Risk Identification of Data Science Projects: A Literature Review

Maike Holtkemper, Maria Potanin, Alexander Oberst and Christian Beecks

University of Hagen, Department of Mathematics and Computer Science, Chair of Data Science, Germany

Abstract

While cost, time, and resources are considered to have a high impact on data science projects, risks are the key for the successful implementation of a project. The correct handling of risks has been proven to increase a project's chances of success. Therefore, awareness of existing and emerging risks as well as their assessment play a major role in data science project management. In 2021, 87 percent of data science projects fail and are thus not implemented successfully. The path of successful implementation of data science projects in companies is more complex and uncertain than for conventional projects, in particular the identification and availability of the necessary skills in a company before and during the project. Regarding software engineering projects, the analysis and evaluation of potential risks is well-known, but for data science projects potential risks have not yet been analyzed at a larger scale. To identify the potential reasons for failure of data science projects, an in-depth understanding of potential risks and measures for mitigation these risks is inevitable. In this paper, we conduct a systematic literature review on risks of data science projects and present the top fifteen risks as first findings. Furthermore, we compare the identified risks to the major risks of software engineering projects to highlight similarities and differences between these two disciplines. The findings of our literature review can be used in guiding the development of future risk assessment systems for complex data science projects.

Keywords
Data Science, Projects, Risk Identification

1. Introduction

Data science projects have become an integral part of today's companies. Leveraging data has proven essential in the ongoing competition for market share and competitiveness [38]. Instead of the service or product itself, the collected data is often considered as valuable asset, which can be used for developing innovative business models such as "Smart Services" [30]. In a representative study conducted by MIT [35] of 2602 managers, executives, and data professionals, it has been demonstrated that companies were able to gain a competitive advantage, optimize existing processes and increase in the development of innovative business models through the use of data and analytics. In addition, the study also reveals that these companies were able to supplement human skills through the use of smart machines, thus reducing time-consuming tasks [35]. However, the path to successful implementation of data science projects in companies is more complex and uncertain than for conventional projects, "a multidimensional challenge requiring specialised tools, processes and methodologies" [26], in particular the identification and availability of the necessary skills in a company before and during the project [11]. In

2013, "55 percent of big data projects don't get completed, and many more fail their objectives" [19]. In 2015, Gartner [14] states that "through 2017, 60 percent of big data projects will fail to go beyond piloting and experimentation, and will be abandoned". In 2021, 87 percent of data science projects fail and do not get deployed [53]. To address this challenge, data science process models [18] have been developed to assist during the realization of a data science project. While process models in software projects are commonly used and well-evolved over time, they gain in importance for data science [26]. A comparison between seven data science process models shows, that four of seven highlight aspects of project management [26], but pre- and post related project tasks such as risk management are poorly represented.

Risk management can not only help to create an awareness for existing risks at the beginning of a data science project, but also to make the project team sensitive to emerging risks during the project duration. In this way, challenges can be recognized and addressed at an early stage. Therefore, the first step in risk management is the identification of risks to assure the project success [47]. For software engineering projects, several risk factors have already been identified. Keil et al. [23] published a framework for identifying software project risks and in the last twenty years, critical factors and a top ten risk lists for software project failure have been published [1, 4, 13, 15, 24, 44]. Compared to software development, data science is still quite young and has only gained importance in recent years [36], so that there is considerably less literature available to date in the area of risk management and project risks, which focus on data science-specific aspects such as cross-disciplinary competences of team members.

This paper aims to investigate the risks of data science projects. First, a literature review on risks of data science projects will be performed and second, the identified risks will be categorized. Therefore, the literature review aims to answer the following research question (RQ1): *What are the major risks of data science projects?* The top fifteen identified risks will be then described and compared to the most common risks of software engineering projects, which leads to the second research question (RQ2): *What are the similarities and differences between the risks of a software engineeering project and those of a data science project?*

The paper is structured as follows: First, the background is described in Section 2. Second, the methodology of the data collection is described in Section 3. Third, the findings are presented and discussed in Section 4. Finally, the future work as well as the limitations are given in Section 5.

