

# IMPORTANCE OF PROCESS MINING FOR BIG DATA REQUIREMENTS ENGINEERING

Sandhya Rani Kourla, Eesha Putti and Mina Maleki

Department of Electrical and Computer Engineering and Computer Science,  
University of Detroit Mercy, Detroit, MI, 48221, USA

## ABSTRACT

*Requirements engineering (RE), as a part of the project development life cycle, has increasingly been recognized as the key to ensure on-time, on-budget, and goal-based delivery of software projects. RE of big data projects is even more crucial because of the rapid growth of big data applications over the past few years. Data processing, being a part of big data RE, is an essential job in driving big data RE process successfully. Business can be overwhelmed by data and underwhelmed by the information so, data processing is very critical in big data projects. Employing traditional data processing techniques lacks the invention of useful information because of the main characteristics of big data, including high volume, velocity, and variety. Data processing can be benefited by process mining, and in turn, helps to increase the productivity of the big data projects. In this paper, the capability of process mining in big data RE to discover valuable insights and business values from event logs and processes of the systems has been highlighted. Also, the proposed big data requirements engineering framework, named REBD, helps software requirements engineers to eradicate many challenges of big data RE.*

## KEYWORDS

*Big data, requirements engineering, requirements elicitation, data processing, knowledge discovery, process mining*

## 1. INTRODUCTION

Requirements engineering (RE) is the first and most critical phase of the Software Development Life Cycle (SDLC). RE is the branch of engineering concerned with the real-world goals for, functions of, constraints on systems, and the relationship of these factors to precise specifications of system behavior [1-4]. Software requirements engineer (SRE) is responsible for translating stakeholders' needs, desires, and wishes of a software system into precise and formal software requirements specification. SREs need to communicate effectively and frequently with stakeholders to elicit, analyze, model, and manage their requirements. Doing this at the early stage of the software development helps ensure requirements are meeting stakeholders' needs while addressing compliance and staying on schedule and within budget [3].

In contrast, it has been indicated in many researches that performing improper and careless RE is one of the main sources of the time-consuming rework, inadequate deliveries, budget overruns, and consequently failures of the projects [3, 5-6]. The HMS Titanic (1912), the Mars Climate Orbiter (1999), the Apollo 13 (1970), Space Shuttle Challenger (1986), and Space Shuttle Columbia (2002) projects are some examples of high-profile projects broke down due to poor requirements engineering [6]. Also, according to PMI's Pulse of the profession (2018), 35% of companies proclaimed that the primary reason for their project failures is inaccurate requirements gathering [7]. As a consequence, performing a proper RE at the early stages of project development is critical in the success of projects, especially for big data projects.

In a newly released study, International Data Corporation (IDC) forecasts that the big data technology and analytics software market, which in 2018 reached \$60.7 billion worldwide, is expected to grow at a five-year compound annual growth rate (CAGR) of 12.5% [8]. Big data is a technology that deals with data sets that are too large or complex to handle with traditional data processing techniques for capturing, storing, analyzing, searching, sharing, transferring, visualizing, querying, and updating of data. The main characteristics of big data technologies are 5V's: volume, velocity, variety, volatility, and variability [4, 9-11]. Volume refers to the massive amount of data; Velocity refers to the high growth rate of incoming data that needs to be processed and analyzed; Variety refers to many different forms of data; Volatility refers to the duration which data is valid and should be stored in the repository; Variability refers to data whose context changes invariably. For big data projects, it is essential to define the targets of the project at earlier stages. Time and cost reduction, finding insights from statistics, and optimizing selection making are the most common goals that make an organization to appreciate big data projects [4, 10]. RE phase in big data projects' development helps in achieving these goals and finding the business values associated with projects; These values help stakeholders to understand the importance of the project and its value in the market [12].

Big data RE is data-centric; it means there are lots of potential information, hidden patterns, and knowledge that can be extracted and discovered from a large amount of historical and actual data existing in big data projects. Knowledge discovery from big raw data is a process of collecting, cleaning, transforming, storing, applying mining algorithms, and discovering values from the raw data [13]. Data processing is a subset of knowledge discovery and deals with the processing of the collected data. Since big data is associated with a large volume of data, employing traditional data processing methods does not help much in discovering valuable insights and knowledge from data. Process mining is an advanced technology that not only can use for discovering business values from data but also can recognize the risks, deviations, bottlenecks, and inefficiencies associated with big data [14]. Process mining helps SRE to elicit, prioritize, and validate requirements from big data using execution logs, process discovery, and conformance techniques [15].

