailed) value is .000 which is less than 0.05 which suggest when one variable change other also changes positively. When we consider hypothesis, we can reject null hypothesis and accept H2: There is a relationship between trained workforce and big data usage in Sri Lanka. From all the information it could be determined that trained workforce is necessary for big data implementation.

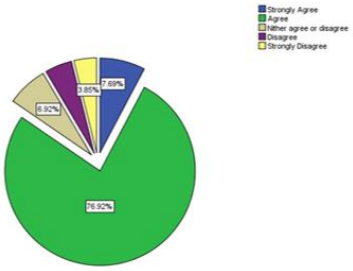

Figure 3 - Opinions about trained workforce and big data usage in Education sector

Information from the questionnaire suggest that 76.92% implies that training is required for big data adaptation. Furthermore, only 8.47% disagree that training is necessary for big data implementation. Majority agree that training is a vital for big data implementation.

Regression analysis trained workforce on big data usage

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.552	.090		6.167	.000
	ICT Trained workforce at the education sector has effect on big data usage	.655	.038	.834	17.109	.000

a. Dependent Variable: There are significant challenges on implementation on big data usage in educational industry

Table 6 – Regression on trained workforce and big data usage in Education sector

Regression equation to identify whether training is necessary in future for big data usage is determined as follows: Y=0.655X+0.552. From the equation it could be determined that even in the future people believe that trained workforce is necessary for big data usage.

Relationship between applied policies on big data usage in Sri Lankan education

Correlations

		Applied Policies in the education sector has effect on big data usage	There are significant challenges on implementation on big data usage in educational industry
Applied Policies in the education sector has effect on big data usage	Pearson Correlation	1	.777**
	Sig. (2-tailed)		.000
	N	130	130
There are significant challenges on implementation on big data usage in educational industry	Pearson Correlation	.777**	1
	Sig. (2-tailed)	.000	
	N	130	130

**. Correlation is significant at the 0.01 level (2-tailed).

Table 7 - Correlation between policies and big data usage in Sri Lankan Education sector

From the information it could be determined that Pearson R value is 0.777 which indicates that high positive correlation. Additionally, sig (2-tailed) value is .000 which is lesser than 0.05 this indicates when policies changes big data usage also changes relatively. Moreover, we can reject null hypothesis and accept H3: there is relationship between policies and big data usage in Sri Lanka. Further, it could be determined lack of policies leads to lack of big data usage and also better policies leads to higher big data usage.

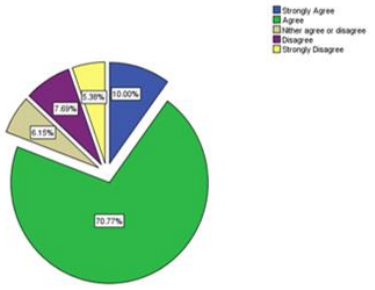

Figure 4 - Opinions about applied policies and big data usage in Education sector

Information obtained from the questioner suggest that 70.77% agree that proper policies should be implemented in order for effectively promote big data usage. Moreover, 13.07% disagree that proper policies are necessary for encourage big data usage. Overall, majority decides that proper policies are crucial to encourage big data usage.

Regression analysis policies on big data usage

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.809	.092		8.834	.000
	Applied Policies in the education sector has effect on big data usage	.520	.037	.777	13.981	.000

a. Dependent Variable: There are significant challenges on implementation on big data usage in educational industry

Table 8 – Regression on policies and big data usage in Sri Lankan Education sector

From the table above we can arrive at the regression equation as follows: Y=0.520X+0. 809. This regression equation indicates that people believe when applied policies increases big data usage also increases. It also predicts that in future employees believe proper policies is necessary for improve big data adaptation.

Relationship between ICT Education on big data usage in Education sector

Relationship between ICT Education on big data usage in Education sector

Correlations

		ICT Education is required in the education sector for the big data usage	There are significant challenges on implementation on big data usage in educational industry
ICT Education is required in the education sector for the big data usage	Pearson Correlation	1	.835**
	Sig. (2-tailed)		.000
	N	130	130
There are significant challenges on implementation on big data usage in educational industry	Pearson Correlation	.835**	1
	Sig. (2-tailed)	.000	
	N	130	130

**. Correlation is significant at the 0.01 level (2-tailed).

Table 9 - Correlation between ICT education and big data usage in Education sector

Pearson R value 0.835 indicates high positive correlation between ICT education and big data usage. Additionally, Sig (2-tailed) value 0.000 is lesser than 0.05 suggests when changes in ICT education impacts on big data usage. It can reject null hypothesis and accept H4: there is relationship between ICT and big data usage. Overall, data indicates that lack of ICT education lead to lack of big data usage and vice-versa.

