Computer Standards & Interfaces 80 (2022) 103585

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

A process assessment model for big data analytics

Mert Onuralp Gökalp [a,*], Ebru Gökalp [b,c], Kerem Kayabay [a], Selin Gökalp [a], Altan Koçyiğit [a], P. Erhan Eren [a]

[a] Middle East Technical University, Informatics Institute, Ankara, Turkey
[b] University of Cambridge, Department of Engineering, Cambridge, United Kingdom
[c] Hacettepe University, Computer Engineering Department, Ankara, Turkey

ARTICLE INFO

Keywords:
Big data
Data analytics
Software development
Software process improvement
Software process assessment

ABSTRACT

Today, business success is essentially powered by data-centric software. Big data analytics (BDA) grasp the potential of generating valuable insights and empowering businesses to support their strategic decision-making. However, although organizations are aware of BDAs' potential opportunities, they face challenges to satisfy the BDA-specific processes and integrate them into their daily software development lifecycle. Process capability/maturity assessment models are used to assist organizations in assessing and realizing the value of emerging capabilities and technologies. However, as a result of the literature review and its analysis, it was observed that none of the existing studies in the BDA domain provides a complete, standardized, and objective capability maturity assessment model. To address this research gap, we focus on developing a BDA process capability assessment model grounded on the well-accepted ISO/IEC 330xx standard series. The proposed model comprises two main dimensions: process and capability. The process dimension covers six BDA-specific processes: business understanding, data understanding, data preparation, model building, evaluation, and deployment and use. The capability dimension has six levels, from not performed to innovating. We conducted case studies in two different organizations to validate the applicability and usability of the proposed model. The results indicate that the proposed model provides significant insights to improve the business value generated by BDA via determining the current capability levels of the organizations' BDA processes, deriving a gap analysis, and creating a comprehensive roadmap for continuous improvement in a standardized way.

1. Introduction

Gartner predicts that 90% of organizations will define data as a critical asset in their corporate strategies and analytics as an essential capability by 2022 [1]. Big data analytics (BDA) is an emerging field that integrates mathematics, statistics, and computer science to extract valuable knowledge and insights from an exponentially expanding volume of data. It can be defined as collecting, storing, and analyzing big data statistically and quantitatively to build explanatory, predictive, and prescriptive models to drive business decisions and actions. It provides an immense potential of changing everything from how businesses operate to how we live. The rapid technological advancements in collecting, storing, and processing vast amounts of data force businesses to perform BDA software development processes efficiently and effectively. In a survey [2], 84% of the responders indicated that the utilization of BDA software improved their competitive positioning in their business environment. Thus, a growing number of companies focus on initializing or improving their BDA software development initiatives to empower their businesses in strategic data-driven decision making and to attain a competitive edge. Accordingly, the number of BDA software development projects is increasing, and data scientists are becoming part of software development teams as software developers, analysts, and testers [3].

Though BDA is regarded as an innovative technological breakthrough by both practitioners and researchers, organizations face problems in generating business value from their BDA investments. It is forecasted [1] that 80% of BDA software development projects will not be able to produce any business value through 2020 because these projects are not managed and scaled by following a standardized and systematic approach. In a recent survey [4], 90.4% of the respondents indicate that process challenges are the most significant barrier to implement and deploy BDA software products and services to integrate

* Corresponding author.
E-mail address: gmert@metu.edu.tr (M.O. Gökalp).

https://doi.org/10.1016/j.csi.2021.103585
Received 22 January 2021; Received in revised form 18 July 2021; Accepted 26 September 2021
Available online 30 September 2021

big data into their businesses. Moreover, studies show that a structured model to initiate BDA adoption can increase return on investments and overall BDA performance [5]. Thus, a standardized approach is needed to guide organizations to assess the current capabilities and provide a roadmap for improvement to evolve their BDA efforts.

Structural approaches, such as process capability maturity assessment models, frameworks, and standards, have been the prominent tools that guide organizations striving for higher quality and excellence in their operations. There are well-known capability maturity assessment models in the software domain, including Capability Maturity Model Integration (CMMI) [6] and Software Process Improvement and Capability Determination (SPICE) [7]. They share the common goals of capability maturity assessment models, including improving software quality, reducing expenses and inconsistencies, and increasing employee productivity and involvement. Although they aim to improve generic software development lifecycle processes, they mainly lack embracing the requirements of contemporary BDA software development practices as BDA poses inherently unique characteristics, processes, and development activities, including data collection, data cleansing, data management, and resource planning and execution planning [8]. Even though both CMMI and SPICE provide a basis for developing an integrated and well-defined standard model to identify specific requirements of BDA-centric software development, there is still a research gap in the literature to assist organizations in developing high-quality BDA software products and services [8].

To address these concerns, we aimed to develop a standardized, repeatable, consistent, and objective BDA process capability assessment model. Our primary goal is to guide organizations by assessing their current BDA process capabilities to reveal their weaknesses and accordingly provide an extensive guideline for improvement in a structured and repeatable way. To this end, we based our model on the ISO/IEC 330xx standards, which replaces and extends the ISO/IEC 15504 standard family, also known as SPICE [7]. The resulting model defines six capability levels from not performed to innovating and six BDA-specific software development processes: business understanding, data understanding, data preparation, model building, evaluation, as well as deployment and use. We evaluated the applicability, validity, and usability of the proposed model by a multiple-case study. To sum up, this study contributes to the growing body of knowledge by demonstrating the applicability of the ISO/IEC 330xx standard series in the BDA domain.

The rest of the paper is organized as follows. In Section 2, we give background information regarding BDA and capability maturity assessment models in the context of this study. Section 3 explains the proposed BDA process capability model and its development methodology. In Section 4, the case study development and assessment stages are detailed. We discuss the results and findings of the case studies in Section 5. Lastly, the paper is concluded, and directions for future research are given in Section 6.

2. Background

In this section, BDA and capability maturity assessment model concepts are explained, and existing studies related to process capability assessment models in the BDA domain are reviewed.

2.1. Big data analytics

The proliferation of emerging technologies and concepts such as smart devices, the Internet of Things (IoT), Web 2.0, and social media has paved the way for gathering a vast amount of complex and real-time data called big data. Big data refers to collecting large, complex, unstructured, and continuous data from a large number of usually disparate data sources [9]. It essentially focuses on efficiently and effectively collecting, storing, and analyzing data on distributed architectures. The exponential growth in data size, speed, and changing data characteristics urges businesses to adopt emerging data and analytics strategies to sustain their competitive edge.

Big data offers a huge potential for businesses to improve their operational efficiencies by generating valuable insights that can be used to enhance their decision-making process, develop new business models, and drive new revenue streams. Nevertheless, not to be devoid of these possible advantages brought by BDA, organizations need to invest in their digital transformation by collecting, storing, and analyzing big data. However, organizations are oblivious to how to integrate BDA into their daily software development lifecycle despite its disruptive impact. According to a recent survey [4], 73.4% of the organizations indicate that business adoption of big data is still one of the most challenging problems and process challenges are the most significant barrier to implement and deploy BDA software products and services. To this end, organizations need to have an inclusive and structured model to assess and improve their current BDA processes capabilities to manage their software projects more effectively and improve the efficacies of resulting products.

2.2. Capability maturity assessment models

Capability Maturity Assessment Models assist organizations in assessing and realizing the value of emerging capabilities and technologies [10–12]. They describe the fundamental practices, activities, and patterns in assessing current process capabilities and provide directions for improvement by considering the historical experiences and competencies of practitioners. Hence, they help organizations to minimize risks and costs, and maximize productivity and product quality.