In this paper, which is an extended version of [16], we have featured the importance of process mining in big data RE process to improve the business insights and values of the big data projects. This paper focuses on how data is prepared and how the processed data is converted to valuable information through the knowledge discovery process and process mining. Moreover, the main components of the proposed big data requirements engineering framework, named REBD, have been described. This framework explains the detailed steps for carrying out RE activities on different software projects development for improving their success rates.

The rest of the paper is organized as follows. After reviewing related works in section 2, section 3 describes big data requirements engineering activities. Section 4 briefs the knowledge discovery steps and the role of process mining in extracting business values. Section 5 explains the proposed REBD framework, and section 5 wraps up the research.

## **2. RELATED WORK**

Requirements engineering in the context of big data applications is a hot research topic that attracted the attention of researchers in recent years. An analysis of the state of the art of big data RE research studies shows that little research has been conducted in this area by 2018 [10]. The investigation areas of this research included the phases of the RE process, type of requirements, application domains, RE research challenges, and solution proposals intended by RE research in

the context of big data applications. In the following, some of the related work for big data RE will be presented.

To understand the context of the problem for the big data software applications, an empirically derived RE artifact model has been proposed in [17], which is equivalent to domain models. The proposed model can capture the fundamental RE components and the relationships concerning the development of big data software applications. “How can the requirements of a project be identified and analyzed to determine the suitability of a combined big data technology application?”. To answer this question, a new process model has been proposed, detailed, and evaluated in [1]. This model considers the compound requirements, Knowledge Discovery in Databases (KDD), and a big data classification framework to identify the most relevant requirements of big data projects. Identified requirements for building big data applications must address big data characteristics (5V’s) in terms of quality requirements. Quality requirements, also known as quality attributes, are those functional or nonfunctional requirements that used to measure the system’s performance, such as reliability and availability. A new approach is proposed in [9] to ensure that the big data characteristics have adequately addressed in the specification of quality requirements.

Even though there are many requirements elicitation techniques in traditional RE, the researchers believe that more efficient tools and techniques should be employed to identify all requirements, business values, and knowledge from big data. Machine learning, deep learning, natural language processing, and data mining are some data analysis technologies that can use to discover requirements and valuable knowledge from big data [18-20]. To formalize the overall knowledge discovery process (KDP) within a common framework, [13] introduced the concept of a process model. This model helps organizations to understand the KDP better and provides a roadmap to follow while planning and executing the project. This, in turn, results in cost and time savings, better understanding, and acceptance of the results of such projects. KDP is primarily used to plan, work through, and reduce the cost of any given project. Also, actionable use case (AUC) diagram is one of the efficient tools introduced in [12] that allows identified business values unleashed from data to be depicted in the data scientists’ models, together with their roles in the software and their interactions with other software components. The AUC diagrams enhance the users’ experience, optimize the systems’ utility, and consequently maximize profit.

Process mining is another efficient method to discover business values from the existing big data [14-15, 21-22]. The research done in [14], explains how to turn the event logs into valuable insights, and the result of this research are used to identify and understand the deviations, risks, bottlenecks, and inefficiencies involved. Process mining helps SRE to elicit, prioritize, and validate requirements from big data using execution logs, process discovery, and conformance techniques [15]. One of the challenges in process mining is monitoring and controlling the existing big data in the organizations. Because of this, [22] introduced a concept of load balancing for distributing the workloads uniformly to support business goals in process mining. [21] explained the perspective of processing mining in an organizational context by conducting an assessing literature review and analyzing 58 papers from the literature on process mining to synthesize the existing knowledge on business value realization from process mining. The capability of process mining to discover valuable insights from event logs and processes of the system helps SRE to eradicate many challenges of traditional RE.

### **3. BIG DATA REQUIREMENTS ENGINEERING**

As mentioned earlier, RE is one of the essential phases of the project’s life cycle, and the failure of the RE leads to the failure of the project. So, software requirement engineers are responsible

for conducting the RE activities in a very well-mannered way, which can resolve many conflicts between stakeholders as well as among requirements. There are three different types of processes in the big data RE model proposed in [12]: processes drove by software requirements engineer (SRE), processes drove by data scientists (DS), and processes drove jointly by software requirements engineers and data scientists (SRE/DS). Table1 demonstrates the RE activities for the big data products, including the RE activities, the responsible person for executing each activity, and their artifacts. Moreover, the column “Included in” indicates whether each activity is carried out in the traditional RE model (TRE) and/or big data RE model (BDRE).

