



Interpretability of Machine Learning Solutions in Industrial Decision Engineering

Inna Kolyshkina¹  and Simeon Simoff² 

¹ Analytikk Consulting, Sydney, Australia
inna@analytikk.com

² Western Sydney University, Penrith, NSW 2751, Australia
s.simoff@westernsydney.edu.au
<http://www.analytikk.com/>
<https://www.westernsydney.edu.au/>

Abstract. The broad application of machine learning (ML) methods and algorithms in diverse range of organisational settings led to the adoption of legislation, like European Union’s General Data Protection Regulation, which require firm capabilities to explain algorithmic decisions. Currently in the ML literature there does not seem to be a consensus on the definition of interpretability of a ML solution. Moreover, there is no agreement about the necessary level of interpretability of such solution and on how this level can be determined, measured and achieved. In this article, we provide such definitions based on research as well as our extensive experience of building ML solutions for various organisations across industries. We present CRISP-ML, a detailed step-by-step methodology, that provides guidance on creating the necessary level of interpretability at each stage of the solution building process and is consistent with the best practices of project management in the ML settings. We illustrate the versatility and effortless applicability of CRISP-ML with examples across a variety of industries and types of ML projects.

Keywords: Interpretability in machine learning · Machine learning methodology · Data science methodology · Level of interpretability · Model interpretability · Project management

1 Introduction and Background to the Problem

The rapid increase in the range and diversity of data-driven algorithmic decision engineering has led to the sharp increase of the need for a consistent and comprehensive methodology and process that govern the development, deployment, utilisation and evaluation of the engineering outcomes. Decision engineering, whether in data science (DS) and data analytics (DA) projects or in autonomous systems, like computer-based recommenders and advisers, rely on machine learning (ML) and artificial intelligence (AI) systems, hence, requires

interpretability/explainability of system behavior and decision making outcomes. Interpretability in AI/ML depends on two connected aspects: (i) development of interpretability solutions for AI/ML algorithms and (ii) development of consistent and comprehensive methodology/framework for data science projects, which minimises the risk of project failures and guarantees achieving the necessary (for the project) level of ML/AI system interpretability. Guidotti et al. [13] provide a systematic overview of the current state-of-the-art in (i). Our paper is focused on the development of methodology, which addresses (ii). The rationale supporting such focus is built on the following major arguments: (a) high proportion of data science project failures - an indicator of the need for a consistent and comprehensive methodology and process for ML/AI projects; (b) emerging requirements for sufficient explainability of ML systems - this puts pressure on creation of frameworks/ methodologies which can ensure the sufficient explainability of ML systems; and (c) lack of standard methodology - contemporary methodologies do not include standard consistent components which ensure interpretability through the project. Further we elaborate each of these arguments.

1.1 High Proportion of Data Science Project Failures

Recent reports estimate that between 70% and 85% of data science/ML/AI projects fail. NewVantage survey [1] noted that 77% of businesses see big data and AI initiatives as a big challenge for business. Gartner research [24] argues that 80% of analytics insights will not deliver business outcomes through 2022. McKinsey research [8] reports that 92% of big companies are not successful in using analytics in the organisation. In the past three years the percentage of firms identifying themselves as being data-driven has declined from 37.1% in 2017 to 31.0% in 2019 [1], which is counterintuitive to the expected impact of AI technologies on decision making. Key reasons for these failures are linked to the lack of proper process and methodology in requirements gathering, establishing realistic project timelines, task coordination, communication, and suitable project management framework [1, 29]). Improved methodologies are needed as the existing ones do not cover many important aspects and tasks [17]. Further, studies have shown that the recent biased focus on the tools and systems has limited the ability to gain value from organisational analytic effort [22] and that data science projects need to increase their focus on methodology, including process and task coordination ([12]). Practitioners agree with this view [11].