The Capability Maturity Assessment Model approach provides a common framework to assess process capabilities and promotes incremental improvements through standardization and optimization [10]. It affects the overall employee involvement, overall costs, and product quality. There are widely adopted capability maturity models within the software industry, such as CMMI [6] and SPICE, which is superseded by a set of ISO/IEC standards, ISO 330xx standard series [7]. These frameworks have been utilized by hundreds of software organizations worldwide because of their proven benefits, including improving software quality, reducing expenses and inconsistencies, and increasing employee productivity and involvement [13]. Thus, adopting them to emerging domains, including financial and physical resource management [14], agile software development [15], green IT [16], Industry 4.0 [17], IT Management [18], and Automotive [19] attracted significant attention from both industry and academia. However, state-of-the-art CMMI and SPICE do not cover BDA-specific aspects such as unique development processes and pertinent lifecycle phases. Although there is a recent study [20] investigating the applicability and usability of the ISO/IEC 330xx standard in the BDA domain, the literature is surprisingly scarce in adapting the capability maturity assessment approach to the BDA domain.

2.3. Big data analytics frameworks

The BDA lifecycle comprises a set of processes from understanding the business requirements to derive actionable insights and deploying data products. There are various well-known frameworks or process models in the data mining and BDA domains, including SEMMA [21], Cross-Industry Standard Process for Data Mining (CRISP-DM) [22], Knowledge Discovery (KDD) [23], and Team Data Science Processes (TDSP) [24].

SEMMA defines five steps: Sample, Explore, Modify, Model, and Assess. The sample step is to select a portion of available data for further investigation. The explore step analyzes relationships, trends, and anomalies. The modify step corresponds to transforming data into a form that is suitable for modeling. The alternative models that may serve as the desired outcome are generated in the model step. Finally, the competing models are evaluated for usefulness and reliability. Similarly, the KDD framework includes data cleaning, integration, selection, transformation, data mining, and evaluation steps for knowledge discovery and mining. CRISP-DM is a well-known analytics framework in the literature. It defines six steps: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The TDSP framework has similar steps: business understanding, data acquisition and understanding, modeling, and deployment, but the TDSP framework includes some Scrum [25] elements such as backlog and sprints.

These well-known analytics frameworks may differ in low-level details, but they mostly share a common understanding of the data analytics lifecycle and essentially highlight six-core processes: Business Understanding, Data Understanding, Model Building, Evaluation, and Deployment and Use. However, these frameworks mainly lack providing structured and well-defined analytics processes to assist organizations by providing extensive guidance and facilitate improvement. Most of the BDA processes are managed ad-hoc at a low-level process capability, and their output work products and results are not repeatable [26]. Software engineering literature does not provide any techniques and frameworks to improve BDA-centric software development [3]. Because BDA poses inherently unique characteristics, processes, and development methodologies. To close this research gap, this study aims to extend these well-known frameworks by explicitly defining BDA processes and developing a BDA process reference model based on the ISO/IEC 330xx standards suite to manage, assess, and improve the BDA processes.

2.4. Related works

We carried out a literature review to identify existing capability maturity assessment models in the BDA domain. In doing so, we searched Web of Science (WoS) and Scopus databases with (*"big data" OR "data analytics" OR "business analytics" OR "business intelligence") AND ("maturity model" OR "capability model" OR "assessment model")* keywords. We initially retrieved 92 studies; 25 studies that appear only in WoS, 48 studies that appear only in Scopus, and 19 studies that appeared in both. Then, we reviewed and evaluated these studies according to two inclusion criteria: 1) published in a conference proceeding or in a journal to indicate an academic approach and 2) proposed a capability maturity assessment model or a roadmap for BDA. We also reviewed the references and citations of the selected papers to extend the literature review results. Therefore, we identified thirteen relevant studies within the scope of this research. These studies are listed in

- *Fitness for Purpose:* The level of fitness in terms of assessing the process capabilities of BDA.
- *Completeness:* The level of completeness of dimensions in addressing all or a subset of major dimensions of BDA.
- *Granularity of the Dimensions:* The level of detail in terms of attributes of each dimension.
- *Definition of the Measurement Attributes*: The level of detail in terms of the measurement attributes of the model.
- *Description of the Assessment Method:* The level of detail regarding the assessment method.
- *Objectivity of the Assessment Method:* The level of objectivity of the process capability assessment method.

As a result of the evaluation of the strengths and weaknesses of the identified thirteen models based on the predefined criteria, it was found out that some studies investigate the application of the data analytics maturity approach in specific domains: M1 for airline network planning, M2 and M4 for health informatics, M8 for zakat institutions, and M9 for financial sector companies. These models investigate data analytics from a domain-specific viewpoint, but they do not entirely address the data analytics requirements for general purposes. Besides, these studies do not explicitly define measurement methods and attributes. Moreover, they do not validate the objectivities of their models. M10, M11, M12, and M13 propose a capability maturity model in the business intelligence domain. M10 and M11 develop business intelligence capability maturity models with five capability maturity stages. M12 and M13 mainly focus on business intelligence maturity indicator identification. Nevertheless, these studies do not provide a complete model, nor do they specify any comprehensive detail about assessment processes and attributes. Hence, it is possible to conclude that there is no theoretically grounded model to assess organizations' business analytics capabilities among these studies. The remaining studies, M3, M5, M6, and M7, focus on proposing a maturity model for the big data domain. Although M7 and M9 broadly cover BDA processes, they have limitations related to the definitions of measurement attributes, descriptions of their assessment methods, and objectivities of their assessments as they are not developed based on a well-established process capability maturity assessment model, and they do not give enough detail to facilitate model application or action plan generation for process capability improvement.

Apart from the academic literature, the grey literature, including reports and whitepapers, also presents some studies [41, 42] in the context of assessing and improving BDA capabilities. Since they were proposed by a technology vendor or a consulting organization, they have some shortcomings: not following an unbiased academic view or a well-accepted standard to develop an objective model as well as not provide any empirical results about development, assessment methods, and attributes since the proposed models are their products to sell organizations, which hinders their acceptance in practice.

To sum up, there is a growing research interest in BDA in recent years. However, there is a scarcity of research in the literature on developing capability maturity assessment models specifically for the BDA domain. Existing studies mainly lack proposing validated models with sufficiently detailed assessment attributes and methods. Besides, they are far away from following a theoretical development approach to provide well-documented, complete, consistent, and unambiguous standard models. As a result, there is limited understanding of how organizations should assess their BDA processes, derive a gap analysis to understand the weaknesses and strengths of their BDA processes, and create a comprehensive roadmap for improvement suggesting process capabilities they should focus on future investments to create higher business value. To this end, this study aims to develop an objective, consistent, impartial, and repeatable capability assessment model for the BDA domain to address these critical research gaps in the literature. The theoretical background of this development is explained in the following section.

3. The big data analytics process capability assessment model

In this study, we propose a BDA process capability assessment model that can serve as a base for improving the BDA processes. The proposed model pursues a structured and standardized approach by assessing relevant BDA processes to perform improvement initiatives in a consistent, repeatable manner, assessed by adequate metrics with guidance on what to do for improvement in BDA. The proposed model is developed by customizing the baseline provided in the ISO/IEC 330xx family of standards according to the specific requirements of the BDA. The primary reasons for the selection of the ISO/IEC 330xx are its well-established and commonly accepted structure. It consists of technical standards, including a reference model for capability maturity model development [40]. It also provides the requirements for capability maturity model design, planning, and execution of process capability maturity assessments and the application of process improvements based on process assessments.