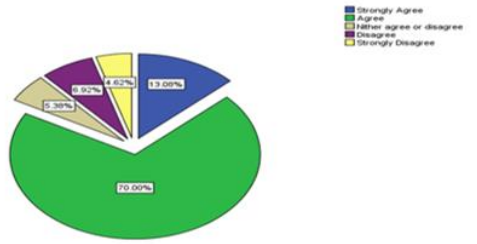

Figure 5 - Opinions about ICT education and big data usage in Education sector

70% of respondents agree that ICT education is required to improve big data usage in education sector in Sri Lanka whilst 11.54% disagree on that statement. Overall, majority accepts that education is a requirement for big data usage in Sri Lanka.

Regression analysis ICT education on big data usage

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.735	.079		9.261	.000
	ICT Education is required for big data usage in education	.572	.033	.835	17.173	.000

a. Dependent Variable: There are significant challenges on implementation on big data usage in educational industry

Table 10 - Regression on ICT education and big data usage in Education sector

From the regression analysis the equation that can be arrived is: Y=0.572X+0. 735. From the equation it can be predicted that people believe when ICT education increases the big data usage also increases.

Relationship between security and standards on big data usage in Sri Lankan Education sector

Correlations

		Lack of Security and standards in the education sector has effect on big data usage	There are significant challenges on implementation on big data usage in educational industry
Lack of Security and standards in the education sector has effect on big data usage	Pearson Correlation	1	.972**
	Sig. (2-tailed)		.000
	N	130	130
There are significant challenges on implementation on big data usage in educational industry	Pearson Correlation	.972**	1
	Sig. (2-tailed)	.000	
	N	130	130

**. Correlation is significant at the 0.01 level (2-tailed).

Table 11 - Correlation between security and standards on big data usage in Education sector

Pearson R value is 0.972 which indicates very strong positive correlation between security standards and big data usage. Additionally, Sig. (2-tailed) value is 0.000 which is lesser than 0.05 that suggest security standards improve big data usage also changes positively. Hence, we can reject null hypothesis and accept where H5: There is relationship between Security Standards and big data usage. Finally, analysis depicts lack of higher security and standards leads to Lack of big data usage and also high security and standards leads to higher big data usage.

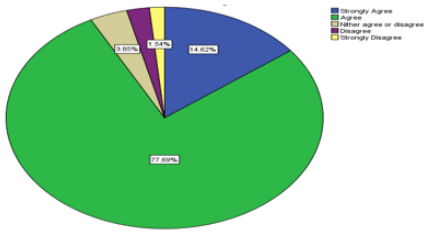

Figure 6 - Opinions about security and standards on big data usage in Education sector

Above pie chart illustrates 77.69% agreed that lack of security and standard in the education sector lead to lack of big data usage whereas only 3.84% disagreed that security and standards are essential for big data usage.

Regression analysis of security and standards on big data usage

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.117	.042		2.782	.006
	Lack of Security and standards in education has effect on big data usage	.945	.020	.972	46.697	.000

a. Dependent Variable: There are significant challenges on implementation on big data usage in educational industry

Table 12 - Regression on security and standards on big data usage in Education sector

From the information we can derive at the equation Y=0.945X+0.117. It can be understood from the equation that employees in education sector Sri Lanka believe that in future they might consider security and standards as a necessity for big data usage.

DISCUSSION

This study clearly identified that infrastructure, trained workforce, security standards, policies and education of employees are required for big data utilization. Firstly, it is identified that ICT infrastructures such as computers, processors, storage facilities and network

