Application of Methodologies and Process Models in Big Data Projects

Rosa Quelal[1][a], Luis Eduardo Mendoza[1,2][b] and Mónica Villavicencio[1][c]
[1]ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL,
Facultad de Ingeniería en Electricidad y Computación, Campus Gustavo Galindo Km. 30.5 Vía Perimetral,
P.O. Box 09-01-5863, Guayaquil, Ecuador
[2]Processes and Systems Department, Simón Bolívar University, Valle de Sartenejas, P.O. Box 89000, Caracas, Venezuela

Keywords: Agile Methodologies, Big Data, Systematic Literature Review, Text Mining.

Abstract: The concept of Big Data is being used in different business sectors; however, it is not certain which methodologies and process models have been used for the development of these kind of projects. This paper presents a systematic literature review of studies reported between 2012 and 2017 related to agile and non-agile methodologies applied in Big Data projects. For validating our review process, a text mining method was used. The results reveal that since 2016 the number of articles that integrate the agile manifesto in Big Data project has increased, being Scrum the agile framework most commonly applied. We also found that 44% of articles obtained from a manual systematic literature review were automatically identified by applying text mining.

1 INTRODUCTION

Big Data projects have been used in different economic sectors. Therefore, it is necessary to study how Big Data projects are planned and executed in order to reach their expectations —execution time (Al-Jaroodi et al., 2017), return on investment, and value for client (Chen et al., 2016). To do so, we performed a Systematic Literature Review (SLR) related to methodologies and process models applied in Big Data projects.

From the SLR, we notice that there is an emerging interest in applying software engineering process models to Big Data initiatives (Al-Jaroodi et al., 2017; Kumar, 2017); i.e., we observe a growth of publications related to both concepts (software engineering and big data). Therefore, in order to improve our literary review process, which involves the continuous incorporation of emerging publications related to these concepts, we validate our manual SLR process with a text mining method.

The rest of this article contains the following: section 2 describes the research methodological framework; section 3 shows the results obtained; section 4 presents the limitations and future work, and section 5 concludes the article.

2 METHODOLOGICAL FRAMEWORK

To conduct this SLR, we used the guidelines proposed by Kitchenham and Petersen (Kitchenham and Charters, 2007; Petersen et al., 2015), including: the formulation of research questions; the search process; inclusion and exclusion criteria; data extraction; data analysis and classification; and quality evaluation.

2.1 Research Questions

The purpose of this research is to identify which methodologies and process models have been applied in Big Data projects. Hence, the research questions are:

[a] https://orcid.org/0000-0002-9292-8441
[b] https://orcid.org/0000-0002-5081-7559
[c] https://orcid.org/0000-0002-2601-2638

Quelal, R., Mendoza, L. and Villavicencio, M.
Application of Methodologies and Process Models in Big Data Projects.
DOI: 10.5220/0007729602770284
In *Proceedings of the 21st International Conference on Enterprise Information Systems (ICEIS 2019)*, pages 277-284
ISBN: 978-989-758-372-8
Copyright © 2019 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

1. What Agile Methodologies have been applied in Big Data projects?
2. What Non-agile Methodologies have been applied in Big Data projects?
3. What kind of Big Data projects are applying either agile or non-agile methodologies?

2.2 Search Processes

To answer the first question, a manual search process was carried out on the following databases: Science Direct, Google Scholar, Springer and Scopus. The process was conducted four times: November 2016, April 2017, October 2017, and May 2018. For the Science Direct, Google Scholar, and Springer databases, the search was run on all article content; and for Scopus, the search was run on titles, keywords, and abstracts.

The query used was:

```
("scrum" or "extreme programming" or "agile Data" or "crystal" or "Kanban" or "agile software") and ("big data" or "data science" or "analytic") and ("case study")
```

To answer the second question, a similar process was carried out. It was executed once, in July 2018 on the following databases: Science Direct, Google Scholar and Springer.

The query used was:

```
("waterfall" or "spiral") and ("big data" or "data science" or "analytic") and ("case study")
```

To answer the third question, we classified the articles obtained in both queries according to the kind of big data projects (i.e., paper S28 applied Text Mining in Social Media).