Table 1. A Requirement Engineering Model for Big Data Software

Process	Performed by	Output	Included in
Requirements Elicitation	SRE/DS along with Customers	List of requirements	TRE and BDRE
Data Acquisition	DS	Filtered data used for knowledge discovery	BDRE
Requirements Analysis	SRE	Technical specifications and SRE models	TRE and BDRE
Data Analysis	SRE	Extracted values and DS model	BDRE
Use case Consolidation	SRE /DS	Combined SRE and DS model	BDRE
Requirements Modelling	SRE	Finalized AUC diagrams	TRE and BDRE
Requirements Validation	SRE	Validated requirements	TRE and BDRE
Requirements Specification	SRE	SRS report	TRE and BDRE

From the table, it is clear that compared to the traditional RE, big data RE contains a few extra steps. The main activities of RE are described below [2-3,12].

### 3.1. Requirements Elicitation

Requirement elicitation as a critical activity in the requirement development process is the practice of uncovering and gathering the stakeholders' needs and desires from the system in terms of functional and nonfunctional requirements. To have a productive elicitation and dig all of the stakeholders' requirements, SRE should be able to communicate effectively with all stakeholders. One of the major problems that influence the elicitation process is the social issue. Conflicts between different cultures and work styles influence the efficiency of the requirements elicitation process. Figuring out social scenarios and considering those in the SRS document can improve the RE process. Social issues rely on a few other factors, like technical, financial, and domain issues, which can disturb the RE process [23]. Multiple techniques such as brainstorming, card sorting, laddering, interviews, prototyping, domain analysis, Joint Application Development (JAD), and Quality Function Deployment (QFD) can be employed to conduct requirements elicitation. SREs select the elicitation techniques based on different factors, including project characteristics, stakeholders' characteristics, organizational characteristics, analyst characteristics, and software development process [24]. The flaw in the requirements elicitation will be carried out to the successive RE activities and then to the following development phase resulting in negotiating the process. As a consequence, spending enough time to investigate and

choose the best requirements elicitation techniques will save the cost and time for rectifying errors at the later stage.

### **3.2. Data Acquisition**

This activity can be explained as collecting, gauging, and fine-tuning the data before storing it in the data repository [12, 25]. Big data acquisition is expensive and characteristics-driven; It is administrated by 5Vs (volume, velocity, variety, variability, and veracity) [25]. Hence, the traditional collecting methods cannot make data acquisition. The importance of data acquisition is to gather data from the distributed information source and desires to store the collected information in a scalable, proficient big data storage. Data acquisition makes use of three components: (a) Protocols for gathering information, (b) Framework which is used to collect information based on the defined protocols, and (c) Technologies which allow continuous storage of the retrieved data [25]. Hence, data scientists use different data acquisition methods to discard useless data and keep important ones.

### **3.3. Requirements Analysis**

Requirement analysis is the practice of defining the system boundaries and analyzing elicited and documented requirements to make sure that they are clear, concise, understandable, unambiguous, consistent, and complete. Also, requirements agreement, which is a process of resolving the conflicts in the requirements derived from different sources, is part of this activity. A successful requirement analysis establishes the first step towards developing an excellent system that meets the needs of various stakeholders and expected to reduce the cost of resolving unnecessary conflict issues between the stakeholders [26]. Use cases and user stories are often useful tools for this purpose. Moreover, task analysis, Rich pictures, and video recording instruct SRE to analyze the requirements.

### **3.4. Data Analysis and Value Discovery**

In this activity, the acquired big data is analyzed by data scientists to reveal information and discover business values. These business values are capable of making high profits for companies. In an IT industry, there will be many channels in the process operation which carries the information. This information gathers for deriving effective value through data analysis [27]. Data analysis is a time-consuming task, and failing in an efficient analysis will lead to making wrong decisions. As a consequence, before analyzing the data, the quality of data should be improved to ensure the deriving of accurate information. The quality of data is measured using multiple dimensions, resulting in the evaluation of accuracy, timeliness, and other factors of data [27]. Different data analysis technologies such as machine learning, deep learning, natural language processing, data mining, text analytics, and predictive analytics can be employed to analyze data and extract business values from it [12, 18-20]. Moreover, techniques like process mining [15] and tools like smart grids [28] can be employed to accelerate the process of knowledge discovery from big data.