2. Background

Based on Muhlbauer, risk can be defined as result in potential failure or loss [32]. In the context of a project, risk can be defined as an event that is uncertain with negative impact or positive objectives in case of occurrence [33]. This led to the standard definition of risk. According to standard ISO 31000:2018, risk is defined as the "effect of uncertainty on objectives" [21]. An effect is described as deviation from the expected and can be positive, negative, or both, and can address, create, or result in opportunities and threats [21]. Objectives can have different aspects and categories and can be applied at various levels [21]. Regarding risk, important terms are sources, events, consequences and likelihood [20]. Risk sources identify where risks can originate, risk events denote the concrete realization of a risk, risk consequence implicates the

potential outcome of the risk event and risk likelihood is a qualitative assessment that describes how likely a risk will occur. The related term risk management then explains the coordination and controlling of risk activities within the organizations, which includes risk assessment, risk treatment, risk acceptance and risk communications. [20, 21, 22]

Risk assessment is the overall process of risk identification, risk analysis, and risk evaluation. The purpose of risk identification is to find, recognize, and describe risks that prevent or help an organization from achieving its objectives. By identifying risks, the fundamental questions are what, where, when, why, and how a potential risk could happen and then categorizing risk elements [21]. According to the same standard, risk analysis is a systematic process for understanding risk and its characteristics. It involves a detailed view of uncertainties, sources of risk, consequences, probability, events, scenarios, controls and their effectiveness. An event can have multiple causes and consequences, which in turn affect multiple objectives. In terms of analysis techniques, these can be qualitative, quantitative in nature, or a combination of both. Risk analysis provides input to risk assessment and decisions about whether and how a risk should be treated. Risk assessment involves the comparison of the results of the risk analysis with the established risk criteria to determine where additional measures are required [20, 21]. According to Grassi et al. [17], risk evaluation is the most important task in risk assessment.

Although risk assessment is an important part of software project management, it is still at its infancy in data science projects and thus highly underestimated and often not performed at all. As Kutzias et al. [26] demonstrated, recent data science process models such as CRISP-DM or DASC-PM emphasize project management, but neglect pre- and post project-specific tasks, which also include risk assessment. In software projects, many risks occur during the process of creating the software, where the risk may lie, for example, in understanding the requirements, integrating modules or feasibility [37]. Risk identification in the design phase is crucial because "if senior manager fail to detect such risks, it is possible that such projects may collapse completely" [31]. To create awareness of risk assessment within data science projects, this paper focuses on identifying potential risks by means of a literature review. The methodology is described in the following section.

3. Research Method

Based on the recommendations of Webster and Watson [52], a literature review is conducted, including keyword and backward search. According to the procedure of vom Brocke et al. [49], the process is documented. The procedure of the literature review process is summarized in Figure 1.

The keyword investigation is focused on risks of data science projects. In the first step, the scientific literature databases Scopus, IEEE Explore, EBSCOhost and ACM Digital Library were searched by defined search queries (cf. Table 1).

As a first result, 354 papers were found. In a second step, the duplicates were removed, so that 314 paper remained. In a third step, the abstracts of these remained paper were evaluated according to their content, leading to 57 paper which were considered to be relevant. In a fourth step, the entire content of these papers was evaluated by extracting named risks. As a result from the keyword search 24 papers (cf. Table 2) and the therein mentioned 143 risks
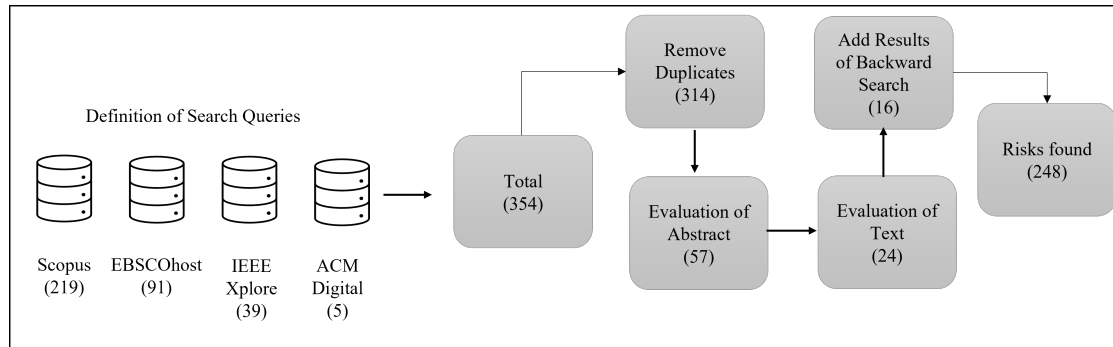
Figure 1: Literature Review Process