2.3 Inclusion and Exclusion Criteria

Although the first "Agile Manifesto" was published in 2001, this field of research shows fluctuations over time, with increases in publications in 2005, 2010 and 2013 (Batra and Dahiya, 2016; Campanelli and Parreiras, 2015). On the other hand, research related to Big Data and its application starts increasing in 2012 (Gandomi and Haider, 2015; Wamba et al, 2015). Given these facts, we consider a range of years from 2012 to 2017 to perform the queries. However, traditional software engineering methodologies and process models exist long time ago; that is why, we perform the second search process without an initial year and until 2017.

Additionally, the language for the search process was restricted to "English", and the type of publication to "Scientific Articles".

The inclusion criteria considered in the first search process are:

- Agile methodologies.
- Big Data projects of any kind.
- Type of research reported: "Case Study".

The inclusion criteria considered in the second search process are:

- Non-agile methodologies and software process models.
- Type of project Big Data.
- Type of research used "Case Study".

2.4 Data Extraction

From the first search process (Agile Methodologies), we found the following amount of articles:

- Science Direct: 170.
- Springer: 96
- Scopus: 11
- Google Scholar: 34

And from the second search process (Non-agile Methodologies), the followings:

- Science Direct: 3
- Springer: 47
- Google Scholar: 13

In both search processes, the articles listed in Google Scholar correspond to the ones not obtained from the other databases.

By reviewing the list of the articles from both searches, we found an intersection of 9 articles [S1, S39, S57, S69, S162, S169, S198, and S247]. Therefore, the extraction process resulted in 365 articles, which are included in Appendix A.

2.5 Data Analysis and Classification

The data analysis and classification were carried out based on the defined inclusion criteria and classification steps, as follows:

1. Reading the abstracts.
2. Searching for each criterion within the complete content of the articles.
3. When necessary, reading the whole article.
4. Classifying articles by criteria.
5. Classifying articles by research type.

Below, an example of each step:

(Step 1) While reading the abstract of the article [S272], we identify a Big Data project and a "Framework" type research report. However, the use of agile methodologies was missing despite the existence of the word "Scrum".

(Step 2) When searching for the word "Scrum" into the article, we find a strange coincidence, which warns us for the need to perform a more in-depth review.

(Step 3) When reading the whole article, we realize that the word "SCRUM" stands for Spatio-Chronological Rugby Union Model, without any relation to agile methodologies.

(Step 4) We classify the article as Big Data projects without Agile Methodologies, and

(Step 5) The type of research as a "Framework".

In summary, step 1 allows us to detect articles other than "Case Studies". Steps 2 and 3 to identify other types of research such as: interviews, literature reviews, systematic mappings, case studies, frameworks, and conceptual models. Steps 4 and 5 to classify the articles. With this procedure, we verify all criteria and avoid discarding articles. Appendices B and C contain the list of articles by search criteria.

2.6 Quality Evaluation

The quality evaluation is performed in two ways: (1) by following the guidelines proposed by Kitchenham and Petersen (Kitchenham and Charters, 2007; Petersen et al., 2015), defining a procedure for each step of the guidelines, and (2) by using a text mining method to validate the manual search processes.

The chosen text mining method was topic classification, specifically the Latent Dirichlet Allocation (LDA) algorithm from the Python's Gensim library. We chose this algorithm because it is one of the most used in similar contexts (Chuang et al., 2012). The process was developed in 3 stages.

Stage 1: Topic modelling process for 365 articles. The steps followed are:

1. Collect the 365 articles obtained from the extraction process.
2. Convert documents from .pdf to .txt format.
3. Tokenize documents; convert words into data to be analysed.
4. Remove stopwords and punctuation marks.
5. Select 30 articles randomly to generate the corpus, dictionary and models for three topics.
6. Classify 365 articles according to the generated models.