facilities are considered vital for education sector to use big data facilities. The data clearly supports that without proper infrastructure, big data implementation is difficult by indicating Pearson R value 0.810 and Sig.2 (tailed) value is 0.000 which is less than 0.5. Furthermore, education sector employees in Sri Lanka would consider that infrastructure is a necessity in future for big data usage according to regression equation: Y=0.546X+0.765. Secondly, it is understood that training the workforce in education sector of Sri Lanka is necessary in order to work with big data and also to encourage big data usage. When workers are trained properly, they will understand how to work with appropriate systems effectively without impacting the whole process. With the Pearson R value identified as 0.834 and sig(2-tailed) value 0.000 which is lesser than 0.5 this suggest that there is high positive correlation between train workforce and big data usage. Furthermore, with the regression equation Y=0.655X+0.552, it could be predicted that education sector employees in Sri Lanka feel that training workforce is a must to encourage big data usage. Thirdly, it is recognized that proper policies need to be implemented by government to encourage big data usage among education sector employees in Sri Lanka. For instance, simple rules and policies need to be implemented to data recording, storing, processing and accessing data so that education sector employees would be motivated to use data. From the data analyzed Pearson R value is identified as 0.777 and Sig (2-tailed) value 0.000 this implies that there is high positive correlation between policies and big data usage. Moreover, education sector employees believe that in future they need simple and supportive policies to encourage big data usage according to regression equation: Y=0.520X+0.809. Moreover, ICT knowledge and skills in Sri Lanka is considered at a lower level, hence it will heavily impact on managing big data. From the information analyzed Pearson R value is at 0.835 and Sig (2-tailed) value is 0.000 which suggest high correlation between ICT education and big data usage in Sri Lanka. This implies knowledge and skills are highly required to implement big data. Even for the future education sector employees will consider knowledge and skills as a requirement for them to implement big data. Finally, education sector employees in Sri Lanka believe that security and standards are a necessity and constantly need to be updated in protecting systems against harmful viruses to effectively implement big data usage. According to opinions given, there is tendency of stealing valuable data and also misuse of data so security and standards are considered highly important for big data usage in Sri Lanka. This is clearly suggesting that from the analyzed data where it shows Pearson R value 0.972 and Sig (2-tailed) value 0.000. Most importantly, according to the regression equation Y=0.945X+0.117 education sector employees perceive that security standards are necessary for big data usage in the future as well.

CONCLUSION

This research mainly focuses on identifying challenges that impact on big data usage in education sector in Sri Lanka. By referring to previous literatures, conceptual model was developed based on identified challenges of big data adaptation. To test whether these challenges essentially impact the big data usage, a quantitative analysis technique was used. To collect the data well-structured online questionnaire was distributed among 130 employees. Obtained information is analyzed through IBM SPSS software to find the correlation and regression. This analysis revealed that Sri Lankan authorities need to focus on infrastructure, trained workforce, security

standards, policies and education of employees in order to improve the big data usage. If required actions are not taken to improve the identified challenges, there will be a future trend of lack of big data usage in education sector of Sri Lanka. Moreover, there will be further consequences such as lack of big data usage for decision-making and impact profitability of education sector in Sri Lanka. Additionally, world is becoming globalized rapidly and there will be so much competition from other educational institutes in the world. Thus, big data usage would allow them to effectively understand and adopt to the market changes. Nonetheless, failure to use big data would cause most of the educational institutes in Sri Lanka to drive away from the international market. Finally, Sri Lankan economy is heavily depended on education sector for generating knowledgeable workforce that would contribute to the economy. If proper attention is not given by authorities to improve big data usage, the efficiency and productivity of education sector would hinder. Therefore, it is highly advisable that effective big data usage should be encouraged by overcoming the challenges identified in the research.

RECOMMENDATION

Based on the findings and conclusion, infrastructure, training, policies, ICT education and security should be improved to encourage big data usage. Firstly, Infrastructure can be improved by use of software like Apache Hadoop which is a free open source software that could be used to store and process ample amount of data conveniently. Additionally, government should provide network facilities to all regional and rural areas. Government should also encourage local innovators by providing funding to develop analytic tools for data storing and processing. Moreover, government should develop a system where all educational institutes could purchase necessary infrastructure required for big data implementation at a lesser rate. Secondly, training of workers can be enhanced by providing a system where education sector employees could connect with industry professionals regularly to receive training about big data usage. Another way to increase big data usage is by training the amount of data scientist in the society. When amount of data scientist increases, other parties could obtain the guidance about the complex areas of big data usage. Thirdly, education institutes should enhance policies to encourage big data usage. Education system can change from manual systems to technological systems to record, store and analyze data. Technological systems would support big data usage. Next, Sri Lankan government should establish institute and regulatory frameworks to ensure the privacy and security of sensitive data due to the higher significance of big data utilization. Proper framework might gain trust among top managers of educational institutes to implement big data, mainly because they do not have to worry about the misuse of data and viruses that cause harm to sensitive data. Finally, education of employees needs to be enhanced to encourage use of big data. Government can incorporate computer science, statistics, and mathematics into Sri Lankan educational curriculum from ordinary level to university level. Moreover, implement strategic partnerships with private and public institutions with expertise in big data tools and techniques which allows to facilitate use of big data. Therefore, it could be highly understood necessary precautions and steps need to be taken earliest in order to encourage and motivate educational institutes in Sri Lanka to properly utilize big data in their daily operational activities.

