Journal of Theoretical and Applied Information Technology
30th November 2016. Vol.93. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BIG DATA ANALYTICS IMPLEMENTATION FOR VALUE DISCOVERY: A SYSTEMATIC LITERATURE REVIEW

[1]CECILIA ADRIAN, [2]FATIMAH SIDI, [3]RUSLI ABDULLAH, [4]ISKANDAR ISHAK, [5]LILLY SURIANI AFFENDEY, [6]MARZANAH A. JABAR

[1,3,6]Department of Information System and Software Engineering, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia

[2,4,5]Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia

Corresponding Author: fatimah@upm.edu.my

ABSTRACT

The growing number of big data technologies and analytic solutions has been developed to support the requirement of big data implementation. The capability of analyzing big data becomes critical issues in the big data implementation because the traditional analytics tools are no longer suitable to process and analyze the massive amount and different types of data. In the recent years, technological issues and challenges on big data adoptions have been actively conducted globally. However, there are still lacking of studies on how big data implementation can derive and discover values for better decision making. The intent of this review is to investigate the capability components for Big Data Analytics (BDA) implementation towards value discovery. Based on this investigation, it was found that the capability components that may impact value discovery is formulating big data framework that includes the enabler technology and processing and using sufficient analytic techniques for analysing big data.

Keywords: *Big Data Analytics Implementation, Capability Components, Processing, Analytics Techniques, Value Discovery*

1. INTRODUCTION

Today, data has grown huge and uncertain that forced organizations to invest more in the new Information Technology (IT) infrastructures and skillful data scientists to oversee the big data issues. Likewise, the rapid growth of big data characteristics in several domains such as electronic commerce, finance institution, insurance, medical and government has created better opportunity and gain insight into the knowledge discovery and value creation [1]. The advancement of Big Data Analytics (BDA) also stimulate organizations to grasp business value opportunities. The initial understanding of Big Data was reported by [2] who defined its characteristics as the 3Vs (Volume, Velocity, and Variety) that make it difficult for an organization to manage, analyze, and extract value using the current or traditional methods and systems. In addition, the definition of Big Data has been improved to 4Vs due to the value by the advancement with additional characteristics 'Value' [3]. Manyika et al. [4] described that big data can create significant values for the world economy by enhancing the productivity and competitiveness of the companies and public sectors and creating a substantial economic surplus for consumers. They further suggested that these value creations were generated from Big Data insights which include:

- Creating transparency;
- Enabling experimentation to discover needs, exposed variability and improve performance;
- Segmenting populations to customize actions;
- Replacing/supporting human decision making with automated algorithms; and
- Innovating new business models, products and services.

BDA was derived from an analysis execution of a huge data in order to unmask the valuable patterns or information [5]. It has a major impact in providing insights into the advance decision-making in the comprehensive statistical and analytical manner to the organization [6]. The creation of BDA involves the process of using algorithms that are running on the powerful

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

supporting platforms to discover the hidden pattern and unknown potential big data.

In general, organizations would gain potential value from data analytics in one of these three techniques: descriptive, predictive and prescriptive analytics [7], [8]. The processes of BDA involve data selection, pre-processing, analysis, and visualization, and result in interpretation [9]. Thus far, BDA is a very dynamic research field and its application is still in its infancy. The objective of this paper is to investigate the capability components of BDA implementation and its relation to derive the value discovery for the big data organization. It also facilitates the researchers on the potential future works related to the value discovery in BDA implementation.

This paper is organized as follows: Section 2 contains the review methodology of the systematics review process and followed by Section 3 which presents the extracted information analyzed to answer the research questions. Section 4 contains the open research issues. Finally, Section 5 serves as the conclusion.

2. REVIEW METHODOLOGY

The review strategy is subject to the System Literature Review (SLR) guidelines for Software Engineering by [10] and [11]. The SLR consists of three stages that include planning and conducting the review and formatting the report as shown in Figure 1. Furthermore, the review protocol comprises the structuring of research questions conducting search strategy, study selection, data synthesis and report writing.

