

SIKS-CWI-UU Symposium on "**Uncertainty** in Data-Driven Systems" in Utrecht

INTRODUCTION

On Januari 28 2019 a SIKS-CWI-UU Symposium will be organized following the PhD-defense of Emma Beauxis-Aussalet. All SIKS-members are cordially invited to attend a series of talks on "**Uncertainty** in Data-Driven Systems", presented by experts in the field. Especially SIKS-PhD-students working on the topics of the symposium are strongly encouraged to participate. The talks will take place on Monday January 28th, starting at 14:00, in the same building as the PhD defence (Domplein 29, Utrecht).

PROGRAM

14:00 - 14:30 *(Un)certain Anomalies in Income Data*
Dr. Ralph Foorthuis - UWV

This presentation gives an overview of the data and quality aspects of the Dutch alliance on income data and taxes (Loonaangifteketen). Several approaches for verifying the data quality by means of anomaly detection will be discussed (e.g., rule-based and data-driven), as well as the degree to which one can be certain with regard to the results.

14:30 - 15:15 *Interpretable Models, Learning with Reject Option, and Learning with Drift*
Prof. Barbara Hammer - Bielefeld University

Machine learning technologies have revolutionized many domains such as vision or language processing, yet many models and in particular the majority of mathematical substantiations are yet restricted to the classical setting of batch learning (i.e. data are given prior to training), stationary distributions (i.e. data characteristics do not change during their lifetime), and optimization of the classification error (rather than optimizing strategies, when to best abstain from a classification in unclear cases). In the talk, we will have a glimpse on machine learning technologies which, by design, provide interpretability and open up avenues how to extend them to reject options and learning from possibly non-stationary streaming data.

Break

15:30 - 16:15 *Data Analysis of Error-Prone Observations*
Dr. Daniel Oberski - Utrecht University

Nothing is perfect, especially data. Measurement error, specifically, can distort practical predictions and scientific conclusions, and a number of solutions to this problem have been suggested in the statistical literature. I review these suggestions, and discuss some of our recent work to simplify data analysis of error-prone observations for applied researchers.

16:15 - 17:00 *Population size estimation where linking registers results in incomplete covariates*
Prof. Peter van der Heijden - Utrecht University

An overview of population size estimation where linking registers results in incomplete covariates, with an application to mode of transport of serious road casualties, where misclassification occurs.

[1] Statistics and Visualizations for Assessing Class Size **Uncertainty**, Emma Beauxis-Aussalet, PhD Thesis, Utrecht University.

<http://tinyurl.com/Emma-PhD-190128>