Early Indicators of Project Abandonment in Industry-Academia Collaborations: Developing an Assessment Framework for Industrial Data Science Projects

Anna Maria Binder de Serdio[1,2(✉)] [0000-0003-3466-3745], Dirk Stegelmeyer[2] [0000-0001-7918-3732] and Fatima Sajid Butt[2,3] [0000-0002-9111-7305]

[1] School of Computing & Engineering, University of Huddersfield, UK
apprise@fra-uas.de
[2] Institut für interdisziplinäre Technik, Frankfurt University of Applied Sciences, Germany
[3] Escuela Superior de Ingeniería, Universidad de Cádiz, Spain

Abstract. This paper presents a novel assessment framework aimed at identifying early indicators of potential project abandonment in industry-academia collaborations (IAC) focusing on industrial data science projects. As these collaborations become more integral to leveraging academic research for industrial applications, understanding the early stages of project development is crucial for minimizing resource wastage and ensuring project continuation. Our research involved analyzing 13 industrial data science case studies, with a particular focus on those that were discontinued in their initial phases. By synthesizing findings from these case studies with literature on success factors of IAC, we propose an assessment framework that includes a set of success factors tailored to avoid project abandonment. This assessment framework is grounded in a methodological approach that integrates established project management success factors and direct industry experience, ensuring practical relevance and applicability. The developed assessment framework consists of 14 success factors, each described with specific industry-relevant aspects. This paper not only contributes to academic literature by offering a structured method to anticipate early project abandonment but also serves as a practical guide for researchers in acquiring and executing data science projects within the industry-academia context.

Keywords: Industry-Academia Collaboration, Industrial Data Science, Innovation ecosystem.

1 Introduction

IAC are increasingly intended by governments, industry, and academic institutions to facilitate the translation of scientific research into industrial applications [1]. These partnerships are particularly relevant with the advent of Industry 4.0 and makes the projects in areas such as industrial data science viable and important. However, the way in which these data science projects are to be carried out remains a major challenge.

Although process models are available as guidance, many of these projects are still abandoned in the initial phases.

This paper focuses on the early phases of data science projects that involve in a later stage the (pre-) processing of industrial data and the application of machine learning techniques. Data science projects typically involve collaboration between university researchers, industrial partners, and potential end-users of industrial equipment, for example the one presented in [2]. Our study incorporates the experience from 13 industrial data science projects carried out in the research group of the authors.

The aim of this paper is to develop an assessment framework to reflect on the initial phases of IAC projects in industrial data science from a researchers' perspective. This evaluation framework aims to minimize the waste of resources by identifying early indicators that projects are likely to be abandoned. The research question we seek to answer is: What is a possible assessment framework for anticipating the abandonment of data science projects with industry partners during the early phases? To address this, we utilize a methodological framework inspired by [3], enhanced by aspects of [1] and the inclusion of the 13 cases as practical examples to ensure a comprehensive and well-founded analysis of the research subject. The structure of this paper is organized as follows: Section 2 defines the early phases of industrial data science projects. Section 3 explains the methodology we use. Section 4 outlines the development of our assessment framework. Section 5 discusses the results. Finally, Section 6 makes suggestions for further research.

2 The Early Stage in Industrial Data Science Projects

One frequently used methodology in data science projects is the Data Mining Methodology for Engineering Application (DMME), a specialized extension of the CRoss-Industry Standard Process for Data Mining (CRISP-DM), tailored specifically for engineering contexts proposed in [4] and [5] respectively.

Since this paper focuses on the early phases of data science projects, this section introduces the first phases of the DMME process model that can be seen in Fig. 1.

The first phase is business understanding. This phase is dedicated to understanding the project objectives and requirements from a business perspective. Furthermore, this knowledge is converted into a data science problem. A resulting preliminary plan, which describes the objectives to be achieved, is created [5].

The second phase is about technical understanding and conceptualization. The goal of this phase is to turn business objectives into clear technical goals. Existing knowledge about relevant physical and process interactions are collected. Additionally, a data science strategy and a technical plan for data collection must be designed. Finally, a testing and experimental plan to evaluate the findings is developed [4].

We limit our analysis to the first two phases of DMME, that is until data is exchanged the first time.

The DMME helps inexperienced engineers conduct data science projects. It breaks down the data science workflow into smaller, manageable steps. The outcomes of these steps can then be assessed and result in applicable findings [4].

