# Preface

These proceedings contain the papers accepted to the 1$^{\text{st}}$ ECML/PKDD Workshop on Statistically Sound Data Mining, which took place at the French Institute for Computer Science (INRIA) in Nancy, at the opening of the *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (ECML/PKDD) on the 15$^{\text{th}}$ of September 2014.

The aim of the workshop was to strengthen collaboration between data mining and statistical researchers, by providing a forum to debate problems faced by both communities, including statistical soundness of data mining methods, scalability issues for standard statistical techniques and comparisons between different types of approaches (e.g., frequentist and Bayesian).

The workshop program consisted of oral presentations of accepted papers, an invited talk and a plenty of discussion among participants. The four accepted full papers concerned important questions on evaluating models vs. learners with cross-validation, significance testing of data mining results, multiple testing problem in the context of genome-wide association studies, and the use of $U$-statistics for networked data. In addition, there was an invited talk by Professor Charles Elkan from the University of California, San Diego, entitled "Massive, sparse, efficient multilabel learning". The workshop was finished by group works on open problems using a problem-based learning method.

We want to thank the authors, the program committee and the organizers of the conference for their help and support. We also thank the INRIA Nancy for their support organising the ECML/PKDD conference and helping organising the logistics of this workshop.

### Program Committee

The program committee was composed of the following 10 members, in addition to the editors of this book:

- Niall Adams, Imperial College London, UK
- Peter Flach, University of Bristol, UK
- Stephane Lallich, Universite Lyon 2, France
- Daniel Lawson, University of Bristol, UK
- Jiuyong Li, University of Southern Australia
- Cecile Low-Kam, Montreal Heart Institute, Canada
- Siegfried Nijssen, KU Leuven, Belgium
- Chedy Raissi, INRIA, France
- Jan Ramon, KU Leuven, Belgium
- Nikolaj Tatti, Aalto University, Finland

Wilhelmiina Hämäläinen
Academy of Finland
whamalai@cs.uef.fi

François Petitjean
Monash University
francois.petitjean@monash.edu

Geoffrey I. Webb
Monash University
geoff.webb@monash.edu