Chair of Explainable AI-Based Business Information Systems
Prof. Dr. Ulrich Gnewuch

Master's Thesis: Uncertainty Representation in Machine Learning Applications

Supervisor: M.Sc. Philipp Hansen (philipp.hansen@uni-passau.de)
Start date: at the next possible date

Motivation and Goals

In the field of Explainable Artificial Intelligence (XAI), several methods have been proposed to address the 'black box' characteristic of complex Machine Learning (ML) models, arising from numerous parameters and nonlinearities (Watson et al., 2023). These XAI methods primarily focus on explaining the ML model's predictions as such, rather than the underlying uncertainty (Watson et al., 2023). However, quantifying this uncertainty is important (Mehdiyev et al., 2024; Watson et al., 2023), as relying on incorrect model predictions poses a risk to informed decision-making (Watson et al., 2023).

In light of this, the goal of the master's thesis is twofold: Firstly, the current state of research concerning uncertainty measures in ML/XAI applications shall be reviewed. Secondly, one type of uncertainty quantification shall be implemented in an XAI prototype and evaluated regarding, e.g., its effect on trust.

Required Skills

- Strong interest in (X)AI & Mathematics
- Good English skills
- Ideally, prior coding experience

Starting Literature (Topic)

Mehdiyev, N., Majlatow, M., & Fettke, P. (2024). Communicating Uncertainty in Machine Learning Explanations: A Visualization Analytics Approach for Predictive Process Monitoring. In L. Longo, S. Lapuschkin, & C. Seifert (Eds.), *Communications in Computer and Information Science* (Vol. 2155, pp. 420–438). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-63800-8_21

Thuy, A., & Benoit, D. F. (2024). Explainability through uncertainty: Trustworthy decision-making with neural networks. *European Journal of Operational Research*, 317(2), 330–340. https://doi.org/10.1016/j.ejor.2023.09.009

Watson, D. S., O'Hara, J., Tax, N., Mudd, R., & Guy, I. (2023). Explaining Predictive Uncertainty with Information Theoretic Shapley Values. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.), *Advances in Neural Information Processing Systems* (Vol. 36, pp. 7330–7350). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2023/hash/16e4be78e61a3897665fa01504e9f452-Abstract-Conference.html

Starting Literature (Method)

Vom Brocke, J., Hevner, A., & Maedche, A. (2020). Introduction to Design Science Research. In J. vom Brocke, A. Hevner, & A. Maedche (Eds.), Design Science Research. Cases (pp. 1–13). *Springer International Publishing*. https://doi.org/10.1007/978-3-030-46781-4_1

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75-105. https://doi.org/10.2307/25148625

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), xiii–xxiii. https://www.jstor.org/stable/4132319