REFERENCES

Ajzen, I. (1988). Attitudes, Personality, and Behavior. The Dorsey Press, Chicago.

Ali-ud-din Khan, M., Uddin, M.F. and Gupta, N. (2014). Seven Vs of Big Data: Understanding Big Data to extract Value. In 2014 Zone 1 Conference of the American Society for Engineering Education (ASEE Zone 1), pp. 3-5.

Barroso, L., Clidaras, J. and Hölzle, U. (2013). The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines, Second edition. Synthesis Lectures on Computer Architecture, 8 (3), pp. 1-154.

Broeders, D., Schrijvers, E., van der Sloot, B., van Brakel, R., de Hoog, J. and Hirsch Ballin, E. (2017). Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data. Computer Law & Security Review, 33 (3), pp. 309-323.

Campbell, A.V. (2007). The Ethical Challenges of Genetic Databases: Safeguarding Altruism and Trust. Kings Law, 18 (2), pp. 227–45.

Dan, S and Roger, C. (2010). Privacy and consumer risks in cloud computing. Computer Law and Security Review, Vol 26, pp. 391-397.

Daniel, B. (2014). Big Data and analytics in higher education: Opportunities and challenges. British Journal of Educational Technology, 46 (5), pp. 904-920.

Davis, F. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13 (3), pp. 319.

Davis, F., Bagozzi, R., & Warshaw, P. (1992). Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. Journal of Applied Social Psychology, 22 (14), pp. 1111–1132.

Department of Census and Statistics Sri Lanka. (2018). Computer Literacy Statistics – 2018 (First six months). pp. 1-4.

Economist Report (2008). The future of higher education: how technology will shape learning. A report from the Economist Intelligence Unit.

Eileen McNulty, E. (2016). Understanding Big Data: The Seven Vs. [online] Data Economy. Available at:

Eynon, R. (2013). The rise of Big Data: what does it mean for education, technology, and media research?. Learning, Media and Technology, 38 (3), pp. 237–240.

Ferguson, A. (2017). Policing Predictive Policing. Wash Univ Law Rev. pp. 211–68.

Goodhue, D. and Thompson, R. (1995). Task-Technology Fit and Individual Performance. MIS Quarterly, 19 (2), pp. 213.

Hilbert, M. (2013). Big Data for Development: From Information - to Knowledge Societies. SSRN Electronic Journal.

Jayasree, M. (2013). Data Mining: Exploring Big Data Using Hadoop and Map Reduce. International Journal of Engineering Science Research - IJESR, 04 (1).

Kaisler, S., Armour, F., Espinosa, J. A., Money, W. (2014). Big Data: Issues and Challenges Moving Forward. 46th Hawaii International Conference on System Sciences, 46 (1).

Kernochan W. (2013). 4 barriers to big data success and ways to overcome them. Enterprise Apps Today.

Laney D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety [Online]. META Group. Pp. 1-4.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Angela Hung Byers, Mckinsey Global Institute and Al, E. (2011).

Petkovics, I., Tumbas, P., Matković, P. and Baracskai, Z. (2014). Cloud computing support to university business processes in external collaboration. 11 (3), pp. 181-200.

Protecting data privacy in health services research. (2000). Washington, D.C.: National Academy Press.

Shacklock, X. (2016). From Bricks to Clicks. The potential of data and analytics in higher education (Report). Higher Education Commission: UK.

Shapiro, C. and Varian, H.R. (2010). Information rules : a strategic guide to the network economy. Boston, Mass.: Harvard Business School Press.

Shitut, N. (2017). 5 Skills You Need to Know to Become a Big Data Analyst. [Online]

Thompson, R., Higgins, C., & Howell, J. (1991). Personal Computing: Toward a Conceptual Model of Utilization. MIS Quarterly, 15 (1), pp. 124–143.