### **3.5. Use Case Consolidation**

Consolidation is merging the work and building models by software requirements engineers and data scientists. This consolidation is represented as an actionable use case diagram. The AUC diagram for big data software proposed in [12] is a merging of the data scientist model and the traditional use case. Once the values are defined and presented, the consolidated model is ready to be presented and discussed with the customers. SRE and DS together analyze AUC diagrams

to discover the values and document them. AUC contains (a) actors who can be a manager, user, system, event triggers, or website, (b) actionable intelligence that is the specific information, value, data needed to come to a decision [29], (c) data scientist model that contains actionable intelligence gathered by the data scientist, algorithms, analysis methods, and data sources for gathering actionable intelligence, and a “Rely on” relationship between actionable intelligence and data scientist model, (d) relationships between actor and actionable intelligence, and (e) relationships between data source, analysis, and algorithm.

### **3.6. Requirements Modeling**

The requirement modeling intends to mold raw requirements into technical representations of the system. This activity guides SRE to archive all requirements during the specification process. Proper requirements representation eases the communication of requirements and conversion of those requirements into system architecture and design. Insights of the requirements are hard to understand when the modeling of the requirements is done improperly. The feature of the requirements modeling is to figure out how the business requirements are connected to the information technology requirements and backward. Modeling helps in promoting business-IT alignment by understanding, analyzing, and structuring the way business requirements are hooked to the information technology requirements and the way around [30]. To represent the analyzed functional and nonfunctional requirements in any systems, various modeling techniques can be employed, such as use cases, user stories, natural languages, formal languages, and a variety of formal, semiformal, and informal diagrams [3].

### **3.7. Requirements Validation**

Requirements validation is the practice of checking and revising requirements to ensure that the specified requirements meet customer needs. Requirements validation also involves checking that (a) the system does not provide the functions that the customer does not need, (b) there are no requirements conflicts, and (c) time and budget constraints will meet. Validation establishes advanced support for the requirements and ensures their good qualities. The main purpose of this step is to identify and accumulate the problems before the stakeholders are dedicated to meet the requirements. An improper requirements validation at the early stages of the project leads to the irrelevant design and out graded product quality because of invalid and error-prone requirements. Validation should be a part of all activities in RE. Systematic manual analysis of the requirements, test-case generation, comparative product evaluation tools, or some of the requirements elicitation techniques can be used for requirements validation [3].

### **3.8. Requirements Specification**

Requirements specification is the process of documenting the stakeholders' needs in terms of precise and formal software requirements specification. A software requirement specification (SRS) is a document that uses as a foundation of software development and acts as a contract between the software stakeholders and the developers. As a consequence, preparing an accurate SRS report at the end of the RE phase is critical in the success of projects. User requirements are used as the input for the first specification draft of the system, its goals, and missions. Preparing a precise requirements specification is not easy if the developing system is a novel system and complex, and when the stakeholders have more knowledge of the acquiring system. SRS is prepared by SRE and stakeholders together. SRS contains the mission, scope, goals of the project, software and hardware requirements of the project, as well as functional and nonfunctional system requirements in terms of formal, semiformal, and formal models created in the previous processes. In addition to the requirements matrices, the big data SRS report contains

business values extracted from data and AUC diagrams. Moreover, to ensure that big data characteristics are appropriately addressed, SRS should contain the quality attribute matrices. The quality attribute matrix is designed by intersecting a big data characteristic with a quality attribute and then identifying the system's quality requirements that apply to that intersection during big data RE process, as explained in [9].

#### **4. KNOWLEDGE DISCOVERY FROM BIG DATA**

As mentioned earlier, traditional RE methods are user-centric; focus on requirements apparent to users. However, big data RE methodologies should be data-centric as well because there are lots of potential information and hidden patterns that can be extracted from data. As a consequence, data processing and knowledge discovery are essential jobs in big data RE.

Knowledge discovery from big raw data in RE refers to the finding of useful data patterns, valuable information, and hidden business values from the massive amount of existing data. Knowledge discovery concerns the entire knowledge extraction process, including (a) how data are stored and accessed, (b) how to use efficient and scalable algorithms to analyze massive datasets, (c) how to interpret and visualize the results, and (d) how to model and support the interaction between humans and machines[31].