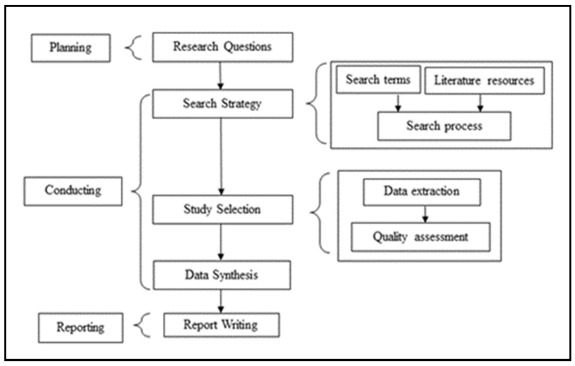

Figure 1: Review Protocol Phases

2.1 Research Questions

In the planning stage, the criteria and scope of the research questions (RQ) for this study are being layout based on the PICOC criteria in [12].

Table 1: Scope for Structuring Research Questions

Criteria	Scope
Population	Big data organization
Intervention	Capability components for the BDA implementation
Comparison	Not applicable
Outcomes	Issues related to BDA implementation
Context	Reviewed of any studies of the BDA

The structure of research questions (RQs) was designed based on the identified criteria and scope as mentioned in Table 1. The formulation of RQs are as follows:

- RQ1: How many papers are focusing on BDA implementation?
- RQ2: What is the focus area of the BDA studies?
- RQ3: What are the types of BDA applications?
- RQ4: What are the capability components of BDA implementation towards value discovery?

2.2 Search Strategy

In the second stage; conducting the review encompasses search strategy, study selection, and data synthesis. The search strategy utilized in this review consists of the search terms, literature resources and search process.

2.2.1 Search terms

The authors built the search terms using the following steps:

- 1st step: Derivation of big data analytic applications (search string) from the research questions.
- 2nd step: Identification of alternative spellings and synonyms for the search string. The initial search strings are big data analytics, then followed by the alternative search string such as big data analytics applications and big data analytics implementation.
- 3rd step: Executing searching using operator Boolean OR and Boolean AND to incorporate primary keyword and/or synonyms of each keyword to search a comprehensive literature review of journal papers on Big Data Analytics. The search string was executed into the 'search field provided' in the digital library/indexing services.

Journal of Theoretical and Applied Information Technology
30th November 2016. Vol.93. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.2.2 Literature resources

The review referred to five electronic database resources including the IEEE Explore, Scopus, Springer Link, Science Direct and ACM Digital Library to extract data based on the title and abstract information. The relevant papers were stored in the reference management tools for selection.

2.2.3 Inclusion and exclusion criteria

The inclusion criteria were all papers must be related to the studies on BDA within a period of publication from January 2011 until December 2015, and only papers written in the English language were accepted for reviewed. The exclusion criteria were papers not related to the scope of the research questions, out of period range and not written in the English Language.

2.3 Study Selection

The study selection was divided into three stages as demonstrated in Figure 2. Firstly, the digital search was conducted and selection processes were based on the title which resulted in a total of 650 relevant papers to be reviewed. In the next stage, the abstract and brief contents of the selected papers were evaluated. Duplicated and irrelevant papers were rejected, leaving only 140 relevant papers. Filtering was continued in the third stage by applying the quality assessment criteria. Finally, 20 papers (14 percent of 140 papers) were accepted for the data synthesis of evidence after conducting the exclusion criteria and screening of the detailed abstracts and full text.

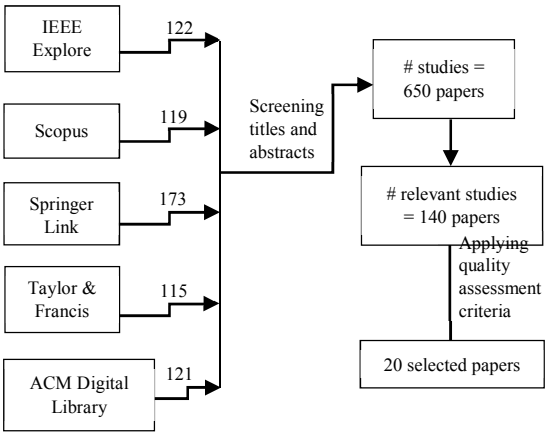

Figure 2: Digital Search and Selection Procedure

The quality assessment was formulated to evaluate the completeness of the papers and advantageous for data extraction. These questions are presented in Table 2. Each question has only three optional answers: Yes=1; partially=0.5; and No=0.