Stage 2: Topic modelling process for 18 articles. The steps followed were:

1. Collect the 18 articles obtained from the manual review (i.e., sections 2.1 to 2.5).
2. Convert documents from .pdf to .txt format.
3. Tokenize documents; convert words into data to be analysed.
4. Remove stopwords and punctuation marks.
5. Generate the corpus, dictionary and models for three topics, with the 18 articles.
6. Classify 18 articles according to the generated models.

Stage 3: Compare results from Stage 1 and 2.

3 RESULTS

By applying the methodology, we obtained 374 articles, 311 from the first search process and 63 for the second. Since 9 articles appear in both searches, we perform the analysis and discussion of results of 365 articles.

3.1 First Query Results

The classification of articles gives the following results: 20% (62 articles) refer to Agile Methodologies and 38% (117 articles) to Big Data projects. Additionally, 66% (206 articles) corresponds to Case Studies. From the 311 papers obtained in the search, only 14 include the criteria: "Case Studies", "Agile Methodologies", and "Big Data". Figure 1 shows a time line of papers per search criterion. From 2016, the number of articles that integrate Agile Methodologies and Big Data concepts have increased.

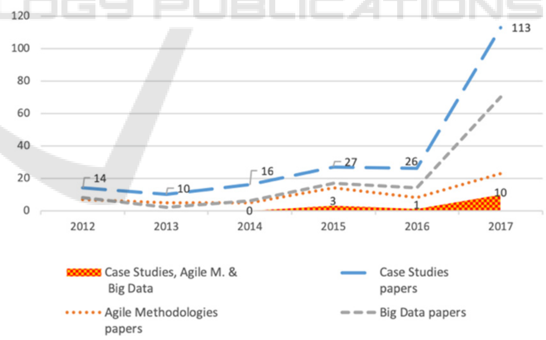

Figure 1: Classification of papers by search criteria.

Table 1 summarizes the information of the 14 articles; four of them [S30, S40, S57, and S241] report using Agile Methodologies in Big Data projects without specifying which one.

SCIENCE AND TECHNOLOGY PUBLICATIONS

Table 1: Papers about case studies of Agile Methodologies used in Big Data projects.

ID	Year	Reference	Data Base	Agile Methodologies	Big Data	Context
S30	2015	Berry, N. M., Prugh, W., Lunkins, C., Vega, J., Landry, R. and Garciga, L. (2015). Selecting video analytics using cognitive ergonomics: A case study for operational experimentation. Procedia Manufacturing, 3, 5245-5252.	ScienceDirect	Not specified	Video Intelligence	Software development
S146	2015	Kalenkova, A. A., van der Aalst, W. M., Lomazova, I. A. and Rubin, V. A. (2017). Process mining using BPMN: relating event logs and process models. Software and Systems Modeling, 16(4), 1019-1048.	Springer	Kanban, Scrum	Process mining techniques using event logs	Academic - Software development
S158	2015	Komenda, M., Schwarz, D., Švancara, J., Vaitsis, C., Zary, N. and Dušek, L. (2015). Practical use of medical terminology in curriculum mapping. Computers in biology and medicine, 63, 74-82.	ScienceDirect	eXtreme Programming (XP)	Academic Curriculum management	Medical
S57	2016	Cope, B. and Kalantzis, M. (2016). Big Data comes to school: Implications for learning, assessment, and research. AERA Open, 2(2), 2332858416641907.	GoogleScholar	Not specified	Educational data mining	Academic - scholastic
S28	2017	Baur, A. W. (2017). Harnessing the social web to enhance insights into people's opinions in business, government and public administration. Information Systems Frontiers, 19(2), 231-251.	Springer	Scrum	Text Mining in Social Media	Automotive industry
S40	2017	Bucksch, A., Das, A., Schneider, H., Merchant, N. and Weitz, J. S. (2017). Overcoming the law of the hidden in cyberinfrastructures. Trends in plant science, 22(2), 117-123.	ScienceDirect	Not specified	Analysis of images of plants	Scientific - Biology
S52	2017	Chrimes, D. and Zamani, H. (2017). Using Distributed Data over HBase in Big Data Analytics Platform for Clinical Services. Computational and Mathematical Methods in Medicine, 2017.	GoogleScholar	Agile Data Science	Analysis of hospital data about 9 billion patients	Hospital health
S241	2017	Ryan, P. J. and Watson, R. B. (2017). Research Challenges for the Internet of Things: What Role Can OR Play? Systems, 5(1), 24.	GoogleScholar	Not specified	Analysis of data from IoT	Academic - Scientific Operations Research
S245	2017	Saltz, J. and Crowston, K. (2017, January). Comparing data science project management methodologies via a controlled experiment. In Proceedings of the 50th Hawaii International Conference on System Sciences.	GoogleScholar	Scrum, Kanban	Algorithms, data mining and machine learning to geographic information	Academic - University
S246	2017	Saltz, J. (2017). Acceptance Factors for Using a Big Data Capability and Maturity Model.	GoogleScholar	They analyse agile methodologies in Big Data projects	Different projects	Business
S247	2017	Saltz, J., Shamshurin, I. and Connors, C. (2017). Predicting data science sociotechnical execution challenges by categorizing data science projects. Journal of the Association for Information Science and Technology.	GoogleScholar	Scrum	Different types of efforts in data science	Business
S269	2017	Su, Y., Luarn, P., Lee, Y. S. and Yen, S. J. (2017). Creating an invalid defect classification model using text mining on server development. Journal of Systems and Software, 125, 197-206.	ScienceDirect	Scrum	Data mining techniques	Software development
S292	2017	Vidgen, R., Shaw, S. and Grant, D. B. (2017). Management challenges in creating value from business analytics. European Journal of Operational Research, 261(2), 626-639.	GoogleScholar, ScienceDirect	eXtreme Programming (XP), Scrum	Different types of efforts in data science	Business
S300	2017	Woodside, A. G. and Sood, S. (2017). Vignettes in the two-step arrival of the internet of things and its reshaping of marketing management's service-dominant logic. Journal of Marketing Management, 33(1-2), 98-110.	GoogleScholar	Scrum	Analysis of data from IoT to support marketing	Business