1.2 Requirements for Sufficient Explainability of ML Systems

In parallel with the above discussed tendencies, there is pressure on creation of frameworks/methodologies, which can ensure the sufficient explainability of the output of the ML systems. Whilst some ML systems (for instance, decision tree and rule induction algorithms) offer methodologically transparent means supporting interpretability/explainability of their output, there is a class of the so-called 'black box' ML models, such as deep neural nets, tree and network

ensembles and support vector machines, which do not provide embedded interpretability. There have been a number of cases where this class of models demonstrated the lack of fairness and poor accuracy [20, 25]. In high-stake situations, systems in which the inner workings are not transparent can be unfair, unreliable, inaccurate and even harmful [6, 25]. This view is reflected in legislations like the European Union's General Data Protection Regulation (GDPR) [2], though there are also warnings to policymakers to be aware of potential impact of legislations like GDPR on AI and emerging algorithmic economy. These developments increase the pressure on creation of frameworks and methodologies, which can ensure the sufficient explainability of AI and ML solutions. A report by AI Now Institute [23] recommended standardising the AI and ML system-building process and incorporating relevant algorithmic impact assessments into the processes the organisations already use. Many organisations and major technology developers are following this recommendation [4].

1.3 The Lack of Standard Methodology

Though having a good methodology is important for the project success, so far there is no formal standard for methodology in the data science projects [26]. CRISP-DM methodology [28], created in the late 1990s, is considered the de-facto standard [5, 14]. It is industry-, tool- and application agnostic [17]. It is not fully meeting the needs of data science community, and its usage appears to be decreasing [26]. While various extensions of the methodology, including IBM's ASUM-DM and Microsoft's TDSP were proposed, at this stage none of them has become the standard. Many CRISP-DM extensions are fragmented and either propose additional elements into the data analysis process, or focus on organisational aspects without the necessary integration of domain-related factors [21]. Finally, while methodologies from related fields, like the agile approach used in software engineering, are being considered for use in data science projects, there is no full clarity on whether they are fully suitable for the purpose [15], therefore we did not include them in the scope of this paper.

1.4 Opportunities in Creating Interpretability-Related Methodologies

Recent state-of-the-art reviews related to interpretability [7, 9, 19] as well as more algorithm-focussed reviews [13, 16, 18] report that: (i) interpretability of AI and ML solutions and the underlying models is not well defined; (ii) the work related to interpretability is scattered throughout a number of disciplines, including AI, ML, human-computer interaction (HCI), visualisation, cognition; and (iii) current research seems to address a particular category or technique instead of the overall concept of interpretability. Similarly, while there is a number of suggested approaches to measuring interpretability [18], there is no consensus neither on measuring or evaluating the *level of interpretability* nor on the best type of explanation metric [9]. Currently there is confusion about the interpretability notion [19], including a lack of clarity about how the many proposed interpretation

approaches can be evaluated and compared against each other, how to choose a suitable interpretation method for a given business issue and audience as well as limited guidance on how interpretability can actually be used in data-science life cycles. The lack of consensus gives an opportunity to create a comprehensive methodology, which takes into account different perspectives and aspects of interpretability (comprehensibility), such as predictive accuracy, bias, noise, sensitivity, faithfulness, specificity, local interpretability, global interpretability and domain specifics.

2 Methodology of Establishing and Building the Necessary Level of Interpretability of ML Business Solution

2.1 The Necessary Level of Interpretability of an ML Solution

In line with interpretability in Google's responsible AI practices [4] and expanding on [10] approach, we introduce the concept of *necessary level of interpretability* (NLI) of a business ML solution as the combination of the degree of accuracy of the underlying algorithm and the extent of understanding of the inputs, inner workings, the outputs, the user interface and the deployment aspects of the ML solution that is required to achieve the project goals. If this level is not achieved, the solution will be inadequate for the purpose. This level needs to be established and documented at the initiation stage of the project as part of requirements collection. We then describe a ML system as *sufficiently interpretable or not* based on whether or not it achieved the required level of interpretability.