Table 2: Quality Assessment Criteria

No.	Questions	Answer
1	Is the aim of the study clearly stated?	
2	Is the reporting of the papers coherent?	'Yes = 1'
3	How detail has the study process been documented?	OR 'Partially = 0.5'
4	Is the proposed study clearly described?	OR 'No = 0'

Based on the quality assessment criteria, the results for the final selected papers consist of 20 studies are shown in Table 3. 15 papers (75 percent) were very good and 5 papers (25 percent) were good. Not any of the papers were rated as fair and poor in quality.

Table 3: Quality Assessment Results

Quality Scale	Poor (>=1)	Fair (>=2)	Good (>=3)	Very Good (=4)	Total
# of papers	0	0	5	15	20
Percentage (%)	0	0	25	75	100

2.4 Data Synthesis

Data Synthesis is the process of extracting information and addressing the answers to the research questions. The 20 selected papers were further assessed with respect to the formulated of research questions in the planning stage. Data related to RQs are reported in the results and discussion in Section 3.

3. RESULTS AND DISCUSSION

This section presents the reporting stage to discuss the findings of the review. The results of answering the research questions are as the following:

3.1 Papers Focusing on Big Data Analytics Implementation (RQ1)

The review was conducted based on five (5) years of publications (January 2011 – December 2015) which revealed that 75% of the publications were conference publication and the rest 25% were journal publications (Figure 3).

Journal of Theoretical and Applied Information Technology
30th November 2016. Vol.93. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

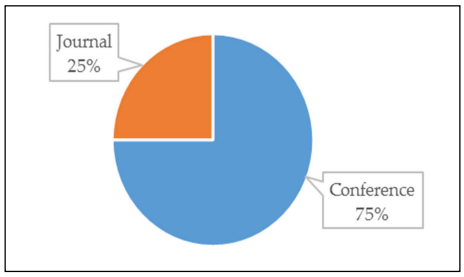

Figure 3: Distribution of Papers by Publication Type

Generally, there was a significant increase in the number of publications on BDA between 2011 until 2015 (Figure 4). The highest number of publications was produced in 2014 with nine (9) papers. This was followed by 2015 with 7 papers, 2013 with two (2) papers and only one (1) paper was published in 2011 and 2012, respectively.

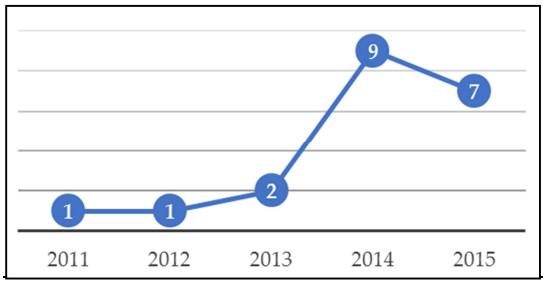

Figure 4: Distribution of Papers by Publication Year

3.2 The Focus Area of BDA Implementation Studies (RQ2)

The review shows that the focus areas of the BDA implementation from the most to the least popular topics are as follows: BDA challenges, BDA process, data mining techniques, BDA trends and business value from BDA (Table 4).

Table 4: Research Focus Area

Focus Area	Papers
BDA Challenges	[7], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22]
BDA Process	[7], [9], [16], [17], [18], [19], [23], [24]
Data Mining Techniques	[13], [18], [20], [24], [25], [26], [27]
BDA Trends	[7], [8], [14], [15], [28]
Business Value from BDA	[22], [29]

3.3 Type of Big Data Analytics Applications (RQ3)

Overall, 14 papers discussed on generic applications which involved various types of analytics applications (Table 5). A study by [7] discussed six (6) types of big data applications organized by the emerging of big data characteristics due to the evolution of data analytics which creates a high impact to the data value: the structure data, text, web, multimedia, network and mobile analytics. This is in line with [8] who also discussed that big data application evolution emerged the analytics research into four (4) critical technical areas such as the data, text, web and network analytics. The BDA applications in [19] and [29] focused on various domains which include the electronic commerce/business and information technology (IT) domains.