As delineated in Figure 1, The proposed model comprises two dimensions: process and capability. The process dimension covers process definitions of BDA lifecycle processes. Business Understanding, Data Understanding, Model Building, Evaluation, and Deployment and Use. A

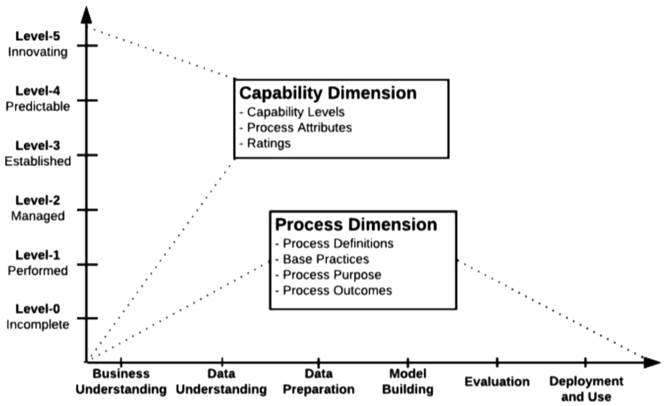

Fig. 1. Big data analytics process capability assessment model.

BDA Process Reference Model including process definitions for the BDA domain is developed based on the ISO/IEC 330xx standard series. The capability dimension includes the same capability levels, Process Attributes (PAs), generic practices, and generic practice indicators defined in ISO 33020 [43]. Applicability of the capability dimension that we employed to various processes across different domains has been shown by several studies [17,19,44–46].

3.1. Model development methodology

The proposed BDA process capability assessment model is grounded on a theoretical development methodology proposed by Becker et al. [47]. The development process includes the following steps:

Problem definition, problem relevance, and comparison of existing models step focuses on understanding the current state of the existing literature to validate problem definition and identify problem requirements. In this step, we reviewed the existing literature to understand the current state and limitations of the capability maturity assessment researches in the context of BDA. As a result of this review, it was figured out that the existing studies are in their early stages, and they mainly lack a theoretical framework to define objective capability measurement attributes, assessment methods and demonstrate the validity and applicability of the proposed models.

Define objectives of the model and determination of the development strategy step is devoted to the identification of business requirements of the model to understand primary processes and attributes for the model development. We first analyzed existing BDA frameworks and process models to determine BDA requirements. Then, we performed an exploratory case study [20] to validate the problem, define the model objectives, and determine the development strategy.

Iterative Model development step focuses on developing and validating the proposed model to demonstrate the final version of the model. We conducted an exploratory case study [20] to evaluate the applicability of the BDA process capability maturity models in the industry and discover improvement opportunities for further developments. Accordingly, the final version of the model was achieved as a result of iterations.

Demonstration of the final model and validation step is devoted to documenting and evaluating the final version of the model. In this step, we followed the multiple-case study research approach to evaluate the validity, applicability, and usability of the proposed BDA process capability assessment model, as we detailed in Section 4. We conducted case studies in different industries, countries, and organizational sizes to improve the generalizability of the results. We also documented and shared the final version of the model and the assessment results with the case study participants to collect feedback.

3.2. Process dimension

BDA process capability assessment model that covers the process definitions of the six BDA processes: Business Understanding, Data Understanding, Data Preparation, Model Building, Evaluation, Deployment and Use.

Business Understanding process involves identifying the business domain, its needs, objectives, and requirements from the data viewpoint. This process' fundamental purpose is to identify organizational gaps and weaknesses and employing a BDA problem to get rid of these weaknesses. As a result of the successful implementation of this process, objectives, scope, business success criteria, and performance indicators of a BDA software project are defined. Project requirements, limitations, assumptions, risks, and corrective action plans for the defined risk should be clearly defined. Moreover, a sketch project should be prepared to analyze the cost and potential outcomes, and business values of the project before deciding on the investment for the BDA software project.

Data Understanding process comprises tasks related to gathering relevant data from the sources and carrying out preliminary analysis by means of statistical and visualization techniques. The primary objectives of this process are to collect the data required, verify data quality, and discover relations that may be useful in model development. Organizations need to define their data limitations, policies, quality requirements, and quality metrics in this process. Thus, they can determine the strengths and weaknesses of the available data to decide whether the projected business goal can be reached or not.

Data Preparation process transforms data into a form that is suitable for building models. This process comprises selecting, cleaning, constructing, integrating, and formatting data steps. Data Preparation is one of the most challenging, influential, and time-consuming processes in BDA. Organizations collect big data from diverse sources, and these data may be incomplete, inconsistent, and erroneous leading to low performance, low-quality outputs, and the inability to extract valuable data patterns. Thus, this process covers practices for data transformation, outlier detection, feature extraction, and integration of data sources to build a complete and consistent data set for the model building process.

Model Building process assesses how organizations infer valuable knowledge and draw conclusions for their businesses by utilizing several data analytics approaches such as descriptive analytics, diagnostics, predictive analytics, and prescriptive analytics. As a result of the successful implementation of the model building process, organizations need to define their modeling requirements, objectives, and assumptions. They need to assess possible BDA techniques and tools to design a conceptual model for implementation. Moreover, they also should define success metrics and accordingly test designs to verify the output model.

Evaluation process aims to validate the success of the big data products and services developed in the model building process before deployment and use. This process includes determining key performance indicators, evaluating the continuity and reliability of the data sources utilized in the data products or services. It also includes collecting feedback from stakeholders to define possible future improvements in the output product.

Deployment and Use process covers preparing and allocating a software environment and information technologies infrastructure to deploy and use developed data products or services in daily organizational processes. Big data applications need emerging distributed and scalable software and hardware technologies for deployment and use. In this process, organizations need to define their deployment and use requirements and standards to determine how the system should work. This may include defining non-functional requirements: scalability, reliability, availability, maintainability, recoverability, and serviceability. This process also covers preparing a plan for monitoring and maintaining the deployed models or services.

These processes are defined in detail in the BDA Process Reference Model. The process definitions include measurable objectives of a process: Project Purpose, Process Outcomes, Base Practices, and Work Products, as defined in the ISO/IEC 33004:2015-*Process assessment-Requirements for process reference, process assessment, and maturity models* [40]. They also comprise definitions and relationships among tasks, activities, and output work products to achieve the process purpose and realize the process outcomes. These process definitions constitute the core of the proposed model as the BDA process capabilities of an organization are assessed according to these definitions by determining the level at which these processes are performed. As an example, we present the process definition of the Data Understanding process in Appendix-A.

3.3. Capability dimension

The capability dimension, applicable to any process, is adapted from the standard of ISO/IEC 33020-*Process assessment—Process measurement framework for assessment of process capability* [43]. According to the standard, the capability level represents the organization's ability to satisfy defined PAs. The defined PAs progress through the improvement of the capability of any process. The proposed model constitutes six capability levels ranging from Level 0 to Level 5, and each capability level includes at least one PA to address a specific strength of the assigned capability level to support continuous improvement, as delineated in Figure 2. As an organization improves its process capability, the BDA software processes and output work products become standardized, better defined, and more consistently implemented.

Level 0 – Incomplete: In this capability level, the organization does not successfully establish any of the defined base practices.

Level 1 – Performed: This capability level assesses the PA 1.1. Process Performance, which structurally defines processes, their base practices, and output work products. To achieve Level 1, organizations should at least largely implement base practices to demonstrate achievement in BDA. In this capability level, processes' outputs are mostly not standardized, inconsistent, and reactive.

Level 2 – Managed: At this level, the organizations are expected to understand the importance of BDA in their daily business processes and attempt to advance these processes to extend generated business value.

Level 3 – Established: This level comprises PA 3.1. Process Definition and PA 3.2. Process Deployment. PA 3.1. Process Definition assesses organizations' ability to establish a standardized and repeatable process definition for each process and PA 3.2. Process Deployment evaluates how organizations manage and maintain defined processes.