Obviously, this level will differ from one project to another depending on the business goals and agreed measures of interpretability. If individuals are directly and strongly affected by the solution-driven decision - for example, in medical diagnostic or legal settings - then both the ability to understand and trust the internal logic of the model as well as the ability of the solution to explain individual predictions are extremely important. In other cases, when a ML solution is used in order to inform business decisions about policy, strategy or interventions aimed to improve the business outcome of interest, then it is the need to understand and trust the internal logic of the model that is of most value and individual predictions are not the focus of the stakeholders. For example, in one of our projects an Australian state organisation wished to establish what factors influenced the proportion of children with developmental issues and what interventions can be undertaken in specific areas of the state in order to reduce that proportion. The historical, socioeconomic and geographic data provided for the project was aggregated at a geographic level of high granularity.

In other cases, for example, in the case of an online purchase recommender solution, the overall outcome, such as increase in sales volume, may be of higher importance, than the interpretability of the model. Similar requirements of solution interpretability were in a project where an organisation owned assets that were located in remote areas and were often damaged by birds or animals nests.

The organisation wished to lower their maintenance cost and planning by identifying as soon as possible the assets where such nests were present instead of doing expensive examination of each asset. This was achieved by building a ML solution that classified Google Earth images of the assets into those with and without nests. In this project it was important to identify as accurately as possible a proportion of assets with nests on them, while misclassifying an individual asset image was not of great concern.

2.2 CRISP-ML Methodology

The proposed methodology of building interpretability of a ML system is based on our methodology CRISP-ML. It is an updated version of CRISP-DM and is industry-, tool- and application-agnostic. It seamlessly accommodates modern ML techniques and creates the NLI through the whole ML solution creation process. In order to explain how to ensure that the NLI of ML system is achieved in the project, we elaborate its seven stages, summarised in Fig. 1. We illustrate key concepts with real-world examples/mini case studies.

The Project Initiation and Planning Stage. Interpretability Matrix

Objectives and Importance. This stage is crucial for the overall project success [3] and for the system interpretability building. It covers *the activities needed to start up the project*, including (i) the identification of project sponsor/key stakeholders and preparation of project charter – a document that outlines project objectives, scope, high-level deliverables, assumptions, constraints, and risks – after being signed off it serves as a reference of authority for the future; and (ii) the *planning activities* such as collecting requirements; agreeing upon initial data to use; preparing a detailed scope statement; estimating effort, duration and costs; assessing and responding to risks; developing communications documents, project schedule and plan and finally, obtaining project sponsor’s approval to proceed with the project.

Establishing the Necessary Level of Interpretability. NLI is established as part of requirements collection. It is driven by the project objectives, and also influenced by domain specifics, stakeholder requirements, project constraints, industry regulator requirements to name the key factors. Proper requirements collection (i.e. determining and documenting conditions or tasks that must be completed to meet the project objectives) is crucial to the project success [3]. As part of requirement gathering we work with key stakeholders to determine NLI of the solution. Typically, this may require that the relevant stakeholders have a clear understanding of the (i) *data inputs used* - are they reliable, of suitable quality and representative of the real-world data; (ii) *solution outputs* - are they consistent with the project goals in terms of accuracy, format, ease of understanding for the end users, level of potential business insight, and are they valid from the ML and business perspectives; (iii) *format* they should be provided to the