Six (6) papers emphasized specific analytics applications such as using Enterprise Information Systems (EIS), automated self-tuning (Starfish) and churn prediction analytics as shown Table 5.

Table 5: Type of BDA Applications

Type of Applications	Papers
Generic	[7], [8], [9], [13], [15], [16], [17], [18], [19], [21], [22], [23], [24], [29]
Specific	[14], [20], [25], [26], [27], [28]

3.4 The Capability Components in BDA Implementation towards Value Discovery (RQ4)

Most researchers simplified the capability components for the BDA into framework and technlogy, processing and analytics techniques as indicate in Table 6.

Table 6: BDA Capability Components

Capability components	Papers
Framework and Technology	[7], [8], [9], [13], [15], [16], [23], [17], [19], [24], [28]
Processing	[7], [14], [17], [15], [19], [21] [29]
Analytic Techniques	[20], [22], [24], [25], [26], [27]

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A. Framework and Technology

Big Data Framework is a strategic management for big data implementation in various domains based on the organization needs. A number of studies on developing the BDA framework were discussed, depending on the needs of the organization or the domain used. Chandarana and Vijayalakshmi [16] pointed that Big Data Framework was based on the technology domain perspective by comparing the Apache Hadoop, Project Storm and Apache Drill based on owner, workload, source code, low latency, and complexity. They concluded that the Apache Hadoop is suitable for workload or batch processing where time is not a critical factor. Project Storm, on the other hand, is well suited for the data stream analysis or real-time processing, while the Apache Drill is best for the interactive and ad-hoc analysis.

Tekiner and Keane [23] proposed Big Data Framework for the technology domain as a solution for data management to enable organizations to gain a competitive advantage by enhancing data processing. In addition, [17] suggested that defining the Big Data architecture and solutions in the technology domain would resolve the existing challenges and known issues or problems with big data by introducing Big Data Analytics Framework in the cloud base infrastructure services, which comprises of five (5) components:

- Data Models, Structures, Types (Data formats, non-relational/relational file, file systems);
- Big Data Management;
- Big Data Analytics and Tools;
- Big Data Infrastructure (BDI); and
- Big Data Security.

B. Processing

BDA is the process of using algorithms running on the powerful supporting platforms to discover the hidden pattern and unknown potential big data [7]. The BDA processes begin with the collection of data generated from various sources in the form of various types of data unstructured, structured, transaction, sensor, image, video and social media. The data were captured, stored and processed. The analytics outputs resulted in the unlocked value of information by visualizing and highlighting knowledge discovered during the exercise. In general, the BDA process model encompassed data selection, gathering (a.k.a pre-processing), analysis and data visualization and result in the interpretation or action [9]. The stream and batch processing are the two (2) types of big data processing. The characteristics of Stream Processing [7] are:

- Stream of new data input.
- Infinite or unknown of data size.
- Store no data.
- Single limited amount of memory for hardware.
- A single or few passes over data processing and a few seconds or milliseconds.

Meanwhile, the characteristics of Batch Processing [7] are:

- Chunks of data input.
- Finite or known data size.
- Stored complex data.
- Multiple CPUs and memories (hardware) and Processed in multiple data rounds.
- More or longer processing time.