Level 4 – Predictable: In this capability level, we assess how organizations establish quantitative objectives to monitor and control their process performance with PA 4.1. Process Control and organizations' ability to employing effective, practical, and quantitative process performance measures to reduce variation in process performance with PA 4.2. Process Analysis. At this level, each BDA process is evaluated and monitored against a quantitative and statistical measure to manage and operate effectively and efficiently within defined measurable limits.

Level 5 – Innovating: In this capability level, we assess PA 5.1. Process Innovation to evaluate organizations' ability to identify potential improvements based on process performance and PA 5.2. Process Self-Optimization Implementation to assess how organizations learn from their quantitative process analysis to improve their processes continuously. At this capability level, organizations continuously strive to improve the range of their process capability, process performance, and quality of the output work products by innovative measures, methods, tools, and technologies.

The capability level of a process is determined based on the ratings of the PAs. The rating scale percentage values used for rating PAs according to ISO/IEC 33020 [43] are defined in Table 2. A process is defined to be at capability level x if all PAs lower than level than x are rated as "Fully Achieved (F.A)", and PAs at level x are rated as "Largely Achieved (L.A)" or "Fully Achieved (F.A)".

4. Materials and methods

To check the usefulness and applicability of the proposed BDA process capability assessment model, the case study approach, which is referred to as *"the most common qualitative method used in information systems"* [48], is used in this study. It is a highly suitable method to seek answers to the research questions and develop a solution. We conducted two case studies in two different organizations of different sizes, countries, and industries. We designated capability levels of the BDA processes performed in the organizations as primary measures of these case studies. To assess the capability levels of the organizations' BDA processes, we conducted face-to-face meetings with process owners, stakeholders, data scientists, and executive members of the organizations. We also collected organizational information-gathering documents and observations to eliminate bias in the assessment results.

4.1. Multiple-case study design

This multiple-case study research aims to evaluate the applicability and usability of the proposed model and to guide scholars and practitioners in how to conduct a process capability assessment with the proposed model. The primary motivations for this multiple-case study are the following research questions;

- *RQ-1:* How applicable and useful is the proposed ISO/IEC 330xx based model to be used to identify the current state of an organization's BDA process capabilities?
- *RQ-2:* What are the strengths and weaknesses of the proposed BDA process capability assessment model?

The case study researches may raise several validity concerns in conducting and analyzing the results. Thus, the potential validity threats need to be examined in the planning phase to take preventive action at the early stages. As suggested by Yin [49], we analyzed potential validity

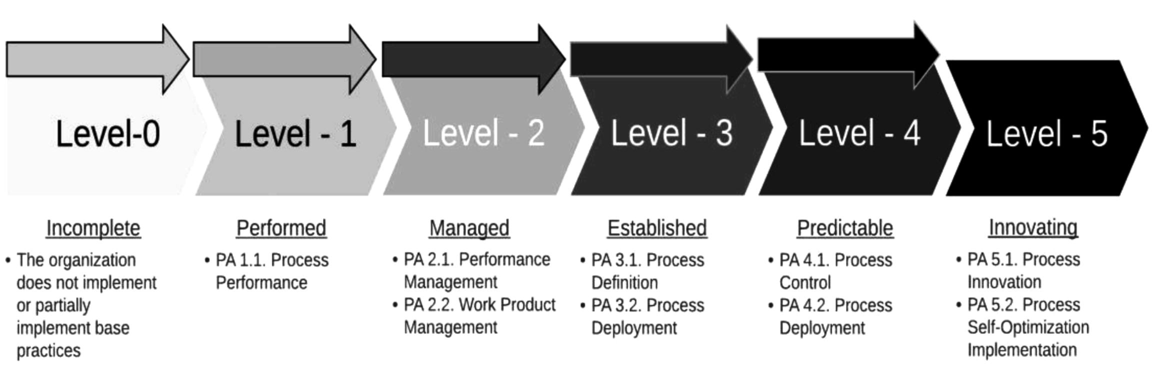

Fig. 2. Capability levels of BDA processes.

threats in four categories as follows:

The possible *Construct Validity* threat for these case studies is identifying the correct source of evidence and participants to gather objective judgments and eliminate bias. Thus, we conducted face-to-face interviews with both management and technical business units' participants, including process owners, stakeholders, data scientists, and executive members. Moreover, we collected data from multiple information sources, including interviews, observations, and information-gathering documents, to assess each process capability objectively.

Internal Validity threats concern the relationships between input variables and produced outputs to understand the factors affecting the results. To mitigate internal validity threats, we prepared an assessment report that includes process capability levels, evidence that indicates the weaknesses in each process, and suggestions to improve the current capability of each process. This report was shared with interview participants, and we discussed the results in a meeting to determine and eliminate any bias in assessment findings for each organization.

External Validity and *Reliability* threats concern the generalizability of the case study results and evaluate if the study is valid in its own setting or applicable in other settings as well. We employed a literal replication logic of the assessment process with two different organizations.

4.2. Data collection

In the multiple-case study, we collected data in two different organizations of different sizes, countries, and industries.

The first organization develops BDA software to improve the operational efficiency of a company that operates in the energy industry in Europe and Asia. We will refer to this case as Case - 1 for the purpose of confidentiality. In Case – 1, we held two different two-hour semi-structured meetings with process owners and executive members to collect direct evidence about the BDA processes of these organizations.

The second organization develops BDA software to improve customer satisfaction, optimize operational costs and resources. We held a three-hour meeting with the data scientist, IT manager, and stakeholders to collect data. We also collected and investigated the organization's system and software requirement specification documents, technical and architectural documentations, and reports for BDA software development. These interviews were recorded during the meeting with the consent of the participants. We will refer to this case as Case - 2 for the rest of the paper.

These organizations are selected in this study since both have an experience with BDA, and they aim to achieve widespread adoption of BDA software development processes. They are willing to advance their BDA software development process capabilities to drive growth and attain a competitive edge. We identified appropriate participants by contacting organizations and presenting the scope of the research. In both organizations, the participants are primarily responsible for digital transformation, BDA, information technologies, and management processes. The proposed BDA capability model, assessment and data collection procedure, assessment plan, and capability levels are explained in detail to each interview participant before conducting assessments. The notes taken during the interviews and organization process frameworks and documents collected were investigated offline in detail. Final assessment scores were identified with this offline investigation.

4.3. Data analysis: process capability assessment

The process capability assessment activities are delineated in Figure 3. ISO/IEC 33002-*Requirements for performing process assessment* [50] was followed to ensure assessment activities, such as assessment planning, data collection, ratings, and reporting, are performed in a standardized manner. The first step of the process capability assessment was preparing an assessment plan to explicitly designate assessment interview participants, assessment schedule, and the assessment reports' delivery date. After that, the final assessment plan schedule was shared with interview participants to arrange a discussion session. In the assessment interviews, we gathered information, process documents, and observations to rate each PA. The assessment results were presented and shared as a final report, including capability levels of each process, evidence and justifications that specify the shortcomings and strengths of the processes, and a roadmap that includes separate guidelines and recommendations for each process to advance their process capability to the next level.

These case studies aim to assess the six BDA processes; Business Understanding, Data Understanding, Data Preparation, Model Building, Evaluation, and Deployment and Use. These processes were assessed according to the guidelines in ISO/IEC 3300xx standards family to eliminate bias and provide consistent results. The Level-1 assessments were carried out with the focus of checking if the base practices indicated in the BDA Process Reference Model are performed according to the corresponding PA 1.1. Process Performance Attribute. This Level-1 assessment can be conducted owing to the developed BDA process reference model. The developed process definitions for six BDA processes were used for the first level assessment. For the assessment of PAs from PA 2.1 to PA 5.2, the generic practices and generic practice indicators provided by *ISO/IEC 33020-2019: Process assessment — Process measurement framework for assessment of process capability* [43] were used.