##### **4.1. Knowledge Mining Process**

Figure 1 depicts the knowledge discovery process from existing raw big data. Exploration of data followed by data preparation is defined as the knowledge discovery process. Before processing data and extracting knowledge, data should be preprocessed. Data preparation includes the following steps:

- **Data Accumulation:** In this step, the raw data is gathered for mining and discovering some valuable information out of it. Different tools such as web data collection tools, encyclopedia, google forms can be employed to gather the data.
- **Data Categorization:** Data categorization is the process of analyzing the data and split them up into different categories. Understanding the data types and assigning data to the target categories is the main goal of data categorization.
- **Data Cleaning:** Raw data are usually noisy, inconsistent with lots of missing values. Data cleaning is defined as identifying incomplete, incorrect, inaccurate, or irrelevant parts of the data and then replacing, modifying, or deleting the dirty data. After categorizing data into different datasets, it is crucial to filtering the unwanted data and keeping the important data and locking the sensitive data.
- **Data Reshaping:** Data reshaping is the process of converting the refined data into useful patterns that can help to understand the values of data more. Data reshaping is done to fit well-defined output that can be understood by a broad range of audiences.
- **Data Storage:** After cleaning and reshaping data, it should be stored for further analysis. Data warehouses are used for storing the preprocessed data.
- **Data preparation** is necessary for improving data quality and removing unwanted errors. Improved data quality will help in finding useful knowledge and helps in understanding the valuable insight and business values from data.

Data exploration allows data scientists to understand the datasets and explore the hidden business values. For this, the preprocessed clean data should be analyzed to investigate the parts which are more useful to discover knowledge. The prepared data is analyzed by the data scientist, to check which process mining algorithm is best suited for the analyzed data and is applied to it. Out of the mining, the information extracted is converted into business values by data analysts and business analysts together. Data can be explored manually by writing some scripts and/or by using automated tools like data visualization tool. Data exploration provides instructions for applying data mining and process mining to discover knowledge.

As the knowledge discovery process is made from a massive amount of data, its efficiency and accuracy are the critical concerns to take care of. Discovering information from such a huge amount of data is out of scope when the traditional data processing techniques are used. Hence, it is essential to use a proper knowledge discovery technique or approach, such as process mining.

## 4.2. Process Mining

Process mining is a technique used in process management for analyzing business processes based on event logs [21]. A business process is defined as the sequence of tasks or activities performed

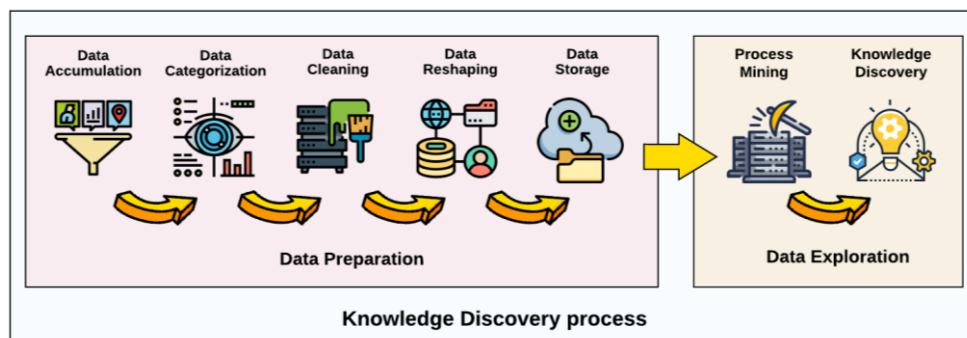


Figure1. Knowledge discovery from big data

by the people or systems in which a specific sequence produces a service. In addition to the identification of business values or insights, process mining helps in recognizing and understanding the bottlenecks, inefficiencies, deviations, and risks from data [14]. Several mining algorithms can be employed to identify patterns, trends, and details present in the event logs of the IT systems such as heuristic miner, fuzzy manner, and genetic miners [32].

A data scientist has to face many challenges when using process mining for big data in value discovery. Some of these challenges are listed below:

- Many of the process discovery algorithms work for the static events in a stagnant fashion, but not pertinent for processing real-time event streams.
- Distribution of keys (load balancing) is another issue regarding the process mining algorithm. Load balancing is a technique for distributing workloads uniformly [22].
- Because of the lack of remote data access support and the lack of information about internal storage, integrity checking has become a crucial aspect. Integrity checking is to ensure the validity and accuracy of data coming from different logs.



- Big data application’s massive amount of data will result in poor quality. The low-quality data leads to the accessibility and integrity issue by restricting the user or organization with limited access to the data.

Data quality can be improved by using the primary process mining analysis. This analysis reveals the vital processes which are not tracked by information systems. This process discovery will help in unfolding the new tracking capabilities such as radio frequency identification (RFID). This tracking power will facilitate advanced analysis, which leads to identify the broader insights.

