

# State of Knowledge Discovery Process Models and Frameworks

Martins Jansevskis<sup>1</sup>, Kaspars Osis<sup>1</sup>

<sup>1</sup>Vidzeme University of Applied Sciences

## INTRODUCTION

The first knowledge discovery process models were developed in the late 1990s, almost three decades ago. Based on several surveys, researchers F. Martínez-Plumed et al. (Martinez-Plumed, F. et al., 2021) and Rotondo A., Fergus Q. (Rotondo, A., Quilligan, F., 2020) argue that CRISP-DM (Chapman, et al., 2000) remains the default standard for developing data acquisition and retrieval projects. In almost thirty years, the industry and technology has evolved, and data science is now the leading term. The knowledge discovery process has changed significantly since the inception of the CRISP-DM model. An area in which the CRISP-DM model does not work well enough is data-driven products, and most products nowadays are in fact data driven. The amount and complexity of data in applications suggests that data processing requires significant technical work on management and infrastructure. In the CRISP-DM model, data is included as a static unit in the middle of the process (Martinez-Plumed, F. et al., 2021), which means that the knowledge retrieval process needs to be viewed in the context of the knowledge retrieval framework in which the process is applied.

## MATERIALS AND METHODS

To identify the state of knowledge discovery process models and frameworks, the authors adopted a systematic literature review approach on knowledge discovery process models and knowledge discovery frameworks to answer two research questions. Q1: what kinds of process models are available and what is the state of knowledge discovery process models? And Q2: what are the design principles that characterise knowledge discovery frameworks? To achieve this objective, research articles addressing process models and frameworks were analysed.

## RESULTS

The knowledge discovery process models developed in the mid 90s are still being used in organisational data mining projects. Most data retrieval algorithms and tools stop at creating and delivering models that meet technical requirements. Models are being developed, but entrepreneurs are either not interested in them or do not know what to do next to add value to their business decisions. Knowledge discovery in organisations is mostly a closed process for solving optimisation problems starting with problem definition, framework or model development to the discovery of workable models designed to provide functioning business insights that can be linked to or integrated with business processes and systems. Obtaining information and hidden correlations from data has a growing trend in information systems; in order to provide better services to end users and support decision-making processes, as well as to acquire valuable knowledge, it is necessary to integrate and analyse the generated data sets from different domains. Multiple innovative knowledge discovery frameworks are being analysed in the research paper.

## DISCUSSION

The heterogeneity of the definition of knowledge and the perception of its concept creates wide possibilities for interpretation. One of the knowledge definitions according to Ikujiro Nonaka and Hirotaka Takeuchi's (Nonaka, I., Takeuchi, H., 1995) theory of knowledge creation is the acquisition of new knowledge applied to its usefulness to a particular organisation by relating it to the social context in which the knowledge is created and used. Knowledge discovery process models are still widely used within organisations and there are multiple knowledge discovery framework proposals for various fields. The necessity and corresponding technological requirements for knowledge discovery frameworks remain open for discussion.

Recommended future research directions involve the following points:

- Technical requirements of knowledge discovery frameworks that define the requirements of components.
- Security requirements for knowledge discovery frameworks.
- Personal data protection requirements for knowledge discovery.

**KEYWORDS:** Knowledge discovery, Process models, Knowledge frameworks

## REFERENCES

Martinez-Plumed, F. *et al.* (2021) 'CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories', *IEEE Transactions on Knowledge and Data Engineering*, 33(8), pp. 3048–3061.

Rotondo, A. and Quilligan, F. (2020) 'Evolution Paths for Knowledge Discovery and Data Mining Process Models', *SN Computer Science*, 1(2), p. 109. doi: 10.1007/s42979-020-0117-6.

Chapman, P., *et al.* (2000) 'CRISP-DM 1.0: Step-by-step data mining guide'. Available at: <https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf>

Nonaka, I. and Takeuchi, H. (1995) *The knowledge-creating company: how Japanese companies create the dynamics of innovation*. New York: Oxford University Press.