Furthermore, Sun et al. [14] proposed an ontology of Big Data, which can be divided into three (3) levels: bottom level includes Big Data and Data Analytics; middle level is divided into big data descriptive analytics, big data predictive analytics and big data prescriptive analytics; while the top level includes big data analytics. Hansmann and Niemeyer [9] stated that there was no common understanding of how to characterize the elements of the Big Data concept. Therefore, they proposed a study on methodologically enriched literature review by deriving the characteristic dimensions from the existing definitions of Big Data such as Data, IT-Infrastructure, Method and Applications perspectives. Each dimension was compared with the generic process model consists of the Data Selection, Gathering/Pre-processing/Storing and Analysis and Result Visualization and Interpretation/Action. The IT-dimension has become popular in the publications, and hardware advances have played a major role in realizing the distributed software platforms needed for the BDA implementation [17].

The evolution of BDA applications has contributed huge and valuable socio-economic impacts on mankind such as in health and human welfare, nature and natural processes, Government and the public sector, electronic commerce, business and economic systems, social networking and the Internet, and also computational and experimental processes [15].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

C. Analytics Techniques

Predictive analytics is an analytics advancement processes which is used to make a prediction on future events in BDA. Predictive analytics is able to provide big data insight to business intelligence and identifies the meaningful patterns of big data to predict future events and access to various options. It was noted from this review exercise that predictive analytics uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze the current data to make a prediction about the future trend.

In the recent years, mining big data has opened problems to methodological and practical issues related to algorithm design and performance aspects [13], [24]. This was supported by [30] that issues of big data mining were related to heterogeneity (or variety), scale (or volume), speed (or velocity), accuracy and trust, privacy crisis, interactive and garbage mining. The availability of Big Data advancement empowered the development of new data mining techniques. According to [25], data mining is very generic and it discovers knowledge from information. A number of studies defined data mining as a process of extraction and analysis of patterns, relationships and useful business information from massive databases sources (big data). The extraction of hidden data from big data required specific analyzing techniques. Some of the BDA commercial tools such as Rapid Miner are based on CRISP-DM (Cross-Industry Standard Process for Data Mining) methodologies and process model approach [31]. Riedel et al. [18] suggested that the 'Smart Data Analytics Method' was meant for high productivity data processing analytics. Meanwhile, [28] suggested a Starfish-self-tuning system for BDA to provide automatic analytics tuning performance to big data analytics practitioners.

Lately, the telecommunication services have invested heavily in the predictive analytics due to the rapid competition in the telco industry. Various predictive modeling using the data mining techniques have been developed for telecommunication churn prediction, such as decision tree model [25] and random forests [32]. Some of them are very accurate in specific churn segmentation, while others might not predict other hidden data patterns well. According to [26], customers tend to acquire expensive process and it costs up to five times as much to make a sale to a new customer as it does make an additional sale to an existing customer, therefore one of the major reasons for predicting churn is that it reduced costs to retain existing customers than to acquire new customers. The predictive analytics can be applied to any type of unknown data, whether it is past, present or future related data. The capability of predictive analytics to identify meaningful patterns of Big Data to predict future events is supported by [8].

3.5 Discussion

Findings were based on the scope of BDA implementation capability components that include the BDA framework and technology, processing and its used in various types of BDA applications in the same or different domains and also analytic techniques. The outcome of this review shows that most of the BDA application tools offer value creation from the processing of the 3Vs big data characteristics and the predictive analytics as it is widely used and developed for the use of current technology demands. The challenges of the predictive analytics are analyzing the unstructured data in a huge amount in order to unlock the hidden data from the massive and heterogeneous data from various sources. Therefore, it would be sufficient if the capability of the big data analytics components is included in all of the data analytics types such as the descriptive, predictive and prescriptive analytics for better value business creation through the advancement of Big Data.

The creation of big data value discovery model was to facilitate ideas or insight to better decision making. The process began with data input from various sources of raw data. Then was the data selection process, where data were selected based on the purpose of analytics processes. The pre-processing of the data would be conducted earlier to clean and transform the raw data into meaningful Big Data. The next process was analyzing big data using big data analytics methods or techniques of analyzing to create value for information and describe or predict trends. The process continues for data visualization and interpretation.