Table 1
Search Terms

Search string	Scopus	EBSCOhost	IEEE Xplore	ACM Digital
("Data Science Projects" OR "Big Data Projects" OR "Machine Learning Projects") AND ("Risk*" OR "Challenge*")	219	91	39	5
Total				354

were considered relevant for a further backward search. In a fifth step, the references of the 24 relevant papers from the keyword search were searched by further papers regarding risks of data science projects. As a result from the backward search, 16 papers (cf. Table 3) were added to the total amount of relevant papers as well as additional 105 identified risks. As a final result, a total of 248 risks of data science projects were examined by the aforementioned literature review process.

Table 2
Findings of Keyword Search

Title	Reference	Year	Search Type
Demystifying Data Science Projects: A Look on the People and Process of Data Science Today	Aho et al.	2022	keyword
On the Application of SCRUM in Data Science Projects	Kraut and Transchel	2022	keyword
Toward a Conceptualization of Big Data Value Chain: From Business Problems to Value Creation	Louati and Mekadmi	2022	keyword
On the Experiences of Adopting Automated Data Validation in an Industrial Machine Learning Project	Lwakatare et al.	2021	keyword
The Risk Management Process for Data Science: Gaps in Current Practices	Lahiri and Saltz	2022	keyword
Risks of Data Science Projects- A Delphi Study	Varela	2022	keyword
Analyzing a Data Science Online Practitioner Community: Trends and Implications for Data Science Project Management	Tacheva et al.	2022	keyword

Title	Reference	Year	Search Type
Evaluating Data Science Project Agility by Exploring Process Frameworks Used by Data Science Teams	Lahiri and Saltz	2023	keyword
Managing and Composing Teams in Data Science: An Empirical Study	Aho et al.	2021	keyword
A survey study of success factors in data science projects	Martinez et al.	2021	keyword
Nine Questions to Evaluate a Data Science Team's Process: Exploring a Big Data Science Team Process Evaluation Framework Via a Delphi Study	Saltz	2022	keyword
Don't Be Afraid of Failure—Insights from a Survey on the Failure of Data Science Projects	Aßmann	2023	keyword
An iterative and incremental data preprocessing procedure for improving the risk of big data project	Lai and Leu	2017	keyword
Bad big data science	Haug	2016	keyword
Identifying critical issues in smart city big data project implementation	Barham and Daim	2018	keyword
Big data and business analytics: Trends, platforms, success factors and applications	Ajah and Nweke	2019	keyword
Evolutionary Computing Environments: Implementing Security Risks Management and Benchmarking	Malik and Singh	2020	keyword
Privacy, security and legal challenges in big data	Abdullah	2018	keyword
Significance and challenges in big data: A survey	Jothi et al.	2016	keyword
The need for an enterprise risk management framework for big data science projects	Saltz and Lahiri	2020	keyword
Significance and Challenges of Big Data Research	Jin et al.	2015	keyword
Big data project success - A meta analysis	Koronios et al.	2014	keyword
Five Reasons Why Your Data Science Project is Likely to Fail.	Preimesberger	2019	keyword
Inadequate infrastructure halting big data projects.	Connolly	2015	keyword
Total Findings of Keyword Search			24

Table 3

Findings of Backward Search

Title	Reference	Year	Search Type
Towards an Improved ASUM-DM Process Methodology for Cross-Disciplinary Multi-organization Big Data and Analytics Projects	Angée et al.	2018	backward
The Age of Data: What You Need to Know About Fundamentals, Algorithms, and Applications	Aust	2021	backward
Achieving Agile Big Data Science: The Evolution of a team's Agile Process Methodology.	Saltz and Shamshurin	2019	backward