5. Findings and discussion

This section presents the application of the BDA process capability assessment model to guide scholars and practitioners in conducting a process capability by providing detailed assessment results with evidences. Intending to find answers for RQ-1 and RQ-2 raised in Section 4, we detailed the follow-up interview process to discuss applicability and usability and identify strengths and weaknesses of the proposed BDA process capability assessment model.

5.1. Assessment results

According to the collected evidence and findings during the interviews, the achievements of the PAs were rated by using the rating

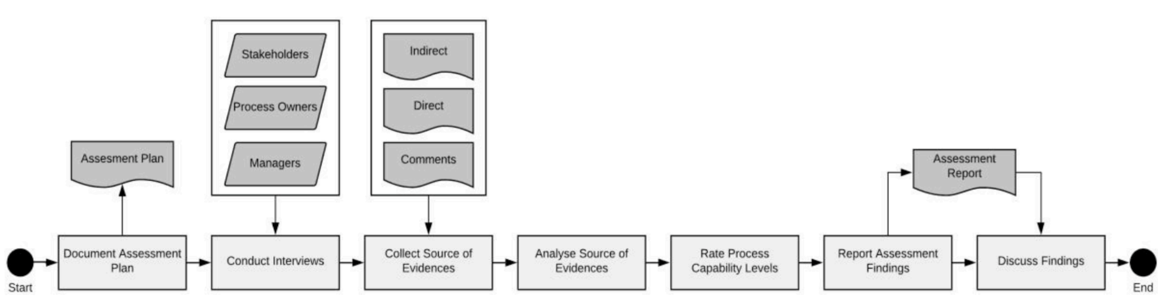

Fig. 3. Process capability assessment activities.

scale given in Table 2, and correspondingly capability levels of the BDA processes were determined based on *ISO/IEC 33020-2019: Process assessment — Process measurement framework for assessment of process capability* [43]. As a result, the capability level of the BDA process of "Business Understanding" was determined as Level – 2, and other BDA processes were determined as Level – 1 in Case – 1, as seen in Table 3. All BDA processes capability levels were determined as Level-3 in Case – 2, as given in Table 4.

As shown in Table 3, in Case – 1, only the business understanding process satisfies the requirements of being rated as level 2. This is mainly because Case – 1 does not have defined processes, including process responsibilities, authorities, key activities, the interaction between process practices, and work product templates. Case – 1 also has shortcomings in (Table 1) documenting the development process, designing a plan for testing intermediate and final BDA models, and preparing a monitoring and maintenance plan for deployment. Therefore, Case – 1 could not perform BDA processes robustly except the business understanding process.

As shown in Table 4, Case – 2 implements the defined BDA processes in a managed fashion by establishing, control, and maintaining process definitions and output work products. Thus, Case – 2 satisfies the requirements of the capability level – 2 in all of the BDA processes. Case – 2 also satisfies the requirements of capability Level – 3 in all of the BDA processes. However, Case – 2 has some problems in defining sequence and interaction among BDA processes to work as integrated processes and determining measures to monitor the efficiency and effectiveness of the processes. As Case – 2 could not fully perform Level – 3 PAs, we did not assess capability level – 4 PAs.

The capability level assessment results for both cases are summarized in Figure 4. As it can be seen from the figure, Case – 1 and Case – 2, are at different BDA process capability levels. Case – 1 has understood the importance of BDA for their businesses to gain competitive advantage and initiated some BDA projects across the organization. However, they perform and manage BDA projects in an ad-hoc manner, and BDA processes are poorly controlled, unpredictable, and reactive. This is mainly because they have a limited budget, limited human resources, and a distributed organizational structure that hinders managing and performing BDA processes more robustly. Thus, Case – 1 should first focus on business and data understanding processes to achieve higher BDA capability. On the other hand, Case – 2 has started its BDA journey three years ago with strong support from the top management. Accordingly, it has a clear business strategy, roadmap, and technology and people investments to build the right environment for BDA. Correspondingly, they have well-established and managed BDA processes. Case - 2 can quickly achieve higher capability in BDA by concentrating on quantitative management of BDA processes.

The proposed model also provides suggestions and a roadmap for organizations to improve the capability levels of their BDA processes. The participant organizations can improve their BDA process capabilities by strengthening their weak PAs. For example, Case – 1 can move the business understanding process to the next level by fully implementing PA 2.1 and PA 2.2. Similarly, Case-2 can improve its BDA processes by fully implementing PA 3.1. and PA 3.2. Accordingly, a set of improvement suggestions was generated for each case based on the BDA process capability assessments. The suggestions are exemplified in Table 5.

Table 1
Existing related works.

Model ID	*Paper Title*
M1 [27]	Towards a maturity model for big data analytics in airline network planning
M2 [28]	A health data analytics maturity model for hospitals information systems
M3 [29]	A Conceptual Framework for Assessing an Organization's Readiness to Adopt Big Data
M4 [30]	Measuring the maturity of business intelligence in healthcare
M5 [31]	Big Data Maturity Model - A Preliminary Evaluation
M6 [32]	How Can SMEs Benefit from Big Data? Challenges and a Path Forward
M7 [33]	How organisations leverage Big Data: a maturity model
M8 [34]	Big Data Maturity Model for Malaysian Zakat Institutions to Embark on Big Data Initiatives
M9 [35]	A Data Analytics Maturity Model for Financial Sector Companies
M10 [36]	Developing a capability maturity model for enterprise intelligence
M11 [37]	Towards a Business Analytics Capability Maturity Model
M12 [38]	Business intelligence maturity: the economic transitional context within Slovenia
M13 [39]	Using quantitative analyses to construct a capability maturity model for business intelligence

According to ISO/IEC 33004 standard, assessment results of a capability maturity assessment model should be "*complete, clear, unambiguous, objective, impartial, consistent, repeatable, comparable, and representative*" [40]. Correspondingly, we evaluated the existing thirteen models based on the following criteria:

Table 2
Rating scale percentage values according to ISO/IEC 33020 [43].

Not Achieved (N.A)	0% to ≤15% achievement	There are no or only very limited indications of PA fulfillment.
Partially Achieved (P. A)	> 15% to ≤ 50% achievement	There is evidence that the PA is partially implemented. However, some processes remain unpredictable.
Largely Achieved (L. A)	> 50% to ≤ 85% achievement	There is a significant achievement of the defined PA in the assessed process, but process performance still has some weaknesses.
Fully Achieved (F.A)	> 85% to ≤ 100% achievement	The measured PA is implemented completely.

Table 3
Capability level assessment results of the case – 1.

	Level – 1	*Level - 2*		*Level - 3*		
Processes	PA 1.1	PA 2.1	PA 2.2	PA 3.1	PA 3.2	*Results*
Business Understanding	F.A.	L.A.	L.A.	P.A.	N.A.	*Level – 2*
Data Understanding	F.A.	P.A.	N.A.	-	-	*Level – 1*
Data Preparation	F.A.	P.A.	N.A.	-	-	*Level – 1*
Model Building	L.A.	N.A.	N.A.	-	-	*Level – 1*
Evaluation	L.A.	P.A.	N.A.	-	-	*Level – 1*
Deployment and Use	L.A.	P.A.	N.A.	-	-	*Level – 1*

Table 4
Capability level assessment results of case – 2.