## 5. CONCEPTUAL REBD FRAMEWORK

In this section, we describe our proposed big data requirements engineering framework, REBD, as depicted in Figure 2. This framework explains the planned approach for carrying out big data projects for improving the success rate. The purpose of this conceptual framework is to eradicate challenges like deciding whether to perform big data requirements engineering or traditional requirements engineering, knowledge discovery, and balancing the efforts to have the successful execution as described in the following.

Initially, when the project starts to develop, it is crucial to derive and analyze its characteristics to find whether the project falls under the category of big data or traditional projects. Doing this at the earliest phase of the project will save a considerable amount of time spent on the project lifecycle. To identify the category of the project, we have to figure out whether the recognized characteristics are big data characteristics or not. As mentioned earlier, the main characteristics of big data projects are 5V’s: volume, velocity, variety, volatility, and variability. Next, a comprehensive quantitative classification model proposed by [1] is used to verify the project type. This model combines the recognized big data characteristics to figure out the project type.

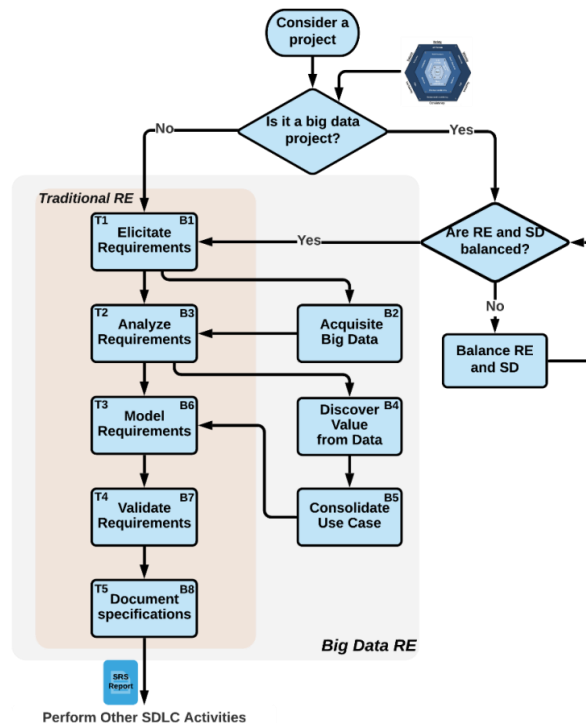


Figure 2. Requirements Engineering for Big Data (REBD) framework

If the project is being considered in the group of traditional projects, traditional RE activities should be carried out in the order specified in Figure 1 by T1 to T5 labels. However, if the project is found to be in the group of big data projects, then it should be checked whether RE and SD are balanced or not.

Balancing RE and SD determines the challenges which are related to RE and SD activities. Challenges related to RE is identifying 5V's, which has been conducted already, and the challenges associated with SD are designing, coding, testing, and maintaining a big data end-user application. SD challenges are very much concerned because it may lead to the failure of the project, even after spending a lot of time in RE activities. To resolve this issue, a model has been proposed by [33], which contains a separate unit for doing the research needed for RE and SD. Obviously, anything without proof of concepts (PoC) and research may cause a shortage of time. The research and PoC, which give the optimal solution, should be used in the project practice. This ensures the balance between RE and SD, and the development of novel and better big data end-user applications. Once there is a solution for all challenges, then big data RE activities should be carried out in the order specified in Figure 1 by B1 to B8 labels.

After gathering engineered requirements documented in the SRS report, other processes of the software development life cycles, including design, implementation, testing, deployment, and maintenance, will start to produce software with the highest quality and lowest cost in the shortest time possible.

## 6. CONCLUSION

One of the main reasons behind the failure of many projects is improper requirements engineering and failing to capture the stakeholders' requirements, which leads to time-consuming rework, inadequate deliveries, and budget overruns. So, performing requirements engineering is a crucial phase of the project's development lifecycle, especially for big data projects. As big data is one of the new and emerging technology, developing big data projects and improving their success rates and market value is an excellent achievement for companies. One of the outcomes of big data RE process is discovering the business values and insights from the existing big data. There is a lot of knowledge, valuable information, business values, and quality features hidden in a large amount of historical and actual raw data existing in big data projects. To analyze these data and discover hidden requirements, data scientists need to collaborate with SRE in big data RE activities. However, employing traditional data processing is not enough to discover valuable information from data because of the 5V's characteristics of big data. To analyze these data and discover hidden requirements, data scientists need to use advanced technology like process mining.