3.2 Second Query Results

The classification process reveals that 60% (38 articles) refer to Non-agile Methodologies, 22% (14 articles) to Big Data projects, and 83% (52 articles) to case studies. Only five articles include the three criteria (Non-Agile Methodology, Big Data, and Case Studies); however, three of them [S322, S344, and S354] do not specify the methodology used –see Table 2. As it can be observed, paper S247 appears in both Tables (1 and 2). In Figure 2, we observe that papers related to Big Data start increasing in 2012, and papers integrating Non-agile Methodologies with Big Data projects in 2014.

3.3 Answering Research Questions

What Agile Methodologies have been applied in Big Data projects? Table 1 provides the answer to this question, showing that Scrum, Kanban, XP and Agile Data Science are the Agile Methodologies used in Big Data projects reported from 2015 to 2017.

What Non-agile Methodologies have been applied in Big Data projects? In our review, we identify that CRISP-DM, SEMMA, and KDDM (Knowledge Discovery via Data Analytics) are Non-agile Methodologies used in big data projects.

Figure 3 summarizes the agile and non-agile methodologies used in Big Data projects from 2012 to 2017.

What kind of Big Data projects are applying either agile or non-agile methodologies? From the reviewed articles, we observe a variety of Big Data projects in which Agile and Non-agile Methodologies have been applied, such as image intelligence, process mining, text mining, association rule mining, geographic information analysis, machine learning algorithms, Internet of Thinks (IoT), etc. These projects were developed in different contexts: academic, industrial, scientific, business, banking, among others (see the last column of Tables 1 and 2).

Table 2: Papers about case studies of Traditional Methodologies used in Big Data projects.