The outcome of BDA processes will create knowledge for Big Data Value Discovery (BDVD) as shown in Figure 5. From BDVD, the top management of business organizations will be equipped with knowledge, ideas or insights that can assist them in the decision-making process with a better understanding of the current problems, fast actions and able to forecast future opportunities.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

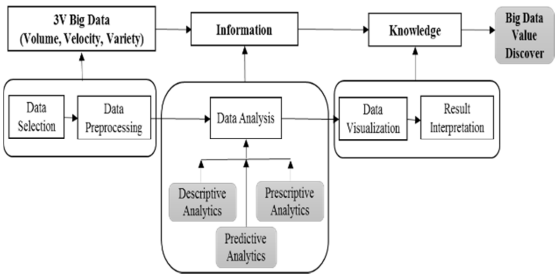

Figure 5: The process of Big Data Value Discovery

4. OPEN RESEARCH ISSUES

From this investigation, the researchers have identified several research issues from big data implementation. Among them were the needs of assessment of business value and corporating the big data value discovery with knowledge management. There is a lack of study (such as case study, experiments or survey) in BDA in relation to its applications in the small and medium-sized organizations. In addition, there was an urgency to enhance the data mining techniques for the hidden and unstructured/semi-structured data as this will provide better insight and valuable information towards improving decision making and creating business opportunities.

The data mining technique is an opportunity for further researches in Big Data phenomenon. The knowledge discovery studies define data mining as a process of extraction and analysis of patterns, relationships and useful information from massive databases. The Big Data issues are related to its characteristics as the growth of mobile apps and Internet of Things (IoT) created great challenges to the existing data mining techniques. The traditional data mining algorithms need to be modified for parallel/distributed using MapReduce way of codings.

5. CONCLUSION

Big data is considered a 'boomed' research area for the academic and technology development. The most popular research areas are big data technologies, tools and challenges. This review found that not many academic papers were related to deriving value discovery from BDA implementation. It was also noted that the capability components may affect the value discovery from big data implementation that include formulating big data model or framework to describe enabler technology and identify the analysing process and using sufficient analytic techniques for big data. Therefore, it can be concluded that there are many potential areas of studies in value discovery from implementing the advancement of the big data phenomenon.

REFERENCES:

[1] S. Fosso Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, "How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study," *Int. J. Prod. Econ.*, Vol. 165, pp. 234–246, 2015.

[2] D. Laney, "3D Data Management: Controlling Data Volume, Velocity, and Variety," 2001.

[3] Gartner, "Big Data," *Glossary*, 2012. [Online]. Available: http://www.gartner.com/it- glossary/big-data/it-glossary/big-data/.

[4] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, and C. A. H. B. Roxburgh, "Big data: The next frontier for innovation, competition, and productivity," *McKinsey Global Institute*. McKinsey Global Institue, 2011.

[5] A. Hafiz, O. Lukumon, B. Muhammad, A. Olugbenga, O. Hakeem, and A. Saheed, "Bankruptcy Prediction of Construction Businesses: Towards a Big Data Analytics Approach," in *2015 IEEE First International Conference on Big Data Computing Service and Applications*, 2015, pp. 347–352.

[6] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Q.*, Vol. 36, No. 4, pp. 1165–1188, 2012.

[7] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial," *IEEE Access*, Vol. 2, pp. 652–687, 2014.

[8] P. S. S. Laxmi and P. S. Pranathi, "Impact of Big Data Analytics on Business Intelligence-Scope of Predictive Analytics," *J. Curr. Eng. Technol.*, Vol. 5, No. 2, pp. 856–860, 2015.

[9] T. Hansmann and P. Niemeyer, "Big Data - Characterizing an Emerging Research Field Using Topic Models," in *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, 2014,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

pp. 43–51.

[10] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," 2007.

[11] C. Okoli and K. Schabram, "A Guide to Conducting a Systematic Literature Review of Information Systems Research," *Sprouts Work. Pap. Inf. Syst.*, Vol. 10(26), pp. 1–51, 2010.

[12] M. Petticrew and H. Roberts, *Systematic Reviews in the Social Sciences: A Practical Guide*. Blackwell Publishing, 2006.