In this paper, we have aimed to describe the importance of process mining in the big data RE process. In addition to improve the business value, process mining helps in determining risks, deviations, inefficiencies, and bottlenecks of the systems. This paper explains the detailed steps of the knowledge discovery from data preparation to the data exploration process. Moreover, the detailed steps for carrying out big data RE activities in the proposed REBD framework has been explained. This conceptual framework helps to increase the success rate of project development. Although process mining increases the market value and digs deeper into the data by making sure that the complete set of collected data is covered, the framework, along with process mining, should be applied practically for the big data projects to validate its efficiency and the market value. This research will be extended to discover better tools for knowledge discovery, which can be used in various industries. Also, the management of big data quality requirements will be investigated in the future in a preferred way.

## REFERENCES

- [1] M. Volk, N. Jamous, and K. Turowski, "Ask the right questions: requirements engineering for the execution of big data projects," presented at the 23rd Americas Conference on Information Systems, SIGITPROJMGMT, Boston, MA, Aug. 2017, pp 1-10.
- [2] J. Dick, E. Hull, and K. Jackson, Requirements Engineering, 4th edition, Springer International Publishing, 2017.
- [3] P.A. Laplante, Requirements Engineering for Software and Systems, 3rd Edition. Auerbach Publications (T&F), 20171024. VitalBook file, 2017.
- [4] D. Arruda, "Requirements engineering in the context of big data applications," ACM SIGSOFT Software Engineering Notes, 43(1): 1-6, Mar. 2018.
- [5] A.G. Khan, et al., "Does software requirement elicitation and personality make any relation?" Journal of Advanced Research in Dynamical and Control Systems. 11. 1162-1168, 2019.
- [6] T.A. Bahill and S.J. Henderson, "Requirements development, verification, and validation exhibited in famous failures," Systems Engineering, 8(1): 1–14, 2005.
- [7] Pulse of the Profession 2018: Success in Disruptive Times, 2018.
- [8] C. Gopal, et al., "Worldwide big data and analytics software forecast, 2019–2023," IDC Market Analysis, US44803719, Sept. 2019.
- [9] I. Noorwali, D. Arruda, and N. H. Madhavji, "Understanding quality requirements in the context of big data systems," presented at the 2nd International Workshop on Big Data Software Engineering (BIGDSE), Austin, USA, May 2016, pp. 76-79.
- [10] D. Arruda and N.H. Madhavji, "State of requirements engineering research in the context of big data applications," presented at the 24th International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ), Utrecht, The Netherlands, Mar. 2018, pp 307-323.
- [11] M. Volk, D. Staegemann, M. Pohl, and K. Turowski, "Challenging big data engineering: positioning of current and future development," presented at the 4th International Conference on Internet of Things, Big Data and Security (IoTBDs), Heraklion, Greece, May. 2019, pp 351-358.
- [12] H.H. Altarturi, K. Ng, M.I.H. Ninggal, A.S.A. Nazri, and A.A.A. Ghani, "A requirement engineering model for big data software," presented at the 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, Mar. 2017, pp 111-117.
- [13] K.J. Cios, W.Pedrycz, R.W.Swiniarski, L. Kurgan, Data Mining: a Knowledge Discovery Approach. Springer, 2010.
- [14] W. van der Aalst, Process Mining in Action: Principles, Use Cases and Outlook, Springer, 1st ed. 2020.
- [15] M. Ghasemi, "What requirements engineering can learn from process mining," presented at the 1st International Workshop on Learning from other Disciplines for Requirements Engineering (D4RE), Banff, Canada, Aug. 2018, pp 8-11.
- [16] S. Kourla, E. Putti, and M. Maleki, "REBD: A Conceptual Framework for Big Data Requirements Engineering," 7th International Conference on Computer Science and Information Technology (CSIT 2020), Helsinki, Finland, June 2020.
- [17] D. Arruda, N.H. Madhavji, and I. Noorwali, "A validation study of a requirements engineering artefact model for big data software development projects," presented at the 14th International Conference on Software Technologies (ICSOFT), Prague, Czech Republic, Jul. 2019, pp 106-116.
- [18] B. Jan, et al., "Deep learning in big data analytics: a comparative study," Journal of Computer Electrical Engineering, 75(1): 275-287, 2019.
- [19] A. Haldorai, A. Ramum, and C. Chow, "Editorial: big data innovation for sustainable cognitive computing," Mobile Netw Application Journal, 24(1): 221-226, 2019.
- [20] R.H. Hariri, E.M. Fredericks, and K.M. Bowers, "Uncertainty in big data analytics: survey, opportunities, and challenges," Journal of Big Data, 6(1), 2019.
- [21] J. Eggers and A. Hein, "Turning Big Data Into Value: A Literature Review on Business Value Realization From Process Mining," presented at the Twenty-Eighth European Conference on Information Systems (ECIS2020), Marrakesh, Morocco, Jun. 2020.
- [22] K.K Azumah, S. Kosta, and L.T. Sørensen, "Load Balancing in Hybrid Clouds Through Process Mining Monitoring," Internet and Distributed Computing Systems, Springer, Cham, 2019, pp 148-157.