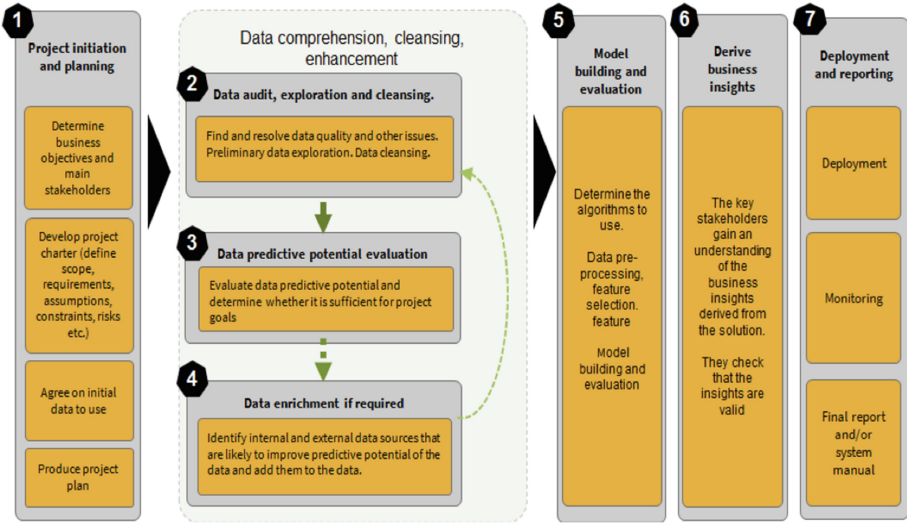


Fig. 1. Conceptual framework of CRISP-ML methodology

end user, e.g. tables, visualisations, graphs, infographics and other representations; (iv) *high-level modelling approach*, its validity and whether it is proven and likely to work in this industry; (v) *implementation process* of the solution in the organisational systems, and how it should be audited, monitored and updated.

For example, in a project in workers compensation insurance that aimed to identify cases likely to become expensive, the objectives included building a ML system that would: (i) explain what factors and to what extent were influencing the outcome of interest i.e. claims cost; (ii) allow the organisation to derive business insights that will help make data-driven accurate decisions regarding what changes can be done to improve the outcome i.e. reduce the likelihood of an expensive claim by a specified percentage; (iii) be accurate, robust and able to work with real-world organisational data; and (iv) have easy-to-understand outputs that would make sense to the executive team and end users (case managers) and that the end users could trust. The established interpretability requirements in this project included: (i) having trustworthy, quality data inputs, representative of the organisational data that the solution would be deployed on; (ii) the outputs should be provided as business rules that were easy to understand for end users and to deploy on organisational data; (iii) the high-level algorithmic approach had to be easily understood by the executive team and the BI team who would monitor its performance; (iv) explain at least 80% of variation in the data, be valid from the ML point of view; and (v) its outputs needed to make sense to the domain experts.

Creating Project Interpretability Matrix as Part of Requirements Collection. The next step is to establish what needs to be done by each stakeholder at each

project stage in order to ensure that NLI is achieved. For this, we create the *interpretability matrix* (IM), whose *rows* show CRISP-ML stages, and *columns* represent key stakeholders. In each IM cell we need to document what needs to be done by each stakeholder at each project stage to ensure that NLI of the solution is achieved. Completed IM becomes part of the business requirements document; the activities it outlines are integrated into the project plan and are performed, updated and monitored along with the project plan as needed. For example, Fig. 2 shows a very high-level IM for the above mentioned insurance project. The green, yellow and white colour background indicate, respectively, high, medium and low level of involvement of a stakeholder group.

Project stages	Key Stakeholders			
	Executive team	DE team	IT team	Modelling team
Project initiation and planning	Establishing project requirements, understanding domain aspects; agreeing on the initial data to use			
Data audit. Quality and sanity checks	Low	clarify domain-related aspects and provide sanity checks	Low	Data analysis; Data predictive potential evaluation and enrichment. Algorithm development and evaluation
Evaluation of data predictive potential	Low		Low	
Data enrichment if required	Low		Low	
Model building and evaluation	Low		Low	
Deriving business insights	Evaluating business insights from modelling team's report		Low	Preparing a report and a presentation for ET
Deploying. Monitoring. Updating	Low		Deployment	Preparing a solution manual

Fig. 2. Example of a very high-level interpretability matrix for the insurance project. (Color figure online)

Entries to the Interpretability Matrix at Each Stage of CRISP-ML

Further we discuss typical entries to the interpretability matrix (IM) at each stage of CRISP-ML, and illustrate them with real-world examples. Usually, in our experience key stakeholders for ML system projects are the executive team (ET); the data provider (DP) team which is often a part of the organisational IT team; the domain experts (DE) and the modelling team (M). These abbreviations will be used in the below descriptions along the stages IM.

Stage 1. Figure 3 provides details of IM content related to this stage.

Stages 2–4. Stages 2, 3 and 4 in Fig. 1 are mainly data-related and form the data comprehension, cleansing and enhancement mega-stage. Further we consider the content of interpretability matrix for each individual stage.

