

# Contents

## **Part I AREA Statistics and Data Analysis: Classification, Cluster Analysis, Factor Analysis and Model Selection**

<b>On Limiting Donor Usage for Imputation of Missing Data via Hot Deck Methods .....</b>	<b>3</b>
Udo Bankhofer and Dieter William Joensen	
<b>The Most Dangerous Districts of Dortmund .....</b>	<b>13</b>
Tim Beige, Thomas Terhorst, Claus Weihs, and Holger Wormer	
<b>Benchmarking Classification Algorithms on High-Performance Computing Clusters.....</b>	<b>23</b>
Bernd Bischl, Julia Schiffner, and Claus Weihs	
<b>Visual Models for Categorical Data in Economic Research .....</b>	<b>33</b>
Justyna Brzezińska	
<b>How Many Bee Species? A Case Study in Determining the Number of Clusters .....</b>	<b>41</b>
Christian Hennig	
<b>Two-Step Linear Discriminant Analysis for Classification of EEG Data .....</b>	<b>51</b>
Nguyen Hoang Huy, Stefan Frenzel, and Christoph Bandt	
<b>Predictive Validity of Tracking Decisions: Application of a New Validation Criterion.....</b>	<b>61</b>
Florian Klapproth, Sabine Krolak-Schwerdt, Thomas Hörstermann, and Romain Martin	
<b>DD<math>\alpha</math>-Classification of Asymmetric and Fat-Tailed Data.....</b>	<b>71</b>
Tatjana Lange, Karl Mosler, and Pavlo Mozharovskyi	

<b>The Alpha-Procedure: A Nonparametric Invariant Method for Automatic Classification of Multi-Dimensional Objects</b> .....	79
Tatjana Lange and Pavlo Mozharovskyi	
<b>Support Vector Machines on Large Data Sets: Simple Parallel Approaches</b> .....	87
Oliver Meyer, Bernd Bischl, and Claus Weihs	
<b>Soft Bootstrapping in Cluster Analysis and Its Comparison with Other Resampling Methods</b> .....	97
Hans-Joachim Mucha and Hans-Georg Bartel	
<b>Dual Scaling Classification and Its Application in Archaeometry</b> .....	105
Hans-Joachim Mucha, Hans-Georg Bartel, and Jens Dolata	
<b>Gamma-Hadron-Separation in the MAGIC Experiment</b> .....	115
Tobias Voigt, Roland Fried, Michael Backes, and Wolfgang Rhode	
 <b>Part II AREA Machine Learning and Knowledge Discovery: Clustering, Classifiers, Streams and Social Networks</b>	
<b>Implementing Inductive Concept Learning For Cooperative Query Answering</b> .....	127
Maheen Bakhtyar, Nam Dang, Katsumi Inoue, and Lena Wiese	
<b>Clustering Large Datasets Using Data Stream Clustering Techniques</b> ....	135
Matthew Bolaños, John Forrest, and Michael Hahsler	
<b>Feedback Prediction for Blogs</b> .....	145
Krisztian Buza	
<b>Spectral Clustering: Interpretation and Gaussian Parameter</b> .....	153
Sandrine Mouysset, Joseph Noailles, Daniel Ruiz, and Clovis Tauber	
<b>On the Problem of Error Propagation in Classifier Chains for Multi-label Classification</b> .....	163
Robin Senge, Juan José del Coz, and Eyke Hüllermeier	
<b>Statistical Comparison of Classifiers for Multi-objective Feature Selection in Instrument Recognition</b> .....	171
Igor Vatolkin, Bernd Bischl, Günter Rudolph, and Claus Weihs	
 <b>Part III AREA Data Analysis and Classification in Marketing</b>	
<b>The Dangers of Using Intention as a Surrogate for Retention in Brand Positioning Decision Support Systems</b> .....	181
Michel Ballings and Dirk Van den Poel	
<b>Multinomial SVM Item Recommender for Repeat-Buying Scenarios</b> .....	189
Christina Lichtenthäler and Lars Schmidt-Thieme	