	Level – 1	*Level - 2*		*Level - 3*		
Processes	PA 1.1.	PA 2.1	PA 2.2	PA 3.1.	PA 3.2.	*Results*
Business Understanding	F.A.	F.A.	F.A.	L.A.	F.A.	*Level – 3*
Data Understanding	F.A.	F.A.	F.A.	F.A.	L.A.	*Level – 3*
Data Preparation	F.A.	F.A.	F.A.	F.A.	L.A.	*Level – 3*
Model Building	F.A.	F.A.	F.A.	L.A.	L.A.	*Level – 3*
Evaluation	F.A.	F.A.	F.A.	L.A.	L.A.	*Level – 3*
Deployment and Use	F.A.	F.A.	F.A.	F.A.	L.A.	*Level – 3*

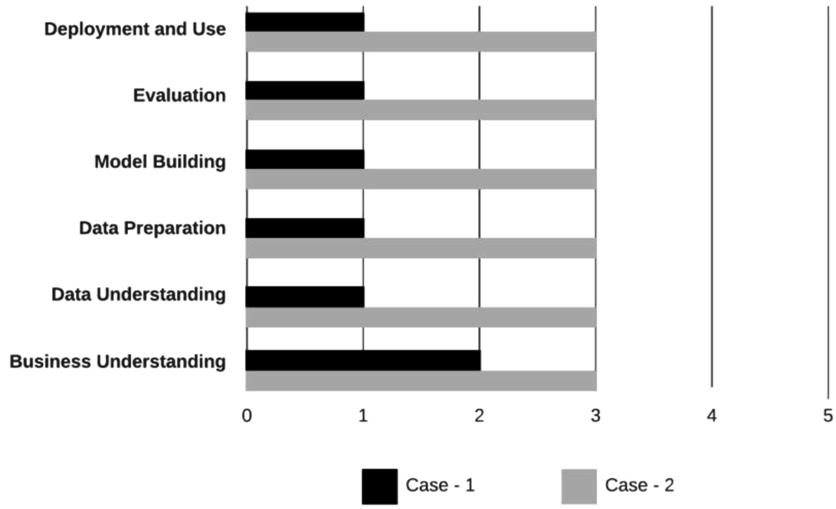

Fig. 4. Assessment results.

Table 5
Suggestions for improvement.

Process Name	Suggestions for Capability Improvement
Business Understanding *For Case – 1*	• Define a terminology glossary relevant to industrial terms and background information about big data analytics projects to enable common understanding. • Establish and maintain a draft project plan to analyze the project needs better and to estimate required IT and Human resources and budget to be needed throughout the project. • Collect and analyze data to assess the progress and achievement of the business understanding process
Data Understanding *For Case – 1*	• Evaluate and verify existing data quality issues and potential data anomalies. • Develop hypotheses that may be identified as necessary implications of the explored data. • Review hypotheses and findings with domain experts and stakeholders to collect feedback for further analyses.
Data Preparation *For Case – 1*	• Identify data selection requirements, standards, and policies. • Clean, transform and validate data quality and standardization issues to improve the success of big data analytics. • Evaluate and maintain enriching existing datasets by constructing, deriving, or generating new attributes.
Model Building *For Case – 2*	• Define the process (Responsibilities and authorities, key activities, sequence and interaction, milestones, work product templates) to perform it more robustly. • Define the interaction and sequence among BDA to integrate system processes. • Document development steps, source code, findings, and testing results to store in a knowledge management system to get benefit in similar projects.
Evaluation *For Case – 2*	• Evaluate the continuity and reliability of data sources. • Define measures to monitor the effectiveness of the evaluation process. • Identify the required hardware and software infrastructure to perform a standard evaluation process.
Deployment and Use *For Case – 2*	• Provide sufficient process infrastructure to support the performance, reliability, and availability. • Assign and communicate roles, responsibilities, and authorities for performing the defined process. • Allocate IT resources to prepare deployment and use environment.

5.2. Discussion

Following the completion of the BDA process assessments, we also shared assessment results and rating scales of each process in detail to compare participant organizations. With the aim of finding answers for the RQ-1 and RQ-2, we detailed the follow-up interview process to discuss applicability and usability and identify strengths and weaknesses of the proposed model. After presenting these assessment results and possible suggestions as a final report to the organizations, we held an additional interview with the same participants of the first interview to discuss assessment results and answer defined RQs. In these meetings, we discussed and evaluated the benefits, usability, and applicability of the proposed BDA capability assessment model with some open-ended structured questions. The 5-point Likert-scale was used, and the questions and medians of the responses are as follows;

- Are the assessment of the BDA process capability and the provided roadmap for improvement helpful for your organization? *Median: 4*
- Do you think that implementing these suggestions will increase your process performance? *Median: 4*
- Do you think that the language and terminology in the questions are easy to understand? *Median: 5*

The participants stated that the assessment reports were well prepared, and the provided results reflect the current situation of their companies. An executive member of Case – 1 stated that the assessment results are fully compatible with their own findings and the provided suggestions for improvement include both previously discovered and undiscovered issues and weaknesses. Both organizations indicated that the provided roadmap and suggestions for process improvements are applicable, and they will implement these suggestions in their BDA processes. They also stated that the language and terminology utilized during the interviews were easy to understand.

We also discussed the validity of the proposed BDA capability assessment model based on the criteria defined in Section 2, for developing capability maturity assessment models in Table 6.

6. Conclusions and future work

BDA offers promising business opportunities for organizations to improve their operational efficiency, develop a new business model, and drive new revenue streams and growth. Organizations need to adopt BDA processes and utilize data products in their daily business activities to seize these opportunities. However, this is not a trivial task. Many organizations are looking for ways to integrate BDA into their daily processes and improve BDA development capabilities. Thus, this research focuses on developing a standardized BDA capability assessment model based on the well-accepted standard, ISO/IEC 330xx, to assess organizations' current BDA process capabilities, derive a gap analysis, and present a comprehensive roadmap for improvement in a standardized way.

Table 6
Evaluation of the proposed model based on the predefined criteria.

Criteria	Analysis of the proposed model
Audience	The audience of the proposed model is practitioners and managers in organizations that develop BDA software, and researchers focus on the adoption and management of BDA.
Purpose and Scope - *Fitness* for Purpose	The proposed model aims to guide businesses in evaluating their existing BDA software development processes capabilities and revealing their weaknesses to provide an extensive guideline for improvement in a structured and repeatable way at both the organization and project levels. We propose a generic model that can be used in the assessment of BDA-centric software development projects in different industries and organizations.
Processes - *Completeness of Processes*	The proposed BDA capability model covers the six processes, extends, and comprises the well-accepted big data analytics frameworks by both practitioners and scholars.
Capability Levels - *Definition of Measurement Attributes* - *Description of Assessment Method* - *The objectivity of the Assessment method*	The proposed model constitutes six capability levels from Level 0 to Level 5 adapted from the ISO/IEC 330xx. Each capability level includes at least one well-defined measurement attribute to address a specific strength of the assigned capability level to support continuous improvement. We followed the assessment method provided by the ISO/IEC 330xx for objective assessment results. The capability levels are rated according to a rating scale provided by the ISO/IEC 330xx for the objective, consistent and unbiased results.
Verification and Validation	We validated the applicability and usefulness of the proposed model by conducting a multiple-case in two different organizations. The proposed model is also validated in follow-up discussion meetings with the case study participants from both management and technical backgrounds.

The main contributions of this study are four folds; (1) providing a review of the available capability maturity assessment models from a specific BDA perspective, (2) proposing a novel BDA process capability assessment model based on a well-accepted ISO/IEC 330xx standard, (3) extending the boundaries of SPICE, the ISO/IEC 330xx by demonstrating its applicability in the BDA domain (4) the observation of the feasibility of the proposed model through a multiple-case study in two different organizations. The multiple-case study results indicate that the proposed model is suitable and applicable in assessing the current BDA processes capabilities of an organization. It also provides valuable knowledge and insights to build an organizational roadmap to improve BDA processes' capabilities and effectiveness.