Title	Reference	Year	Search Type
Comparing Data Science Project Management Methodologies via a Controlled Experiment	Saltz et al.	2017	backward
SKI: An Agile Framework for Data Science.	Saltz and Suthrland	2019	backward
Exploring Project Management Methodologies Used Within Data Science Teams.	Saltz et al.	2018	backward
Progressive Data Science: Potential and Challenges.	Turkay et al.	2018	backward
Addressing barriers to big data	Alharthi et al.	2017	backward
Data-intensive applications, challenges, techniques and technologies: A survey on Big Data	Chen and Zhang	2014	backward
Beyond the hype: Big data concepts, methods, and analytics	Gandomi and Haider	2015	backward
Data science: challenges and directions.	Cao	2017	backward
Critical success factors for managing data science projects within agile methodology	Limesha	2021	backward
Big-data/analytics projects failure: a literature review	Reggio	2020	backward
Data Management Risks: A Bane of Construction Project Performance	Tanga et al.	2022	backward
A Critical Quality Measurement Model for Managing and Controlling Big Data Project Risks	Lai et al.	2018	backward
Top Ten Lists of Software Project Risks: Evidence from the Literature Survey	Arnuphaptrairong	2011	backward
Total Findings of Backward Search			16
Total Findings of Keyword and Backward Search			40

4. Results

As described in the previous section, a total of 248 risks emerging in data science projects have been identified through our literature review process. Since these risks can have different wordings depending on the literature source, the next step is to classify the risks into categories. As a result, we obtain 29 risk categories of varying frequency. These categories are summarized in Table 4.

As can be seen in the table above, "Insufficient project management" is the risk category comprising the most frequently mentioned risks. This category includes risks such as "poor task communication", "poor time management", "lack on focus on process and team coordination" or "plan cost overrun and schedule delay". In literature, these project management risks dominate over technical issues. Saltz et al. [40] criticizes for example that "data science projects need to focus on people, process and technology" and that in most cases immature processes risk are among others responsible for a project failure [42]. The risk category with the second most frequently mentioned risks is "Data security and privacy", which includes security concerns such as "cyber attacks" or "data privacy" concerns. The third risk category, named "Poor data availability, quality, and timeliness", includes risks such as "bad data quality", "broken data", "limited data access" or "timeliness of data". In literature, data quality issues are mentioned in a manifold way: in case of data cleansing, impact on model accuracy or high complex but faulty data sets [46]. Risks, which are categorized in "Lack of data science competence/skills", were also named quite often and reflect the challenge for companies to find skilled data scientists, data analysts or machine learning engineers on the labor market to handle successful data

Table 4
Risk categories

No.	Risk category	Frequency
1	Insufficient project management	87
2	Data security and privacy	26
3	Poor data availability, quality, and timeliness	20
4	High complexity	18
5	Lack of data science competence/ skills	15
6	Poor technical development/deployment practices	14
7	Insufficient data and information management	11
8	Model accuracy	9
9	Poor communication with customer/stakeholders	7
10	Organizational culture	6
11	Poor user management	5
12	Poor customer expectation management	4
13	Poor requirement management	4
14	Poor team management	3
15	Data inconsistency and incompleteness	2
16	New technology	2
17	Uncertainty about project outcome	2
18	Data ownership unclear	2
19	Poor domain knowledge	1
20	Insufficient documentation	1
21	Poor maintenance planning	1
22	Insufficient infrastructure	1
23	Poor data verification	1
24	Publication bias	1
25	Operational risks	1
26	Market risks	1
27	Political risks	1
28	Data dependency risks	1
29	No interaction with analytics-based program	1
	Total	248

science projects [12, 48]. Our findings are summarized in Figure 2, answering the first research question (RQ1).

