Bachelor/Master Project

Feature and Scenario-Driven Evaluation of Selected Data Science Reference Frameworks/Models

Term: Summer Term 2022

Language: English/German

Motivation:

Decision-making processes in organizations are nowadays often supported by information technology used to process and analyze large amount of data. Data-driven disciplines under the umbrella term 'data science' (e.g., data mining, machine learning, knowledge management) provide various approaches, reference models and frameworks that could be used for data analysis projects. Examples of such frameworks are the well-known cross-industry standard process for data mining (CRISP-DM), knowledge discovery in databases (KDD), or the data science process model (DASC-PM). Each framework has different foci and thus, is more or less suitable to support some specific scenarios.

Description

The main aim of this bachelor/master project is to conduct a comparative evaluation (both feature-based as well as scenario-driven) of selected data science/data analysis reference frameworks/approaches (among others, CRISP-DM, KDD, DASC-PM), identify their strengths and weaknesses, as well as their suitability to support data science projects in organizational context. To this aim the students should:

(1) Make themselves familiar with the existing data-analysis reference frameworks, their main assumptions, core concepts and phases.

(2) Select a real-world scenario to be used for the needs of comparative evaluation.

(3) Design an evaluation framework as well as an evaluation protocol.

(4) Conduct the evaluation of selected approaches/frameworks.

(5) Critically analyze obtained results and draw conclusions.

Institut für Informatik und Wirtschaftsinformatik (ICB)

Lehrstuhl für Wirtschaftsinformatik und Unternehmensmodellierung

Dr. Monika Kaczmark-Heß
Tel.: 0201 / 183 - 4330
monika.kaczmarek-hess@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

Expected Outcomes

A report describing conducted state-of-the-art analysis, cf. point (1) of the project description, the selected scenario, proposed evaluation framework and protocol, a detailed analysis of the evaluation results, and final assessment of the approaches and their suitability to support conducting data science projects in organizational context. In addition, a final presentation of the project results is expected.

Introductory Literature

- Cao, L. 2017."Data Science: A Comprehensive Overview", ACM Computing Surveys, (50:3), pp. 1-42.
- Schulz, Michael; Neuhaus, Uwe; Kaufmann, Jens; Badura, Daniel; Kuehnel, Stephan; Badwitz, Wolfgang; Dann, David; Kloker, Simon; Alekozai, Emal M.; and Lanquillon, Carsten, "Introducing DASC-PM: A Data Science Process Model" (2020). ACIS 2020 Proceedings. 45.
- Azevedo, A., and Santos, M.F. 2008. "KDD SEMMA and CRISP-DM: A Parallel Overview", Proc. Int'l Assoc. Development of the Information Soc. European Conf. Data Mining, pp. 182-185
- Wirth, R., Hipp, J. 2000. "CRISP-DM: Towards a Standard Process Model for Data Mining", Proc. 4th Int. Conference on Practical Applications of Knowledge Discovery and Data mining, pp. 29-39.

Application Procedure:

Please apply via email to the supervisors. Please attach a short letter of motivation (approximately one A4 page) and a transcript of records ('Leistungsnachweis'). You can apply individually or in a group of **2-4 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 April 2022, 23:59 h