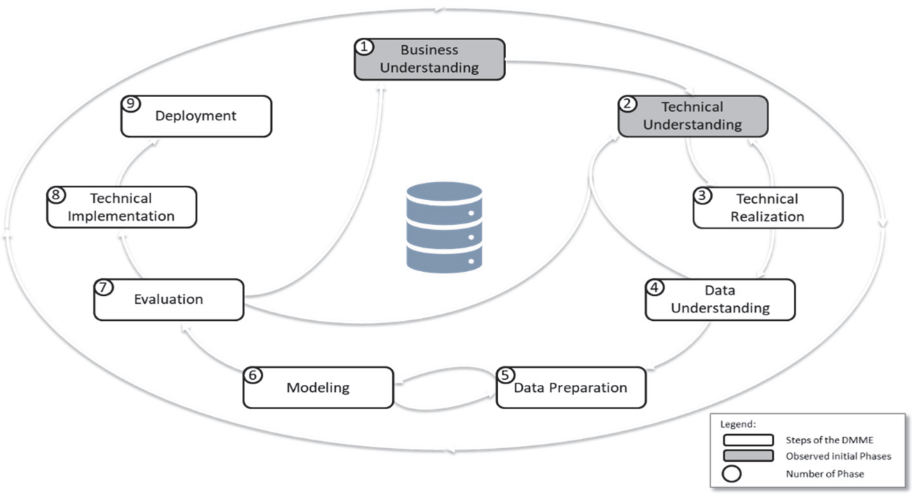

Fig. 1. Phases of the DMME [6]

Our study examines 13 industrial data science cases from which eight were halted or abandoned in their initial phases (see Table 1).

Table 1. Industrial Case Studies.

	Company	Result	Success factor Met	Success factor Unmet
A	Coating machines	success [7]	4, 5, 8, 14, 16, 17, 19, 21, 26	15
B	Dying machines	abandoned in phase 3	3, 4	1, 5, 10, 16, 19
C	Forklifts	abandoned in phase 1	4, 26	6, 8, 16, 19
D	Spindles for machine tools	abandoned in phase 1		1, 5, 16
E	Automated agricultural machines	abandoned in phase 1	14	1, 3, 4, 6, 10, 16
F	Coating machines	success [2]	1,3, 6, 14, 15, 16, 17, 19, 21	4, 5
G	Beverage dosage systems	abandoned in phase 2	4, 5, 15	1, 6, 14, 16, 26
H	Industrial shredders	abandoned in phase 1		1, 4, 5, 16, 26
I	Control valve	halted in phase 1	16	4, 5, 15
J	Chemical Infrastructure provider	halted in phase 1	16	3, 4, 5
K	Machine Tools	abandoned in phase 1		5, 16, 21
L	IT consultancy	started phase 1	1, 21, 26	
M	Railway maintenance	started phase 1	1, 8, 17, 21, 26	

Since the full strength of the DMME lies in the reusability of previous projects based on the lessons learned, we want to use our experience from the initial phases of our industry projects and contribute to the literature on data science project success [4, 7].

With this short introduction to the initial phases of the DMME, the research question should be specified.

3 Methodology

In the context of this study, we have used the framework proposed by [3], which is based on literature research, as the methodological basis. This framework offers 25 success factors for projects in the construction industry [3].

Furthermore, the authoring team conducted a brainstorming session at the beginning of the research process to formulate 12 aspects that could anticipate the abandonment of data science projects with industrial partners in the early phases. Therefore, our experience with several industrial projects was utilized.

Then, we used aspects resulting from a literature review presented by [1] on works on IAC. This review analyzed nine papers to derive a list of more than 30 aspects for successful IAC. The more than 30 aspects of [1] and the 12 aspects of the initial brainstorming were then related to the 25 success factors of [3]. Finally, we looked at which success factors were observable in the respective industrial cases and whether these were met or not.

4 Development of Assessment Framework

This paper's research question is: What is a possible assessment framework for anticipating the abandonment of data science projects with industry partners during the early phases? In order to develop such an assessment framework, we only take externally observable variables in the industrial context since it is not possible to look into the internal processes of our industry partners in the early phases.

As a start, seven aspects of Kettunen et. al [1] were eliminated, since they were recommendations for academia and thus did not serve as criteria for assessing industry.

The aspects mentioned by [1] and resulting from the brainstorming session were related to the success factors of [3]. The result is that for 12 of the 25 success factors, no aspects could be found because neither the academic literature nor the industrial case study reports consider them to be relevant. Therefore, we delete them as success factors for IAC.

Some aspects were observed by literature and brainstorming but could not be related to a success factor of the framework. Therefore, the factor "subject matter co-creation mindset and attitude" (26) was added to the framework.

Table 2. shows the resulting assessment framework consisting of 14 success factors. Five of the success factors mentioned in [3] were reformulated to describe better the intended meaning in the IAC context. With the development of the assessment framework the research aim is achieved. The result deserves to be discussed in the following section.