ID	Year	Reference	Data Base	Methodologies	Big Data	Context
S247	2017	Saltz, J., Shamshurin, I. and Connors, C. (2017). Predicting data science sociotechnical execution challenges by categorizing data science projects. Journal of the Association for Information Science and Technology.	GoogleScholar	Crisp-DM	Different types of efforts in data science	Business
S322	2015	D'Souza, M. J., Kashmar, R. J., Hurst, K., Fiedler, F., Gross, C. E., Deol, J. K. and Wilson, A. (2015). Integrative biological chemistry program includes the use of informatics tools, GIS And SAS software applications. Contemporary issues in education research (Littleton, Colo.), 8(3), 193.	GoogleScholar	Not specified	Different techniques Applied to GIS data	Biology
S340	2016	Li, Y., Thomas, M. A. and Osei-Bryson, K. M. (2016). A snail shell process model for knowledge discovery via data analytics. Decision Support Systems, 91, 1-12.	ScienceDirect	Crisp-DM, SEMMA, KDDM, KDDA	Different techniques	Business
S344	2017	Mafereka, M. and Madikane, N. (2017) Data Management is key to Banks'success.	GoogleScholar	Not specified	Different techniques	Banking
S354	2016	Saltz, J., Shamshurin, I. and Connors, C. (2016, July). A Framework for Describing Big Data Projects. In International Conference on Business Information Systems (pp. 183-195). Springer, Cham.	GoogleScholar	Not specified	Different techniques	

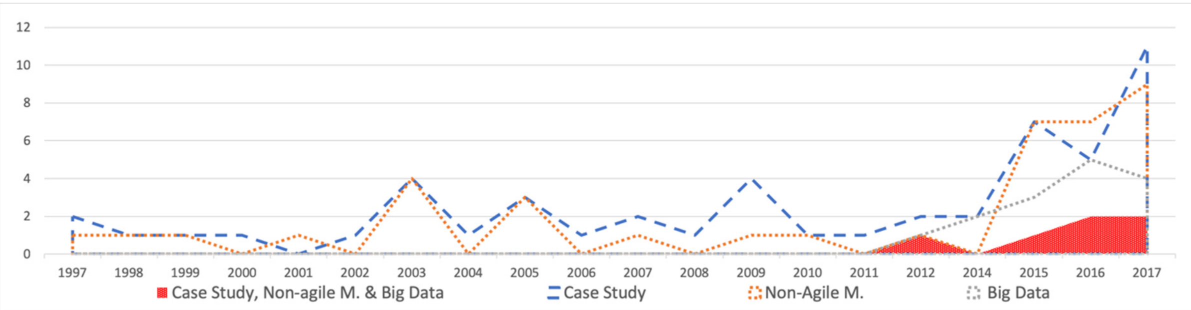

Figure 2: Classification of papers by search criteria.

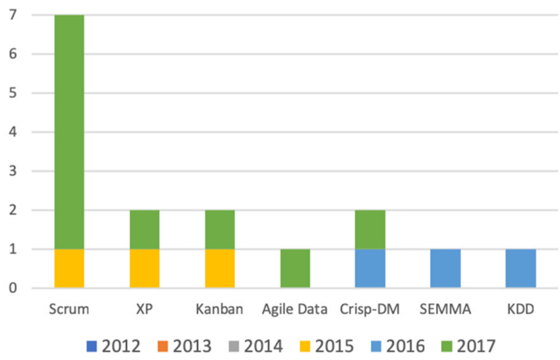

Figure 3: Methodologies used in Big Data projects.

3.4 Quality Evaluation

To perform the quality evaluation, we develop the process explained in section 2.6. Table 3 shows the topics obtained from mining the 365 papers, with the model generated from 30 aleatory papers. Each topic represents a set of words that match the analyzed documents. For example, the topic 1 is focused on *software development*, topic 2 *models and simulation*, and topic 3 on *products*, *process* and *manufacture*.

Table 3: Topic model generated from 365 papers.

Topic 1	Topic 2	Topic 3
`use`	`model`	`data`
`model`	`use`	`product`
`develop`	`figure`	`use`
`system`	`simulation`	`manufacture`
`software`	`electronic`	`grind`
`information`	`image`	`machine`
`social`	`result`	`system`
`manage`	`time`	`process`
`waterfall`	`value`	`technology`
`data`	`cell`	`manage`

Table 4 shows the topic modelling generated from the 18 articles obtained from the whole manually performed SLR. Topic 1 is about *testing*, topic 2 about *business analytic* and *research*, and topic 3 about *big data* and *data science*.