	ET	DE	IT/DP/Deployers	M
Stage 1. Project initiation and planning	Stage aim: Identify project sponsor and key stakeholders, prepare the project charter; collect requirements; agree upon initial data to use; prepare a detailed scope statement; estimate effort, duration and costs; do risk analysis; develop communication plan etc.			
	Work together with the other stakeholders to establish and/or gain in-depth understanding of the project objectives; scope; assumptions, etc. Gather requirements, specifically establish and document the required interpretability level including the minimal required accuracy and understanding of inputs, outputs, inner workings and implementation aspects.			
	Gain a high-level understanding of the approach for the ML solution building process and potential outcomes. Help modellers to gather requirements, establish assumptions, constraints etc.		Gain in-depth understanding of requirements and interpretability level Gain a high-level understanding of a) domain aspects relevant to the project goals achievement; b) the data agreed to be used.	

Fig. 3. CRISP-ML: Stage 1 - typical IM content related to this stage.

Stage 2. Data audit, exploration and cleansing play a key role in the development of stakeholder trust in the approach and ultimately in the solution, if achieving user trust in the solution is part of the established NLI for the project. Figure 4 demonstrates the typical content of IM at this stage. This stage is important in any project where interpretability is of high priority, because wrong data values may slide in unnoticed and skew the outputs. For example, in a project aiming to establish what drives morbidity of pregnant women with diabetes and their children, the data on the age of the mother had records of 99 yo. Domain experts clarified that ‘99’ was a code for ‘Age Unknown’.

Stage 3. Figure 5 demonstrates the typical content of the interpretability matrix related to the evaluation of the predictive potential of the data. This stage is often either omitted or not stated explicitly in other processes/frameworks (for example, in CRISP-DM), however it is crucial in terms of achieving NLI because it establishes whether the information in the data is sufficient for achieving the project goals (for example, for explaining the outcome of interest). At this stage, in-depth data exploration and preliminary modelling is performed, where several advanced and powerful non-supervised and supervised ML techniques are used to explore the data, establish the most promising strategies for feature engineering/data transformations and modelling and assess whether the initially identified data and other resources are sufficient for achieving the business goals. The choice of the ML techniques is tailored for each project; detailed description of them and the process of assessment of the predictive potential is beyond the scope of this paper. Techniques used for estimating predictive potential include

Stage 2. Data audit, exploration and cleansing.	Stage aim: Demonstrate that the data inputs into the solution are of adequate quality, make sense to DEs and are representative of the real-world data that the solution will be deployed on.			
	ET	DE	IT/DP/Deployers	M
	Low	Provide domain-related knowledge to help modellers to develop the needed data understanding, conduct data quality assessment and run relevant data "sanity checks".	Provide data-related knowledge to help modellers to develop the needed data understanding and conduct data quality assessment.	Develop the needed data understanding, conduct data quality assessment and run relevant data "sanity checks".

Fig. 4. CRISP-ML: Stage 2 - typical IM content related to this stage.

components of various dimensionality reduction approaches, advanced clustering methods and proven highly-predictive methods such as random forest, boosting methods and deep neural networks).

In our experience, initially identified data often needs to be enriched by external data. For example, in the insurance example the predictive potential of the data containing claim and worker data history was shown to be insufficient for the project objectives. The domain experts suggested to enrich the initial data with the history of what doctors and other health service providers a worker saw, the medicines worker was prescribed (for example, opioids) and some other data. Adding these data significantly improved the model accuracy. Enrichment by additional data is not always needed. Specifically, from our experience, image and free text data often do not require additional information to build an accurate model. For example, in a project where social media data were used to compare customer perception of the four Australian major banks, at stage 3 we established that collected data were enough for project purposes, but additional in-depth feature engineering was required.

Stage 3. Data predictive potential evaluation	Stage aim: assess whether the initially identified data is sufficient for project purposes.			
	ET	DE	IT/DP/ Deployers	M
	Low	Consultation and clarification of any domain-related aspects	Low	Assess data predictive potential using ML methods

Fig. 5. CRISP-ML: Stage 3 - typical IM content related to this stage.

