

Master Project

The Contribution of Artificial Intelligence Platforms to Explainable Machine Learning: A Comparative Assessment

Term: Winter Term 2021/22

Language: English

Motivation

Historically, artificial intelligence (AI) technologies are faced with volatile phases of excitement and disillusionment. The current phase of AI excitement is centered around inductive machine learning (ML) approaches which spans a broad range of different implementations such as logistic regressions, k-means clustering algorithms, or deep learning models. Especially more extensive ML models, however, often lack comprehensibility. For example, it is often not clear how inferences are produced or what the influence of the underlying training data is. Such information is however crucial for its sophisticated use in organizations, e.g., with regards to accountability and traceability. A frequent example is discrimination in HR models. The research field commonly denoted as “explainable AI” (XAI) or “interpretable ML” addresses these challenges.

The specification and implementation of ML models is supported through numerous libraries, frameworks, and platform solutions. Various sophisticated platform solutions not only support the development of ML models through reusable code fragments but intend to contribute to the explainability of these models.

Description

This master’s project shall analyze and evaluate existing ML platform solutions against requirements postulated by XAI research. It is therefore required to develop an analysis framework and use cases to guide the examination. Open data sources such as those provided by Kaggle can support the definition of feasible use cases. Based on the specified use cases and corresponding ML models, the respective solution should be developed with a selected number of ML platforms. A broad population of vendors offer free ML platforms with diverse features aimed at increasing the explain ability of ML models. Among them are: Google’s Tensorflow (especially TensorBoard and Playground), H2O.ai, Dataiku, DeepCognition, Azure Machine Learning, Oracle Skater, and Amazon SageMaker. The number and range of use cases and platforms can vary based on the quantity of group members.

**Institut für Informatik und
Wirtschaftsinformatik (ICB)**

**Lehrstuhl für
Wirtschaftsinformatik und
Unternehmensmodellierung**

Pierre Maier

Tel.: 0201 / 183 - 4563

pierre.maier@uni-due.de

Mario Nolte

Tel.: 0201 / 183 - 4088

mario.nolte@uni-due.de

R09 R04 H41

Universitätsstraße 9

45127 Essen

www.umo.wiwi.uni-due.de

The range of tasks can include the following:

- elaboration of analysis framework based on XAI research
- specification and definition of use cases with corresponding ML models
- identification and selection of ML platforms
- implementation of use cases on platforms
- assessment of XAI support by platform

Expected Outcomes

ML use cases and proposed solutions should be developed. The specified ML models need to be executable and pre-trained models have to be provided. The results can be captured in a single project report with respective access to the developed solutions. The project report must encompass the presentation of use cases and ML-based solutions. Furthermore, a structured listing and evaluation of XAI methods as offered by the platforms should be included.

Introductory Literature

- Murphy KP (2012) Machine Learning: A Probabilistic Perspective. The MIT Press, Cambridge, MA, London.
- Hall P, Gill N (2019) An Introduction to Machine Learning Interpretability. Sebastopol, CA, O'Reilly
- Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D (2018) A Survey of Methods for Explaining Black Box Models. ACM Computing Surveys 51 (5)
- Bauer K, Hinz O, van der Aalst W, Weinhardt C (2021) Expl(AI)n It to Me – Explainable AI and Information Systems Research. Business & Information Systems Engineering 63 (2): 79-82
- Samek W, Müller KR (2019) Towards Explainable Artificial Intelligence. In: Samek W, Montavon G, Vedaldi A, Hansen LK, Müller KR (eds.) Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, pp. 5-22. Springer, Cham.
- Weller A (2019) Transparency: Motivation and Challenges. In: Samek W, Montavon G, Vedaldi A, Hansen LK, Müller KR (eds.) Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, pp. 23-40. Springer, Cham.
- Zerill K, Knott A, Maclaurin J, Gavaghan C (2019) Transparency in Algorithmic and Human Decision-making: Is There a Double Standard? Philosophy & Technology 32 (4): 551-683

Application Procedure:

Please apply via email to the supervisor. Please attach a short letter of motivation (approximately one A4 page) and a recent performance record ('Leistungsnachweis'). You can apply individually or in a group of 2-3 participants (in this case each person should still send a separate e-mail, however point to the other members of the group).

Application deadline: 20 October 2021, 23:59