Table 2. Success factors of the assessment framework.

Nr.	Success factor	Explanation; Example
1	Effective communication between stake-holders	Transparent information, collaboration at different hierarchical levels
3	Company's technical domain know how	Domain know-how; data collection capacity
4	Scope and work definition	Formulated RQ that should be answered with the data science project
5	Clarity of the project mission	Balance between the expectations of practitioners and researcher's
6	Planning efforts, resources provision including domain experts	Availability of capacity of domain experts during the project execution
8	Effective scheduling	Fast pace and rhythm of joint interaction occasions
10	Adequacy of plans and specifications	Reasonable perception of the time span of the data science project
14	Team's commitment to the project	Qualification and dedication, attitude towards co-creation, project champion
15	Relation to use cases and users	Relationship and access to data sources or other cooperating companies
16	Top management support	Regular meetings, active management
17	Project manager capabilities and commitment	Willingness to appoint a project manager to actively drive the project forward from industry side
19	Team motivation (rewards and incentives)	Clarification and consideration of the different governing variables and objectives of academia and industry
21	Staff qualification and dedication, attitude towards co-creation	Cooperation rather than competing with each other
26	Mindset and attitude towards subject matter co-creation	Understanding industry/ academia context; data dictionary provided by industry Awareness that academia is not a software development company

5 Results and Discussion

We used the work of [3] as a starting framework. The context is projects in civil construction. Since industrial data science projects differ, success factors such as 11 (effective procurement and tendering methods), 18 (effective Site Management), 20 (effective technical review), 22 (completion of design at the construction start), 23 (political conflicts and corruption) and 24 (harsh climate conditions and environment) did not have any related aspects to our industrial context and thus are not applicable.

Researchers must assess the likelihood of project success before they invest resources, especially time. Success factor 2 (company's financial strength), 7 (adequate risk analysis), 12 (adequate project management techniques), 13 (control system) and 25 (unforeseen conditions) did not have aspects related to our brainstorming session or the aspects mentioned by [1]. Those aspects are very difficult to assess by researchers before they invest resources and were therefore eliminated in our assessment framework.

Success factor 1 (effective communication between stakeholders) entails aspects such as transparent communication, collaboration at different levels between companies and universities with frequent opportunities to meet, communication about company strategy and technology [1]. The authors of this paper experienced difficulty in accessing the company's governing strategy regarding the industrial problem at hand despite frequent discussions in cases B and H.

Success factor 3 (company's technical domain know how) includes the domain knowledge. In case J for example, the major chemical infrastructure provider did not have sufficient knowledge about the solid fuel plants process to turn business objectives into clear technical goals.

Aspects of success factor 4 (scope and work definition) include lean research approaches such as business cases (defined by industrial organizations with business impact), identifying the "right" research problem, ensuring practicality and applicability, conducting costbenefit analysis, industrial challenge vs. actual problem [1]. In cases E, F, H, I, and J for example it was not clear to the industrial partner what the business case is, coined by the question: what do you do next after the research aim is achieved?

An example for a success factor 5 (clarity of the project mission) is a match between practitioners and researcher's expectations [1]. In case B for example the industrial partner expected the researchers to develop a working commercial software, which a university neither can do nor wants to do.

Success factor 6 (planning efforts, resources provision including domain experts) includes a realistic plan. For example, expect funding organizations linear up-front research proposals and plans (waterfallish) [1].

Success factor 8 (effective scheduling) aspects are for example close collaboration realized with applied agile methodologies (scrum) 6-month sprints [1].

Success factor 10 (adequacy of plans and specifications) refers to the industrial partner having a realistic timeline. In case E, a startup company, the owners expected deliverable results by an unrealistic date, because the next financing stage was due.

Success factor 14 (team's commitment to the project) can be expressed by aspects such as artifact and asset sharing, a champion as the main driver of the collaboration on the industry-side, practitioners learning to appreciate research rigor requires time and continuous reflection efforts, industrial companies having limited resources (especially time) for academic research prepared to do "extra work", attitude towards co-creation. [1] In case G, for example, the industry partners team, including the management director, was positive toward a rewarding project, but in the company of 50 no champion was prepared to drive the project and it died slowly.

Success factor 15 (relation to use cases and users) is only applicable, if the industrial partner is not the user and thus has not access to the data. For example, in case F, the coating machine manufacturer had good relation to the customer, the optical lens manufacturer using the coating machines. In case G a beverage dosage system manufacturer asked its trusting customers and a sparkling wine producer teamed up. In the case I, the control valve manufacturer did not find a customer willing to share the data the valves offer. This is because industries like the chemical companies fear leak of information.