Stage 4. Figure 6 shows a typical content of IM if it was determined at stage 3 that the initially identified data or other resources are not sufficient for the project purposes and the data have to be enriched. In practice this involves additional analysis, usually, data enrichment by adding new data, less often by in-depth feature engineering of the existing data. Additional internal and external data sources are identified, the new data is extracted, audited, cleansed and added to the previously used data. Then predictive potential of the enriched data is again assessed by applying the same ML methods as in stage 3.

Stage 4. Data enrichment	Stage aim: additional internal and external data sources are identified, the new data is extracted, audited, cleansed and added to the previously used data. Then predictive potential of the enriched data is again assessed. This step is repeated until the necessary level of predictive potential is achieved.			
	ET	DE	IT/DP/Deployers	M
	Low	Help identify the data sources suitable for enrichment and clarify any domain-related aspects	Help identify the data sources suitable for enrichment and clarify any data-related aspects	Access the new data, conduct data enrichment and return to stage 3 to evaluate whether the enriched data is sufficient for the project purposes

Fig. 6. CRISP-ML: Stage 4 - typical IM content when data enrichment is required.

This step is repeated until the necessary level of predictive potential is achieved or, if it has been established that achieving it is impossible, this finding is further discussed with the key stakeholders and the relevant decisions are made. Thorough planning at Stage 1 minimises the risk of that occurring. In the insurance example described above, data enrichment was a key step. The fact that the model showed that the cost of a claim can be significantly dependent on the providers a worker visited, built further trust in the solution because it confirmed the domain experts hunch that they previously had not had enough evidence to prove.

Stage 5. Figure 7 shows a typical content of IM for the model building and evaluation stage. To achieve NLI, modellers have to choose the appropriate technique(s) that will balance the required outcome interpretability with the required accuracy and with other requirements/constraints (e.g. the needed functional form of the model and/or algorithm). The ML techniques to be used for modelling are selected taking into account the predictive power of the model, its suitability for the domain and the task, and NLI. The data is pre-processed, and modelled and model performance is evaluated. Detailed description of the process of algorithm choice and model assessment is beyond the scope of this paper and will be covered in a separate publication.

Stage 5. Model building and evaluation	Stage aim: select ML techniques to be used for modelling; preprocess the data, build the model/models; evaluate model performance/choose the best model			
	ET	DE	IT/DP/Deployers	M
	Low	Clarify any domain-related aspects	Low	Select the ML techniques to be used for modelling taking into account interpretability requirements; preprocess the data (including feature selection and feature engineering); build the model and evaluate its performance. Choose the model whose performance level is as required by the necessary interpretability level established at stage 1.

Fig. 7. CRISP-ML: Stage 5 - typical IM content indicating how NLI influences the strategy of choosing modelling techniques by the modelling team.

In the insurance example, the solution output had to be produced in the form of business rules. Therefore the feature engineering methods and modelling algorithms used included rule-based techniques such as decision trees and association rule-based methods. In another example, a large Australian asset-owning organisation needed a ML solution that would help them to proactively optimise asset maintenance planning and cost and asset failure risk reduction as well as to justify funding requests to industry regulator. The regulator specifically requested that the solution be delivered in the linear model form. Such a requirement towards the model type is common in some areas. For example, in credit risk assessment certain models have to be in the logistic regression format. Often there is no constraint on the model functional form. For example, in the above mentioned image classification project, we simply used the most accurate model we could build, which turned out to be a convolutional neural network. Some other techniques used in stage 5 include boosted regression trees, random forest, LASSO methods and deep neural networks.

Stage 6. Figure 8 shows how the IM reflects the role of interpretability in the formulation of business insights necessary to achieve the project goals and in helping the ET and end users to understand the derived business insights and to develop trust in them. DE team might also have a medium to low level of involvement for clarification of any domain-related aspects.

For example, in the insurance project modellers and DEs prepared a detailed presentation for the ET explaining not only the learnings from the solution but also the high-level model structure and its accuracy. In the image processing project, on the other hand, the presentation was focussed on the results and their accuracy rather than on model inner workings.

