Data Science in Government Agencies: The Challenge of Deployment and Operation

Sandra de Kruif, Guido Ongena & Marlies van Steenbergen
HU University of Applied Sciences Utrecht, Netherlands.
E-mail: sandradekruif@hotmail.nl, guido.ongena@hu.nl, marlies.vansteenbergen@hu.nl

Abstract Despite the numerous business benefits of data science, the number of data science models in production is limited. Data science model deployment presents many challenges and many organisations have little model deployment knowledge. This research studied five model deployments in a Dutch government organisation. The study revealed that as a result of model deployment a data science subprocess is added into the target business process, the model itself can be adapted, model maintenance is incorporated in the model development process and a feedback loop is established between the target business process and the model development process. These model deployment effects and the related deployment challenges are different in strategic and operational target business processes. Based on these findings, guidelines are formulated which can form a basis for future principles how to successfully deploy data science models. Organisations can use these guidelines as suggestions to solve their own model deployment challenges.

Keywords: data Science model, data science deployment, informatio value chain, government case study.

University of Maribor Press

DOI https://doi.org/10.18690/um.fov.4.2022.29
ISBN 978-961-286-616-7

1 Introducton

Expectations of the benefits of data science models are high. For instance, data science models can enable business model changes (e.g., Chen, Schütz, Kazman, & Matthes, 2017). Propelled by increased accessible infrastructure and computing power, and the acquisition of more volumes of data accumulating into big data, big data analytics is thought to be one of the most valuable strategic business sources in the coming years (McAfee & Brynjolfsson, 2012).

Literature reports on exemplary successful case studies of Facebook (Thusoo, et al., 2010), LinkedIn (Sumbaly, Kreps, & Shah, 2013), and Twitter (Lin & Kolcz, 2012). Although the private sector has been ahead in development of data science models (Ransbotham, Gerbert, Reeves, Kiron, & Spira, 2018), scholars also state the value of these models to promote the public good by governmental agencies (Kim, Trimi, & Chung, 2014; Desouza, Dawson, & Chenok, 2020). Nevertheless, there are challenges that must be overcome to bring data science models to fruition.

Data science model deployment is a major challenge on which little research has been done. A recent survey revealed that only 23% of the respondents had at least one data science project in production (Castellanos, Pérez, Varela, Villamil, & Correal, 2019). This deployment failure is often caused by a lack of knowledge on how to deploy data science models and immature deployment procedures (Brethenoux, Vashisth, & Hare, 2018). To fill this knowledge gap, this study sets out to explore how governmental agencies can overcome data science model deployment challenges that result from the interdependencies between model deployment and its context, i.e., the target business process, the model itself and the model development process. To do so, we examine five model deployment cases in a Dutch government organization. The aim of this research is to formulate guidelines that can form a basis for future principles how to successfully deploy data science models. Organizations can use these guidelines as tools or suggestions to solve their own model deployment challenges.

The remainder of this article is structured as follows: section 2 provides a short context description and a summary of previous research on data science model deployment. Section 3 describes the research method and is followed by a summary of the main research findings in section 4. Section 5 discusses the findings and identifies similarities and differences to the extant literature. Section 6 provides conclusions and describes research limitations.

2 Related work

Data science is conducted in the context of the information value chain. We draw upon the information value chain by Abbasi, Sarker and Chiang (2016), which is a reference model for the collection, validation and storage of data and subsequent creation of information and knowledge. The created knowledge is then used to support decision making in order to take actions. Figure 1 shows where data science model development, model deployment and model use are located within the information value chain.

