

Association for Information Systems

AIS Electronic Library (AISeL)

ICIS 2022 TREOs

TREO Papers

12-12-2022

Domain expertise afore and beyond KDD in data preparation projects. Developing field protocols

Sana Skandrani

Karlstad Business School, sana.skandrani@kau.se

Muhammad Murtaza Ali

Karlstad Business School, muhammadmurtaza.ali@kau.se

Follow this and additional works at: https://aisel.aisnet.org/treos_icis2022

Recommended Citation

Skandrani, Sana and Ali, Muhammad Murtaza, "Domain expertise afore and beyond KDD in data preparation projects. Developing field protocols" (2022). *ICIS 2022 TREOs*. 71.

https://aisel.aisnet.org/treos_icis2022/71

This material is brought to you by the TREO Papers at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2022 TREOs by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

TREO

Technology, Research, Education, Opinion

Domain expertise afore and beyond KDD in data preparation projects

Developing field protocols

Sana Rouis Skandrani sana.skandrani@kau.se

Muhammad Murtaza Ali muhammadmurtaza.ali@kau.se

Researchers and practitioners in information systems and artificial intelligence (AI) are facing today a twofold situation: First, despite the improvement made throughout the last two decades (with a decrease to 45% preparation time in analytics projects (Woodi, 2020) from 80%), there is still a need for considerable improvement of the data preparation process. Second, recent studies on data preparation for Machine Learning and AI have iterated discussion, and investigation on how relevant it would be to integrate domain expertise. The last discussion comes at two levels: in the data preparation process, when identifying the data sources to use for the Knowledge Discovery in Databases (KDD) process, but also when integrating experts knowledge in assessing the final AI and ML models that result from the analysis (Fügener et al., 2021).

In order to help reducing the time spent on data preparation, but also to guarantee high quality ingested data, AI models and results, we presume that there is a need to establish a protocol that presents and describes key steps to an effective data preparation and analysis process. Nevertheless, observing that researchers' first conclusion are disparate across fields, we assume that it is necessary to have different protocols for different fields. For instance, in fields such as archeology, there is a need to use local, national and international standards and repositories and references. In other fields such as industry, there is no need of such references. However, earlier references and expertise used in practical research in the health sector have not resulted in effective machine-learning models (Lebovitz et al., 2021). This shows the need for the protocol to handle exceptions as well.

References

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will Humans-In-The-Loop Become Borgs? Merits and Pitfalls of Working with Ai. *MIS Quarterly*, 45(3), 1527–1556. <https://doi-org.bibproxy.kau.se/10.25300/MISQ/2021/16553>

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is Ai Ground Truth Really True? The Dangers of Training and Evaluating Ai Tools Based on Experts' Know-What. *MIS Quarterly*, 45(3), 1501–1525. <https://doi-org.bibproxy.kau.se/10.25300/MISQ/2021/16564>

Woodie, A. (2020), Data Prep Still Dominates Data Scientists' Time, Survey Finds. <https://www.datanami.com/2020/07/06/data-prep-still-dominates-data-scientists-time-survey-finds/>. Last accessed: October 11, 2022