<b>Predicting Changes in Market Segments Based on Customer Behavior ...</b>	199
Anneke Minke and Klaus Ambrosi	
<b>Symbolic Cluster Ensemble based on Co-Association Matrix versus Noisy Variables and Outliers .....</b>	209
Marcin Pelka	
<b>Image Feature Selection for Market Segmentation: A Comparison of Alternative Approaches .....</b>	217
Susanne Rumstadt and Daniel Baier	
<b>The Validity of Conjoint Analysis: An Investigation of Commercial Studies Over Time .....</b>	227
Sebastian Selka, Daniel Baier, and Peter Kurz	
<b>Solving Product Line Design Optimization Problems Using Stochastic Programming .....</b>	235
Sascha Voekler and Daniel Baier	
 <b>Part IV AREA Data Analysis in Finance</b>	
<b>On the Discriminative Power of Credit Scoring Systems Trained on Independent Samples .....</b>	247
Miguel Biron and Cristián Bravo	
<b>A Practical Method of Determining Longevity and Premature-Death Risk Aversion in Households and Some Proposals of Its Application .....</b>	255
Lukasz Feldman, Radoslaw Pietrzyk, and Pawel Rokita	
<b>Correlation of Outliers in Multivariate Data .....</b>	265
Bartosz Kaszuba	
<b>Value-at-Risk Backtesting Procedures Based on Loss Functions: Simulation Analysis of the Power of Tests .....</b>	273
Krzysztof Piontek	
 <b>Part V AREA Data Analysis in Biostatistics and Bioinformatics</b>	
<b>Rank Aggregation for Candidate Gene Identification .....</b>	285
Andre Burkovski, Ludwig Lausser, Johann M. Kraus, and Hans A. Kestler	
<b>Unsupervised Dimension Reduction Methods for Protein Sequence Classification .....</b>	295
Dominik Heider, Christoph Bartenhagen, J. Nikolaj Dybowski, Sascha Hauke, Martin Pyka, and Daniel Hoffmann	

<b>Three Transductive Set Covering Machines .....</b>	<b>303</b>
Florian Schmid, Ludwig Lausser, and Hans A. Kestler	
 <b>Part VI AREA Interdisciplinary Domains: Data Analysis in Music, Education and Psychology</b>	
<b>Tone Onset Detection Using an Auditory Model.....</b>	<b>315</b>
Nadja Bauer, Klaus Friedrichs, Dominik Kirchhoff, Julia Schiffner, and Claus Weihs	
<b>A Unifying Framework for GPR Image Reconstruction .....</b>	<b>325</b>
Andre Busche, Ruth Janning, Tomáš Horváth, and Lars Schmidt-Thieme	
<b>Recognition of Musical Instruments in Intervals and Chords .....</b>	<b>333</b>
Markus Eichhoff and Claus Weihs	
<b>ANOVA and Alternatives for Causal Inferences.....</b>	<b>343</b>
Sonja Hahn	
<b>Testing Models for Medieval Settlement Location.....</b>	<b>351</b>
Irmela Herzog	
<b>Supporting Selection of Statistical Techniques .....</b>	<b>359</b>
Kay F. Hildebrand	
<b>Alignment Methods for Folk Tune Classification.....</b>	<b>369</b>
Ruben Hillewaere, Bernard Manderick, and Darrell Conklin	
<b>Comparing Regression Approaches in Modelling Compensatory and Noncompensatory Judgment Formation .....</b>	<b>379</b>
Thomas Hörstermann and Sabine Krolak-Schwerdt	
<b>Sensitivity Analyses for the Mixed Coefficients Multinomial Logit Model .....</b>	<b>389</b>
Daniel Kasper, Ali Ünlü, and Bernhard Gschrey	
<b>Confidence Measures in Automatic Music Classification .....</b>	<b>397</b>
Hanna Lukashevich	
<b>Using Latent Class Models with Random Effects for Investigating Local Dependence .....</b>	<b>407</b>
Matthias Trendtel, Ali Ünlü, Daniel Kasper, and Sina Stubben	
<b>The OECD's Programme for International Student Assessment (PISA) Study: A Review of Its Basic Psychometric Concepts.....</b>	<b>417</b>
Ali Ünlü, Daniel Kasper, Matthias Trendtel, and Michael Schurig	
<b>Music Genre Prediction by Low-Level and High-Level Characteristics ...</b>	<b>427</b>
Igor Vatulkin, Günther Rötter, and Claus Weihs	

**Part VII LIS WORKSHOP: Workshop on Classification  
and Subject Indexing in Library and Information  
Science**

<b>Using Clustering Across Union Catalogues to Enrich Entries with Indexing Information .....</b>	<b>437</b>
Magnus Pfeffer	
<b>Text Mining for Ontology Construction .....</b>	<b>447</b>
Silke Rehme and Michael Schwantner	
<b>Data Enrichment in Discovery Systems Using Linked Data .....</b>	<b>455</b>
Dominique Ritze and Kai Eckert	
<b>Index .....</b>	<b>463</b>
<b>Author Index .....</b>	<b>469</b>

Data Analysis, Machine Learning and Knowledge  
Discovery

Spiliopoulou, M.; Schmidt-Thieme, B.; Janning, R. (Eds.)

2014, XXI, 470 p. 120 illus., 32 illus. in color., Softcover

ISBN: 978-3-319-01594-1