With regard to the second research question (RQ2), software engineering is concerned with the development of software and thus, a software engineering project includes the design, implementation and testing of the software. In addition, the planning of a software system, the requirement analysis and the maintenance are added to the design process [8, 51]. To structure a software engineering project, process models such as the spiral model are used, which is a risk-driven procedure model for software development and follows the principle of the repeated run of its partial steps: description of the basic conditions with definition of the objectives, evaluation of the identified alternative solutions to mitigate or avoid any risks, development and reflection of an intermediate product and planning of the next iteration [51].Despite this close-meshed approach, not all risks can be reduced.
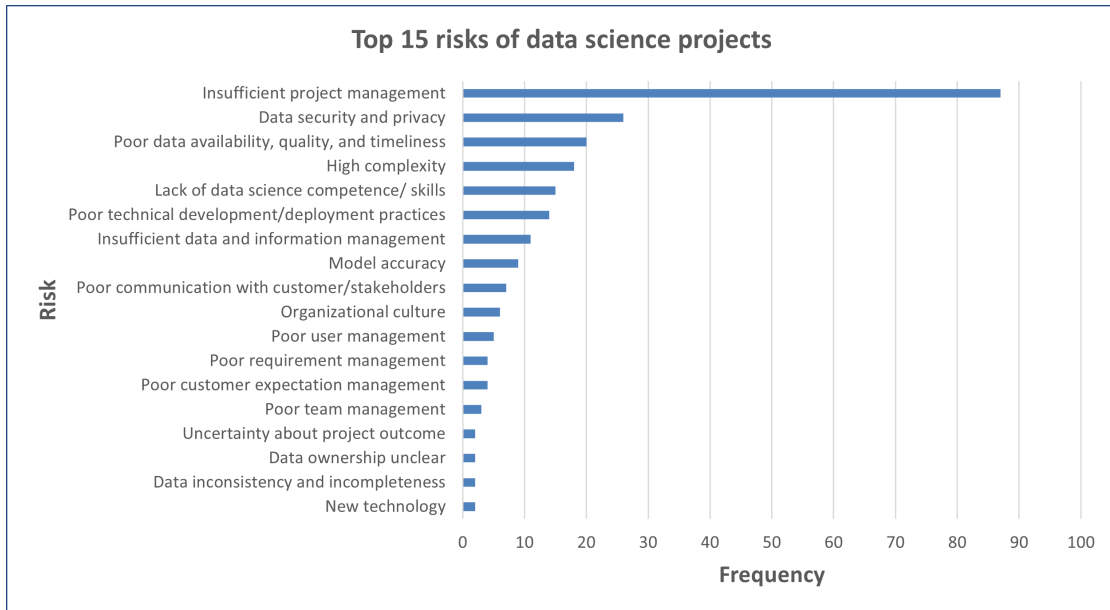
Top 15 risks of data science projects

Risk	Frequency
Insufficient project management	(≈87)
Data security and privacy	(≈26)
Poor data availability, quality, and timeliness	(≈20)
High complexity	(≈18)
Lack of data science competence/ skills	(≈15)
Poor technical development/deployment practices	(≈14)
Insufficient data and information management	(≈11)
Model accuracy	(≈9)
Poor communication with customer/stakeholders	(≈7)
Organizational culture	(≈6)
Poor user management	(≈5)
Poor requirement management	(≈4)
Poor customer expectation management	(≈4)
Poor team management	(≈3)
Uncertainty about project outcome	(≈2)
Data ownership unclear	(≈2)
Data inconsistency and incompleteness	(≈2)
New technology	(≈2)

Figure 2: Literature Review Process

The top ten risks in software engineering projects are insufficient requirement management, lack of management commitment and lack of project management methodology [1, 4, 13, 15, 16, 24, 44]. Ghazali et al. [16] conducted a literature review, which categorized the identified risks in "management risks", "people risks" and "technology risks" and conclude that management risks are the highest risks compared to people and technology risks. These management risks include, for example, "project milestones undefined", "requirements change", "lack of agile progress tracking mechanism" as well as "lack of resources", "failure to manage end-user expectation" or "lack of management commitment" [16]. With regard to data science project risks, the insufficient project management, which is comparable to Ghazali's management category, is also the highest risk to occur. We can conclude that the risk of insufficient project management techniques on software engineering projects as well as on data science projects is high and should not be underestimated.

Regarding Ghazali's category "people risks", risks such as "lack of necessary skill-set", "experience and training problem", "lack of team work" or "unmotivated team member" address the risk of an insufficient team management, which is also a common problem of data science projects [16]. In addition, the goal of a software engineering project is to develop a product that is useful to the end-user and if the end-user has difficulties using the final product, then it's a considerable risk. Therefore, frequent testing is vital for software engineering projects. Compared to data science projects, a successful customer expectation management is also crucial for the project outcome.

The last category of Ghazali's literature review, "technical risks", includes risks such as "lack of key technology", "inappropriateness of technology and tools" or "processor management insufficient", which appear less frequent in data science projects [16]. Compared to Ghazali's

"technical risks", the risk of "Poor data availability, quality, and timeliness" is one of the highest risks to occur in data science projects and can contribute significantly to the failure of the project.

