Data Science Day 2024

Wednesday, 8 May 2024 - Wednesday, 8 May 2024

Auditorium in the main building of the Friedrich Schiller University Jena

Programme

Event Schedule: Data Science Day Jena 2024
 1:00 PM - 1:05 PM
Welcome and Introduction

Speaker: Joachim Giesen
 1:05 PM - 1:50 PM

Keynote: Digitalisation strategy of the University Jena [english]
Speaker: Christoph Steinbeck, *Vice-President for Digitalisation*
Abstract:
The digital transformation provides the Friedrich Schiller University Jena with opportunities for profiling as well as for a variety of innovations and, at the same time, challenges it to adopt new ways of working and modernization. Digitalization should be the subject and instrument of interdisciplinary research, requires the further development of teaching content, enables new ways of imparting knowledge and is a basis for service-oriented and modern administration. The strategic development of digitalization at the University of Jena is to focus on the needs and requirements of university members and is in line with the university's strategy processes for *research, teaching, promotion of young scientists* and *internationalization*.
 1:50 PM - 2:20 PM

Coffee Break & Poster Session
 2:20 PM - 2:45 PM

Presentation: Kleine Moleküle: Was uns tötet, was uns heilt [german]

Speaker: Sebastian Böcker
Abstract:
Im Juli/August 2022 verendeten die Hälfte der Fische in der Oder, und es begann eine hektische Suche nach dem verantwortlichen Gift. Gifte wirken in geringen Konzentrationen und sind analytisch nur äußerst schwer nachzuweisen. Außerdem sind viele Gifte noch gänzlich unbekannt, töten uns aber trotzdem. Oft handelt es sich bei Giften um sogenannte "kleine Moleküle". Andererseits sind fast alle Medikamente ebenfalls kleine Moleküle, und der Übergang ist fließend: So nutzen wir Gifte von Pilzen als Antibiotika gegen bakterielle Infektionen.

In meinem Vortrag werde ich beleuchten, wie man einem kleinen Molekül "ansieht", ob es giftig oder heilsam ist, und wie ich die Messdaten kleiner Moleküle interpretieren kann, die noch nie zuvor gemessen wurden. In beiden Fällen sind es Methoden des maschinellen Lernens, die es uns erlauben, von bekannten kleinen Molekülen auf unbekannte zu generalisieren.
 2:45 PM - 3:10 PM

Presentation: Data-driven designation of nitrate-polluted groundwater areas in Germany [english]
Speaker: Alexander Brenning
Abstract:
Implementing the EU Nitrates Directive in Germany requires the designation of nitrate-polluted groundwater areas using geostatistical or deterministic regionalization methods. This study evaluates methods for national-scale designation, identifying limitations of currently applied approaches, outlining suitable state-of-the-art geostatistical approaches, and exploring spatial random forest techniques. Regression-kriging is an established geostatistical technique that fulfills regulatory requirements while accounting for spatial heterogeneity, being applicable to the unbiased identification of exceedance regions based on point measurements of groundwater nitrate concentration. Empirical comparisons highlight biases in traditional approaches, emphasizing the

importance of considering local prediction uncertainty and exceedance modeling. The potential and challenges of integrating geostatistical techniques with machine learning models in regulatory contexts are discussed.
⬜ 3:10 PM - 3:25 PM

Coffee & Networking Break ⬜
⬜ 3:25 PM - 3:50 PM

Presentation: Symbolic Regression [english]
Speaker: Paul Kahlmeyer
Company: Friedrich Schiller University Jena
Abstract: TBA
⬜ 3:50 PM - 4:15 PM

Presentation: Safe AI for Computer Vision of Automated Vehicles [english]
Speaker: Stefan Milz
Company: Spleenlab
Abstract:
The presentation will show the current challenges for AI in the area of safety-critical scenarios for real-time robot automation. In Detail use cases like Precise Automated Landing for Delivery or Near-field Automotive Perception will be investigated and example applications will be given including cutting edge sensor technologies. A methodology is presented how AI should be certified using operational domain definition, scenario based analytics, requirement traceability throughout neural networks. These principles are generically applicable in all field of automation (drones, cars) The presentation will open eyes for a new certification of AI using large scale data models and motion based functions.
⬜ 4:15 PM - 5:15 PM

Company Exhibition ⬜
- ZEISS Group
- Jena Digital e. V.
- Accenture Technology Solutions GmbH
- d-fine GmbH
- Xceptance Software Technologies GmbH
- Alpha Analytics UG & Co. KG
- Spleenlab GmbH
- Navimatix GmbH
- Inverso GmbH
- DAKO GmbH
- NFDI
- ELLIS unit Jena
- zedif: Kompetenzzentrum Digitale Forschung
- Projekt Data Literacy Jena (DaLiJe)
⬜ 5:15 PM - 6:00 PM

Capstone Presentation: Backends Unveiled: The Hidden Heroes of Data Science [english]

Speaker: Alex Breuer
Abstract:
Data science is a dynamic field that thrives on extracting insights from vast amounts of data. While high-level algorithms and models often take center stage, the unsung heroes in the background — the software backends — enable efficient computation and accelerate data science workflows. This

presentation explores how the backends seamlessly bridge the gap between high-level data science code and low-level execution.

The grand challenge faced by all backends lies in the diversity of workloads and hardware. Imagine a scenario with N distinct machine learning models. Simultaneously, we target M different hardware architectures, ranging from CPUs and GPUs to specialized accelerators like TPUs. At first glance, the solution appears simple: map each workload to every platform. However, this approach quickly becomes infeasible as the number of implementations grows to N*M.

One solution to this problem is to find an effective Intermediate Representation (IR). The IR acts as an intermediary language that abstracts away workload specifics and hardware details. If the IR is composed of K << N primitives, for example, matrix multiplications or element-wise operations, this step reduces the problem from N*M to a more manageable K*M, where K << N. We will discuss existing IR approaches and conclude the presentation by discussing recent contributions to the open-source backend ecosystem made by my group at Friedrich Schiller University Jena.
□ 6:00 PM - 8:00 PM

Get-together □

Enjoy the event!