- [23] S. Ramachandran, S. Dodda, and L. Santapoor, "Overcoming Social Issues in Requirements Engineering," in Meghanathan N., Kaushik B.K., Nagamalai D. (eds) *Advanced Computing*, Springer, Berlin, Heidelberg, 2011, pp 310-324.
- [24] M. Batra and A. Bhatnagar, "Requirements Elicitation Technique Selection: A Five Factors Approach," *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249 – 8958, 8(5C):1332-3141, India, May 2019.
- [25] K. Lyko, M. Nitzschke, and AC. Ngonga Ngomo, "Big data acquisition," in *New Horizons for a Data-Driven Economy*, Springer, Cham, 2016, pp 35-61.
- [26] M. Suhaib, "Conflicts Identification among Stakeholders in Goal Oriented Requirements Engineering Process," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 2019.
- [27] C. Yao, "The Effective Application of Big Data Analysis in Supply Chain Management," *IOP Conference Series: Materials Science and Engineering* Mar. 2020.
- [28] X. Han, X. Wang, and H. Fan, "Requirements analysis and application research of big data in power network dispatching and planning," presented at the 3rd Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing Shi, China, Oct. 2017, pp 663-668.
- [29] J. Kelly, M.E. Jennex, K. Abhari, A. Durcikova, and E. Frost, "Data in the Wild: A KM Approach to Collecting Census Data Without Surveying the Population and the Issue of Data Privacy," in *Knowledge Management, Innovation, and Entrepreneurship in a Changing World*, IGI Global, 2020, pp. 286-312.
- [30] D. Pandey, U. Suman, and A. Ramani, "A Framework for Modelling Software Requirements," *International Journal of Computer Science Issues*. 8(3):164, 2011.
- [31] O. Marbán, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," in book *Data Mining and Knowledge Discovery in Real Life Applications*, IntechOpen, 2009.
- [32] O. Dogan, J.L. Bayo-Monton, C. Fernandez-Llatas, and B. Oztaysi, "Analyzing of gender behaviors from paths using process mining: A shopping mall application," *Sensors*, 19(3), 557, 2019.
- [33] NH. Madhavji, A. Miranskyy, and K. Kontogiannis, "Big picture of big data software Engineering: with example research challenges," presented at the IEEE/ACM 1st International Workshop on Big Data Software Engineering (BIGDSE), Florence, Italy, May 2015, pp 11-14.

## AUTHORS

**Sandhya Rani Kourla** received her Bachelor's degree in Computer Science and Software Engineering from Kuvempu University, Davangere, India, in 2011. She is currently pursuing her Master's degree majoring in Computer Science and Software Engineering from the University of Detroit Mercy, MI, USA. Before joining Detroit Mercy, she worked as a software engineer in Mindtree Ltd, Bangalore, India. She is skilled in Requirements Engineering, Software Engineering, Software development, Agile software development, and Manual testing.



**Eesha Putti** is a Master student in the Management Information system at the University of Detroit Mercy, Michigan, USA. She received her B.Tech in Computer Science from Manav Bharati University, Shimla, India. Prior to this, she had participated in several Computer Science Fairs and had developed a skill relevant to Computer Science and Software engineering. She has an aspiration to exile further in the field of Big Data, Data Base Management Systems, and Cloud related areas.



**Mina Maleki** received her Bachelor's degree in computer engineering from Azzahra University, Tehran, Iran, in 2002, her Master in computer engineering and information technology from Amirkabir University of Technology, Tehran, Iran, in 2006, and her Ph.D. in computer science from the University of Windsor, Canada, in 2014. She is currently working as an Assistant Professor of Software Engineering and Computer Science at the University of Detroit Mercy, MI, USA. Her research interests are mainly focused on software engineering, machine learning, text, and big data mining.