5. Conclusion

This paper reports the results of a literature review on risks of data science projects. Through this literature analysis, relevant sources were collected and potential risks identified and categorized (RQ1). At the end, 354 sources were found, whereas in 40 papers 248 relevant risks were documented. After cleaning the results, 248 risks were categorized by main term and the most frequently mentioned risk categories were presented. The results show the need for a more detailed risk assessment to assist the project manager during the project duration. Risks, which are summarized by the term "insufficient project management", can be addressed through a frequent risk sensitivity analysis to highlight, for example, upcoming time schedule challenges right at the beginning to avoid project failure. Risks, such as "data security and privacy" risks, as well as "poor data availability, quality, and timeliness", which effect the project outcome should be aware as early as possible in the project. Therefore, a risk assessment, which evaluates the identified risks of a data science project at the beginning and during the project period is vital.

Furthermore, the similarities and differences between the risks of a software engineering project and those of a data science project (RQ2) have been described. The comparison between the risks of a software engineering project and those of a data science project show that there are similarities between both disciplines, especially regarding the risks of a insufficient project and team management. Regarding the technical risks, data science projects have a particularly high risk of failure if, for example, the data is not available or of poor quality, while software engineering projects fail less frequent due to technical risks.

The limitations of this literature analysis lie on the one hand in the definition of the search queries and on the other hand in the naming and assignment of risks to the individual categories, since these are always shaped by subjective criteria such as personal level of knowledge and experience.

Regarding the risk assessment, the first step of risk identification was successfully performed in this paper. As an outlook to future work, these categorized risks form the basis for the development of a method for automated risk assessment of data science projects. Among others, discrete multi-attribute decision-making (MADM) methods and conitnuous multi-objective decision-making (MODM) methods [10], are considered for this purpose.

6. Acknowledgement

References

[1] Addison, T., Vallabh, S. (2002, September). Controlling software project risks: an empirical study of methods used by experienced project managers. In Proceedings of the 2002 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology (pp. 128-140).

[2] Alharthi, A., Krotov, V., Bowman, M. (2017). Addressing barriers to big data. Business Horizons, Volume 60, Issue 3, 2017, Pages 285-292.

[3] Angée, S., Lozano, S., Montoya-Munera, E., Ospina Arango, J., Tabares, M. (2018). Towards an Improved ASUM-DM Process Methodology for Cross-Disciplinary Multi-organization Big Data and Analytics Projects: 13th International Conference, KMO 2018, Žilina, Slovakia, August 6–10, 2018, Proceedings.

[4] Arnuphaptrairong, T. (2011). Top Ten Lists of Software Project Risks: Evidence from the Literature Survey. In: Proceedings of the International MultiConference of Engineers and Computer Scientists 2011 Vol I, IMECS 2011, March 16 - 18, 2011, Hong Kong.

[5] Asay, M. (2017). 3 ways to massively fail with machine learning (and one key to success)", https://www.techrepublic.com/article/3-ways-to-massively-fail-with-machine-learning-and-one-key-to-success/.

[6] Aßmann, J., Sauer, J., Schulz, M. (2023). Don't Be Afraid of Failure—Insights from a Survey on the Failure of Data Science Projects. In Apply Data Science: Introduction, Applications and Projects (pp. 65-76). Wiesbaden: Springer Fachmedien Wiesbaden.

[7] Aust, H. (2021). The Age of Data: What You Need to Know About Fundamentals, Algorithms, and Applications / Das Zeitalter der Daten: Was Sie über Grundlagen, Algorithmen und Anwendungen wissen sollten. Springer, Berlin.

[8] Balzert, H. (2000). Lehrbuch der Software-Technik: Software-Entwicklung, 2. Auf., Heidelberg, Spektrum Akademischer Verlag, 2000.

[9] Cao, L. (2017). Data science: challenges and directions. Communications of the ACMVolume 60Issue 8August 2017pp 59–68https://doi.org/10.1145/3015456

[10] Djenadic, S.,Tanasijevic, M., Jovancic, P., Ignjatovic, D., Petrovic, D., Bugaric, U. (2022). Risk Evaluation: Brief Review and Innovation Model Based on Fuzzy Logic and MCDM. Mathematics 2022, 10, 811. https://doi.org/10.3390/math10050811

[11] Dukino, C., Kutzias, D., Link, M. (2022). Roles and competences of data science projects. The Human Side of Service Engineering, Vol. 62., AHFE International, pp. 250-255.

