

## HDS-LEE Seminars

### Jülich Plant Science Seminar (JPSS)

#### **Prof. Kristian Kersting**

Technical University of Darmstadt, Department of Computer Science

talks about

#### **The Automatic Data Scientist**

Have you ever tried to stand on another machine learning/ data science researcher's work and not been able to repeat their empirical finding? Most likely, you are not alone. A 2016 survey presented in the journal Nature (Baker. Nature 533(7604):452-4) argues that about "70% of researchers have tried and failed to reproduce another scientist's experiments." And reproducing machine learning/ data science results is seldom straightforward either, as noted e.g. by Henderson et al. at AAAI 2018. Thus, the democratization of machine learning and data science does not mean dropping the data on everyone's desk and saying, "good luck"! It means making machine learning and data science methods usable in such a way that people can easily instruct machines to have a "look" at data and help them to understand and act on it.

This is also the vision of high-level programming languages for machine learning/ data science, as I shall argue in the talk. High-level descriptions using relations, quantifiers, loops, functions, and procedures provide clarity and succinct characterisations of the machine learning/ data science problem at hand and improve the credibility of past and future data-driven research. Putting deep probabilistic learning onto the stack, it may even help the domain expert to "make sense" of her data with minimal expert input. Moreover, putting the expert into the loop, she may even remove "Clever Hans"-like moments, making use of confounding factors within datasets, from data science.

**Date: Tuesday March 17, 2020**

Time: starting at 10:30

Place: Seminar room of IBG-2 (R406), Building 06.2, Forschungszentrum Jülich