Stage 6. Deriving business insights.	Stage aim: Executives (and end users if relevant) gain an understanding of the business insights derived from the solution, develop an understanding that the insights are valid and valuable and build trust in these insights.			
	ET	DE	IT/DP /Deployers	M
	Gain an understanding of the business insights derived from the solution, develop an understanding that the insights are valid and valuable and build trust in these insights.	Clarification of any domain-related aspects	Low	Help the executives and end users to develop the trust and understanding. Usually, they prepare presentations or reports to achieve that.

Fig. 8. CRISP-ML: Stage 6 - typical content of IM related to this stage.

Stage 7. Figure 9 shows the shift of responsibilities for ensuring the achieved interpretability level is maintained during the future use of the solution. At this stage, a deployment is conducted if required and monitoring/updating process and schedule is prepared, based on the developed technical report.

This stage and the related interpretability aspects differ significantly depending on the project goals. We illustrate this diversity with some of our projects. In the childhood development project no deployment was required, but a report and a visualisation of the solution was needed. In the nest identifying project no deployment was required, but the list of assets likely to have nests on them was needed as well as a brief report and the model code. In the insurance and

Stage 7. Deployment and reporting.	Stage aim: Deployment is conducted if required and monitoring/updating process and schedule is prepared. A technical report is created.			
	ET	DE	IT/DP/ Deployers	M
	Low	Clarification of any domain-related aspects	Gain understanding of any information regarding the solution that is necessary for its deployment and maintenance. Deployment and related activities	Write technical report, manual or other documentation as required explaining the aspects needed to ensure the necessary interpretability level Provide consultation about model specifics to the deployer team

Fig. 9. CRISP-ML: Stage 7 - activities ensuring the achieved interpretability level is maintained during the future utilisation of the solution.

asset management examples deployment was needed, as well as a full technical report, a solution manual and an updating and monitoring recommendations.

2.3 Conclusions

This article addresses the problem of providing companies with capabilities to explain algorithmic decision engineering. We introduced a definition of interpretability of an end-to-end business ML solution, the necessary level of interpretability of such solution and a methodology (CRISP-ML) of achieving it. CRISP-ML integrates interpretability aspect into the overall framework instead of just at the modelling stage. It requires to take more than the algorithm accuracy into consideration when deciding what the ‘best’ model is by pushing questions about use and interpretability up front. Further, it defines the responsibilities of different stakeholders to ensure that this is done.

CRISP-ML is an extension of CRISP-DM, which enables organisations to (i) establish shared understanding across all key stakeholders about the solution and its use; (ii) build stakeholder trust in the solution outputs; and (iii) get buy-in from all relevant parts of the organisation. It allows the end users to confidently interpret the solution results and make successful evidence-based business decisions. If needed, they can explain these decisions to any external party. We successfully applied this methodology in commercial projects across a variety of industries including banking, insurance, utilities, retail, FMCG, public health and transport to name some areas. It effortlessly accommodates the diversity of industry specifics as well as variety of organisational goals, ML techniques and data types. While comparing the effectiveness of this methodology to other approaches is beyond the scope of this paper, future work includes experimental assessment similar to the one performed in [27].

References

1. Big Data and AI executive survey. Technical report, NewVantagePartners LLC (2019)
2. General Data Protection Regulation (GDPR). Official Journal of the European Union L 119/1 (2016). <https://gdpr-info.eu/>
3. PMBOK® Guide - Sixth Edition. Project Magament Institute (2017)
4. Google AI: Responsible AI Practices - Interpretability. <https://ai.google/responsibilities/responsible-ai-practices/?category=interpretability>. Accessed 5 Aug 2019
5. Ahmed, B., Dannhauser, T., Philip, N.: A Lean Design Thinking Methodology (LDTM) for machine learning and modern data projects. In: Proceedings of 2018 10th Computer Science and Electronic Engineering (CEECE), pp. 11–14. IEEE (2018)
6. Dawson, D., et al.: Artificial intelligence: Australia’s ethics framework. Technical report, Data61 CSIRO, Australia (2019)
7. Doshi-Velez, F., Kim, B.: Towards a rigorous science of interpretable machine learning. arXiv e-prints [arXiv:1702.08608](https://arxiv.org/abs/1702.08608), February 2017