First, we evaluate the *search process*, and the *inclusion and exclusion criteria process*. For this evaluation, we compare the words used for the manual queries (query 1 and query 2 from page 2 – section 2.2) with the words generated by the topic model (see Table 4). The matching five words appear in light blue: `data`, `analytic`, `software`, `big`, and `science`.

Table 4: Topic model generated from manual SLR.

Topic 1	Topic 2	Topic 3
`defect`	`data`	`data`
`test`	`system`	`project`
`develop`	`research`	`process`
`use`	`use`	`model`
`material`	`analytic`	`cid` *
`project`	`model`	`team`
`software`	`business`	`use`
`model`	`manage`	`big`
`manuscript`	`develop`	`science`
`function`	`inform`	`system`

* Compound's ID number, used in [48]

This fact makes us to infer that a simpler query (i.e. five words) can lead us to the same result for the search process.

With respect to the *data extraction process*, if we compare the topics generated from the 365 papers (see Table 3) with queries 1 and 2, we identify three matching words: `software`, `waterfall`, and `data`.

The topic model algorithm gives to each analyzed paper a percentage of affinity with every topic. For example, the paper S247 (Predicting data science sociotechnical execution challenges by categorizing data science projects) has 57% of affinity with Topic 1, 21% with Topic 2 and 22% with Topic 3.

The higher the percentage, the stronger the relationship between the content of the article and the topic is. As an example, Table 5 presents the percentage of affinity for ten papers.

Table 5: Example of affinity between papers and topics.

Paper	Topic 1	Topic 2	Topic 3
S241	21%	57%	22%
S242	19%	18%	63%
S243	20%	61%	19%
S244	19%	21%	60%
S245	57%	22%	21%
S246	57%	21%	22%
S247	57%	21%	22%
S248	18%	21%	61%
S249	18%	19%	63%
S250	20%	61%	19%

Next, we sort and give a ranking per topic for each paper, where the higher percentage is ranking 1 and the lower percentage is ranking 3, we group and count how many papers belong to each group. For example, the paper S245 corresponds to the group where topic 1 is ranked 1 (57%), topic 2 is 2 (22%), and topic 3 is 3 (21%). Others examples are the papers S244 and S248 where the topic 1 is ranked 3, topic 2 is 2, and topic 3 is 1.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

Table 6 shows the summary of the percentage of the 365 papers ranked per topic. The ranking position 1 represents which topic was assigned the highest percentage. For example, the percentage of documents whose ranking is 1 for topic 1 is 16%, whereas the percentage of documents whose ranking is 3 for topic 3 is 32%.

Table 6: Complete Topic Classification.

		Ranking			
		1	2	3	Total
Topic	1	16%	44%	40%	100%
	2	35%	37%	28%	100%
	3	49%	20%	32%	100%
	Total	100%	100%	100%	

The topic classification for the 18 resulting papers presented in Tables 1 and 2, is shown in Table 7. It can be noticed that 44% of them belong to topic 1, ranking 1.

Table 7: Topic classification of resulted papers.

		Ranking			
		1	2	3	Total
Topic	1	44%	39%	17%	100%
	2	28%	22%	50%	100%
	3	28%	39%	33%	100%
	Total	100%	100%	100%	

From this sample, if we would like to use topic modelling to reduce the number of articles to review, we can say that reviewing only 16% of the total of articles generated in the search, we could find 44% of the results sought. In other words, we could get 8 (44% of 18) papers of our interest, avoiding reading 307 papers (i.e. only reading 58 out of 365 papers — 16% of the total number of articles).

3.5 Final Remarks

According to the sample presented in previous section, we believe it is possible to use topic modelling to reduce the number of articles to read, filtering the papers more representatives to our research. Although it is necessary to perform more tests to improve the technique and increase the percentage of success; the results presented here demonstrate the benefits of using a topic mining process.

Finally, with respect to the data analysis and classification process, the manual SLR generates information such as: methodologies used, Big Data projects type, and the context or industry where they were developed. However, this level of details was not possible to obtain with a topic mining process. To automate the classification process, we can try Fuzzy techniques and supervised processes.