Despite the contributions of this research, there are a number of limitations to be addressed as part of future work. The main limitation of this research is the generalizability of the case study results. We performed a multiple-case study in two different organizations to check the applicability and usefulness of the proposed approach. In future studies, we plan to conduct literal replication of the multiple-case study with organizations in different industries and sizes to demonstrate efficacy and validate the generalizability of the proposed model. We are also planning to develop a self-assessment software tool to ease the assessment process and enable organizations to assess their own BDA capabilities.

CRediT authorship contribution statement

Mert Onuralp Gökalp: Conceptualization, Methodology, Validation, Writing – original draft. **Ebru Gökalp:** Conceptualization, Methodology, Supervision. **Kerem Kayabay:** Conceptualization, Formal analysis. **Selin Gökalp:** Investigation, Visualization, Formal analysis, Writing – review & editing. **Altan Koçyiğit:** Supervision, Writing – review & editing. **P. Erhan Eren:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Appendix- A

Table A1

Table A.1
Data understanding process definition.

Process Name	Data Understanding
Process Purpose	The Data Understanding process aims to analyze data needs sources and verify data quality and analyze data. It is the ability to understand the data with statistical and mathematical thinking
Process Outcomes	As a result of the successful implementation of this process: 1) Data needs, standards, limitations, and policies are defined; 2) Internal and external possible data sources are listed; 3) Data quality requirements and quality metrics are defined; 4) A data description report is prepared; 5) Data quality is assessed and verified against defined quality metrics; 6) A data exploration report is established; 7) Hypotheses are developed and tested to infer some novel findings; 8) Data sources, quality, hypotheses are shared and reviewed with related stakeholders.
Base Practices (BPs)	**BP1: Define Data Needs and Sources.** Define data needs and correspondingly determine internal and external data sources to meet these needs. Moreover, specify standards, limitations, and policies in data source selections. [Outcomes: 1, 2]**BP2: Define Data Quality Requirements and Metrics.** Define data quality requirements to meet organizational needs. These requirements may include business policies, regulatory compliance, data exchange formats, and industry standards. Define metrics to report and analyze data quality. [Outcome: 3]**NOTE 1:** This may include; ○ specifying the level of data granularity, including data value, data element, data record, and data table. ○ defining data quality indicators including, but are not limited to; Accuracy, Completeness, Consistency, Precision, Privacy, Timeliness, and Validity. **BP3: Prepare a Data Description Report.** Prepare a data description report for available data sources. [Outcome: 4] **BP4: Assess and Verify Data Quality.** Evaluate and verify existing data quality issues and potential data anomalies. [Outcome: 5]**BP5: Prepare a Data Exploration Report.** Prepare a data exploration report to understand the data with statistical and mathematical thinking. [Outcome: 6]**NOTE 2:** This may include; ○ Relationships between attributes ○ Performing simple aggregations ○ Simple statistical analyses **BP5: Develop and Test Hypotheses to Infer Findings.** Develop hypotheses that may be identified as necessary implications of the explored data. Moreover, infer findings and observations about the dataset to summarize the essential points, explain the dataset. [Outcome: 7]**BP6: Review Findings with Domain Experts.** Review findings with domain experts and collect feedback for further analyses. [Outcome: 8]
Output Work Products	Definition of data needs, standards, and policies. → [Outcome: 1] Internal and external data sources. → [Outcome: 2] Data description report → [Outcomes: 1, 2, 3, 4] Data exploration report → [Outcomes: 5, 6] Data understanding document → [Outcome: ALL]