[12] Eberhard, B., Podio, M., Pérez Alonso, A., Radovica, E., Avotina, L., Peiseniece, L., Sendon, M.C., Gonzales Lonzano, A., Solé-Pla, J. (2017). Smart work: The transformation of the labour market due to the fourth industrial revolution. International Journal of Business and Economic Sciences Applied Research, Vol. 10, Issue 3.

[13] Elzamly, A., Hussin, B. (2015). Modelling and evaluating software project risks with quantitative analysis techniques in planning software development. Journal of computing and information technology, 23(2), 123-139.

[14] Gartner (2015). Gartner says business intelligence and analytics leaders must focus on mindsets and culture to kick start advanced analytics. Gartner web site, https://www.gartner.com/en/newsroom/press-releases/2015-09-15-gartner-says-business-intelligence-and-analytics-leaders-must-focus-on-mindsets-and-culture-to-kick-start-advanced-analytics.

Last visted on 19/07/2023.

[15] Georgiev, V., Stefanova, K. (2014). Software development methodologies for reducing project risks. Economic Alternatives, 2, 104-113.

[16] Ghazali, N., Fauzi, S., Gining, R., Sobri, W., Suali, A. (2020). Visualizing Software Risks in Software Engineering Projects using Risk Sensitivity Analysis Approach. Journal of Physics: Conf. Ser. 1529 022074.

[17] Grassi, A., Gamberini, R., Mora, C., Rimini, B (2009). A fuzzy multi-attribute model for risk evaluation in workplaces. Safety Science, Vol 47, Issue 5, 707–716.

[18] Haertel, C., Pohl, M., Nahhas, A., Staegemann, D., Turowski, K. (2022). Toward a Lifecycle for Data Science: Literature Review of Data Science Process Models. PACIS 2022 PROCEEDINGS.

[19] Survey: What IT Teams Want Their CIOs to Know About Enterprise Big Data. https://www.prnewswire.com/news-releases/survey-what-it-teams-want-their-cios-to-know-about-enterprise-big-data-188190311.html, Last visited on 19/07/2023.

[20] ISO Guide 73:2009 (2009). Risk Management—Vocabulary. International Standards Organisation: Geneva, Switzerland.

[21] ISO 31000:2018 (2018). Risk Management—Guidelines. International Standards Organisation: Geneva, Switzerland.

[22] ISO/IEC 31010:2019 (2019). Risk Management—Risk Assessment Techniques. The International Organization for Standardization and The International Electrotechnical Commission: Geneva, Switzerland.

[23] Keil, M., Cule, P. E., Lyytinen, K., Schmidt, R. C. (1998). A framework for identifying software project risks. Communications of the ACM, 41(11), 76-83.

[24] Khanfar, K., Elzamly, A., Al-Ahmad, W., El-Qawasmeh, E., Khalid, A., Abuleil, S. (2008). Managing Software Project Risks with the Chi-Square () Technique. International Management Review, 4(2), 18-29.

[25] Kraut, N., Transchel, F. (2022). On the Application of SCRUM in Data Science Projects. 7th International Conference on Big Data Analytics (ICBDA).

[26] Kutzias, D., Dukino, C., Kett, H. (2021). Towards a Continuous Process Model for Data Science Projects. In C. Leitner, W. Ganz, D. Satterfield, C. Bassano (Eds.), Lecture Notes in Networks and Systems. Advances in the Human Side of Service Engineering, Vol. 266, pp. 204–210. Springer International Publishing.

[27] Lahiri, S., Saltz, J. (2022). The Risk Management Process for Data Science: Gaps in Current Practices. Proceedings of the 55th Hawaii International Conference on System Sciences.

[28] Lai, ST., Leu, FY. (2018). A Critical Quality Measurement Model for Managing and Controlling Big Data Project Risks. In: Barolli, L., Xhafa, F., Conesa, J. (eds) Advances on Broad-Band Wireless Computing, Communication and Applications. BWCCA 2017. Lecture Notes on Data Engineering and Communications Technologies, vol 12. Springer, Cham.

[29] Limesha, G. (2021). Critical success factors for managing data science projects within agile methodology (Doctoral dissertation).