[13] A. Cuzzocrea, "Big data mining or turning data mining into predictive analytics from large-scale 3vs data: The future challenge for knowledge discovery," in *4th International Conference MEDI 2014*, 2014, Vol. 8748, pp. 4–8.

[14] Z. Sun, F. Pambel, and F. Wang, "Incorporating Big Data Analytics into Enterprise Information Systems," in *Information and Communication Technology*, Vol. 9357, Korea: Springer, 2015, pp. 300–309.

[15] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," *J. Parallel Distrib. Comput.*, Vol. 74, no. 7, pp. 2561–2573, 2014.

[16] P. Chandarana and M. Vijayalakshmi, "Big Data Analytics Frameworks," in *2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, 2014, pp. 430–434.

[17] Y. Demchenko, C. De Laat, and P. Membrey, "Defining Architecture Components of the Big Data Ecosystem," in *2014 International Conference on Collaboration Technologies and Systems (CTS)*, 2014, pp. 104–112.

[18] M. Riedel, A. S. Memon, and M. S. Memon, "High productivity data processing analytics methods with applications," in *37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2014, May, pp. 289–294.

[19] S. Sruthika and N. Tajunisha, "A Study on Evolution of Data Analytics to Big Data Analytics and Its Research Scope," in *2015 IEEE International Conference on Innovations in Information Embedded and Communications Systems (ICIIECS)*, 2015, pp. 1–6.

[20] L. Ye, C. Qiu-Ru, X. Hai-Xu, L. Yi-Jun, and Y. Zhi-Min, "Telecom customer segmentation with K-means clustering," in *7th International Conference on Computer Science & Education (ICCSE)*, 2012, pp. 648–651.

[21] N. Mohamed and J. Al-Jaroodi, "Real-Time Big Data Analytics : Applications and Challenges," in *Proceedings - 2014 IEEE International Conference on High Performance Computing & Simulation (HPCS)*, 2014, pp. 305–310.

[22] D. Arora and P. Malik, "Analytics : Key to go from generating big data to deriving business value," in *2015 IEEE First International Conference on Big Data Computing Service and Applications*, 2015, p. 7.

[23] F. Tekiner and J. A. Keane, "Big Data Framework," in *IEEE International Conference on Systems, Man, and Cybernetics*, 2013, pp. 1494–1499.

[24] C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos, "Big Data Analytics: A Survey," *J. Big Data*, Vol. 2, No. 1, pp. 1–32, 2015.

[25] D. M. Balasubramanian and M. Selvarani, "Churn Prediction in Mobile Telecom System Using Data Mining Techniques," *Int. J. Sci. Res. Publ.*, Vol. 4, No. 1, pp. 2250–3153, 2014.

[26] N. Kamalraj and a Malathi, "A Survey on Churn Prediction Techniques in Communication Sector," *Int. J. Comput. Appl.*, Vol. 64, No. 5, pp. 39–42, 2013.

[27] M. R. Khan, J. Manoj, A. Singh, and J. Blumenstock, "Behavioral Modeling for Churn Prediction: Early Indicators and Accurate Predictors of Custom Defection and Loyalty," in *International Congress on Big Data (BigData Congress)*, 2015, pp. 677–680.

[28] H. Herodotou, H. Lim, G. Luo, N. Borisov, and L. Dong, "Starfish: A Self-tuning System for Big Data Analytics," in *5th Biennial Conference on Innovative Data Systems Research (CIDR)*, 2011, Vol. 11, pp. 261–272.

[29] Meetali, "From Big Data to Big Values: A Big Science Leading to a Revolution," in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2015, pp. 56–59.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[30] D. Che, M. Safran, and Z. Peng, "From Big Data to Big Data Mining: Challenges, Issues, and Opportunities," in *International Conference on Database Systems for Advanced Applications*, 2013, pp. 1–15.

[31] V. Kotu and B. Deshpande, "Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner". Morgan Kaufmann, 2014.

[32] U. Yabas, H. C. Cankaya, and T. Ince, "Customer Churn Prediction for Telecom Services," in *IEEE 36th Annual Computer Software and Applications Conference*, 2012, pp. 358–359.