References

[1] Gartner, "Our Top Data and Analytics Predicts for 2019," 2019. https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/(accessed Nov. 19, 2020).
[2] Deloitte, "The Analytics Advantage We're just getting started," 2013. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Deloitte-Analytics/dttl-analytics-analytics-advantage-report-061913.pdf (accessed Oct. 14, 2020).
[3] M. Kim, Software Engineering for Data Analytics, IEEE Softw 37 (4) (2020) 36–42, https://doi.org/10.1109/MS.2020.2985775.
[4] NewVantage, "Big Data and AI Executive Survey," 2020. http://newvantage.com/wp-content/uploads/2020/01/NewVantage-Partners-Big-Data-and-AI-Executive-Survey-2020-1.pdf (accessed Nov. 10, 2020).
[5] P. Mikalef, M. Boura, G. Lekakos, J. Krogstie, Big Data Analytics Capabilities and Innovation: The Mediating Role of Dynamic Capabilities and Moderating Effect of the Environment, Br. J. Manag. 30 (2) (2019) 272–298, https://doi.org/10.1111/1467-8551.12343.
[6] C. P. Team, "CMMI for Development, version 1.3," 2010. https://resources.sei.cmu.edu/library/asset-view.cfm?assetID=9661 (accessed Oct. 12, 2020).
[7] ISO/IEC, "ISO/IEC 33001:2015 Information technology – Process assessment – Concepts and terminology," 2015. https://www.iso.org/standard/54175.html (accessed May 12, 2019).
[8] I. Ozkaya, What Is Really Different in Engineering AI-Enabled Systems? IEEE Software 37 (2020) 3–6, https://doi.org/10.1109/MS.2020.2993662, 4. IEEE.
[9] M.O. Gökalp, A. Koçyiğit, P.E. Eren, A visual programming framework for distributed Internet of Things centric complex event processing, Comput. Electr. Eng. 74 (2019) 581–604, https://doi.org/10.1016/j.compeleceng.2018.02.007.
[10] M.C. Paulk, B. Curtis, M.B. Chrissis, C.V Weber, Capability maturity model, version 1.1, IEEE Softw 10 (4) (1993) 18–27, https://doi.org/10.1109/52.219617.
[11] T.C. Lacerda, C.G. von Wangenheim, Systematic literature review of usability capability/maturity models, Comput. Stand. Interfaces 55 (2018) 1339–1351, https://doi.org/10.1016/j.csi.2017.06.001.
[12] D. Proença, J. Borbinha, Formalizing ISO/IEC 15504-5 and SEI CMMI v1.3 – Enabling automatic inference of maturity and capability levels, Comput. Stand. Interfaces 60 (2018) 13–25, https://doi.org/10.1016/j.csi.2018.04.007.
[13] A. Tarhan, O. Turetken, H.A. Reijers, Business process maturity models: A systematic literature review, Inf. Softw. Technol. 75 (2016) 122–134, https://doi.org/10.1016/j.infsof.2016.01.010.
[14] E. Gökalp, O. Demirörs, Model based process assessment for public financial and physical resource management processes, Comput. Stand. Interfaces 54 (2017) 186–193, https://doi.org/10.1016/j.csi.2016.11.011.
[15] Ö. Özcan-Top, O. Demirors, Application of a software agility assessment model – AgilityMod in the field, Comput. Stand. Interfaces 62 (2019) 1–16, https://doi.org/10.1016/j.csi.2018.07.002.
[16] J.D. Patón-Romero, M.T. Baldassarre, M. Rodríguez, M. Piattini, Green IT Governance and Management based on ISO/IEC 15504, Comput. Stand. Interfaces 60 (2018) 26–36, https://doi.org/10.1016/j.csi.2018.04.005.
[17] E. Gökalp, U. Şener, and P. E. Eren, "Development of an assessment model for industry 4.0: Industry 4.0-MM," in *Communications in Computer and Information Science*, 2017, vol. 770, pp. 128–142, doi: 10.1007/978-3-319-67383-7_10.
[18] B. Barafort, A. Shrestha, S. Cortina, A. Renault, A software artefact to support standard-based process assessment: Evolution of the TIPA® framework in a design science research project, Comput. Stand. Interfaces 60 (2018) 37–47, https://doi.org/10.1016/j.csi.2018.04.009.
[19] S.I.G. Automotive, Automotive SPICE process assessment model, Final Release, v4 4 (2010) 46.
[20] M.O. Gökalp, K. Kayabay, E. Gökalp, A. Koçyiğit, P.E. Eren, Towards a Model Based Process Assessment for Data Analytics: An Exploratory Case Study, Communications in Computer and Information Science 1251 CCIS (2020) 617–628, https://doi.org/10.1007/978-3-030-56441-4_46.
[21] SAS, SEMMA (2017). https://documentation.sas.com/?docsetId=emref&docsetTarget=n061bzurmej4j3n1jnj8bbjjm1a2.htm&docsetVersion=14.3&locale=en. accessed Nov. 25, 2020.
[22] R. Wirth, CRISP-DM : Towards a Standard Process Model for Data Mining, Proc. Fourth Int. Conf. Pract. Appl. Knowl. Discov. Data Min. (24959) (2000) 29–39.
[23] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, Knowledge Discovery and Data Mining: Towards a Unifying Framework, Int Conf Knowl. Discov. Data Min. 96 (1996) 82–88.
[24] Microsoft, "Team Data Science Process," 2018. https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/(accessed Jan. 10, 2020).
[25] K. Schwaber, SCRUM Development Process," in *Business Object Design and Implementation*, Springer, 1997, pp. 117–134.
[26] J. Saltz, I. Shamshurin, K. Crowston, Comparing Data Science Project Management Methodologies via a Controlled Experiment, in: Proc. 50th Hawaii Int. Conf. Syst. Sci, 2017, https://doi.org/10.24251/hicss.2017.120.
[27] I. Hausladen, M. Schosser, Towards a maturity model for big data analytics in airline network planning, J. Air Transp. Manag. 82 (2020), 101721, https://doi.org/10.1016/j.jairtraman.2019.101721.
[28] J.V. Carvalho, Á. Rocha, J. Vasconcelos, A. Abreu, A health data analytics maturity model for hospitals information systems, Int. J. Inf. Manage. 46 (2019) 278–285, https://doi.org/10.1016/j.ijinfomgt.2018.07.001.
[29] C.M. Olszak, M. Mach-Król, A conceptual framework for assessing an organization's readiness to adopt big data, Sustain 10 (10) (2018) 3734, https://doi.org/10.3390/su10103734.
[30] L. Gastaldi, et al., Measuring the maturity of business intelligence in healthcare: Supporting the development of a roadmap toward precision medicine within ISMETT hospital, Technol. Forecast. Soc. Change 128 (2018) 84–103, https://doi.org/10.1016/j.techfore.2017.10.023.
[31] S.M. Drus, N.H. Hassan, Big Data Maturity Model – a Preliminary Evaluation, in: 6th International Conference on Computing and Informatics, ICOCI, 2017, pp. 613–620.
[32] S. Coleman, R. Göb, G. Manco, A. Pievatolo, X. Tort-Martorell, M.S. Reis, How Can SMEs Benefit from Big Data? Challenges and a Path Forward, Qual. Reliab. Eng. Int. 32 (6) (2016) 2151–2164, https://doi.org/10.1002/qre.2008.
[33] M. Comuzzi, A. Patel, How organisations leverage: Big Data: A maturity model, Ind. Manag. Data Syst. 116 (8) (2016) 1468–1492, https://doi.org/10.1108/IMDS-12-2015-0495.
[34] H. Sulaiman, Z.C. Cob, N. Ali, Big data maturity model for Malaysian zakat institutions to embark on big data initiatives, in: 2015 4th International Conference on Software Engineering and Computer Systems 2015, ICSECS, 2015, pp. 61–66, https://doi.org/10.1109/ICSECS.2015.7333084. *Virtuous Software Solutions for Big Data*.
[35] J. Perales-Manrique, J. Molina-Chirinos, P. Shiguihara-Juarez, A Data Analytics Maturity Model for Financial Sector Companies, in: SHIRCON 2019 - 2019 IEEE Sciences and Humanities International Research Conference, 2019, pp. 1–4, https://doi.org/10.1109/SHIRCON48091.2019.9024885.
[36] J. Huffman and L. E. Whitman, "Developing a capability maturity model for enterprise intelligence," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 2011, vol. 44, no. 1 PART 1, pp. 13086–13091, doi: 10.3182/20110828-6-IT-1002.02641.
[37] R. Cosic, G. Shanks, S. Maynard, Towards a business analytics capability maturity model, in: ACIS 2012 : Proceedings of the 23rd Australasian Conference on Information Systems, 2012, pp. 1–11.
[38] T. Lukman, R. Hackney, A. Popovič, J. Jaklič, Z. Irani, Business intelligence maturity: The economic transitional context within Slovenia, Inf. Syst. Manag. 28 (3) (2011) 211–222, https://doi.org/10.1080/10580530.2011.585583.
[39] D. Raber, R. Winter, F. Wortmann, Using quantitative analyses to construct a capability maturity model for Business Intelligence, in: Proceedings of the Annual Hawaii International Conference on System Sciences, 2012, pp. 4219–4228, https://doi.org/10.1109/HICSS.2012.630.
[40] ISO/IEC, "ISO/IEC 33004: Information technology - Process assessment - Requirements for process reference, process assessment and maturity models," 2015. https://www.iso.org/standard/54178.html (accessed Dec. 10, 2020).
[41] B. F. Halper and D. Stodder, "A Guide to Achieving Big Data Analytics Maturity (TDWI Benchmark Guide)," 2016. https://tdwi.org/whitepapers/2018/01/aa-all-ms-a-guide-to-achieving-big-data-analytics-maturity.aspx (accessed May 06, 2019).
[42] V. Dhanuka, Hortonworks Big Data Maturity Model, Hortonworks (2016). http://hortonworks.com/wp-content/uploads/2016/04/Hortonworks-Big-Data-Maturity-Assessment.pdf. accessed May 06, 2019.
[43] ISO/IEC, "ISO/IEC 33020:2015 - Information technology - Process assessment - Process measurement framework for assessment of process capability," 2015. https://www.iso.org/standard/54195.html (accessed Dec. 10, 2020).
[44] H.W. Jung, K.F. Ting, An empirical evaluation of ISO/IEC 15504-5 capability measures: Reflective or formative, Comput. Stand. Interfaces 53 (2017) 123–130, https://doi.org/10.1016/j.csi.2017.03.002.
[45] M.O. Gökalp, K. Kayabay, E. Gökalp, A. Koçyiğit, P.E. Eren, Assessment of process capabilities in transition to a data-driven organisation: A multidisciplinary approach, IET Softw (2021), https://doi.org/10.1049/sfw2.12033.
[46] Ö. Tanriöver, O. Demirörs, A process capability based assessment model for software workforce in emergent software organizations, Comput. Stand. Interfaces 37 (2015) 29–40, https://doi.org/10.1016/j.csi.2014.05.003.
[47] J. Becker, R. Knackstedt, J. Pöppelbuß, Developing Maturity Models for IT Management, Business & Information Syst. Eng. 1 (3) (2009) 213–222, https://doi.org/10.1007/s12599-009-0044-5.
[48] J. Creswell, Qualitative, quantitative, and mixed methods approaches. Research design, Sage publications, 2013, pp. 1–26.
[49] R.K. Yin, Case study research: Design and methods (applied social research methods, London Singapore Sage (2009).
[50] ISO/IEC, "ISO/IEC 33002:2015 Information technology — Process assessment — Requirements for performing process assessment," 2015. https://www.iso.org/standard/54176.html (accessed Dec. 10, 2020).