[30] Marquardt, K. (2017): Smart services – characteristics, challenges, opportunities and business models. Proceedings of the International Conference on Business Excellence, Vol. 11, No. 1, pp. 789–801

[31] Mizuno, O., Kikuno, T. (2000). Characterization of Risky Projects based on Project Man-

agers' Evaluation. ICSE '00: Proceedings of the 22nd international conference on Software engineering. pp. 387-395.

[32] Muhlbauer, W. K. (2004). Pipeline risk management manual: ideas, techniques, and resources. Elsevier.

[33] Pasaribu, R., Taufik, T. A. (2021). Risk Management Implementation at XYZ Project Using Failure Mode Effect Analysis and Hybrid Multi Criteria Decision Making. 2nd International Conference on Management of Technology, Innovation, and Project, 2020.

[34] Pilliang, M., Munawar, M. (2022). Risk Management in Software Development Projects: A Systematic Literature Review. Khazanah Informatika Jurnal Ilmu Komputer dan Informatika. 8. 3. 10.23917/khif.v8i2.17488.

[35] Ransbotham, S., Kiron, D. (2017). Analytics as a Source of Business Innovation. MIT Sloan Management Review.

[36] Reggio, G., Astesiano, E. (2020, August). Big-data/analytics projects failure: a literature review. In 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA) (pp. 246-255). IEEE.

[37] Rekha, J. H., Parvathi, R. (2015). Survey on Software Project Risks and Big Data Analytics. 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15). Procedia Computer Science 50, pp. 295-300.

[38] Robinson, E., Nolis, J. (2020). Build a Career in Data Science, New York: Manning Publications Co.ISBN 9781617296246.

[39] Saltz, J. S., Lahiri, S. (2020). The Need for an Enterprise Risk Management Framework for Big Data Science Projects. In: DATA (pp. 268-274).

[40] Saltz, J., Shamshurin, I., Crowston, K. (2017). Comparing Data Science Project Management Methodologies via a Controlled Experiment. 10.24251/HICSS.2017.120.

[41] Saltz, J., Shamshurin, I. (2019). Achieving Agile Big Data Science: The Evolution of a team's Agile Process Methodology. 2019 IEEE International Conference on Big Data, pp. 3477-3485.

[42] Saltz, J., Suthrland, A. (2019). SKI: An Agile Framework for Data Science. 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 3468-3476, doi: 10.1109/BigData47090.2019.9005591.

[43] Saltz, J., Wild, D., Hotz, N., Stirling, K. (2018). Exploring Project Management Methodologies Used Within Data Science Teams. Twentyfourth Americas Conference on Information Systems, New Orleans, 2018, pp. 1-5.

[44] Schmidt, R., Lyytinen, K., Keil, M., Cule, P. (2001) Identifying Software Project Risks: An International Delphi Study, Journal of Management Information Systems, 17:4, 5-36, DOI: 10.1080/07421222.2001.11045662

[45] Tanga O, Akinradewo O, Aigbavboa C, Oke A, Adekunle S. (2022). Data Management Risks: A Bane of Construction Project Performance. Sustainability. 2022; 14(19):12793. https://doi.org/10.3390/su141912793

[46] Turkay, C., Pezzotti, N., Binnig, C., Strobelt, H., Hammer, B., Keim, D., Fekete, J.-D., Palpanas, T. Wang, Y., Rusu, F. (2018). Progressive Data Science: Potential and Challenges.

[47] Varela, C., Domingues, L. (2022). Risks of Data Science Projects-A Delphi Study. Procedia Computer Science, 196, 982-989.

[48] Verma, A., Yurov, K.M., Lane, P. L., Yurova, Y. V. (2019). An investigation of skill requirements for business and data analytics positions: A content analysis of job advertisements.

Journal of Education For Business, Vol. 94, Issue 4, pp. 1-8.

[49] Brocke, J. V., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: ECIS Proceedings 161.

[50] Wallace, L., Keil, M. (2004). Software project risks and their effect on outcomes. Communications of the ACM, 47(4), 68-73.

[51] Weber, P., Gabriel, R., Lux, T., Menke, K. (2022). Basiswissen Wirtschaftsinformatik. Wiesbaden, 4. Aufl., Springer Vieweg, 2022.

[52] Webster, J. and Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly.

[53] Weiner, J. (2021). Why AI/Data Science Projects Fail. Morgan and Claypool Publishers, San Rafael, California.