Success factor 16 (top management support) materialized in case F with a quarterly steering committee meeting. In contrast, in case G the researchers never made it beyond speaking to the R&D employee in charge of machine learning.

The mere nomination of a project manager in the early phase is an example of success factor 17 (project manager capabilities and commitment). We experienced in some cases, that industry partners expected researchers to carry out the project management.

Success factor 19 (teams' motivation, rewards and incentives) can be drilled down to the benefit for the industry partners from a project. Is should be noted that academia and industry have by nature different governing variables and goals [1]. For example, we learned that it is advisable to name the goal of researchers, that is to publish, in the first meeting. In case B it turned out, that the only goal of the R&D employee in charge of the project was personal training in machine learning.

Success factor 21 (staff qualification and dedication, attitude towards co-creation) can be expressed by aspects such as interest in learning from the researchers, work jointly rather than even competing with each other [1]. Generally, we felt in our projects that this success factor was met in many cases. In one case, however, we felt that the industries data science department perceived the researchers as competition, challenging their knowledge in face of the organization. In one case the industrial partners presented a draft of a purchase agreement of research services at first meeting with the price open for discussion and it took various meetings to change the attitude from supplier-customer to partnership.

Success factor 26 (mindset and attitude towards subject matter co-creation) implies that the industry partner understands the subject matter and the content that each partner is responsible for. This includes to be conscious about knowledge exchange vs. technology transfer, the mindset and attitude towards co-creation, importance of industry partners role to document precisely the facts to enable machine learning.

An overview on the (un-)met success factors is given in Table 1. Some success factors could not be assessed with reasonable likelihood before the project failed, therefore not all 14 success factors can be found for each case. Cases with many success factors (A, F, M) seem to be successful which could be seen as a positive evaluation of our proposed assessment framework. A detailed statistical cluster analysis is planned; however, some preliminary conclusions can be drawn. Missing top management support (16) is noteworthy in all abandoned cases. Clarity of the project mission (5) was present in the two successful cases and a problem in many failed cases. Mindset and attitude towards subject matter co-creation (26) seems to be an important success factor, but maybe difficult to assess in the early stage or a negative mindset even be changed to a positive attitude.

Summarizing the findings, we contributed a novel assessment framework for researchers to use in the early phases of IAC projects in industrial data science.

6 Future Directions and Research Opportunities

The novel assessment framework needs to be validated. For example, for new IAC projects we will assess the 14 success factors in regular time intervals and relate those data points to later project success or abandonment.

We have taken the researchers' perspective while compiling this study. Possibly it is of value for industry to look during the early phases from the industry's point of view to facilitate a successful IAC with academia.

To operationalize the findings of this paper, for each of the 14 success factors of the assessment framework measurements should be identified. For example, 14 (team's commitment to the project) could be measured with the time needed to schedule meetings and with yes/no if a project champion is identifiable. 16 (top management support) could be measured by the hierarchy of the highest involved person from industry.

To further put in practice the assessment framework it could be converted in a questionnaire or checklist for easy use during meetings.

References

1. Kettunen, P., Järvinen, J., Mikkonen, T., Männistö, T.: Energizing collaborative industry-academia learning: a present case and future visions. European Journal of Futures Research, vol. 10 (2022). doi: 10.1186/s40309-022-00196-5
2. Butt, F.S., Schäfer, J., Wagner, M.F., Stegelmeyer, D., Gomez-Ullate Oteiza, D.: Application of CRISP-DM and DMME to a Case Study of Condition Monitoring of Lens Coating Machines IEEE International Workshop on Metrology for Industry 4.0 & IoT, 409–414 (2023). doi: 10.1109/MetroInd4.0IoT57462.2023.10180178
3. GUNDUZ, M., YAHYA, A.M.A.: Analysis of project success factors in construction industry. Technological and Economic Development of Economy, vol. 24, 67–80 (2015). doi: 10.3846/20294913.2015.1074129
4. Wiemer, H., Drowatzky, L., Ihlenfeldt, S.: Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model. Applied Sciences, vol. 9 (2019). doi: 10.3390/app9122407
5. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0. Step-by-step data mining guide, vol. (2000)
6. Huber, S., Wiemer, H., Schneider, D., Ihlenfeldt, S.: DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. Procedia CIRP, vol. 79, 403–408 (2019). doi: 10.1016/j.procir.2019.02.106
7. Husenbeth, A., Berens, G., Stegelmeyer, D.: Die Anwendung des maschinellen Lernens – Fokus Interaktionsprozess in Predictive Maintenance Projekten. In: Henke, M. (ed.) 18. InstandhaltungsForum, pp. 78–88 (2019). doi: 10.5281/zenodo.11079251