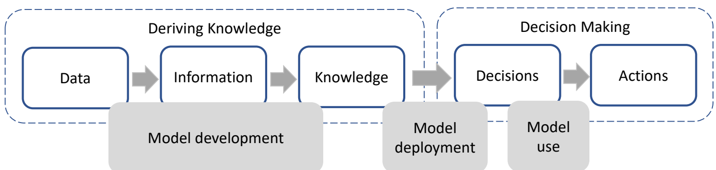

Figure 1: Data science within the information value chain

2.1 Data science model development

Model development takes place in the first phase of the information value chain, deriving knowledge. First many kinds of data are collected, processed and stored, using various combinations of data processing technologies and practices (Chen, Chiang, & Storey, 2012). Then information and knowledge are extracted from the stored data by developing data science models (Provost & Fawcett, 2013; Watson, 2017). The model development process is based on data mining process descriptions (Fayyad, Piatetsky-Shapiro, & Smyth, 1996; Wirth & Hipp, 2000). Data scientists can use all stored data types to develop their statistical, machine learning and other models (Provost & Fawcett, 2013), to produce trends, forecasts, predictions, simulations and other outputs (Watson, 2017). In some cases, model development is done directly on the data stores (Lin & Kolcz, 2012), but in other cases, data scientists work with copies of the stored data (Sumbaly, Kreps, & Shah, 2013; Thusoo, et al., 2010), which are known as analytical sandboxes (Watson, 2017).

2.2 Data science model deployment

Model deployment is situated at the transition between the deriving knowledge phase of the information value chain and the decision-making phase. Model deployment can change business processes. Davenport and Ronanki (2018) state that data science models support individual tasks within target business processes, rather than complete business processes. Such tasks can be automated by incorporating a repeatable data mining process in the target business process (Wegener & Rüping, 2010; Wirth & Hipp, 2000). In addition, the necessary changes in the supporting business applications must be designed (Rupnik & Jaklič, 2009). The production IT infrastructure uses other hardware, standards, programming languages and application frameworks than the analytical sandboxes. The required security levels and processing speeds also differ (John, Olsson, & Bosch, 2020; Jackson, Yaqub, & Li, 2019).

Conversely, during deployment, data science models may need to be adapted or even rebuilt (Sumbaly, Kreps, & Shah, 2013; John, Olsson, & Bosch, 2020). They may be adapted to run without disturbing the surrounding operational applications (Baylor, et al., 2017). During use, machine learning models need sufficient processing speed to immediately provide their results to receiving applications. This could necessitate choosing a faster model with lower output accuracy (John, Olsson, & Bosch, 2020). Additionally, in an automated business process, the data science model must be included in a series of business applications, which necessitates adding application interfaces (Cetinsoy, Martin, Ortega, & Petersen, 2016). Furthermore, the model may have to integrate with reports, or it may have to communicate with users. Thus, it may need a user interface or web interface (John, Olsson, & Bosch, 2020).

Data science models are deployed by software engineers or by users (e.g., Crankshaw, et al., 2017). However, model deployment should be the responsibility of data scientists (Davenport & Malone, 2021). Thus data scientists should start preparing for model deployment during the development phase, by investigating the production IT infrastructure and the necessary changes in the target business process and the business application (Davenport & Malone, 2021; Davenport & Ronanki, 2018). Jackson et al. (2019) and Karamitsos et al. (2020) describe combinations of the data mining process and agile software engineering, to better facilitate model deployment. The data science model deployment process framework described by John et al. (2020) can fit into this 'hybrid agile data science' process defined by Jackson et al (2019). Other scholars describe using a combination of machine learning application development and systems engineering to solve deployment problems (Martínez-Plumeda, Gómeza, & Hernández-Orallo, 2021).

2.3 Data science model use

Once a machine learning model has been deployed and it is in use within a business process, changes in the input data and in the process environment can influence the working of the model. Therefore, deployed machine learning models need to be monitored continuously (John, Olsson, & Bosch, 2020), by business users (Wegener & Rüping, 2010). However, Davenport and Malone (2021) state that data scientists should do the monitoring. When the working of the machine learning model changes, data scientists need to retrain or redesign the model (John, Olsson, & Bosch, 2020).

3 Research method

From the literature can be concluded that the deployment of data science models is complex (section 2) and not well understood so far (section 1). We thus employed a multiple-case study (Dubé & Paré, 2003). A multiple-case study allows for the exploration of the differences and commonalities across cases to predict similar results across cases (Yin, 2018). In this research, five real-world deployments of data science models are described and compared.

3.1 Selected cases

To promote *external validity,* that focusses on the generalizability of the results (Yin, 2018), we replicated the case studies. In the context of deployment, it was decided to conduct a multiple case study from a single governmental agency. The case studies were selected according to the "literal replication logic" (Dubé & Paré, 2003). Table 1 provides short descriptions of cases. Table 2 presents relevant case characteristics.

Table 1: Short case descriptions

#	Description
Case 1	Predictions for the next two to three years about the economic developments and prospects. Policy makers use the predictions in their decision-making.
Case 2	Scenarios of the effects of current risks that may affect groups of institutions or sectors. Policy makers use this information to prevent or mitigate those risks.
Case 3	Text analysis of annual reporting documents. Remarkable sentences are selected and marked, to support risk analists.
Case 4	In-depth assessments of loan data to support decision making by credit risk experts.
Case 5	Detection of outliers in collected datafiles.

Table 2: Characteristics of the cases

	Case 1	Case 2	Case 3	Case 4	Case 5
Aim	Prediction	Scenarios	Text mining	Risk analysis	Outlier detection
Focus	Strategic	Strategic	Operational	Operational	Operational
Business process	Manual	Manual	Manual	Manual	Automated
Data volume	Small	Small	Small	Large	Small
Data type	Structured	Structured	Unstructured	Structured	Structured
Data sources	External and internal	External and internal	External	External	External
Deployment status	Deployed	Deployed	Deployed	Deployed	Not deployed

3.2 Data collection

Our main information sources are in-depth expert interviews with key-informants. Interviewees were data scientists and users. The data collection started in June 2020 and stretched over a period of four months. In total 11 people are interviewed. Each interview lasted approximately 90 minutes and was conducted online during the COVID-19 pandemic.

To promote *internal validity* as defined by Yin (2018), the interview guide was developed based on the topics of data science model deployment itself and of the resulting deployment effects described in the literature (section 2), i.e. changes in the target business process, changes in the model itself, changes in the model development process and the monitoring and adaptation of models in use. A list of topics/open questions was sent to the interviewees prior to the interview, as recommended by Maimbo and Pervan (2005). The questions were tailored according to the case. The expert interviews were semi-structured, and the questions were kept open to allow interviewees to speak freely.

Yin (2018) suggests triangulation to promote *construct validity*. Within the case studies, different data sources were therefore used. Additionally, for each case background information from project documentation, reports and memos, presentations and internal and external organizational communications was collected and summarized.

In order to minimize errors and biases, the *reliability* of the case studies was promoted by establishing a case study database. There, we stored all information about the data collection process, the data itself and the case study results. This helps to provide the same results in repeated trials (Yin, 2018).

3.3 Data analysis

The interviews were recorded and transcribed, coded and then analyzed using ATLAS.ti. We used first-level coding (Miles, Huberman, & Saldana, 2013) to identify and relate similar statements. The interviews were coded bottom-up, without using predefined codes. The codes were summaries of the interview quotes and were placed in code groups per subject and then gathered in main code groups representing main subjects. The code groups were gradually added with every interview. New code groups were added as needed. Eventually fifteen main code groups were defined, each containing two to eleven code groups. One example is the main code group Data Science Application or DSA (i.e. data science model), containing the code groups DSA input data programming, DSA algorithms, DSA output data programming, DSA size, DSA tooling, DSA machine learning basis and DSA traditional basis. Other main code groups represent various data science process steps within the target business process, model version management, encountered deployment challenges and deployment recommendations made by the interviewees. The code groups and main code groups were visualized in networks and the following themes were derived: changes in the data science model, model version management, the model development process, the target business process, incorporation of the model in the target business process, deployment challenges and model deployment recommendations. Then the interview analysis results were compared to and supplemented by the collected background documentation and for each case a case descriptions was created. Finally, the case descriptions were compared to each other and similarities and differences between cases were analysed.

4 Results

In the research cases data science model deployment led to changes in the target business process, changes in the models themselves, maintenance of deployed models in the model development process and a feedback loop between the target business process and the model development process. These four model deployment effects are further described below, with the related challenges and recommendations that were mentioned in the case interviews.

4.1 Adaptations in the business process

In all cases the model deployment resulted in the incorporation of a data science subprocess in the target business process. This data science subprocess automated one of the target business process tasks by using the data science model, and consisted of preparation of the model input data that were collected from the processed and stored data, definition of the model input parameters, limited adaptations of the model algorithms, running the model, validation of the model output data, interpretation of the results and communication of the outcomes to users or other business process steps. Interviewees recommended to make the data science subprocess fixed and stable, and preferably automated.

The strategic business processes in the predictions and scenarios cases contained the data science subprocess only. The operational business processes in the text mining, risk analysis and outlier detection cases contained additional business process steps. The respondents recommended adapting these additional process steps to provide input to and receive output from the data science subprocess, before developing the data science model. They also recommended incorporating the data science subprocess in the target business process before developing the model.

Only the outlier detection case had an automated business process (Table 2).
The business applications were adapted to exchange input and output data with the data science model. The respondents recommended to do this before developing the model. Adapting the business applications was challenging because the data

scientists lacked the necessary software engineering skills. In the interviews cooperation with software engineers was recommended.

In all cases the data science models were developed in a separate sandbox IT environment. The deployed models were run in the sandbox as well, because the production IT infrastructure lacked facilities to run data science models. As a result the operational business processes became distributed over two different IT environments: the data science subprocess was run in the sandbox and the other business process steps were run on the production IT infrastructure. This necessitated model input and output data transport between the two IT environments. This data transport was challenging because the two IT environments were strictly separated for security reasons. Another challenge was that the sandbox did not provide the continuity guarantees that were required for business-critical processes.

4.2 Adaptation of the data science models

As a result of deployment, the data science models in the operational business processes were adapted to achieve the required processing speed (risk analysis case) or to communicate with users (text mining case). Adaptation of the models in the manual strategic business processes (predictions and scenarios cases) was not necessary as the models did not need to exchange data with other applications and because the data scientists also were the users of their models so they did not need a user interface.

4.3 Maintenance of data science models

In all cases with a deployed model, the models could be adapted because of user feedback and new user requirements, new business theories and situations, new input data types, new strategic themes, new available data science algorithms or errors. This model maintenance was considered to be part of the model development process. The models could be adapted at the request of users and on the initiative of small teams of data scientists, who also have business knowledge.

The maintenance was carried out ad hoc (risk analysis case, text mining case), especially for smaller changes, in development projects (predictions case, text mining case) or as part of a medium-term development plan (scenarios case). Having such a plan was recommended, as it helps to keep the model well-structured and maintainable. In the risk analysis case, keeping the model modular and maintainable was a challenge caused by the lack of necessary software engineering skills. As a consequence, getting software engineering skills or cooperating with software engineers was recommended.

4.4 Feedback loop

In all cases, as a result of model deployment a feedback loop was established between the model development process and the target business process. Based on the experience of using the model in the business process, the model was improved in the development process. Subsequently, the model was used in the target business process again. In the cases with an operational business process, obtaining user feedback and cooperation was a challenge, but was strongly recommended as it was regarded as essential to have the deployed models used and improved.

In all cases, the feedback loop was enabled and supported by the establishment of a master model and a master model version management process. The master model was the tested and stable model version that forms the basis for all other model versions. Model improvement started by making a copy of the current master model. Permanently used model adaptations were merged into the master model again. One-off adaptations will be retained, but are not merged into the master model. Establishing a master model and careful management of the master model versions were strongly recommended.

4.5 Impact on strategic and operational business processes

Table 3 summarises the deployment effects that were found in the two cases with a strategic business process, in the two cases with a manual operational business process and in the case with an automated operational business process. In Table 3 can be seen that model deployment had more impacts on the business process and the data science model in the operational business processes than in the strategic business processes.

Table 3: Deployment effects in strategic and operational business processes

Deployment effect	**Strategic business process, manual Cases 1 and 2**	**Operational business process, manual Cases 3 and 4**	**Operational business process, automated Case 5**
Changes in target business process	Data science subprocess	Data science subprocess	Data science subprocess
	-	Changes in other business process steps	Changes in other business process steps
	-	-	Changes in business applications
	Business process in sandbox	Business process in sandbox and IT production	Business process in sandbox and IT production
Changes in data science model	-	Changes in model	- (model not yet deployed)
Model maintenance	In development process	In development process	In development process
Feedback loop	Feedback loop	Feedback loop	Feedback loop

5 Discussion

The research cases provide an overview of possible data science model deployment effects on the target business process, on the model itself and on the model development process.

The *adaptations in the business processes* resulting from model deployment in the research findings are in part consistent with prior literature and partly complementary. The incorporation of the data science subprocess in the target business process is consistent with the implementation of a repeatable data mining process mentioned by Wirth and Hipp (2000). A description of the data science subprocess or repeatable data mining process complements the literature. The data science subprocess contains the same activities as described in the data mining reference processes by Wirth and Hipp (2000). However, the data science subprocess is fixed and stable instead of iterative and flexible. In addition, in the data science subprocess the adaptations of the deployed model are limited to the minimum necessary to achieve the goals of the business process, while the data mining reference process describes the development of a complete data science model (Wirth & Hipp, 2000). The adaptations of the business process steps and business applications in operational business processes are consistent with the conclusions of Rupnik and Jaklič (2009).

The deployment of data science models in a sandbox is not described in the literature, which could be explained by the resulting challenges in the research findings.

The *changes in the data science models* resulting from model deployment in the research findings are consistent with the deployment changes in the models described by prior literature (John, Olsson, & Bosch, 2020; Baylor, et al., 2017; Cetinsoy, Martin, Ortega, & Petersen, 2016).

With regard to *maintenance*, the research results are different to extant literature. In all research cases examined in this study, maintenance of deployed data science models is considered to be part of the model development process. In contrast, previous studies (John, Olsson, & Bosch, 2020) consider model maintenance as part of the model deployment processes.

Concerning the *feedback loop*, our study presents interesting results. In all researched cases, the goals of the feedback loops are model improvements based on user experiences. In the literature, feedback loops are described for machine learning models only, to correct errors in the working of the models that are caused by changes in the input data (John, Olsson, & Bosch, 2020). In the two cases with a machine learning model (Table 2) such a feedback loop was not present. An explanation for this could be that the text mining model is in use for just a short time and the outlier detection model is not deployed yet.

The research findings regarding the *differences in deployment effects in strategic and operational business processes* are an addition to the extant literature.

6 Conclusions and limitations

The resulting *guidelines* derived from the recommendations made by the interviewees are in part consistent with prior literature and in part complementary. Adapting the business process and business applications before building the model is consistent with recommendations in prior literature (Davenport & Ronanki, 2018; Davenport & Malone, 2021; Rupnik & Jaklič, 2009). Working with a mixed team of data scientists, software engineers and users is consistent with recommendations of Davenport & Ronanki (2018) and Jackson et al. (2019). Establising the data science subprocess before developing the model, making the data science subprocess fixed and stable and establishing a feedback loop for model improvement based on user experiences are additional guidelines to the extant literature. So are the establishment of a master model and careful management of the master model versions, and the definition of a medium-term master model development plan.

This study comes with limitations. The case studies were carried out in one Dutch government organization. Therefore, the results may not be transferable to other government organizations and to commercial organizations. The small number of cases may further limit transferability (Yin, 2018). Further research is needed to assess generalizability of the research findings to other organizations. Despite these limitations this study provides valuable insights for both practitioners as well as academics in the effects and challenges of data science model deployment.

References

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. Journal of the Association for Information Systems, 17(2).

Baylor, D., Breck, E., Cheng, H.-T., Fiedel, N., Foo, C. Y., Haque, Z., . . . Koc, L. (2017). TFX: A TensorFlow-Based Production-Scale Machine Learning Platform. KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (pp. 1387–1395).

Brethenoux, E., Vashisth, S., & Hare, J. (2018). How to Operationalize Machine Learning and Data Science Projects. Retrieved from Gartner: https://www.gartner.com/en/documents/3880054/how-to-operationalize-machine-learning-and-data-science-

Castellanos, C., Pérez, B., Varela, C., Villamil, M., & Correal, D. (2019). A Survey on Big Data Analytics Solutions Deployment. In T. Bures, L. Duchien, & P. Inverardi (Ed.), Software Architecture. ECSA 2019. Lecture Notes in Computer Science. 11681. Cham: Springer.

Cetinsoy, A., Martin, F. J., Ortega, J. A., & Petersen, P. (2016). The past, present, and future of machine learning APIs. Conference on Predictive APIs and Apps, (pp. 43-49).

Chen, H. M., Schütz, R., Kazman, R., & Matthes, F. (2017). How Lufthansa Capitalized on Big Data for Business Model Renovation. MIS Quarterly Executive, 16(1).

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 36(4).

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. Harvard business review, 96(1), 108-116.

Davenport, T., & Malone, K. (2021). Deployment as a Critical Business Data Science Discipline. Harvard Data Science Review.

Desouza, K., Dawson, G., & Chenok, D. (2020). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. Business Horizons, 63(2), 205-213.

Dubé, L., & Paré, G. (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. MIS Quarterly, 27(4), 597–636.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37-37.

Jackson, S., Yaqub, M., & Li, C. X. (2019). The agile deployment of machine learning models in healthcare. Frontiers in Big Data, 1(7).

John, M., Olsson, H., & Bosch, J. (2020). Architecting AI Deployment: A Systematic Review of State-of-the-Art and State-of-Practice Literature. International Conference on Software Business (pp. 14-29). Cham: Springer.

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. Communications of the ACM, 57(3), 78-85.

Lin, J., & Kolcz, A. (2012). Large-scale machine learning at twitter. Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data (pp. 793-804). New York, NY: ACM.

Maimbo, H., & Pervan, G. (2005). Designing a case study protocol for application in IS research. PACIS 2005 Proceedings.

McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. Harvard Business Review, 90(10), 60-68.

Miles, M. B., Huberman, A. M., & Saldana, J. (2013). Qualitative Data Analysis: A Methods Sourcebook (3rd ed.). Los Angeles, CA: SAGE Publications.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big data, 1(1), 51-59.

Ransbotham, S., Gerbert, P., Reeves, P., Kiron, D., & Spira, :. (2018). Artificial intelligence in business gets real: Pioneering companies aim for AI at scale. MIT Sloan Management Review.

Rupnik, R., & Jaklič, J. (2009). The Deployment of Data Mining into Operational Business Processes. In J. Ponce, & A. Karahoca, Data Mining and Knowledge Discovery in Real Life Applications (pp. 373-388). IntechOpen.

Sumbaly, R., Kreps, J., & Shah, S. (2013). The big data ecosystem at linkedin. Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data (pp. 1125-1134). New York, NY: ACM.

Thusoo, A., Shao, Z., Anthony, S., Borthakur, D., Jain, N., Sarma, J., . . . Liu, H. (2010). Data warehousing and analytics infrastructure at facebook. SIGMOD '10: Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (pp. 1013–1020). New York, NY: Association for Computing Machinery (ACM).

Watson, H. J. (2017). Preparing for the Cognitive Generation of Decision Support. MIS Quarterly Executive, 16(3).

Wegener, D., & Rüping, S. (2010). On integrating data mining into business processes. International Conference on Business Information Systems (pp. 183-194). Heidelberg: Springer.

Wirth, R., and Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1). London, UK: Springer-Verlag.

Yin, R. K. (2018). Case Study Research and Applications: Design and Methods (6th ed.). Thousand Oaks, CA: Sage.