8. Fleming, O., Fountaine, T., Henke, N., Saleh, T.: Ten red flags signaling your analytics program will fail. Technical report, McKinsey & Company (2018)
9. Gilpin, L.H., Testart, C., Fruchter, N., Adebayo, J.: Explaining explanations to society. CoRR abs/1901.06560 (2019). <http://arxiv.org/abs/1901.06560>
10. Gleicher, M.: A framework for considering comprehensibility in modeling. *Big Data* **4**(2), 75–88 (2016)
11. Goodson, M.: Reasons why data projects fail. KDnuggets, November 2016. <https://www.kdnuggets.com/2016/11/ten-ways-data-project-fail.html>
12. Grady, N.W., Underwood, M., Roy, A., Chang, W.L.: Big data: challenges, practices and technologies: In: NIST Big Data Public Working Group workshop at IEEE Big Data 2014. Proceedings of IEEE International Conference on Big Data 2014, pp. 11–15 (2014)
13. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. *ACM Comput. Surv.* **51**(5), 93:1–93:42 (2018)
14. Huang, W., McGregor, C., James, A.: A comprehensive framework design for continuous quality improvement within the neonatal intensive care unit: integration of the SPOE, CRISP-DM and PaJMa models. In: Proceedings of IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), pp. 289–292 (2014)
15. Larson, D., Chang, V.: A review and future direction of agile, business intelligence, analytics and data science. *Int. J. Inf. Manag.* **36**(5), 700–710 (2016)
16. Lipton, Z.C.: The mythos of model interpretability. *ACM Queue* **16**(3), 30:31–30:57 (2018)
17. Mariscal, G., Marbán, O., Fernández, C.: A survey of data mining and knowledge discovery process models and methodologies. *Knowl. Eng. Rev.* **25**(2), 137–166 (2010)
18. Molnar, C., Casalicchio, G., Bischl, B.: Quantifying interpretability of arbitrary machine learning models through functional decomposition. [arXiv:1904.03867](https://arxiv.org/abs/1904.03867)
19. Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B.: Interpretable machine learning: definitions, methods, and applications. [arXiv:1901.04592](https://arxiv.org/abs/1901.04592)
20. O’Neil, C.: *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishers, New York (2016)
21. Plotnikova, V.: Towards a data mining methodology for the banking domain. In: Kirikova, M., et al. (ed.) Proceedings of the Doctoral Consortium Papers Presented at the 30th International Conference on Advanced Information Systems Engineering, CAiSE 2018, pp. 46–54 (2018)
22. Ransbotham, S., Kiron, D., Prentice, P.K.: Minding the analytics gap. *MIT Sloan Manag. Rev.* **56**, 1 (2015)
23. Reisman, D., Schultz, J., Crawford, K., Whittaker, M.: Algorithmic impact assessments: a practical framework for public agency accountability. Technical report, AI Now Institute, April 2018
24. Roy Schulte, W., et al.: Predicts 2019: data and analytics strategy. Technical report, Gartner Research, November 2018
25. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019)
26. Saltz, J.S., Shamshurin, I.: Big data team process methodologies: a literature review and the identification of key factors for a project’s success. In: Proceedings of 2016 IEEE International Conference on Big Data 2016, pp. 2872–2879 (2016)
27. Saltz, J.S., Shamshurin, I., Crowston, K.: Comparing data science project management methodologies via a controlled experiment. In: HICSS (2017)

28. Shearer, C.: The CRISP-DM model: the new blueprint for data mining. *J. Data Warehouse*. **5**, 13–22 (2000)
29. Stieglitz, C.: Beginning at the end - requirements gathering lessons from a flowchart junkie. In: PMI® Global Congress 2012–North America, Vancouver, British Columbia, Canada. Project Management Institute, Newtown Square, PA (2012)