4 LIMITATIONS AND FUTURE WORK

This article is limited to the search of Agile and Non-agile Methodologies reported in case studies associated with Big Data projects, excluding other kinds of research such as formal experiments or surveys. In addition, the searches were only executed on four databases: Science Direct, Springer, Google Scholar and Scopus. Also, the field of expertise of our research team is mainly oriented to Software Engineering.

As future work, we plan to replicate the whole process with other kind of research studies to evaluate how text mining contributes to the quality evaluation of a SLR process and test Fuzzy techniques to perform a supervised classification of the analysed articles.

Additionally, we are interested in design a framework for developing Big Data projects applying agile principles.

5 CONCLUSIONS

The SLR carried out in this work demonstrates the use of methodologies and process models since the emergence of Big Data projects, increasing the use of Agile Methodologies in this kind of projects from 2015 onwards. The methodologies most commonly reported in publications related to Big Data projects are: Scrum, XP, Kanban, and Crisp-DM.

According to the SLR, the applications of the Big Data started in the scientific and academic fields rather than the industrial and commercial sectors. However, in the last two years, there has been an increase in the number of Big Data projects in the business field, especially in areas such as Marketing and Innovation.

The integration of text mining as part of the quality evaluation of the SLR process has allowed us to test the ability of this technique to optimize this kind of process.

REFERENCES

Al-Jaroodi, J., Hollein, B. and Mohamed, N. (2017). Applying software engineering processes for big data analytics applications development. In: *7th Annual Computing and Communication Workshop and Conference (CCWC)*. [online] Las Vegas: IEEE, pp. 1-7. Available at: https://ieeexplore.ieee.org/document/7868456 [Accessed 2 May 2018].

Batra, D. and Dahiya, R. (2016). Adapting Agile Practices for Analytics Projects. In: *15th AIS SIGSAND Symposium*. [online] Lubbock: AIS, pp. 1-10. Available at: https://pdfs.semanticscholar.org/96e5/0da9309c6e2128d877e7f0842ac45d935410.pdf [Accessed 7 Apr. 2017].

Campanelli, A. and Parreiras, F. (2015). Agile methods tailoring – A systematic literature review. *Journal of Systems and Software*, [online] 110, pp. 85-100. Available at: https://www.sciencedirect.com/science/article/pii/S0164121215001843 [Accessed 8 Nov. 2016].

Chen, H., Kazman, R., Garbajosa, J. and González, E. (2016). Toward Big Data Value Engineering for Innovation. In: *2016 IEEE/ACM 2nd International Workshop on Big Data Software Engineering (BIGDSE)*. [online] Austin: IEEE, pp. 44-50. Available at: https://ieeexplore.ieee.org/document/7811386 [Accessed 5 Oct. 2017].

Chuang, J., Manning, C. D. and Heer, J. (2012, May). Termite: Visualization techniques for assessing textual topic models. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces,* pp. 74-77. ACM.

Gandomi, A. and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, [online] 35(2), pp. 137-144. Available at: https://www.sciencedirect.com/science/article/pii/S0268401214001066 [Accessed 11 May 2016].

Kitchenham, B. and Charters, S. (2007), 'Guidelines for performing Systematic Literature Reviews in Software Engineering' (EBSE 2007-001), Technical report, Keele University and Durham University Joint Report.

Kumar, V. (2017). *Software Engineering for Big Data Systems*. Master Thesis. University of Waterloo.

Petersen, K., Vakkalanka, S. and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, [online] 64, pp. 1-18. Available at: https://www.sciencedirect.com/science/article/pii/S0950584915000646 [Accessed 5 Nov. 2016].

Wamba, S., Akter, S., Edwards, A., Chopin, G. and Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, [online] 165, pp. 234-246. Available at: https://www.sciencedirect.com/science/article/pii/S0925527314004253#! [Accessed 13 Oct. 2016].

APPENDICES

Appendices are available on: https://drive.google.com/file/d/1ajQyGnUf0ONvPHjuosYcS6_tiuIu99MK/view?usp=sharing.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS